Démographie et sociétés

Centre de recherche en démographie
Institut d'analyse du changement dans l'histoire
et les sociétés contemporaines
Université catholique de Louvain

Répartition des décès et table de mortalité : à boire et à manger !

Christophe Vandeschrick

Document de travail 5

juin 2015
www.uclouvain.be/demo

Table des matières

Merci au Professeur Wunsch et à Antoine Pierrard pour leur relecture plurielle et attentive. Ce texte leur doit beaucoup. Merci aussi à Anne-Marie Lefin et Pierre Vandeschrick d'avoir eu la patience de relire un texte sans doute ardu, particulièrement pour des non-démographes.

Résumé

Pour calculer une table de mortalité, un exemple d'analyse de survie en temps discret, il faut poser une hypothèse sur la répartition des évènements entre deux âges successifs. Le plus souvent, la fonction de survie entre deux âges successifs est considérée comme étant linéaire (répartition uniforme des évènements). Si cette hypothèse est souvent largement acceptable, dans certaines circonstances, elle n'est plus appropriée, comme en tout début (ou fin) de vie. Ce texte a pour objectif de proposer des méthodes simples pour revisiter l'hypothèse de répartition uniforme et d'apprécier les conséquences sur les résultats de l'utilisation d'hypothèses de répartition non uniforme. Ainsi le démographe pourra plus facilement choisir une hypothèse appropriée aux données à traiter, notamment en cas d'application de la méthode de la table dans des contextes éloignés de son champ d'application habituel.

Summary

To calculate a life table, an example of survival analysis in discrete time, we need to choose a hypothesis concerning the distribution of events between two successive ages. In general, the survivorship function between two successive ages is considered to be linear (events uniformly distributed). If this hypothesis is often widely acceptable, in some circumstances it is not appropriate, as at the very beginning (and end) of life. In this text, simple methods are proposed to revisit the uniform distribution assumption and to evaluate the consequences of the use of other hypotheses. The demographer will therefore be able to choose more easily an appropriate hypothesis for the data to be analysed, especially when the method has to be applied in an unusual context.

Répartition des décès et table de mortalité : à boire et à manger !

1. Introduction

Pour analyser la mortalité et singulièrement établir une table de mortalité, le démographe doit s'interroger sur la répartition des décès à l'intérieur des surfaces du diagramme de Lexis (diagramme qui est supposé bien connu par le lecteur). Bien souvent, le choix se porte sur l'hypothèse de répartition uniforme des décès. Ce papier vise à revisiter cette hypothèse, à vérifier si les conséquences que l'on en tire sont cohérentes et à envisager d'autres formes de répartition et leurs potentielles conséquences sur les résultats

Le premier point abordera la question de l'âge moyen au décès dans un 1er triangle, en envisageant différentes façons d'arriver au résultat recherché. L'utilité de certaines des méthodes envisagées ne s'affirme pas spécialement en cas de répartition uniforme (si ce n'est à titre de vérification), mais bien en cas d'abandon de cette dernière, ce qui sera envisagé à la fin de ce point 1.

Le deuxième point traitera, pour sa part, de l'âge moyen exact des individus formant l'effectif de 0 an révolu. Avant de démontrer que le résultat habituel (soit 0,5 an exact) n'est en fait pas acceptable, au moins sur un plan théorique, il sera montré qu'une autre hypothèse, concernant la répartition des naissances, doit être aussi posée, ce qui est assez généralement omis.

La lecture de ce texte, plutôt destiné à des étudiant(e)s, suppose de connaitre, outre les bases du calcul intégral, le diagramme de Lexis, le calcul des quotients au départ des données observées et la table de mortalité dans ses fonctions les plus habituelles. Ces notions ne seront pas expliquées ici[1]. Enfin, la vocation pédagogique du texte excuse certaines longueurs ou répétitions dans les développements.

2. Âge moyen au décès dans le 1er triangle

En cas de répartition uniforme des décès, l'âge moyen au décès dans le 1er triangle d'un parallélogramme entre deux âges exacts successifs correspond au 1er âge augmenté de 1/3, l'argument invoqué étant que cet âge correspond à celui du centre de gravité du triangle en cause[2]. Cet argument est éminemment recevable. Toutefois, sans passer par la notion de centre de gravité, Boyarski[3] a proposé une démonstration pour arriver au tiers (point 2.1). Ensuite, ce résultat sera de nouveau établi via un calcul d'abord discret (point 2.2), ensuite infinitésimal (point 2.3). Grâce à la méthode infinitésimale, il sera particulièrement aisé de s'affranchir de l'hypothèse de répartition uniforme des décès (point 2.4). Le point 2.5 reviendra sur les intégrales du point précédent, mais en utilisant la densité de probabilité, et non plus la densité des décès. Finalement, après les conclusions sur l'âge moyen au décès dans le 1er triangle (point 2.6), le cas de la médiane sera envisagé en reprenant une partie des méthodes utilisées pour la moyenne (point 2.7).

[1] Pour un exposé complet de la construction d'une table de mortalité, *cf.* Vallin et Caselli (2001, a et b) ; Preston, Heuveline et Guillot (2001) ; Pressat (1973 et 1985) ; Vandeschrick (2004) ; Vandeschrick (2005) ou Wunsch et Termote (1978).

[2] *Cf.* par exemple Wunsch (1999), p. 16 du chapitre 4.

[3] Boyarski (1945, pp. 214-219). Nous devons cette référence à T. Kucera (qui nous a renseigné la publication de Boyarski), avec la complicité de G. Wunsch (qui nous a mis sur la piste) et de N. Kalmykova (qui nous a fourni une copie du document tout en assurant la traduction des passages essentiels pour le propos). Il nous plait de les remercier vivement ici de conserve.

2.1. L'astucieuse démonstration de Boyarski

Dans la suite, on va se concentrer spécifiquement sur le 1er triangle entre 0 et 1 an exact, malgré le fait que la répartition uniforme soit largement inacceptable à ces âges[4]. Outre sa fréquente utilisation même dans ce contexte précis, il sera aisé de généraliser les conclusions méthodologiques à tous les âges. Selon nos connaissances actuelles, Boyarski serait le premier à avoir proposé une démonstration justifiant que, en cas de répartition uniforme, l'âge moyen à l'évènement (l'émigration dans le cas de Boyarski) dans un 1er triangle valait 1/3. Pour expliquer la vision de Boyarski, nous allons utiliser le diagramme de Lexis tel qu'il est actuellement utilisé alors que Boyarski en utilisait sa version originale[5].

La question à résoudre est donc la suivante : en cas de répartition uniforme des évènements, quel est l'âge moyen à cet évènement pour les individus le subissant dans le 1er triangle de la table, à savoir le triangle limité par les âges de 0 exact et 0 révolu (*cf.* le 1er triangle sur la figure 1.a). Dans la suite de ce texte, l'expression « 1er triangle » se réfère exclusivement à cette surface géométrique.

Figure 1.a. 1er triangle sur diagramme de Lexis

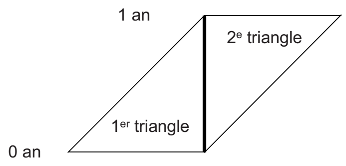
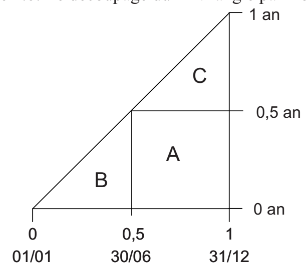


Figure 1.b. Le découpage du 1er triangle par Boyarski



Pour sa démonstration Boyarski découpe le 1er triangle en trois parties (*cf.* figure 1.b) :

- un carré « A » dont le coté vaut 0,5, couvrant des âges allant de 0 à 0,5 an ;
- un triangle « B » de base et de hauteur valant 0,5, couvrant des âges 0 à 0,5 an ;
- un triangle « C » de base et de hauteur valant 0,5, couvrant des âges 0,5 à 1 an.

Désignons par $\bar{y}$ l'âge moyen à l'évènement dans le 1er triangle, soit la valeur à découvrir. Cette valeur sera une moyenne pondérée des âges moyens dans le carré A et les deux triangles B et C, les poids étant les surfaces de ces trois éléments, répartition uniforme des décès oblige. Identifions d'abord l'âge moyen dans ces trois surfaces A, B et C :

- dans le carré A, vu l'hypothèse de répartition uniforme, l'âge moyen à l'évènement dans ce carré vaut 1/4 d'année, ce que Boyarski justifie en p. 215 comme suit : « Compte tenu de la symétrie de la partie III, il est facile de constater que la hauteur en moyenne correspond à la moitié de ac, c'est-à-dire ¼ (car $B_xA_{x+1} = 1$) ». Dans cette citation, la « partie III » est le carré A de la figure 1.b ; « ac » est la hauteur du carré A (soit 0,5) et « B_xA_{x+1} » celle du 1er triangle dans lequel s'inscrit le carré, soit 1. Notons, que, à notre avis, cette justification n'est pas immédiate et semble par ailleurs utiliser, au moins implicitement, le centre de gravité d'un carré, ce qui nous ramène presqu'à la solution passant par le centre de gravité du triangle évoquée en tout début de point 1…

[4] *Cf.* par exemple, Wunsch et Termote (1978), pp. 81-82 et Duchesne *et al.* (2002), pp. xiv-xv et 1-2. Notons aussi que, quand le risque de mourir à la naissance est calculé par transformation d'un taux de mortalité, généralement le caractère non uniforme de la répartition des décès est pris en compte (*cf.* par exemple Wunsch (2002, p. 16)).

[5] À propos de l'histoire du diagramme de Lexis, *cf.* Vandeschrick (1992, 2001 & 2005).

- pour le triangle B, Boyarski part du constat que ce triangle B et le 1er triangle sont semblables, étant entendu que la base de B ainsi que sa hauteur valent la moitié de celles du 1er triangle. Conclusion, l'âge moyen dans ce triangle vaut la moitié de $\bar{y}$;
- pour le triangle C, Boyarski part du même constat : ce triangle C est semblable au 1er triangle. Conclusion, l'âge moyen dans ce triangle vaut la moitié de $\bar{y}$ à laquelle il convient d'ajouter 1/2, étant donné que les âges couverts par ce triangle C vont de 0,5 à 1.

Pour le calcul de la moyenne dans le 1er triangle, que vaut le poids en termes de décès du carré A et des triangles B et C ? Vu l'hypothèse de répartition uniforme des décès, les deux triangles B et C auront le même poids, fixé arbitrairement pour la suite à 1. Le carré se verra attribuer un poids de 2, étant donné que sa surface est elle-même la somme des surfaces des triangles B et C. L'âge moyen recherché peut donc s'exprimer comme suit :

$\bar{y} = \dfrac{\left(2*\frac{1}{4}\right)+\left(\frac{\bar{y}}{2}\right)+\left(\frac{\bar{y}}{2}+\frac{1}{2}\right)}{4}$, où la 1re parenthèse concerne le carré A ; les 2e et 3e, les triangles B et C.

Il suffit de traiter cette équation pour obtenir le résultat recherché :

$\bar{y} = \dfrac{\left(2*\frac{1}{4}\right)+\left(\frac{\bar{y}}{2}\right)+\left(\frac{\bar{y}}{2}+\frac{1}{2}\right)}{4} = \dfrac{\bar{y}+1}{4} \quad \Rightarrow 4*\bar{y} = \bar{y}+1 \quad \Rightarrow 3*\bar{y} = 1 \quad \Rightarrow \bar{y} = \dfrac{1}{3}$, cqfd.

Ce calcul de l'âge moyen se fait au départ de l'âge moyen dans des parties du 1er triangle (ce qui nous semble un avantage, au moins sur le plan pédagogique) et pas sur la recherche du centre de gravité. Malgré ces qualités, nous préférons encore d'autres démonstrations.

2.2. Calcul de l'âge moyen par une voie discrète

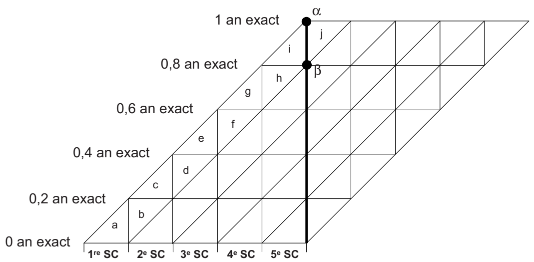
La démonstration de Boyarski montre que l'âge moyen à l'évènement dans un 1er triangle vaut 1/3 en cas de répartition uniforme des décès. Cette méthode marche très bien tant que la répartition uniforme est de mise. Par contre, en cas de répartition non uniforme, il est préférable de suivre d'autres voies. C'est la raison pour laquelle nous allons exposer d'autres méthodes d'estimation de cet âge moyen. L'exposé commence par la méthode discrète qui n'est pas la plus « efficace », mais sans doute la plus « intuitive ». Par ailleurs, cette méthode a pour vertu de suggérer toute la force du calcul infinitésimal pour résoudre de façon économique ce type de question. En effet, la méthode discrète conduit par glissement, en quelque sorte naturel, à poser et résoudre la question en termes infinitésimaux.

2.2.1. Une 1re approche

Pour la démonstration, nous partirons d'une situation avec 5 sous-cohortes, identifiées dans une génération et désignées par « SC » (*cf.* figure 2). Chaque sous-cohorte « x » est numérotée de 1 à 5, « 1 » étant la 1re à gauche et « 5 », la dernière à droite ; « n » désigne pour sa part le nombre de sous-cohortes, soit 5 ici. Pour rappel, sur la figure 2, l'expression « 1er triangle » désigne le triangle à gauche du trait vertical en gras sur lequel figurent les points α et β.

Comme le montre la figure 2, chaque sous-cohorte est composée de 10 petits triangles ; nous supposerons que chacun comporte 1 décès (répartition uniforme), ce qui fait 10 décès par sous-cohorte et 50 pour l'ensemble du parallélogramme. Pour la 1re sous-cohorte (la plus à gauche), 9 petits triangles (de « a » à « i ») sont dans le 1er triangle du parallélogramme, ce qui correspond à 9 décès ; un seul est dans le 2e triangle, soit le petit triangle « j ». L'âge moyen au décès des 9 décès de la 1re sous-cohorte situés dans le 1er triangle est de 0,45 ans ; en effet, la bandelette oblique de cette 1re sous-cohorte dans le premier triangle est délimitée par 0 an exact et 9/10 d'année (soit l'âge central du segment vertical de la 1re sous-cohorte en fin de la 1re année, le segment en gras qui sépare les triangles « i » et « j »), dont la moyenne donne 0,45 an.

Figure 2. Diagramme de Lexis en cas de calcul discret



Cet âge de 0,45 est en fait une approximation. Il est important ici de noter que, à mesure que le nombre de sous-cohortes augmente, l'importance de cette approximation diminue pour disparaitre quand ce nombre tend vers l'infini. En définitive, pour le raisonnement, cette approximation n'est gênante que dans la mesure où le nombre de sous-cohortes est faible !

Revenons à la question de l'âge moyen au décès dans le 1er triangle. D'une manière générale, le nombre de petits triangles de la sous-cohorte x localisés dans le 1er triangle (NPT_1T(x)) se calcule comme suit : $NPT\_1T(x) = (2n) - (2x-1) = 2n - 2x + 1 = (2n+1) - 2x$.

Dans le cas de 5 sous-cohortes, pour « x » = 1, cette formule donne 9 ; pour « x » = 2, 7 ; etc. (*cf.* tableau 1, colonne 2).

Tableau 1. Calcul de l'âge moyen en cas de 5 sous-cohortes (n = 5)

1	**2**	**3**	**4**	**5**
x	**NPT_1T(x)**	$\overline{ys}_x$	$\bar{y}_x$	**2*4**
1	9	0,9	0,45	4,05
2	7	0,7	0,35	2,45
3	5	0,5	0,25	1,25
4	3	0,3	0,15	0,45
5	1	0,1	0,05	0,05
Total	**25**			**8,25**

Sur la figure 2, le segment α-β ferme la bandelette de la sous-cohorte x. L'âge moyen des individus atteignant ce segment (soit $\overline{ys}_x$) est donné par :

$$\overline{ys}_x = \frac{NPT\_1T(x)}{2n} = \frac{(2n)-(2x-1)}{2n} = \frac{(n)-\left(x-1/2\right)}{n}.$$

Cet âge moyen est donc le nombre de petits triangles dans le 1er triangle divisé par 2 fois le nombre de sous-cohortes. Appliquée au cas où x = 1 (2 ; 3…), cette formule donne 0,9, comme explicité plus haut (0,7 ; 0,5…) (*cf.* tableau 1, colonne 3).

L'âge moyen au décès pour les décédés de la sous-cohorte x morts dans le 1er triangle ($\bar{y}_x$) se calcule comme suit : $\bar{y}_x = \frac{0+\overline{ys}_x}{2}$.

Dans cette formule, 0 est l'âge sur le segment horizontal qui marque le début de la sous-cohorte. L'autre terme, additionné à 0, représente l'âge moyen au segment vertical qui ferme la partie de la sous-cohorte dans le 1er triangle. L'âge moyen dans la sous-cohorte x ($\bar{y}_x$) est donc la moyenne entre les âges initial et final de la partie de la sous-cohorte dans le 1er triangle. Comme l'âge initial vaut 0, cet âge moyen est simplement la moitié de l'âge final :

$$\bar{y}_x = \frac{0+\overline{ys}_x}{2} = \frac{\left(\frac{NPT\_1T(x)}{2n}\right)}{2} = \left(\frac{NPT\_1T(x)}{2*2n}\right)$$

De la sorte, on obtient 0,45 si « x » = 1 ; 0,35 si « x » = 2 ; etc. (*cf.* tableau 1, colonne 4).

L'âge moyen au décès dans le 1er triangle ($\bar{y}$) se calcule alors en divisant la somme des NPT_1T(x) multipliés par l'âge moyen au décès $\bar{y}_x$ par la somme des NPT_1T(x), ce qui donne 0,33 (8,25/25 ; *cf.* colonnes 2 et 5 du tableau 1) : $\bar{y} = \frac{\sum_{x=1}^{5}\left[(NPT\_1T(x))*(\bar{y}_x)\right]}{\sum_{x=1}^{5} NPT\_1T(x)}$.

L'application de la même méthode en cas de 1, 2, 100 et 1.000 sous-cohortes donne respectivement un âge moyen au décès dans le 1er triangle de 0,25 ; 0,3125 ; 0,33332500 et 0,33333325. Ce résultat converge vers 1/3 d'année avec l'augmentation du nombre de sous-cohortes.

Si le nombre de sous-cohortes tend vers l'infini, l'âge moyen au décès devrait tendre vers 1/3. Comment démontrer cette proposition ? La 1re solution proposée se fera via un raisonnement discret, tout en faisant le pont avec l'analyse infinitésimale (point 1.2.b). La 2e solution utilisera de façon résolue l'apport de ce dernier type d'analyse pour proposer une solution bien plus rapide (point 1.3).

2.2.2. Calcul de l'âge moyen au décès si le nombre de sous-cohortes tend vers l'infini

Le calcul de l'âge moyen au décès dans le 1er triangle peut s'exprimer comme suit :

$$\bar{y} = \frac{\sum_{x=1}^{n}\left[(NPT\_1T(x))*(\bar{y}_x)\right]}{\sum_{x=1}^{n} NPT\_1T(x)} = \frac{\sum_{x=1}^{n}\left[[(2n+1)-(2x)]*\frac{[(2n+1)-(2x)]}{4*n}\right]}{\sum_{x=1}^{n}[(2n+1)-(2x)]} = \frac{\sum_{x=1}^{n}[(2n+1)-(2x)]^2}{(4n)*\sum_{x=1}^{n}[(2n+1)-(2x)]}.$$

On peut démontrer que cette expression tend vers 1/3 quand n tend vers l'infini (*cf.* annexe 1).

Principale objection de cette démonstration : sa lourdeur. Or, en passant à l'analyse infinitésimale, la solution devient plus facile à établir et surtout à généraliser (notamment en cas de répartition non uniforme), ce qui serait éventuellement possible, mais vraiment fastidieux avec les sommes. Par ailleurs, cette solution infinitésimale n'est que la suite logique du raisonnement avec les sommes.

2.3. Calcul de l'âge moyen au décès par une voie intégrale en cas de répartition uniforme

Après la méthode des sommes, pourquoi passer par ce calcul intégral[6] que d'aucuns jugeront sans doute moins accessible, du moins à priori ? Outre son élégance mathématique, cette procédure permettra de faire sauter à peu de frais le verrou de l'hypothèse de répartition uniforme des décès, hypothèse dont il est très utile de s'affranchir afin d'arriver à une expression la plus générale possible. Avant de traiter spécifiquement la question centrale de l'âge moyen au décès dans le 1^er^ triangle, la surface du triangle sera calculée par voie intégrale, non pas tant pour établir le résultat qui est connu (1/2) que pour se familiariser avec la méthode et surtout faire émerger dans un contexte simple des expressions mathématiques où il suffira d'introduire tel ou tel élément pour calculer aisément des grandeurs intéressantes pour la suite du propos.

2.3.1. En guise de préambule : calcul de la surface d'un triangle par voie intégrale

Supposons que l'axe x supporte le temps et l'axe y, l'âge. Le premier triangle (*cf.* figure 3.a) sera délimité par les points de coordonnées (0,0), (1,1) et (1,0). Sa superficie peut être calculée par voie intégrale de 4 façons différentes.

2.3.1.1. Calcul avec une perspective initiale verticale

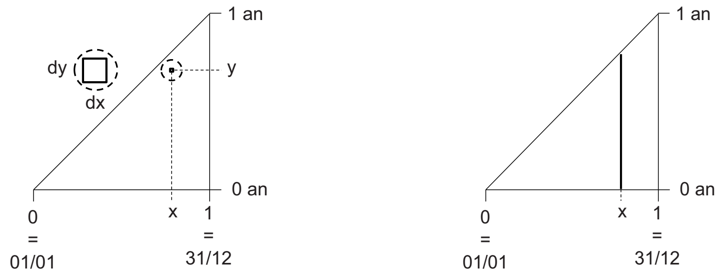
Par voie intégrale, sa surface (S_{T1}) peut se calculer comme suit[7] : $S_{T1} = \int_{x=0}^{x=1}\int_{y=0}^{y=x} dy*dx$,

où « dx*dy » définit la surface élémentaire de hauteur et de largeur infinitésimales (*cf.* figure 3.a, où cette surface élémentaire a été agrandie à gauche du triangle).

L'intégration intérieure se fait sur y. Autrement dit, le calcul estime d'abord la surface d'une bandelette verticale de largeur dx et de hauteur égale à x et constituée par un empilement de surfaces élémentaires carrées dx*dy (*cf.* figure 3.b).

Figure 3. Calcul de la surface du 1er triangle – 4 diagrammes de Lexis

a. le carré élémentaire b. Intégration initiale sur y

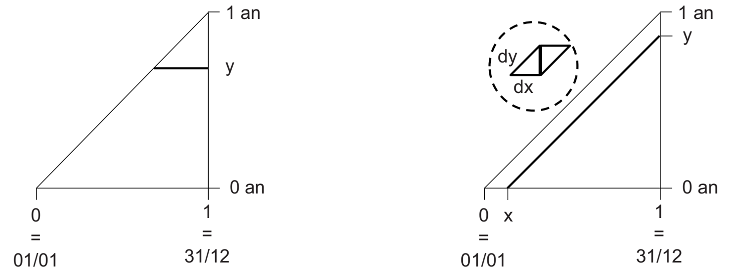


[6] Notons bien que le calcul intégral est souvent utilisé en analyse démographique. Ainsi, à titre d'exemple : Preston *et al.* (2005, p. 69) utilisent le calcul intégral pour quantifier « average person-years lived in the never-married state between ages x and x+n by women marrying in the interval » ou Leridon et Toulemon (1997, p. 49) l'utilisent pour calculer l'âge moyen à la maternité. Ces exemples ont été choisis par la proximité des grandeurs à estimer par rapport au sujet de ce texte.

[7] Si nécessaire, les intégrales ont été vérifiées/exécutées sur le site : http://www.wolframalpha.com/widgets/gallery/view.jsp?id=f5f3cbf14f4f5d6d2085bf2d0fb76e8a

Figure 3. Calcul de la surface du 1^er^ triangle – 4 diagrammes de Lexis (suite)

c. Intégration initiale sur x d. Intégration initiale oblique



Comment justifier les bornes de l'intégration (y = 0 et y = x) ? La bandelette verticale au-dessus de la coordonnée x commence à la valeur 0 de y (ce qui justifie la borne inférieure) et s'arrête à l'hypoténuse du 1^er^ triangle qui a pour équation y = x (ce qui justifie la borne supérieure de l'intégration intérieure[8]). L'expression x*dx (qui résulte du calcul de l'intégrale intérieure) donne la surface de la bandelette verticale. Ce premier résultat sera ensuite intégré sur x entre les valeurs 0 et 1 pour obtenir la surface du triangle (ce qui revient à balayer toute la surface du 1^er^ triangle à l'aide des verticales issues de toutes les valeurs de x) :

$$ST_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} dy*dx = \int_{x=0}^{x=1}\Big[y\Big]_0^x dx = \int_{x=0}^{x=1} x*dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Comme attendu, le résultat est bien de ½. Dans ce premier calcul, l'intégration intérieure visait le calcul d'une bandelette verticale. Il est aussi possible de faire porter cette intégration intérieure sur une bandelette horizontale ou oblique, ce qui supposera d'adapter les bornes des intégrales.

2.3.1.2. Calcul avec une perspective initiale horizontale

Dans le cas d'une intégration initiale portant sur l'horizontale à hauteur de y (*cf.* figure 3.c), l'intégrale du calcul de la surface du triangle s'écrit :

$$ST_1 = \int_{y=0}^{y=1}\int_{x=y}^{x=1} dx*dy = \int_{y=0}^{y=1}\Big[x\Big]_y^1 dy = \int_{y=0}^{y=1} 1-y*dy = \left[y-\frac{y^2}{2}\right]_0^1 = \frac{1}{2}.$$

D'épaisseur dy, elle commence à gauche à un point appartenant à l'oblique d'équation y = x (ce qui justifie la borne inférieure de l'intégrale) et s'arrête à x = 1 (ce qui justifie le choix de la borne supérieure). L'expression (1-y)*dy donne la surface de la bandelette horizontale (qui correspond à un alignement horizontal de surfaces élémentaires carrées dx*dy). L'intégration sur y entre 0 et 1 de cette expression donne la surface du triangle (ce qui revient à balayer toute la surface du 1^er^ triangle avec les horizontales issues de toutes les valeurs de y). De nouveau, comme attendu, le résultat est bien de ½.

[8] Les esprits chagrins (oui, oui, il y en a, même dans mon couloir) pourraient faire état d'un triangle non repris dans cette surface (le dernier, juste avant d'atteindre l'oblique), mais c'est cela tout l'art du calcul intégral… *Mutatis mutandis*, le même commentaire pourrait être reproduit pour les 3 calculs qui suivent. Nous ne le ferons pas.

2.3.1.3. Calcul avec une perspective initiale oblique, version 1

Dans le cas d'une intégration initiale sur une bandelette oblique (*cf.* figure 3.d), l'intégrale du calcul de la surface du triangle peut s'écrire comme suit si l'intégration intérieure se fait sur y :

$$ST_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=1-x} dy*dx = \int_{x=0}^{x=1}\Big[y\Big]_0^{1-x} dx = \int_{x=0}^{x=1}(1-x)*dx = \left[x-\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Attention, dans le calcul qui vient d'être fait, la surface élémentaire est un parallélogramme de hauteur dy et de base dx (*cf.* figure 3.d, où ce parallélogramme élémentaire a été agrandi à gauche du triangle). La bandelette oblique est constituée par l'enchainement de ces parallélogrammes élémentaires pour y variant de 0 à 1-x (par construction, la coordonnée y du point le plus élevé de l'oblique vaut 1-x), enchainement qui forme un parallélogramme de hauteur égale à 1-x et de base égale à dx. Et donc, (1-x)*dx correspond à la surface de ce dernier parallélogramme. L'intégration pour x variant de 0 à 1 de cette dernière expression donne la surface du triangle (ce qui revient à balayer tout le 1er triangle avec les obliques commençant aux différentes valeurs de x). Comme attendu, le résultat est bien de ½.

Des petites surprises du calcul intégral !

Dans le calcul en oblique qui vient d'être proposé, la tentation pourrait être forte de considérer que la surface de la bandelette oblique correspond à sa longueur multipliée par dx. Étant donné que, par construction, le triangle (x,0) ; (0,y) et (1,0) de la figure 4.d est un triangle rectangle isocèle, cette longueur (L) de l'hypoténuse vaut : $L^2=(1-x)^2+(1-x)^2 \;\Rightarrow\; L=\sqrt{2}*(1-x)$.
Dès lors, la surface du triangle, obtenue par intégration de cette expression sur x entre 0 et 1, donnerait une superficie fausse : $ST_1 = \int_{x=0}^{x=1}\sqrt{2}*(1-x)*dx = \sqrt{2}*\left[x-\frac{x^2}{2}\right]_0^1 = \sqrt{2}*\frac{1}{2} = 0{,}7071.$
Ce résultat est en contradiction avec la superficie qui est indiscutablement de ½ (triangle isocèle rectangle dont les côtés adjacents à l'angle droit valent 1). Il faut donc proscrire ce type de calcul, même si, par effet de compensation, il pourrait confirmer nos résultats, notamment en matière d'âge moyen au décès dans le 1er triangle[9]. La bonne équation est bien celle qui a été proposée un peu plus haut dans ce texte.

2.3.1.4. Calcul avec une perspective initiale oblique, version 2

Enfin, toujours dans la perspective oblique, rien n'empêche de partir du principe que l'intégrale intérieure serait faite sur x, et plus sur y. Dans ce cas, le calcul de la surface serait donné par :

$$ST_1 = \int_{y=0}^{y=1}\int_{x=1-y}^{x=1} dx*dy = \int_{y=0}^{y=1}\Big[x\Big]_{1-y}^{1} dy = \int_{y=0}^{y=1} 1-1+y*dy = \int_{y=0}^{y=1} y*dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2}.$$

Par construction, le point initial à gauche de l'oblique a comme coordonnée x = 1 – y et, comme point final, x = 1, ce qui justifie les bornes de l'intégrale intérieure. La surface de la bandelette oblique correspond à celle d'un parallélogramme de hauteur y et de base dy. L'intégration de l'expression y*dy entre les valeurs 0 et 1 de y (ce qui revient à balayer tout le 1er triangle avec les obliques commençant aux différentes valeurs de y) donne le résultat attendu. Rappelons l'erreur de céder à la tentation de faire intervenir la longueur de la bandelette oblique dans le calcul.

[9] *Cf.* Kucera (1986), en pages 90-91 où, à notre avis, c'est exactement ce qui se passe. Sans entrer dans les détails, notons que le calcul de Kucera portait sur le temps passé dans le triangle après l'évènement.

En définitive, 4 équations concurrentes permettent de calculer la surface d'un triangle par voie intégrale. La différence vient de la variable sur laquelle porte l'intégrale intérieure. Ces 4 équations débouchent sur le même résultat[10]. Par choix purement personnel, dans la suite du texte, nous avons privilégié la première équation proposée.

2.3.2. Calcul de l'âge moyen au décès dans le 1er triangle par voie intégrale

Revenons à la question centrale de ce point, à savoir : quel est l'âge moyen au décès des décédés du 1er triangle dans un diagramme de Lexis (figure 3), moyennant l'hypothèse de répartition uniforme et une densité supposée de 1.000 décès par unité de surface, soit 1.000 décès sur l'ensemble du parallélogramme entre 0 an exact et 1 an exact, ce parallélogramme ayant une superficie unitaire ?

Dans un 1er temps, il faut calculer le nombre de décès. Le nombre de décès dans un carré élémentaire (NCE) s'obtient en multipliant :

- la surface dx*dy ;
- par la densité des décès à cet âge y, soit $Dens_y$, qui pour rappel est constante sur tout le triangle vu l'hypothèse de répartition uniforme.

$$NCE = Dens_y * dx * dy = 1.000 * dx * dy.$$

Le nombre de décès dans le triangle, soit NDT_1 (résultat certes évident, mais qu'il est intéressant d'établir par voie intégrale pour la suite des opérations) s'obtient en intégrant cette expression sur x et y. Nous avons opté pour une intégration initiale sur y :

$$NDT_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} 1.000*dy*dx = 1.000\int_{x=0}^{x=1}\Big[y\Big]_0^{x} dx = 1.000\int_{x=0}^{x=1} x*dx = 1.000*\left[\frac{x^2}{2}\right]_0^1 = 1.000*\frac{1}{2} = 500.$$

Comme attendu, le résultat est 500, soit la moitié de la densité vu que cette densité est appliquée sur une surface valant ½.

Dans un 2^{e} temps, il faut calculer l'âge cumulé des décédés du 1er triangle au moment de leur décès. Cet âge cumulé dans le carré élémentaire (ACCE) s'obtient en multipliant :

- la surface dx*dy ;
- par la densité des décès à cet âge y, soit $Dens_y$, ou 1.000 vu les conditions du calcul ;
- par y, qui est, par définition, l'âge au décès des décédés enregistré à l'âge y !

$$ACCE = Dens_y * y * dx * dy = 1.000 * y * dx * dy.$$

L'âge cumulé des décédés dans le triangle, soit $ACDT_1$ s'obtient en intégrant cette expression sur x et y :

$$ACDT_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} 1.000*y*dy*dx = 1.000\int_{x=0}^{x=1}\left[\frac{y^2}{2}\right]_0^{x} dx = 1.000\int_{x=0}^{x=1}\frac{x^2}{2}*dx = 1.000*\left[\frac{x^3}{6}\right]_0^1 = 1.000*\frac{1}{6}.$$

Finalement, l'âge moyen au décès dans le 1er triangle ($\bar{y}$) se calcule comme suit :

[10] À titre d'exemple, Leridon et Toulemon (1997, p. 65) soulignent également que l'ordre de variables sur lesquelles l'intégration s'opère (d'abord sur x, puis sur y ou l'inverse) n'influence pas les résultats.

$$\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{1.000 * \frac{1}{6}}{1.000 * \frac{1}{2}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}\text{, cqfd.}$$

Comme attendu, l'âge moyen des décédés du 1er triangle vaut donc bien 1/3 en cas de répartition uniforme des décès. Par ailleurs, quelle sobriété dans la démonstration, surtout par comparaison avec le fastidieux calcul faisant appel à des sommes ! On remarquera que, étant donné que la densité intervient haut et bas dans la fraction, sa valeur est sans influence sur le résultat.

En définitive, pour toute densité constante k, on aura : $\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{k * \frac{1}{6}}{k * \frac{1}{2}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$.

Vu ces circonstances, le plus simple est d'opter pour une densité unitaire, malgré une certaine forme d'inconfort : comment imaginer une répartition uniforme si un seul décès sur l'ensemble de la surface ? Pourtant, d'un point de vue purement mathématique, ce choix ne pose par ailleurs aucun problème. Avec une densité unitaire, les calculs deviennent :

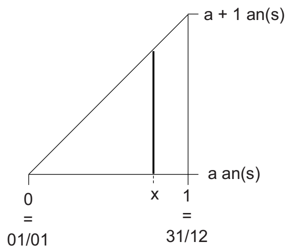
- $NDT_1 = \int_{x=0}^{x=1} \int_{y=0}^{y=x} 1 * dy * dx = \int_{x=0}^{x=1} \int_{y=0}^{y=x} dy * dx = \frac{1}{2}.$
- $ACDT_1 = \int_{x=0}^{x=1} \int_{y=0}^{y=x} 1 * y * dy * dx = \int_{x=0}^{x=1} \int_{y=0}^{y=x} y * dy * dx = \frac{1}{6}.$
- $\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$

La conclusion à propos de l'âge moyen de 1/3 s'accommode de n'importe quelle densité absolue pour autant qu'elle soit *constante* sur la surface. Les nombres de décès et l'âge cumulé des décédés seront modifiés, mais avec des facteurs de multiplication identiques, ce qui explique que l'âge moyen, but du calcul, n'est point modifié selon la valeur de la densité.

Généralisation du calcul à tous les âges

Jusqu'ici, les calculs ont porté sur le 1er triangle, délimité par les âges 0 exact et 0 révolu. Qu'en est-il pour les autres triangles reposant sur base (ou « triangles inférieurs »), mais situés à un autre âge, comme par exemple le triangle inférieur quelconque entre les âges a exact et a révolu (*cf.* figure 4) ?

Figure 4. Calcul à un âge quelconque pour un triangle inférieur



En les appliquant à ce triangle délimité par les âges a exact et a révolu, les calculs deviennent :

- $NDT_1 = \int_{x=0}^{x=1} \int_{y=a}^{y=a+x} dy * dx = \int_{x=0}^{x=1} \Big[y \Big]_a^{a+x} dx = \int_{x=0}^{x=1} \big((a+x)-a\big)dx = \int_{x=0}^{x=1} x * dx = \frac{1}{2}.$

Justification des bornes de l'intégrale :

- pour y : la bandelette verticale au-dessus de x commence à l'âge a et se termine en a + x, vu que son point sommital appartient à la droite d'équation y = a + x.
- pour x : par l'intégrale extérieure, on va balayer toute la surface du triangle à l'aide des verticales au-dessus de toutes les valeurs possibles de x, soit de 0 à 1. En effet, x varie entre 0 et 1, soit le début et la fin de l'année sous observation. Notons que rien n'empêche de prendre par exemple 2005 et 2006 (ou m et m + 1, avec m étant le millésime de l'année concernée) comme bornes : cela ne changera rien aux résultats, mais les rendra simplement moins immédiats…
- $ACDT_1 = \int_{x=0}^{x=1} \int_{y=a}^{y=a+x} y * dy * dx = \int_{x=0}^{x=1} \left[\frac{y^2}{2} \right]_a^{a+x} dx = \frac{1}{2}\left(\frac{1}{3} + a\right).$
- $\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{\frac{1}{2}\left(\frac{1}{3} + a\right)}{\frac{1}{2}} = \frac{1}{3} + a.$

Finalement, l'âge moyen au décès dans un triangle orienté semblablement au premier triangle vaut l'âge exact délimitant le triangle vers le bas augmenté d'un tiers. Appliquée au 1er triangle (entre 0 exact et 0 révolu), cette constatation donne le tiers bien établi par ailleurs. On l'aura compris, le calcul appliqué à un autre triangle que le premier est « compliqué » par l'apparition de l'âge exact initial a en remplacement du 0.

Autant nous trouvons la 1re méthode (la somme avec n tendant vers l'infini) utile pour bien saisir la nature du problème posé d'une façon intuitive ou artisanale, autant le recours à l'analyse infinitésimale ouvre la voie à des solutions très pratiques et complètes, permettant de s'affranchir à peu de frais de l'hypothèse de la répartition uniforme des décès au profit de répartitions non uniformes, objet du point suivant.

2.4. En cas de répartition non uniforme

2.4.1. Répartition non uniforme, mais linéaire

Afin de bien mettre en exergue l'intérêt du calcul infinitésimal, il nous parait très utile de reprendre les questions qui viennent d'être abordées sous hypothèse de répartition non uniforme des décès. Supposons :

- une densité double à 0 an exact par rapport à 1 an exact et une évolution linéaire entre 0 et 1 an exact ;
- une densité de 1.000 à 1 an exact, considérée comme densité de référence.

Dans ces condition, la densité ($Dens_y$) évolue en fonction de l'âge y comme suit :

$$Dens_y = (2 - y) * 1.000.$$

Pour y = 0, la densité vaut 2.000 ; pour y = 1, la densité vaut 1.000. Dans ces conditions, le nombre de décès dans le 1er triangle (NDT_1) se calculera comme suit :

$$NDT_1 = \int_{x=0}^{x=1} \int_{y=0}^{y=x} Dens_y * dy * dx = \int_{x=0}^{x=1} \int_{y=0}^{y=x} 1.000 * (2-y) * dy * dx = 1.000 * \int_{x=0}^{x=1} \int_{y=0}^{y=x} (2-y) * dy * dx$$

$$= 1.000 * \int_{x=0}^{x=1} \left[2y - \frac{y^2}{2} \right]_0^x * dx = 1.000 * \int_{x=0}^{x=1} 2x - \frac{x^2}{2} * dx = 1.000 * \left[2\frac{x^2}{2} - \frac{x^3}{6} \right]_0^1 = 1.000 * \left(1 - \frac{1}{6}\right) = 1.000 * \left(\frac{5}{6}\right).$$

Par définition, les décédés à hauteur de y sont âgés de… y an exact à leur décès (*cf.* figure 3.a). Dès lors, l'âge cumulé des décédés du 1er triangle ($ACDT_1$) peut se calculer en introduisant y dans l'équation :

$$ACDT_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} Dens_y * y * dy * dx = 1.000 \int_{x=0}^{x=1}\int_{y=0}^{y=x} (2-y) * y * dy * dx = 1.000 * \left(\frac{1}{4}\right).$$

Finalement, l'âge moyen dans le 1er triangle ($\bar{y}$) se calcule comme suit :

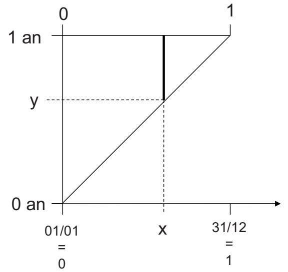
$$\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{1.000 * \left(1/4\right)}{1.000 * \left(5/6\right)} = \frac{\left(1/4\right)}{\left(5/6\right)} = 0{,}3 .$$

La valeur de l'âge moyen est indépendante de la valeur de la densité de référence, qui vaut ici 1.000. On peut donc se passer de cette donnée. La seule chose qui importe : que l'évolution relative de la densité en fonction de l'âge exact entre 0 et 1 an soit connue. Pour la suite et par facilité, nous fixerons la densité de référence à 1.

Généralisation du calcul pour les triangles orientés autrement que le premier

Jusqu'ici, tous les calculs ont été effectués pour des triangles reposant sur leur base (triangle inférieure). Qu'en est-il pour les triangles reposant sur un somment (triangle dit « supérieur ») ? Nous commencerons par le deuxième triangle de la figure 1.a, délimité par les âges 0 an révolu et 1 an exact pour ensuite généraliser les écritures à tous les âges (figure 5).

Figure 5. Calcul pour le triangle supérieur entre 0 an révolu et un an exact



Pour le triangle délimité par 0 révolu et 1 an exact et en supposant une densité unitaire et uniforme des décès, les calculs deviennent :

- $NDT_1 = \int_{x=0}^{x=1}\int_{y=x}^{y=1} dy * dx = \int_{x=0}^{x=1} \left[y\right]_x^1 dx = \int_{x=0}^{x=1} (1-x) * dx = \left[x - \frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$

Justification des bornes de l'intégrale :

- pour y : la bandelette verticale au-dessus de x commence à la valeur de y = x (le début de la verticale appartient à la droite d'équation y = x et se termine à 1, vu que son point sommital appartient à la droite d'équation y = 1.
- pour x : par l'intégrale extérieure, on va balayer toute la surface du triangle à l'aide des verticales au-dessus de toutes les valeurs possibles de x, soit de 0 à 1.
- $ACDT_1 = \int_{x=0}^{x=1}\int_{y=x}^{y=1} y * dy * dx = \int_{x=0}^{x=1} \left[\frac{y^2}{2}\right]_x^1 * dx = \int_{x=0}^{x=1} \left(\frac{1}{2} - \frac{x^2}{2}\right) * dx = \frac{1}{3}$
- $\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{1/3}{1/2} = \frac{2}{3}.$

Pour le triangle quelconque délimité par a révolu et a + 1 exact(s) et en supposant une densité unitaire et uniforme des décès, les calculs deviennent :

- $NDT_1 = \int_{x=0}^{x=1}\int_{y=a+x}^{y=a+1} dy * dx = \int_{x=0}^{x=1} \left[y\right]_{a+x}^{a+1} dx = \int_{x=0}^{x=1} (a+1-a-x) * dx = \int_{x=0}^{x=1} (1-x) * dx = \frac{1}{2}.$

Justification des bornes de l'intégrale :

- pour ce calcul, l'âge 0 sur la figure 5 est à remplacer par « a » et l'âge 1, par « a + 1 » ;
- pour y : vu ce qui vient d'être précisé, la bandelette verticale au-dessus de x commence à l'âge x + a et se termine à a+1 ;
- pour x : pas de changement.
- $ACDT_1 = \int_{x=0}^{x=1}\int_{y=a+x}^{y=a+1} y * dy * dx = \int_{x=0}^{x=1} \left[\frac{y^2}{2}\right]_{a+x}^{a+1} * dx = \frac{1}{2} * \left(a + \frac{2}{3}\right).$
- $\bar{y} = \frac{ACDT_1}{NDT_1} = \frac{\frac{1}{2} * \left(a + \frac{2}{3}\right)}{1/2} = a + \frac{2}{3}.$

Finalement, l'âge moyen au décès dans un triangle orienté semblablement au deuxième triangle vaut l'âge révolu délimitant le triangle à gauche augmenté de deux tiers. Appliquée au 1er triangle (entre 0 exact et 0 révolu), cette constatation donne deux tiers, soit le résultat déjà connu.

Ce résultat obtenu (0,3) reste, somme toute, fort proche de 1/3. *Supposons maintenant une densité 10 fois plus forte à 0 an qu'à un an exact avec évolution linéaire de la densité entre 0 et 1 an exact* (par exemple de 10 à 0 an exact contre 1 à 1 an exact). La densité varie donc comme suit :

$$Dens_y = 10 - 9y .$$

Pour y = 0 (alternativement 0,5 et 1), la densité vaut 10 (alternativement 5,5 et 1). Dans ces conditions, l'âge moyen dans le 1er triangle se calcule comme suit :

$$\bar{y} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x} Dens_y * y * dy * dx}{\int_{x=0}^{x=1}\int_{y=0}^{y=x} Dens_y * dy * dx} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x} (10-9y) * y * dy * dx}{\int_{x=0}^{x=1}\int_{y=0}^{y=x} (10-9y) * dy * dx} = \frac{\frac{11}{12}}{\frac{7}{2}} = 0{,}26190.$$

Sans surprise, la différence par rapport à un tiers se creuse à mesure que la répartition des décès s'éloigne de l'uniformité : au plus les décès se concentrent aux premiers âges de la vie, au plus l'âge moyen au décès diminue, ce qui était hautement prévisible. Dans le point suivant, l'évolution de la densité ne suivra plus une évolution linéaire.

2.4.2. Répartition non uniforme et non linéaire

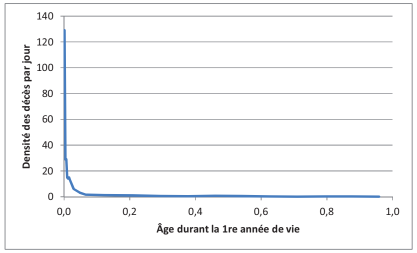
La figure 6 montre la répartition des décès durant la 1re année de vie par jour pour les naissances survenue en 2007 en Belgique[11]. L'âge moyen au décès dans le 1er triangle vaut :

- 0,07937 année, en tablant sur une répartition uniforme des décès à l'intérieur de chaque jour, ce qui revient à prendre le numéro d'ordre du jour et retirer un demi-jour pour obtenir l'âge moyen au décès. Ainsi, pour les décès du 1er jour, il est de 0,5 jour ;
- 0,07904 année, en tablant sur un âge moyen au décès de 0,1 jour pour les décès du 1er jour et en gardant la même hypothèse que précédemment pour les autres jours ;

[11] Les données nécessaires nous ont été fournies par la Direction générale de la Statistique (DGS), que nous tenons à remercier, et tout particulièrement Michel Willems de la DGS. Les données proviennent de l'exploitation par la DGS des bulletins d'état civil (Modèle IIID Bulletin statistique de décès d'un enfant de moins d'un an ou d'un mort-né), avec exploitation de l'heure de la naissance et du décès pour déterminer l'âge au décès de façon précise. Il est à noter que ces données étaient provisoires quand nous avons fait ce calcul. Depuis, les données ont été corrigées. Nous n'avons pas refait l'exercice au départ des données corrigées.

- 0,07896 année, en tablant sur un âge moyen au décès de 0,01 jour pour les décès du 1er jour et en gardant la même hypothèse que précédemment pour les autres jours.

Figure 6. Répartition des décès de la 1re année de vie par jour
Belgique naissances 2007



La densité par jour (donnée initiale) doit être transformée en une densité instantanée pour les besoins du calcul intégral :

- La densité initiale (au moment 0) a été fixée à 228 décès par jour. Il s'agit de la densité au début du 1er jour de vie. Dans les données observées, 129 est le nombre de décès du 1er jour, soit une densité instantanée moyenne sur le 1er jour. Comme les 2 jours suivants comptent chacun 29 décès, nous avons posé l'hypothèse que, en début de 2e jour (soit en fin du 1er jour), la densité instantanée était de 30 décès (hypothèse qui pourrait être modifiée à la hausse). En conséquence, la densité instantanée en début du 1er jour (DI1) est de :

$$129 = \frac{\mathrm{DDI} + 30}{2} \Rightarrow \mathrm{DDI} = (129 * 2) - 30 = 228.$$

- Pour la densité instantanée finale (au moment 1), nous avons opté pour 0,60 décès par jour. Cette densité est d'application sur la plus grande part de la 1re année et est un peu exagérée, car la densité instantanée modélisée passe un peu trop près du point de rencontre des axes (*cf.* figures 6 et 7)

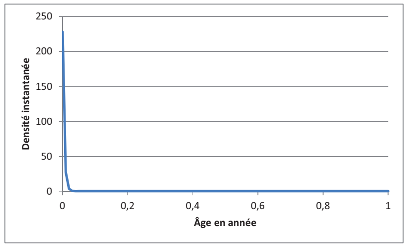
Pour modéliser cette évolution de la densité instantanée des décès, nous avons opté pour la fonction suivante dont l'évolution suit assez bien le schéma présenté en figure 6 :

$$\mathrm{Dens}_y = \left(\left(\frac{\left((2-2y)^{220}\right) * (10-10y)}{2^{217}}\right) * 2{,}8425\right) + 0{,}60.$$

La densité évolue comme indiqué dans le tableau 2 et sur la figure 7.

Tableau 2 et figure 7. Évolution de la densité instantanée en fonction de l'âge

y	Densité
0	228,0000000
0,1	0,6000001
0,2	0,6000000
0,3	0,6000000
0,4	0,6000000
0,5	0,6000000
0,6	0,6000000
0,7	0,6000000
0,8	0,6000000
0,9	0,6000000
1	0,6000000



Dans ces conditions, l'âge moyen au décès dans le 1er triangle se calcule comme suit :

$$\bar{y} = \frac{\mathrm{ACDT}_1}{\mathrm{NDT}_1} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x} \mathrm{Dens}_y * y * dy * dx}{\int_{x=0}^{x=1}\int_{y=0}^{y=x} \mathrm{Dens}_y * dy * dx} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x}\left(\left(\left(\frac{\left((2-2y)^{220}\right)*(10-10y)}{2^{217}}\right)*2{,}8425\right)+0{,}60\right)*y*dy*dx}{\int_{x=0}^{x=1}\int_{y=0}^{y=x}\left(\left(\left(\frac{\left((2-2y)^{220}\right)*(10-10y)}{2^{217}}\right)*2{,}8425\right)+0{,}60\right)*dy*dx}$$

$$= \frac{0{,}104552}{1{,}319730} = 0{,}079222.$$

Les valeurs du numérateur et du dénominateur peuvent surprendre, mais la densité utilisée est exprimée en nombre de décès par jour. Pour retomber sur le nombre annuel de décès, il faut multiplier le dénominateur par 365, ce qui donne 481,70145 décès alors que dans l'observé il y en avait 482. Le même calcul pour le numérateur donne 38,161480 (ces résultats ont été confirmés aux arrondis près en introduisant un facteur « *365 » dans la fonction à intégrer). Par ailleurs l'âge moyen obtenu (soit 0,079222 année) est très proche des valeurs calculées sur les données observées (de l'ordre de 0,079 année selon l'hypothèse retenue pour l'âge moyen au décès durant le 1er jour).

Pour suivre, d'autres facteurs pouvant influencer la densité des décès vont être intégrés au raisonnement. Il s'agit de la saisonnalité des décès et des naissances.

2.4.3. Répartition non uniforme des décès avec, en sus, saisonnalité des naissances et des décès

La densité des décès et, par ricochet, l'âge moyen au décès dans le 1er triangle, peuvent se modifier en fonction non seulement de la variation de mortalité par âge, mais aussi de la saisonnalité des naissances et de celle des décès[12]. Supposons que, *indépendamment des autres effets* :

- la densité des décès évolue selon la fonction : $\mathrm{DensD}_y = 2 - y$,

  soit une densité deux fois plus forte à 0 an qu'à 1 an ; autrement dit, *toutes choses égales par ailleurs,* il y a deux fois plus de décès à 0 an exact qu'à 1 an exact ;
- la saisonnalité des décès se matérialise par la fonction : $\mathrm{SaisD}_x = 1 + 2x$,

  soit une densité de décès 3 fois plus importante en fin d'année qu'au début ; autrement dit, *toutes choses égales par ailleurs*, les décès sont 3 fois plus nombreux au temps 1 qu'au temps 0.
- la saisonnalité des naissances se matérialise par la fonction : $\mathrm{SaisN}_z = 3 - z$, avec $z = x - y$,

  la date de naissance des individus âgés de y à la date x dans le 1er triangle valant x-y, soit z. En définitive, la densité des naissances est de 1,5 fois plus importante au temps 0 (le début de l'année) qu'au temps 1 ; autrement dit, *toutes choses égales par ailleurs*, les naissances sont 50 % plus nombreuses au temps 0 qu'au temps 1. En conséquence, toutes choses égales par ailleurs, les décès seront 50 % plus nombreux parmi la sous-cohorte née au temps 0 par rapport à celle née au temps 1 ! Insistons : les densités de naissances dont question sont relatives. Il faut multiplier cette densité relative par un facteur k pour obtenir la densité absolue des naissances. Exemple : si k vaut 1.000, la densité absolue des naissances sera de 3.000 au temps 0 et de 2.000 au temps 1. Notons que ce facteur k finira par s'éliminer dans les calculs, comme montré dans le point 1.3. Pour la suite de ce point, il s'avère inutile de s'encombrer de l'évolution de la densité absolue des naissances.

[12] Pour une présentation élaborée des calculs incluant les saisonnalités des décès et des naissances, *cf.* Calot et Franco (2002).

Il est bien évident que les saisonnalités employées ne sont guère réalistes. Par exemple, elles ne sont pas cycliques selon les années, la fin d'année marquant une rupture. Vu la nature théorique de l'exercice, on peut toutefois s'en contenter d'autant que l'adoption de fonctions plus réalistes (et notamment cycliques) rend rapidement le calcul ingérable (*cf.* annexe 2).

Calcul de l'âge moyen au décès si les trois effets décrits plus haut sont actifs :

$$NDT_1 = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x} DensD_y * SaisN_z * SaisD_x * dy\right] * dx = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}(2-y)*(3-z)*(1+2x)*dy\right]*dx$$

$$= \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}(2-y)*(3-x+y)*(1+2x)*dy\right]*dx = \frac{201}{40} = 5{,}02500.$$

$$ACDT_1 = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x} DensD_y * SaisN_z * SaisD_x * y * dy\right] * dx = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}(2-y)*(3-z)*(1+2x)*y*dy\right]*dx$$

$$= \int_{x=0}^{x=1}\left(3x^2 + \frac{14}{3}x^3 - \frac{31}{12}x^4 + \frac{1}{6}x^5\right)*dx = \frac{151}{90} = 1{,}67778.$$

$$\bar{y} = \frac{ACDT}{DT} = \frac{1{,}67778}{5{,}02500} = 0{,}33389.$$

Les résultats sont arrondis à la 5^e décimale, ce qui peut entrainer des approximations lors de vérifications. C'est la raison pour laquelle certains résultats ont aussi été exprimés sous forme de fractions, ce qui permet, le cas échéant, de les calculer plus précisément qu'à la 5^e décimale.

Le tableau 3 donne les résultats en nombre de décès, âge cumulé et âge moyen dans le 1er triangle pour différents scénarios, se différenciant selon que les 3 effets sont actifs ou pas (scénarios 1 à 8) ou alors pour des scénarios reprenant des saisonnalités différentes de celles proposées initialement.

Tableau 3. Récapitulatif des calculs dans le cas du 1er triangle

N°	$DensD_y$	$SaisN_z$	$SaisD_x$	Décès	Âge cumulé	Âge moyen
1	$2-y$	$3-z=3-x+y$	$1+2x$	5,02500	1,67778	**0,33389**
2	$k=1{,}5$	$k=2{,}5$	$k=2$	3,75000	1,25000	**0,33333**
3	$2-y$	$k=2{,}5$	$k=2$	4,16667	1,25000	**0,30000**
4	$k=1{,}5$	$3-z=3-x+y$	$k=2$	4,00000	1,37500	**0,34375**
5	$k=1{,}5$	$k=2{,}5$	$1+2x$	4,37500	1,56250	**0,35714**
6	$2-y$	$k=2{,}5$	$1+2x$	4,79167	1,54167	**0,32174**
7	$2-y$	$3-z=3-x+y$	$k=2$	4,41667	1,36667	**0,30943**
8	$k=1{,}5$	$3-z=3-x+y$	$1+2x$	4,62500	1,7125	**0,37027**
9	$k=1{,}5$	$k=2{,}5$	$3-2x$	3,12500	0,93750	**0,30000**
10	$k=1{,}5$	$2+z=2+x-y$	$k=2$	3,50000	1,12500	**0,32143**

Comparaisons :

- Cas 1 et 2 : dans le cas 1, les 3 effets sont actifs alors que dans le cas 2, aucun n'est actif. Le cas 2 montre l'âge moyen en cas de répartition uniforme des décès, avec le résultat attendu d'un tiers. La différence entre les 2 âges moyens résulte de l'action des 3 effets.
- Cas 2 et 3 : par rapport au cas 2, dans le cas 3, l'effet d'âge est actif avec une plus forte concentration des décès aux premiers âges. En conséquence, l'âge moyen **diminue**. Cette diminution se retrouve chaque fois que la différence se marque uniquement au niveau de la présence ou pas de la variation par âge, comme le montrent les comparaisons : cas 4 et 7 ou cas 5 et 6. En outre, il va de soi que, si la concentration des décès devait être plus affirmée vers les âges élevés (vers 1 an, donc), l'âge moyen augmenterait !
- Cas 2 et 4 : par rapport au cas 2, dans le cas 4, la saisonnalité des naissances est active ce qui entraine plus de naissances en début d'année ; l'âge moyen **augmente** car les naissances, étant plus concentrées en début d'année, le nombre de décès survenus près de 1 an se renforce relativement. Cette augmentation se retrouve chaque fois que la différence se marque uniquement en ce qui concerne la présence ou pas de saisonnalité des naissances, comme le montrent les comparaisons : cas 3 et 7 ou cas 5 et 8. Sans surprise, en cas de concentration des naissances plus forte en fin d'année, c'est une diminution de l'âge moyen au décès qui prévaut comme le montre la comparaison des cas 4 et 10.
- Cas 2 et 5 : par rapport au cas 2, dans le cas 5, les décès sont plus nombreux en fin d'année ; l'âge moyen **augmente** car la saison de forte mortalité étant en fin d'année, elle agit à tous les âges alors que si cette période est en début d'année, elle n'agit que sur les seules générations déjà nées à ce moment, ce qui a pour effet de diminuer l'âge moyen, comme le montre la comparaison des cas 2 et 9. L'augmentation consécutive à une plus forte mortalité en fin d'année se constate aussi dans les comparaisons suivantes : cas 3 et 6 ou cas 4 et 8.
- Etc.

La méthode utilisée ici pourrait se généraliser en cas de variation par âge et de saisonnalité non linéaires. Aucune tentative dans cette direction n'est proposée ici pour des raisons pratiques déjà évoquées (*cf.* annexe 2).

La densité des décès peut être utilement remplacée dans les calculs par la densité de probabilité des décès. C'est ce thème qui va être abordé dans le point suivant.

2.5. Âge moyen au décès dans le 1er triangle via les densités de probabilité

Dans le contexte précis du calcul de l'âge moyen au décès dans le 1er triangle, la densité de probabilité donne la probabilité d'observer un des décès comptabilisés dans l'ensemble du triangle en chaque point de ce dernier. Il s'agit bien d'une probabilité liée à une surface infinitésimale. De manière qui peut sembler étonnante, la densité de probabilité en un point peut dépasser l'unité, comme démontré bientôt.

La densité de probabilité montre la répartition des décès du triangle sur l'ensemble de sa surface. Dès lors, par définition, l'intégration de la fonction de densité de probabilité sur l'ensemble de son domaine d'application doit donner 1. Appliquée au cas du 1er triangle, cette propriété se traduit comme suit : la fonction de densité de probabilité intégrée sur l'ensemble du triangle doit donner 1. Que donne cette propriété en cas de répartition uniforme des décès dans le 1er triangle ? Dans ces conditions, la densité de probabilité est elle-même constante. Désignons par « k_{dp} » cette densité de probabilité constante. Dès lors l'équation de l'intégration de k_{dp} sur l'ensemble du 1er triangle doit donner 1 :

$$\int_{x=0}^{x=1}\int_{y=0}^{y=x} k_{dp} * dy * dx = 1 \quad \Rightarrow k_{dp}\int_{x=0}^{x=1}\int_{y=0}^{y=x} dy * dx = 1 \quad \Rightarrow k_{dp} * \frac{1}{2} = 1 \quad \Rightarrow k_{dp} = 2.$$

Une densité de probabilité de 2 appliquée sur toute la surface du 1er triangle (soit sur une surface de ½) donne 1. Nous avons bien ici un cas où la densité de probabilité dépasse l'unité. Wilmoth *et al.* (2007, pp. 72-73) utilisent les densités de probabilité pour faire des calculs qui s'apparentent aux calculs d'âge moyen au décès dans le 1er triangle effectués dans les points précédents. Voici l'équation (E1) proposée en page 73, adaptée pour le 1er triangle en reprenant notre notation :

$$\bar{y} = \int_{x=0}^{x=1}\int_{y=0}^{y=x} 2*y*dy*dx = 2\int_{x=0}^{x=1}\int_{y=0}^{y=x} y*dy*dx = 2\int_{x=0}^{x=1}\left[\frac{y^2}{2}\right]_0^x *dx = 2\int_{x=0}^{x=1}\frac{x^2}{2}*dx = 2*\left[\frac{x^3}{6}\right]_0^1 = \frac{1}{3}.$$

On peut retrouver ce résultat et cette écriture par une autre voie. Si la densité des décès (à bien distinguer de la densité de probabilité) vaut k_d et est constante sur le 1er triangle, le nombre de décès et l'âge cumulé des décédés sont donnés par :

$$NDT_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} k_d*dy*dx = \frac{k_d}{2} \qquad ACDT_1 = \int_{x=0}^{x=1}\int_{y=0}^{y=x} k_d*y*dy*dx.$$

L'âge moyen se calcule en divisant l'âge cumulé par le nombre de décédés qui vaut la densité divisée par 2 :

$$\bar{y} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x} k_d*y*dy*dx}{\int_{x=0}^{x=1}\int_{y=0}^{y=x} k_d*dy*dx} = \frac{\int_{x=0}^{x=1}\int_{y=0}^{y=x} k_d*y*dy*dx}{\frac{k_d}{2}} = \int_{x=0}^{x=1}\int_{y=0}^{y=x}\frac{k_d}{k_d/2}*y*dy*dx = \int_{x=0}^{x=1}\int_{y=0}^{y=x} 2*y*dy*dx.$$

On retombe bien ainsi sur la formule utilisée par Wilmoth *et al.* (2007). La densité de probabilité est équivalente à la densité des décès divisée par le nombre total de décès du triangle. Pour passer de la densité de probabilité à la probabilité d'observer un des décès du 1er triangle (il s'agit d'un des décès du 1er triangle et pas d'un décès tout court), il faut multiplier la première par la surface où elle est d'application.

Le recours aux densités de probabilité offre un avantage en matière d'écriture dans le sens où, pour trouver l'âge moyen au décès, la formule est bien plus compacte et d'un traitement bien plus rapide que la procédure passant par le calcul du nombre de décès et de l'âge cumulés des décédés, mais par contre, elle est, à priori, moins directement lisible que cette dernière procédure, sauf à être accoutumé aux fonctions de densité de probabilité. Dans un processus de formation, nous pensons que la 1re méthode peut être privilégiée dans une 1re approche, quitte à glisser vers les fonctions de densité de probabilité quand le processus est bien compris.

Pour terminer ce point à propos des densités de probabilité, abordons un cas de répartition non uniforme. Supposons une densité de probabilité deux fois plus forte à 0 an qu'à 1 an. Si « a » désigne la densité de probabilité de référence à 1 an, l'intégration sur le triangle qui doit donner 1 s'écrira :

$$\int_{x=0}^{x=1}\int_{y=0}^{y=x}(2-y)*a*dy*dx = 1 \qquad \Rightarrow a\int_{x=0}^{x=1}\left[2y-\frac{y^2}{2}\right]_0^x dx = 1 \Rightarrow a\int_{x=0}^{x=1}\left(2x-\frac{x^2}{2}\right)*dx = 1$$

$$\Rightarrow a*\left[2\frac{x^2}{2}-\frac{x^3}{6}\right]_0^1 = 1 \Rightarrow a*\frac{5}{6} = 1 \quad \Rightarrow a = \frac{6}{5} = 1{,}2.$$

La densité de probabilité vaudra donc 2,4 à 0 an et 1,2 à 1 an, soit une densité double à 0 an.

2.6. Conclusions à propos de l'âge moyen au décès dans le 1er triangle

En cas de répartition uniforme des décès, l'âge moyen au décès dans un 1er triangle vaut un tiers, résultat attesté par la démonstration de Boyarski ou celle passant par le centre de gravité. En outre, cette propriété a aussi été démontrée par des méthodes de sommation en calcul discret et par voie infinitésimale.

Les écritures pour produire les résultats se retrouvent en fait dans les équations pour déterminer le barycentre ou centre de gravité, les poids des surfaces élémentaires correspondant aux nombres ou proportions des décès y localisés[13]. Pourrait-on recourir au calcul du barycentre pour déterminer l'âge moyen dans le 1er triangle ? Certes oui : la coordonnée âge du barycentre est aussi l'âge moyen recherché.

Notons toutefois que, par définition, le barycentre correspond à un point. Dans le cas qui nous occupe, il s'agirait de calculer la situation de ce point dans le 1er triangle du diagramme de Lexis. Pour localiser ce point, il faudra connaitre ses deux coordonnées d'âge (y) et de date (x) (pour être complet, signalons que l'une de ces deux cordonnées pourrait être remplacée par la date de naissance). Or, pour la question visée dans ce texte (l'âge moyen au décès dans le 1er triangle), la connaissance de l'âge moyen suffit, le date moyenne de survenance des décès (ou la date moyenne de naissance des décédés) restant sans intérêt.

Nous avons donc opté pour la méthode qui nous semble la plus transparente pour arriver au résultat recherché, à savoir la division de l'âge cumulé des décédés par leur nombre, sans chercher faire intervenir les notions de barycentres, de centre de gravité, de calcul de moments..., ni à établir un parallélisme avec différentes façons d'exprimer les calculs les concernant (notation vectorielle, notamment), et ce, même si les écritures utilisées dans ce texte se retrouvent dans l'environnement de ces notions.

Le calcul infinitésimal (plus que le calcul discret) offre un avantage important à souligner : il est aisé de s'affranchir de l'hypothèse de répartition uniforme des décès pour explorer d'autres schémas de répartition de la mortalité et en apprécier l'influence sur l'âge moyen au décès. Par ailleurs, il peut être pratique de travailler avec les densités de probabilité qui permettent une simplification des écritures : le calcul d'une seule intégrale suffit pour produire un âge moyen.

Ajoutons que, même si nous avons travaillé uniquement sur triangle, les deux dernières méthodes pourraient aussi s'adapter à des calculs dans des carrés ou des parallélogrammes et servir ainsi à calculer des âges moyens au décès entre 0 et 1 an exact.

2.7. En marge de l'âge moyen, l'âge médian

Dans ce point, nous allons quitter la question de l'âge moyen des décédés du 1er triangle pour aborder un autre paramètre fort utilisé, à savoir la médiane (et autres quantiles, le cas échéant). Pour rappel, par définition, 50 % des observations sont inférieures à la médiane et 50 % lui sont supé-

[13] *Cf.* par exemple, http://fr.wikipedia.org/wiki/Centre_de_gravit%C3%A9.

rieures. Comment déterminer l'âge au décès médian dans le 1er triangle ? La question sera abordée non seulement sous hypothèse de répartition des décès uniforme « à la Boyarski » ou par voie intégrale, mais aussi sous hypothèse de répartition non uniforme, uniquement par voie intégrale.

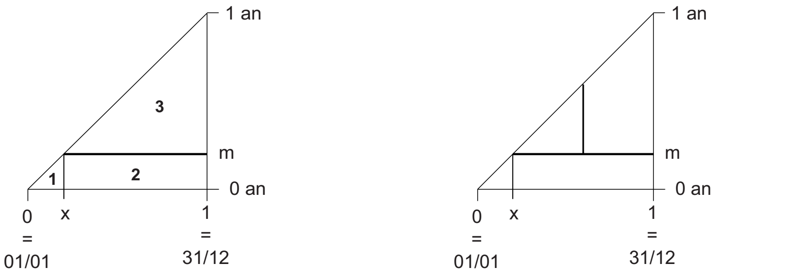
2.7.1. Répartition uniforme des décès

Attaquons d'abord cette question de l'âge médian des décédés du 1er triangle « à la Boyarski ». Sur la figure 8, « m » positionne l'âge médian, encore inconnu à ce stade. Vu les conditions (densité uniforme et unitaire), la moitié des décès qui doivent se produire avant cet âge médian, représente 0,25, soit aussi la moitié de la surface du 1er triangle. La surface sous l'horizontale à hauteur de m se subdivise en deux parties :

- la partie 1 : un triangle rectangle de base et de hauteur valant m ;
- la partie 2 : un rectangle de hauteur m et de base 1-m.

Figure 8. Diagramme de Lexis pour le calcul de la médiane

a. La médiane « à la Boyarski » b. La médiane par voie intégrale



Dès lors, l'équation pour déterminer la valeur de m s'écrit :

$$(1)+(2) = 0,25 \Rightarrow \left(\frac{m^2}{2}\right)+\left(m*(1-m)\right)=0,25 \Rightarrow \left(\frac{m^2}{2}\right)+\left(m-m^2\right)=0,25$$
$$\Rightarrow m-\left(\frac{m^2}{2}\right)=0,25 \Rightarrow m-\left(\frac{m^2}{2}\right)-0,25=0.$$

La racine utile de cette fonction vaut 0,29289, ce qui donne la valeur de l'âge médian. Boyarski aurait-il vraiment suivi cette méthode ? Nous pensons que non. Il aurait sans doute opté pour une voie plus directe et plus simple, en basant le calcul sur le triangle situé au-dessus de la médiane (*cf.* figure 8, la surface (3)). De base et de hauteur valant 1-m, l'équation du calcul devient :

$$\frac{(1-m)^2}{2}=0,25 \Rightarrow m^2-2m+0,5=0 \Rightarrow \frac{m^2}{2}-m+0,25=0 \Rightarrow m-\frac{m^2}{2}-0,25=0.$$

La méthode qui vient d'être suivie est satisfaisante en cas de la répartition uniforme. Par contre, nous nous trouvâmes fort dépourvu quand la répartition non uniforme des décès fut venue. Dans les cas de ce type, il faudra recourir au calcul intégral que nous allons maintenant utiliser, mais en commençant par envisager la répartition uniforme afin de mettre en place les écritures dans un contexte le plus simple possible. Nous allons privilégier le calcul via la surface du triangle au-dessus de

l'horizontale de l'âge médian dans un souci de simplicité[14]. Nous avons opté pour un calcul avec intégrale intérieure sur y (*cf.* figure 8.b). Vu que l'intégrale intérieure se fera sur y et que les bandelettes verticales pour les différentes valeurs de x seront limitées par y = m à y = x et vu que l'intégrale extérieure qui se fera sur x sera limitée par x = m et x = 1, le calcul s'écrira :

$$\int_{x=m}^{x=1}\int_{y=m}^{y=x} dy*dx=0,25 \Rightarrow \int_{x=m}^{x=1}\left[y\right]_m^x dx=0,25 \Rightarrow \int_{x=m}^{x=1}(x-m)*dx=0,25 \Rightarrow \left[\frac{x^2}{2}-mx\right]_m^1=0,25$$
$$\Rightarrow \frac{1}{2}m^2-m+\frac{1}{2}=0,25 \Rightarrow m-\frac{m^2}{2}-0,25=0.$$

Sans surprise, on retombe sur le résultat obtenu en suivant la méthode de Boyarski, mais avec l'avantage de pouvoir généraliser aisément le calcul en cas de répartition non uniforme.

2.7.2. Répartition non uniforme des décès

Supposons une densité des décès variant selon la fonction suivante : $Dens_y=(2-y)$. Dans ces circonstances, nous avons établi précédemment que le nombre de décès dans le 1er triangle était de 5/6. Dès lors, en basant le calcul sur le triangle au-dessus de l'horizontale de l'âge médian (*cf.* la surface 3 de la figure 8), l'équation du calcul de l'âge médian s'établit comme suit :

$$\int_{x=m}^{x=1}\int_{y=m}^{y=x}(2-y)*dy*dx=\frac{5}{12} \Rightarrow \int_{x=m}^{x=1}\left[2y-\frac{y^2}{2}\right]_m^x dx=\frac{5}{12} \Rightarrow \int_{x=m}^{x=1}\left(2x-\frac{x^2}{2}-2m+\frac{m^2}{2}\right)*dx=\frac{5}{12}$$
$$\Rightarrow \left[\frac{2x^2}{2}-\frac{x^3}{6}-2mx+\frac{m^2}{2}x\right]_m^1=\frac{5}{12} \Rightarrow 1-\frac{1}{6}-2m+\frac{m^2}{2}-\frac{2m^2}{2}+\frac{m^3}{6}+2m^2-\frac{m^3}{2}=\frac{5}{12}$$
$$\Rightarrow \frac{5}{6}-2m+3\frac{m^2}{2}-\frac{m^3}{3}=\frac{5}{12} \Rightarrow \frac{5}{12}-2m+3\frac{m^2}{2}-\frac{m^3}{3}=0.$$

La valeur de l'âge médian a été déterminée via une procédure d'itération. Cette procédure a été appliquée entre 0 et 1, soit les seules valeurs envisageables vu la nature de la question à résoudre. Le résultat[15] de cette procédure donne un âge médian de 0,25398 année. Cette valeur est plus faible que dans le cas de la répartition uniforme, ce qui n'est pas une surprise puisque les décès sont plus concentrés près de la naissance avec la répartition non uniforme choisie.

Cet exemple, généralisable à tous les quantiles, montre en quoi le calcul intégral peut aussi nous aider pour l'estimation de paramètres autres que la moyenne. Le point suivant portera de nouveau sur une moyenne, à savoir l'âge moyen des individus intégrant l'effectif à 0 an révolu.

[14] L'annexe 3 reprend le calcul en partant de la désagrégation de la surface sous la ligne horizontale de l'âge médian en un triangle et un rectangle (*cf.* figure 8, surface (1) et (2)).

[15] Avec plus de décimales : 0,25398334941409.

3. L'âge moyen des individus intégrant l'effectif à 0 an révolu

L'effectif à 0 an révolu en fin d'une année regroupe tous les individus âgés de 0 an exact à moins de 1 an exact et donc nés durant cette année. Ils constituent une classe d'âge, celle des 0 à moins de 1 an. Dans ce type de situation, en démographie, et plus généralement en statistique, l'âge moyen des individus de la classe est considéré comme équivalent au centre de la classe, soit à 0,5 an[16]. Si on ajoute l'hypothèse de répartition uniforme des décès, il est vraiment tentant d'accepter cette valeur les yeux fermés.

Pour être complet en matière d'hypothèses, il convient d'ajouter une hypothèse sur la répartition des naissances (point 2.1). Par ailleurs, même en supposant la répartition uniforme des naissances et des décès, ce résultat (0,5 an) n'est pas logique sauf en cas de risque de mourir… nul (point 2.2). Enfin, le mode de calcul mis en place sera testé en cas de répartition non uniforme des décès, ce qui s'avérera relativement aisé grâce au calcul intégral (point 2.3).

3.1. La nécessité de l'hypothèse de répartition uniforme des naissances

L'hypothèse de répartition uniforme des décès est plus que régulièrement citée quand le calcul de la table est expliqué. Généralement, celle portant sur la répartition uniforme des naissances est purement et simplement omise[17]. Pourquoi cet ostracisme ? En effet, la répartition uniforme des naissances ne va pas plus de soi que celle des décès[18]. Par ailleurs, l'influence des ces deux hypothèses sur la validité des méthodes peut, selon les circonstances, prendre autant d'importance.

En cas de répartition uniforme des décès entre 0 et 1 an exact, quel est l'âge moyen des individus constituant l'effectif à 0 an révolu ? En général, on considère que cet âge moyen vaut 0,5 an. Pourquoi poser l'hypothèse de répartition uniforme des naissances en plus de celle affectant la répartition des décès ? Prenons trois exemples :

- 1re situation : imaginons que presque toutes les naissances d'une année soient concentrées le 1er janvier (hypothèse certes extrême que la nature théorique du propos excuse). Ceux qui intègrent de ce fait l'effectif à 0 an révolu ont majoritairement vécu (pratiquement) une année avant d'intégrer cet effectif ; l'âge moyen pour l'effectif à 0 an révolu serait donc très proche de 1 an ;
- 2e situation : si les naissances sont réparties uniformément au cours de l'année ainsi que les décès, ceux qui intègrent l'effectif à 0 an révolu auront 0,5 an, du moins approximativement, comme expliqué plus loin ;
- 3e situation : si chaque mois, le nombre de naissances varie et que les décès sont répartis uniformément, on n'obtiendra pas (nécessairement) en moyenne un âge de 0,5 an de vie parmi ceux qui intègrent l'effectif à 0 an révolu : si, par exemple, les naissances sont plus nombreuses en fin d'année, la moyenne sera forcément inférieure à 0,5 an.

On le voit la simple répartition uniforme des décès n'est pas suffisante ; il faut la compléter avec une hypothèse portant sur la répartition uniforme des naissances. Pourquoi cette dernière est-elle souvent ignorée ? À notre avis, cette situation vient de ce que, dans beaucoup de calculs, elle n'est pas nécessaire au contraire de l'hypothèse sur la répartition des décès. Il n'y a pourtant pas là un

[16] *Cf.*, par exemple, Grais (1994) p. 124, Bailly (1999) p. 142 ou dans un contexte plus démographique : https://monitoringdesquartiers.irisnet.be/static/attachments/statistics/fr/age-moyen/Demo08_FI_20130418_FR.pdf.

[17] Avec de précieuses exceptions : *cf.* notamment Calot et Caselli (1991) ou Calot et Franco (2002).

[18] Pour se faire un avis à propos de la répartition des naissances au cours de l'année, *cf.* Régnier-Loilier et Rohrbasser (2011) ou Calot et Caselli (1991). Par ailleurs, certaines répartitions non uniformes des naissances pourraient contrebalancer exactement la répartition uniforme des décès pour obtenir 0,5 an ; cela ne nous semble toutefois pas du tout une voie prioritaire de réflexion !

argument pour la passer sous silence, notamment dans les circonstances où elle est indispensable aux calculs. Elle mérite alors de retenir autant l'attention que la répartition uniforme des décès.

Venons-en maintenant à la question de l'âge moyen de l'effectif de 0 an révolu en cas de répartition uniforme des décès et des naissances.

3.2. Mécomptes de la répartition uniforme des décès

Comme signalé en introduction de ce point, généralement, on considère que l'âge moyen des individus constituant l'effectif à 0 an révolu est de 0,5 an. Ce 0,5 an est-il justifié ? En cas de répartition uniforme des naissances ainsi que des décès entre 0 exact et 0 révolu, quel est l'âge exact moyen des individus intégrant l'effectif à 0 an révolu ?

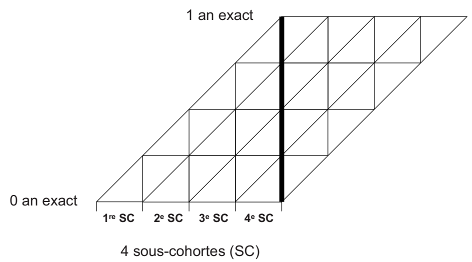
3.2.1. Un calcul discret

3.2.1.1. Un premier exemple exploratoire

Pour étudier cette question, nous allons prendre l'exemple d'un calcul entre âges exacts. Soit la situation suivante : les 400 naissances d'une année se sont réparties uniformément ; en particulier, chaque quart de l'année a enregistré un quart des naissances, soit 100 naissances (*cf.* figure 9). La cohorte annuelle va être décomposée en 4 sous-cohortes (SC), une par quart d'année.

Chaque sous-cohorte est divisée en 8 triangles, soit un total de 32 triangles. Si chacun des triangles compte un décès localisé adéquatement, la répartition uniforme des décès est respectée. L'effectif à 0 an révolu sera de 384 (soit 400-16), ce qui correspond bien à la moyenne entre les effectifs à 0 (soit 400 individus) et 1 an exact (soit 400-32 = 368 individus).

Figure 9. Une cohorte et 4 sous-cohortes



Qu'en est-il de l'âge moyen des individus qui intègrent l'effectif à 0 an révolu, dont la localisation sur la figure 9 est soulignée par un trait plus épais ? *De manière spécifique : vaut-il 0,5 an ?*

Pour la 1re sous-cohorte, il y aura eu 7 décès car 7 triangles avant d'atteindre l'effectif à 0 an révolu. 93 individus de cette 1re sous-cohorte intègrent donc l'effectif à 0 an révolu. Quel est leur âge moyen ? En moyenne, ces 93 survivants ont vécu approximativement – on reviendra plus tard sur ce dernier mot – 10,5 mois, soit la moyenne entre 12 et 9 (12 pour ceux qui sont nés en tout début d'année et 9 pour ceux qui sont nés juste à la fin du 1er quart d'année). Pour les 3 autres sous-cohortes, les mêmes grandeurs sont respectivement 95, 97 et 99 survivants à 0 an révolu et un âge moyen au décès de 7,5 ; 4,5 et 1,5 mois. La moyenne pondérée de l'âge des individus de l'effectif à 0 an révolu donne :

((10,5*93) + (7,5*95) + (4,5*97) + (1,5*99))/384 = 5,92 mois !

Si chaque triangle contenait 10 décès (ce qui correspond à un risque de mourir entre 0 et 1 an exact de 0,80, soit 320 décès divisés par 400 naissances (risque de mourir certes très élevé, mais la nature théorique du propos…)), l'âge moyen des individus à 0 an révolu serait de :

((10,5*30) + (7,5*50) + (4,5*70) + (1,5*90))/240 = 4,75 mois.

Dans les deux cas, on n'obtient pas 0,5 an. La raison de ce résultat est très simple : les sous-cohortes les plus âgées ont des effectifs plus faibles que les sous-cohortes plus jeunes vu que les premières ont éprouvé plus longuement le risque de mourir ! Ainsi, dans le dernier calcul proposé, l'effectif des sous-cohortes varie de 30 à 90, le 1er servant de poids à la valeur 10,5 et le 2e à celle de 1,5. En définitive, même en supposant la répartition uniforme des naissances et des décès, l'âge moyen des individus de l'effectif à 0 an révolu ne peut pas être 0,5 an.

Cette conclusion souligne une approximation dans notre raisonnement : pour déterminer l'âge moyen des sous-cohortes au moment où elles intègrent l'effectif à 0 an révolu, on a accepté l'idée de la valeur centrale (10,5 mois pour la 1re sous-cohorte, par exemple) dont on vient de dire qu'elle n'était pas recevable au niveau de la cohorte. Il va de soi qu'elle ne l'est pas plus au niveau des sous-cohortes ! Toutefois, cela ne remet pas en cause la conclusion principale. En effet, en corrigeant l'âge moyen des sous-cohortes, suivant le mécanisme qui vient d'être décrit, il serait abaissé dans chaque sous-cohorte. En effet, les effectifs des subdivisions de la sous-cohorte seraient de plus en plus importants en se rapprochant de la borne inférieure de la sous-cohorte. Ce mécanisme conduirait donc à une valeur encore plus faible pour l'âge moyen recherché !

Enfin, remarquons bien que l'hypothèse de répartition uniforme des décès suffit pour assurer que

- l'âge moyen au décès des décédés entre 0 et 1 an exact est de 0,5
- l'effectif à 0 an révolu est la moyenne entre les effectifs aux âges exacts de 0 et 1 an !

Ces éléments ne sont nullement remis en cause par la conclusion qui vient d'être proposée à propos de l'âge moyen des individus intégrant l'effectif de 0 an révolu. Dans le point suivant, nous allons analyser la sensibilité de ce résultat aux deux facteurs suivants : le nombre de sous-cohortes et la valeur du quotient.

3.2.1.2. Analyse systématique en fonction du nombre de sous-cohortes et du risque de mourir

Un 1er point va mettre en place des équations pour analyser la façon dont évolue la différence entre l'hypothèse habituelle sur l'âge moyen de l'effectif à 0 an révolu (0,5 an) et un calcul plus juste. Dans un 2e point, ces équations serviront à tester la variabilité de la sensibilité des résultats en fonction du niveau de mortalité (le quotient) et du nombre de sous-cohortes.

α. Les équations

Pour ce faire, repartons de l'exemple où il y a 4 sous-cohortes (n = 4) dont l'effectif initial est de 100 et où il y a un décès dans chaque petit triangle des sous-cohortes (*cf.* figure 9).

Dans ces conditions, nous avons établi une approximation de l'âge moyen en mois ($\overline{ym}$) parmi les individus âgés de 0 an révolu :

$$\overline{ym} = \frac{(10,5*93)+(7,5*95)+(4,5*97)+(1,5*99)}{93+95+97+99} = 5,92 \text{ mois}.$$

Dans cette formule, 93 (95, 97 et 99) est le nombre d'individus de la 1re (2e, 3e et 4e) sous-cohorte intégrant l'effectif à 0 an révolu et 10,5 (7,5 ; 4,5 et 1,5) leur âge moyen quand ils l'intègrent[19]. L'âge moyen est simplement une moyenne pondérée des âges dans les sous-cohortes, les poids étant les effectifs de chaque sous-cohorte intégrant l'effectif de la cohorte à 0 an révolu.

Si « d » désigne le nombre de décès par triangle (soit 1 vu les conditions du calcul) et si chaque sous-cohorte compte 100 individus à la naissance (*idem*), l'effectif de la sous-cohorte x (avec x variant de 1 à 4), dénommé eff_x, se calculera avec la formule suivante :

$$eff_x = 100 - [((2*n)-((2*x)-1))*d].$$

Appliquée respectivement aux 4 sous-cohortes, on obtient bien : 93, 95, 97 et 99.

L'âge central de l'effectif de la sous-cohorte x qui intègre l'effectif à 0 an révolu de la cohorte, dénommé $âge_x$ et exprimé en mois, se calcule comme suit :

$$âge_x = 12 - \left[\left(\left(\frac{12}{n}\right)*(x-1)\right)+\left(\frac{12}{2*n}\right)\right].$$

Appliquée respectivement aux 4 sous-cohortes, on obtient bien : 10,5 ; 7,5 ; 4,5 et 1,5 mois. Dès lors, pour obtenir l'âge moyen en mois de l'effectif de la cohorte à 0 an révolu, on peut utiliser la formule suivante :

$$\overline{ym} = \frac{\sum_{x=1}^{n}\left\langle\left\{12-\left[\left(\left(\frac{12}{n}\right)*(x-1)\right)+\left(\frac{12}{2*n}\right)\right]\right\}*\left\{100-[((2*n)-((2*x)-1)*d)]\right\}\right\rangle}{\sum_{x=1}^{n}\left\{100-[((2*n)-((2*x)-1)*d)]\right\}}.$$

Pour analyser les résultats, nous avons trouvé plus commode d'exprimer les âges moyens en jours ($\overline{yj}$). Faisant résolument fi des années bissextiles, l'âge moyen en mois sera transformé en un âge moyen en jours comme suit :

$$\overline{yj} = \frac{\text{âge moyen en mois}}{12}*365.$$

β. L'analyse de sensibilité

Le tableau 4 reprend les résultats établis selon les formules proposées ci-dessus. L'analyse de sensibilité portera non seulement sur le niveau du risque de mortalité en prenant des risques de 0 ; 0,0001 ; 0,001 ; 0,01 ; 0,10 et 1,00 (ce qui conditionne le nombre de décès par petit triangle), mais aussi sur le nombre de sous-cohortes, en prenant successivement 10, 20, 40, 80, 160, 320 et 640 sous-cohortes[20]. En multipliant les sous-cohortes, on s'approche d'une analyse continue, l'idée étant de minimiser l'approximation venant de l'utilisation de la valeur centrale comme âge moyen des individus de la sous-cohorte intégrant l'effectif de la cohorte à 0 an révolu. À la limite, si le nombre de sous-cohortes tend vers l'infini, l'âge central sera en fait un âge exact et plus aucune approximation ne sera faite à propos de l'âge moyen dans les sous-cohortes (*cf.* point 2.2.2).

La dernière ligne de chaque sous-tableau (« Différence avec le suivant ») permet d'apprécier la différence entre un âge moyen et celui qui est à sa droite. Par exemple, la valeur 0,0002282 du 3e sous-

[19] Nous savons que cette hypothèse privilégiant la valeur centrale n'est pas recevable, mais dans le début du raisonnement, on peut la maintenir. Plus loin, le nombre de sous-cohortes va être augmenté, ce qui en diminuera la portée pratique !

[20] L'effectif à 0 an exact sur lequel porte le risque est sans influence sur le résultat. Il a été fixé arbitrairement à 1.000.

tableau (quand le risque vaut 0,001 et le nombre de sous-cohortes, 10) a été obtenue en faisant 182,469872−182,469644 (cette dernière valeur venant de la colonne avec 20 sous-cohortes). Le résultat est positif indiquant que, quand le nombre de sous-cohortes augmente, l'écart avec 182,5 jours augmente. Dans chaque sous-tableau, à mesure que l'on se déplace vers la droite, la valeur de la différence diminue, mais de façon de moins en moins prononcée : les résultats convergent.

Tableau 4. Âge moyen de l'effectif à 0 an révolu selon le nombre de sous-cohortes et la valeur du quotient de mortalité

	Nombre de sous-cohortes						
q(0) = 0	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	182,5000000	182,5000000	182,5000000	182,5000000	182,5000000	182,5000000	182,5000000
≠ avec le suivant	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	0,0000000	
q(0) = 0,0001	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	182,4969886	182,4969658	182,4969601	182,4969587	182,4969583	182,4969582	182,4969582
≠ avec le suivant	0,0000228	0,0000057	0,0000014	0,0000004	0,0000001	0,0000000	
q(0) = 0,001	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	182,4698724	182,4696442	182,4695871	182,4695729	182,4695693	182,4695684	182,4695682
≠ avec le suivant	0,0002282	0,0000571	0,0000143	0,0000036	0,0000009	0,0000002	
q(0) = 0,01	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	182,1973618	182,1950691	182,1944959	182,1943526	182,1943168	182,1943078	182,1943056
≠ avec le suivant	0,0022927	0,0005732	0,0001433	0,0000358	0,0000090	0,0000022	
q(0) = 0,10	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	179,3302632	179,3062500	179,3002467	179,2987459	179,2983707	179,2982769	179,2982534
≠ avec le suivant	0,0240132	0,0060033	0,0015008	0,0003752	0,0000938	0,0000235	
q(0) = 1	**10**	**20**	**40**	**80**	**160**	**320**	**640**
Âge moyen (jours)	122,2750000	121,8187500	121,7046875	121,6761719	121,6690430	121,6672607	121,6668152
≠ avec le suivant	0,4562500	0,1140625	0,0285156	0,0071289	0,0017822	0,0004456	

Par ailleurs, plus le risque est élevé, plus l'écart par rapport à 182,5 jours est important. Quand le risque vaut 0, l'âge moyen est égal à 182,5 dans toutes les circonstances (1er sous-tableau). Ce résultat était prévisible puisque chaque sous-cohorte a le même poids quand elle intègre l'effectif à 0 an révolu, à savoir son effectif initial, supposé identique pour chacune d'entre elles. En cas de risque unitaire, l'âge moyen est de l'ordre de 122 jours, soit pratiquement 1/3 d'année (1/3 d'année = 121,66667 jours). En définitive, l'âge moyen recherché variera donc entre 0,5 et 1/3 d'année selon la valeur du risque entre les 2 âges concernés.

3.2.2. Une démonstration par recours au calcul intégral

La conclusion qui vient d'être proposée peut être également démontrée grâce au calcul intégral. La procédure qui va être suivie reprend les étapes suivantes :

- calcul de la probabilité d'atteindre l'effectif à 0 an révolu ;
- calcul de l'effectif des individus à 0 an révolu ;
- calcul du temps vécu par ces individus avant d'atteindre cet effectif ;
- finalement, par division de ces deux dernières quantités, calcul de l'âge moyen des individus de l'effectif à 0 an révolu.

3.2.2.1. Probabilité d'atteindre l'effectif à 0 an révolu

Soit $q_{0\to 1}$ le risque de mourir entre 0 et 1 an exact. Vu l'hypothèse de répartition uniforme des décès, que vaut la probabilité de survivre de 0 à l'âge exact y, pour tout y compris entre 0 et 1 ?

Le risque entre 0 et 1 an exact s'obtient en divisant les décès entre 0 et 1 an exact par le nombre d'individus à 0 an exact : $q_{0\to 1} = \frac{d_{0\to 1}}{l_0}$.

Que se passe-t-il pour l'âge de z an exact, par exemple (*cf.* figure 10) ? Vu la répartition uniforme, la proportion des décès se localisant entre les lignes horizontales des âges 0 exact et z exact vaut forcément z. Par exemple, si l'âge z vaut 0,7 an, 70 % des décès se localisent entre les âges 0 et 0,7 an exact. De manière générale : $d_{0\to z} = z * d_{d0\to 1}$.

Dès lors, le risque de mourir entre 0 et z an exact se calcule comme suit :

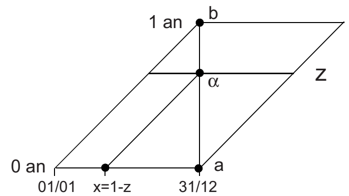
$$q_{0\to z} = \frac{d_{0\to z}}{l_0} = \frac{z * d_{0\to 1}}{l_0} = z * q_{0\to 1}.$$

La probabilité de survivre de la naissance (point x = 1-z) jusqu'à l'âge z (point α) vaut :

$$p_{0\to z} = 1 - z * q_0.$$

Cette probabilité quantifie la partie de la sous-cohorte née en x qui survit en fin de 1re année de vie. Ce résultat est d'application pour toutes les valeurs de y comprises entre 0 et 1. Si « q » désigne le quotient entre 0 et 1 an exact : $p_{0\to y} = 1 - y * q$.

Figure 10. Diagramme de Lexis et calcul intégral



3.2.2.2. Effectif des individus à 0 an révolu

En supposant une densité de naissances uniforme et unitaire[21], l'effectif à 0 révolu (désigné par $l_{\underline{0}}$) se calculera comme suit : $l_{\underline{0}} = \int_{y=0}^{y=1} \left(1 - (q * y)\right) * dy = \left[y - \frac{q * y^2}{2} \right]_0^1 = 1 - \frac{1}{2} q$.

Vu les circonstances, ce résultat était hautement prévisible. En effet, vu la répartition uniforme des décès, la moitié des décès entre 0 et 1 an exact se produisent dans le 1er triangle. L'effectif de survivants à 0 an révolu se calcule donc en retirant la 1/2 des décès entre 0 et 1 an exact de l'effectif initial, ce qui revient à dire que le risque de mourir dans le 1er triangle vaut la moitié du risque entre 0 et 1 an exact. Par ailleurs, vu que la densité de naissances est uniforme et unitaire, l'effectif à 0 an révolu vaut bien 1 moins la moitié du quotient. Malgré son caractère hautement prévisible, ce résultat méritait d'être calculé par voie intégrale. Cela montre notamment que la procédure donne le résultat attendu. En outre, cette procédure de calcul intégral va être déjà utilisée pour finaliser le calcul de l'âge moyen des individus intégrant l'effectif à 0 an révolu.

21 Il n'est pas obligatoire, mais très pratique de prendre un effectif unitaire né à la date indiquée. Cela débouchera certes sur des résultats difficiles à interpréter car s'exprimant en partie d'individu survivant, mais le calcul s'en trouve simplifié. Si cette hypothèse est jugée insupportable, il est loisible d'opter pour un effectif quelconque, ce qui ne modifiera en rien la conclusion, mais rendra les calculs simplement un peu plus longs et se terminant par des simplifications comme celles rencontrées dans le texte !

3.2.2.3. Temps vécu par les individus avant d'intégrer l'effectif à 0 an révolu

Par définition, les individus d'âge y au 31/12 ont vécu un temps égal à y avant d'intégrer l'effectif à 0 an révolu. Le cumul de leurs âges correspond à leur nombre multiplié par y :

$$\text{Âge cum}_y = (1 - y * q) * y = y - y^2 q .$$

Le total du temps vécu pour l'ensemble de l'effectif à 0 an révolu le 31/12 (« Âge tot ») s'obtient en intégrant cette dernière fonction entre 0 et 1 :

$$\text{Âge tot} = \int_{y=0}^{y=1} \left(y - q * y^2\right) * dy = \left[\frac{y^2}{2} - \frac{q * y3}{3}\right]_0^1 = \frac{1}{2} - \frac{1}{3} q .$$

3.2.2.4. Âge moyen des individus de l'effectif à 0 an révolu

Finalement, l'âge moyen de l'effectif à 0 an révolu ($\bar{y}$) s'obtient par le rapport entre le temps vécu par l'ensemble de cet effectif et cet effectif lui-même :

$$\bar{y} = \frac{\int_{y=0}^{y=1} (y - q * y^2) * dy}{\int_{y=0}^{y=1} (1 - (q * y)) * dy} = \frac{\frac{1}{2} - \frac{1}{3} q}{1 - \frac{1}{2} q} .$$

Pour q = 1, l'âge moyen vaut : $\bar{y} = \dfrac{\frac{1}{2} - \frac{1}{3}}{1 - \frac{1}{2}} = \dfrac{1/6}{1/2} = \dfrac{1}{3}$.

Pour q = 0,1, l'âge moyen vaut : $\bar{y} = \dfrac{\frac{1}{2} - \left(\frac{1}{3} * \frac{1}{10}\right)}{1 - \left(\frac{1}{2} * \frac{1}{10}\right)} = \dfrac{\frac{1}{2} - \left(\frac{1}{30}\right)}{1 - \left(\frac{1}{20}\right)} = \dfrac{0{,}5 - 0{,}0333}{1 - 0{,}05} = \dfrac{0{,}4667}{0{,}95} = 0{,}49123$.

Pour q = 0, l'âge moyen vaut : $\bar{y} = \dfrac{\frac{1}{2} - \left(\frac{1}{3} * 0\right)}{1 - \left(\frac{1}{2} * 0\right)} = \dfrac{0{,}5}{1} = 0{,}5$.

Le tableau 5 reprend les résultats pour les mêmes risques que dans le tableau 4. Grand avantage de cette méthode : on obtient directement la valeur vers laquelle convergent les résultats proposés dans le tableau 4. Toutefois, nous avons gardé le 1er mode de calcul car il montre de manière très intuitive le processus mathématique en jeu. Par contre, au niveau des résultats, la 2e méthode est incontestablement plus performante. Sans surprise, cette 2e méthode confirme les résultats obtenus via la 1re, à savoir : si le quotient de mortalité entre 0 et 1 an exact vaut 0, l'âge moyen de l'effectif à 0 an révolu sera de 0,5 an et de 1/3 d'année si ce quotient vaut 1. Un quotient intermédiaire donnera une valeur entre ces deux extrêmes.

Tableau 5. Âge moyen de l'effectif à 0 an révolu – 2e méthode

Risque de mourir	Âge moyen	
	En année	En Jours
0	0,5000000	182,5000000
0,0001	0,4999917	182,4969582
0,001	0,4999166	182,4695681
0,01	0,4991625	182,1943049
0,1	0,4912281	179,2982456
1,00	0,3333333	121,6666667

La procédure qui vient d'être suivie va maintenant être appliquée en cas de distribution non uniforme des décès. Ces résultats pourraient aussi être produits au départ d'un quotient calculé dans le triangle (*cf.* annexe 4).

3.3. En cas de répartition non uniforme

Supposons une densité des décès variant en fonction de l'âge y comme suit : (2 – y)*1, où « 1 » est la densité de référence à l'âge d'un an exact. À 0 an exact, elle vaut 2 et 1, à un an exact, avec évolution linéaire entre 0 et 1 an exact.

La figure 10 montre la localisation des décès survenus à l'âge exact z, soit sur une horizontale à hauteur de cet âge. Le point α est l'intersection entre cette horizontale et la verticale « ab », elle-même localisant l'effectif à 0 an révolu. La longueur de l'horizontale d'âge z vaut 1, à l'image des horizontales dans le parallélogramme pour tous les âges exacts compris entre 0 et 1.

Dès lors, le nombre de décès survenus entre 0 exact et z exact (ND_z) se calcule comme suit :

$ND_z = \int_{y=0}^{y=z} (2 - y) * 1 * 1 * dy$, où

- le 1er « 1 » indique la densité de référence ;
- le 2e « 1 » se justifie par la longueur des horizontales.

Le nombre de décès survenus entre 0 exact et 1 exact (ND_1) se calcule comme suit :

$$ND_1 = \int_{y=0}^{y=1} (2 - y) * dy .$$

La proportion de décès survenus entre 0 exact et z exact (PD_z) est donnée par :

$$PD_z = \frac{\int_{y=0}^{y=z} (2 - y) * dy}{\int_{y=0}^{y=1} (2 - y) * dy} = \frac{\left[2y - \frac{y^2}{2}\right]_0^z}{\left[2y - \frac{y^2}{2}\right]_0^1} = \frac{2z - \frac{z^2}{2}}{2 - \frac{1}{2}} = \frac{2z - \frac{z^2}{2}}{\frac{3}{2}} = \frac{4}{3} z - \frac{z^2}{3} .$$

Si q = le risque de mourir entre 0 et 1 an exact, le risque de mourir entre 0 et z ($q_{0 \to z}$) vaut :

$$q_{0 \to z} = \frac{d_{0 \to z}}{l_0} = \frac{(d_{0 \to 1}) * PD_z}{l_0} = \frac{(d_{0 \to 1})}{l_0} * PD_z = q * PD_z .$$

La probabilité de survie entre 0 et z ($p_{0 \to z}$) se calcule comme suit :

$$p_{0\to z} = 1-(q_{0\to z}) = 1-(q * PD_z) = 1-\left(q*\left(\frac{4}{3}z-\frac{z^2}{3}\right)\right).$$

« $p_{0\to z}$ » quantifie la proportion de survivants parmi ceux qui sont nés en 1-z, et localisés au point α sur la figure 10. En supposant une densité de naissances uniforme et unitaire pour tous les moments de naissance 1-y, l'effectif à l'âge exact z vaut :

$$\text{Effectif}_z = p_{0\to z} = 1-\left(q*\left(\frac{4}{3}z-\frac{1}{3}z^2\right)\right).$$

Cette expression est d'application pour toutes les valeurs de y comprises entre 0 et 1. Dès lors, le nombre de survivants à 0 an révolu ($l_{\underline{0}}$) et la somme des âges des individus intégrant l'effectif à 0 an révolu (Âge tot.), se calculent respectivement comme suit :

$$l_{\underline{0}} = \int_{y=0}^{y=1} 1-\left(q*\left(\frac{4}{3}y-\frac{1}{3}y2\right)\right)*dy = \int_{y=0}^{y=1}\left(1-q\frac{4}{3}y+q\frac{1}{3}y2\right)*dy = \left[y-q\frac{4}{3}\frac{y^2}{2}+q\frac{1}{3}\frac{y^3}{3}\right]_0^1$$

$$= 1-q\left(\frac{4}{6}-\frac{1}{9}\right) = 1-\frac{30}{54}q = 1-\frac{5}{9}q.$$

$$\text{Âge tot.} = \int_{y=0}^{y=1}\left(1-q\frac{4}{3}y+q\frac{1}{3}y2\right)*y*dy = \int_{y=0}^{y=1}\left(y-q\frac{4}{3}y^2+q\frac{1}{3}y3\right)*dy = \left[\frac{y^2}{2}-q\frac{4}{3}\frac{y^3}{3}+q\frac{1}{3}\frac{y^4}{4}\right]_0^1$$

$$= \frac{1}{2}-q\left(\frac{4}{9}-\frac{1}{12}\right) = \frac{1}{2}-q\frac{48-9}{108} = \frac{1}{2}-q\frac{39}{108} == \frac{1}{2}-q\frac{13}{36}.$$

Finalement, l'âge moyen s'obtient en divisant le total des âges par le nombre d'individus :

- pour q = 1 : $\bar{y} = \dfrac{1/2 - 13/36}{1-5/9} = 0{,}31250$;
- pour q = 0,1 : $\bar{y} = \dfrac{1/2 - ((13/36)*0{,}1)}{1-((5/9)*0{,}1)} = 0{,}49118$;
- pour q = 0 : $\bar{y} = \dfrac{1/2 - ((13/36)*0)}{1-((5/9)*0)} = 0{,}50000$, résultat logique vu les conditions.

Supposons maintenant une densité des décès 10 fois plus forte à 0 an exact qu'à 1 an exact et une densité unitaire à 1 an exact. La procédure suivie plus haut donne les résultats suivants :

$$PD_z = \frac{\int_{y=0}^{y=z}(10-9y)*dy}{\int_{y=0}^{y=1}(10-9y)*dy} = \frac{\left[10y-9\frac{y^2}{2}\right]_0^z}{\left[10y-9\frac{y2}{2}\right]_0^1} = \frac{10z-\frac{9}{2}z^2}{10-\frac{9}{2}} = \frac{10z-\frac{9}{2}z^2}{\frac{11}{2}} = \frac{20}{11}z-\frac{9}{11}z^2$$

$$p_{0\to z} = 1-(q_{0\to z}) = 1-(q*PD_z) = 1-\left(q*\left(\frac{20}{11}z-\frac{9}{11}z^2\right)\right).$$

Ce résultat est d'application pour tous les âges exacts entre 0 et 1. Dès lors,

$$l_{\underline{0}} = \int_{y=0}^{y=1} 1-\left(q*\left(\frac{20}{11}y-\frac{9}{11}y^2\right)\right)*dy = \int_{y=0}^{y=1}\left(1-q*\frac{20}{11}y+q*\frac{9}{11}y^2\right)*dy == 1-q*\frac{7}{11}.$$

$$\text{Âge tot.} = \int_{y=0}^{y=1}\left(1-\left(q*\left(\frac{20}{11}y-\frac{9}{11}y^2\right)\right)\right)*y*dy = \left(\frac{1}{2}\right)-\left(q*\left(\frac{20}{33}-\frac{9}{44=}\right)\right) = \left(\frac{1}{2}\right)-\left(q*\left(\frac{53}{132}\right)\right).$$

Finalement, l'âge moyen vaut :

- pour q = 1 : $\bar{y} = \dfrac{1/2 - (53/132)}{1-7/11} = 0{,}27083$;
- pour q = 0,1 : $\bar{y} = \dfrac{1/2 - (0{,}1*(53/132))}{1-(0{,}1*7/11)} = 0{,}49110$;
- pour q = 0 : $\bar{y} = \dfrac{1/2 - (0*(53/132))}{1-(0*7/11)} = 0{,}5$, résultat logique vu les conditions.

Grâce à ces calculs, nous avons pu simuler l'influence de la répartition non uniforme des décès sur l'âge moyen des individus intégrant l'effectif à 0 an révolu. Même si peu d'exemples sont disponibles, on peut constater que

- cet âge n'est point 0,5 an, sauf en cas de quotient de mortalité nul ;
- plus les décès sont concentrés vers le tout début de la vie, plus l'écart avec 0,5 est important : pour q = 1, l'âge moyen vaut 0,49110 en cas de densité 10 fois plus forte à 0 qu'à 1 an, mais 0,49123 en cas de densité 2 fois plus forte à 0 an qu'à 1 an ;
- pour des valeurs, mêmes fortes du risque de mourir (q = 0,1) avec une répartition très éloignée de la répartition uniforme (10-9y), cet âge ne s'écarte guère de 0,5.

Au contraire de ce qui avait été fait au point 1, l'hypothèse de non-linéarité de la variation de densité des décès ne sera pas abordée ici.

3.4. Conclusions du point 3

Donc, la conclusion est solide : en cas de répartition uniforme des décès et des naissances, à 0 an révolu, l'âge moyen des individus n'est pas 0,5 an, sauf en cas de risque de mourir nul !

Sur un plan purement pratique, la différence entre l'hypothèse habituelle (âge moyen = 0,5) et un calcul plus juste est négligeable, sauf quand la mortalité est fort élevée, comme en début de table dans certains pays ou à certaines époques. Par ailleurs, plus la concentration des décès près de 0 an est élevée[22], plus l'écart est grand par rapport à 0,5 an. Toutefois, même si les conditions actuelles de mortalité n'engendrent pas d'impact significatif sur les résultats, au moins dans les pays du Nord, cette constatation ne constitue pas une raison suffisante pour occulter la question, quand on veut comprendre en profondeur les méthodes et surtout en apprécier, en pleine connaissance de cause, les limites.

Nous avons donc montré que, même en cas de répartition uniforme des décès et des naissances, l'âge moyen de ceux qui intègrent l'effectif à 0 an révolu n'est pas 0,5 parce que l'effectif des sous-

22 L'hypothèse alternative (concentration des décès près du 1er anniversaire) n'a pas été testée. Logiquement, elle devrait aboutir à une diminution de l'écart par rapport avec 0,5.

cohortes, quand elles intègrent cet effectif à 0 an révolu, n'est pas constant[23]. Ce déséquilibre affectant le premier effectif révolu, se renforcera dans les suivants, même en cas de répartition uniforme des décès entre 2 âges révolus successifs (*cf.* annexe 5).

À l'inverse, dans ce type de table entre âges révolus, si l'effectif à 0 an révolu est réparti uniformément (ce qui serait hautement surprenant) et que les décès dans la suite sont répartis uniformément (ce qui serait tout aussi surprenant), à chaque âge révolu suivant, l'âge moyen des individus sera l'âge révolu + 0,5. En effet, en se basant sur un raisonnement discret, si l'effectif à 0 an révolu se répartit uniformément, cela implique que les différentes sous-cohortes ont le même effectif. Vu la répartition uniforme des décès au-delà de 0 an révolu, chaque sous-cohorte va perdre le même nombre d'individus par décès d'un âge révolu au suivant et donc l'importance relative des sous-cohortes ne changera pas. En définitive, l'âge moyen sera toujours l'âge + 0,5 an.

Précisons enfin, que les effectifs aux âges exacts ne rencontrent pas le même problème : leur âge moyen sera, par définition, l'âge… exact en cause, avec ou sans répartition uniforme des naissances et des décès[24]. De manière spécifique, en cas de table en âges révolus, les effectifs au centre des parallélogrammes (reposant sur pointe) correspondent à un âge exact. L'âge moyen des individus intégrant cet effectif est cet âge exact lui-même, sans aucune discussion.

4. Conclusions générales

Premier point : *en cas de répartition uniforme, l'âge moyen au décès dans le 1er triangle vaut 1/3*

Le fait que l'âge moyen au décès dans un 1er triangle vaut l'âge exact à la base du triangle plus un tiers peut être démontré par plusieurs voies. Nous trouvons que la démonstration de Boyarski est particulièrement élégante. Par ailleurs, le recours aux sommations ou au calcul infinitésimal permet de s'affranchir à peu de frais de l'hypothèse de répartition uniforme des décès.

Cette conclusion est généralisable à tous les triangles orientés comme le premier (« lower triangle »), moyennant que le correctif suivant : l'âge moyen au décès vaut l'âge exact à la base du triangle + un tiers. Pour le 2e triangle et tous ceux de son espèce (« upper triangle »), l'âge moyen vaut 2/3 augmenté de l'âge révolu les délimitant à gauche.

Deuxième point : *nécessité d'une hypothèse de répartition uniforme des naissances*

Les démographes usent d'abondance de l'hypothèse de répartition uniforme des décès lors de l'établissement de la table de mortalité, voire dans les calculs qui la précèdent pour extraire des données initiales des indices de mortalité par âge. Certaines procédures obligent à poser en plus l'hypothèse de répartition uniforme des naissances. Pourquoi cet ostracisme pour cette dernière hypothèse, opposé à un accueil à bras largement ouverts pour l'autre, alors que toutes deux sont nécessaires pour justifier le résultat de certaines procédures ? Il faut reconnaitre que la présence de l'hypothèse sur les naissances est plus discrète dans le sens où beaucoup de calculs parmi les plus habituels se contenteront de la seule hypothèse de répartition sur les décès.

---

23 C'est par facilité que la démonstration a été faite en partant d'un exemple concernant un parallélogramme entre âges exacts, soit un exemple de table entre âges exacts (*cf.* figure 9). Cela ne remet pas en cause le commentaire qui est fait à propos des tables entre âges révolus. En effet, dans la figure 9, chaque petit triangle du 1er triangle renferme 1 décès ; il y a donc bien répartition uniforme des décès entre 0 exact et 0 révolu. Et donc, la conclusion tirée de cet exemple s'applique aussi à un 1er triangle dans une table entre âges révolus. Pour preuve, *cf.* annexe 5.

24 À propos de ces effectifs aux âges exacts, la logique voudrait que l'on aborde aussi la question de la date d'anniversaire moyenne, soit la question correspondant exactement à celle de l'âge exact moyen d'un effectif à un âge révolu. Nous n'aborderons pas ce point ici. Si le sujet intéresse, *cf.* Calot et Caselli (1991) ou Vandeschrick (2005), pp. 227-229.

Troisième point : *âge moyen de l'effectif à 0 an révolu, remise en cause d'une idée reçue*

Que vaut l'âge exact moyen des individus intégrant un effectif à 0 an révolu sous hypothèse de répartition uniforme des décès et des naissances ? Habituellement, on considère qu'il s'agit de 0,5 an exact. Cette solution est correcte pour autant que le risque de mourir entre 0 et 1 an exact soit… nul. Si ce n'est pas le cas, l'âge moyen est d'autant plus bas que le risque de mourir augmente, avec une valeur minimale de 1/3 d'année en cas de risque de mourir unitaire.

Quatrième point : *modestie des effets versus importance méthodologique*

Ce texte revisite différents aspects du calcul en matière de mortalité. Très souvent, les correctifs qu'il propose sont de piètre ampleur sur les résultats, voire carrément de l'ordre du négligeable. Exemples :

- Selon les conditions de mortalité dans un pays comme la Belgique, le choix de l'âge moyen au décès dans le triangle entre 0 an exact et 0 an révolu entraine au maximum 2 jours de différence dans l'estimation de l'espérance de vie, soit moins d'un millième de pourcent de cette espérance (calculs non présentés dans ce texte).
- Il faut atteindre des risques de mourir supérieur à 0,1 pour que l'âge moyen des individus intégrant l'effectif à 0 an révolu soit inférieur à 0,49 contre le 0,50 retenu habituellement (point 2.2).
- En supposant une densité des décès 2 fois plus forte à 0 an qu'à 1 an, l'âge moyen au décès dans le 1er triangle est de 0,3 au lieu du tiers correspondant à la répartition uniforme (point 1.4). On pourrait trouver cet écart important (1/3 contre 0,3). Toutefois, appliqué à d'autres âges que le 1er, où la répartition uniforme, sans être strictement de mise, est plus largement acceptable, le raisonnement suivi conduirait à la conclusion que l'hypothèse habituelle suffit amplement.

Il ne s'agit pourtant point là d'arguments pour faire l'économie des réflexions qui ont été menées dans ce texte. Tout d'abord, si une information est disponible sur tel ou tel aspect, mieux vaut être en mesure de l'inclure dans le calcul. Ensuite, selon les conditions de mortalité, tel ou tel correctif, n'ayant que de médiocres effets avec la mortalité actuelle d'un pays comme la Belgique, pourrait voir ses conséquences amplifiées dans d'autres circonstances :

- analyse de la mortalité dans des populations à relativement forte mortalité ;
- analyse d'autres phénomènes que la mortalité, comme la mobilité spatiale ;
- analyse portant sur des petites populations ;
- analyse de la mortalité au départ de groupes quinquennaux d'âge ;
- application des méthodes de la démographie en dehors du champ habituel de la démographie, par exemple dans des contextes spécifiques aussi bien en ce qui concerne les populations visées que les phénomènes à analyser ;
- etc.

Un démographe se doit de connaitre au mieux les limites des méthodes qu'il utilise. Via une connaissance approfondie des techniques, le démographe en formation, futurs professionnel dans son domaine, sera en mesure non seulement de décider de leur utilisation ou non face à des données, mais aussi de mesurer la fiabilité des résultats qu'il propose ou, pour le dire autrement, en connaitre les limites. C'est bien l'objectif de ce texte.

Cinquième point : *de l'utilité du calcul intégral*

Même si les temps changent (*cf.* dernier paragraphe de ce 5^{e} point), généralement, dans la phase d'initiation, la formation classique du démographe se base sur le calcul faisant intervenir des données agrégées dans une optique temporelle discrète. Par exemple, le risque de mourir entre 0 et 1 an révolu durant une année s'obtient en divisant le nombre de décès observés dans le parallélogramme durant l'année de référence par l'effectif à 0 an révolu au 1er janvier de cette année (avec correction pour tenir compte des perturbations). À notre avis, c'est à juste titre que l'on procède de la sorte. En effet, il s'agit là d'un accès très aisé aux grandeurs que le démographe est appelé à manipuler. Toutefois, il ne s'agit pas d'une raison pour en rester là et ne pas s'engager dans d'autres voies, dont le calcul intégral. Ce dernier nous a permis, notamment :

- de nous affranchir à peu de frais de l'hypothèse de répartition uniforme des décès pour estimer, par exemple, l'âge moyen au décès dans un 1er triangle (point 1.4) ;
- de mettre en place une procédure aisée pour calculer l'âge moyen des individus intégrant un effectif révolu (point 2.3).

Ces opérations auraient été, si pas impossibles, du moins très fastidieuses à exécuter et, surtout, à généraliser via des calculs discrets. Bref, sans rien nier de l'intérêt de l'analyse démographique classique (données agrégées et calculs discrets), il serait dommage de se priver d'outils comme le calcul intégral pour explorer certaines questions que soulève l'emploi des méthodes de l'analyse démographique. Nous pensons notamment à la transformation d'une table de mortalité entre âges exacts pour obtenir une table entre âges révolus : si la fonction de répartition des décès est connue entre les âges exacts ronds, il sera facile de produire des nombres de décès par triangle, voie royale pour produire une table dans n'importe quel format.

Dans la littérature démographique, il n'est pas rare de rencontrer des ouvrages où une part des explications est proposée via le calcul infinitésimal, qui en définitive correspond mieux au fonctionnement du système démographique[25] : une population évolue en continu, ce qui est mieux rendu par ce type de calcul. Dans ce texte, tel n'était pas l'objectif, même s'il est louable. Ce texte part des méthodes de la démographie classique et montre, notamment, en quoi la compréhension de certains points précis peut être améliorée et approfondie grâce au recours à quelques calculs infinitésimaux, mais sans chercher à établir un ensemble d'équations infinitésimales pour décrire le fonctionnement du système démographique dans son ensemble.

Dans un autre registre, ajoutons que, en démographie, l'émergence, voire la proéminence, de méthodes d'analyse reposant sur des données, non plus agrégées comme dans l'analyse classique, mais bien individuelles, n'annule pas ce qui a été proposé dans ce texte. Loin d'être antagonistes, ces deux types d'approche (via des données individuelles ou agrégées), se devraient d'être complémentaires, notamment pour aller plus en profondeur dans les explications. Toujours est-il que nous pensons que l'analyse classique est un passage obligé, dans l'optique de bien comprendre la nature des grandeurs en jeu et de vérifier que des analyses plus sophistiquées (comme une régression logistique en temps discret) sont adaptées ou non aux circonstances. Ce point ne sera pas davantage développé ici.

[25] *Cf.* notamment Preston *et al.* (2001) ou Brouard (1989). Dans Calot and Franco (2002), l'établissement des formules s'initie par référence au calcul infinitésimal, mais ce dernier laisse la place à un calcul dans une forme plus classique (mais sophistiquée) avant détermination de la forme définitive à appliquer aux données (*cf.* p. 43 et 46, par exemple).

Bibliographie

Bailly P. (1999), *Statistique descriptive*, Grenoble, Presses Universitaires de Grenoble.

Boyarski A. (1945), *Kurs demograficheskoy statitiki*, Gosplanizdat.

Brouard N. (1989), *Mouvements et modèles de population*, Yaoundé, IFORD.

Calot G. et Caselli G. (1991), « Détermination d'une table de mortalité : la conversion des taux en quotients », *Population*, 6, pp. 1.441-1.490.

Calot G. and Franco A. (2002), « The Construction of Life Tables », *in* G. Wunsch, M. Mouchart and J. Duchêne (Editors), *The Life Table. Modelling Survival and Death*, Dordrecht/Boston/London, Kluwer Academic Publishers, pp. 33-78.

Duchesne D., Tully P., Thomas B. et Bourbeau R. (2002), *Tables de mortalité. Canada, provinces et territoires 1995-1997*, Statistique Canada/Division de la statistique de la santé, xix + 70 p. (http://www.prdh.umontreal.ca/BDLC/data/pdfs/84-537-XIF.pdf)

Grais B. (1994), *Statistique descriptive*, Paris, Dunod.

Kucera T. (1986), « Quotients, Rates and Disturbing Events in Demography », *Acta Universitatis Carolinae*, 2, pp. 87-98.

Leridon H. et Toulemon L. (1997), *Démographie. Approche statistique et dynamique des population*, Paris, Economica.

Pressat R. (1973), *L'analyse démographique*, Paris, Presses universitaires de France.

Pressat R. (Sous la direction de) (1985), *Manuel d'analyse de la mortalité*, Gap, Organisation mondiale de la Santé/Institut national d'Études démographiques.

Preston S., Heuveline P. et Guillot M. (2005), *Demography. Measuring and Modelling Population Processes*, Oxford, Blackwell.

Régnier-Loilier A. et Rohrbasser J.-M. (2011), « Y a-t-il une saison pour faire des enfants ? », *Population et Sociétés*, 474, Janvier.

Vallin J. et Caselli G. (2001, a), « Chapitre 11. La table de mortalité d'une génération », *in* G. Caselli, J. Vallin et G. Wunsch, *Démographie : analyse et synthèse. Tome I. La dynamique des populations*, Paris, Éditions de l'Institut National d'Études Démographiques (INED), pp. 165-212.

Vallin J. et Caselli G. (2001, b), « Chapitre 14. L'artifice de la cohorte fictive », *in* G. Caselli, J. Vallin et G. Wunsch, *Démographie : analyse et synthèse. Tome I. La dynamique des populations*, Paris, Éditions de l'Institut National d'Études Démographiques (INED), pp. 271-327.

Vandeschrick C. (1992), « Le diagramme de Lexis revisité », *in Population*, 5, pp. 1241-1262.

Vandeschrick C. (2001), « The Lexis Diagram, a Misnomer », *in Demographic Research*, 4, pp 97 – 124 (www.demographic-research.org).

Vandeschrick C. (2004), *Analyse démographique. Troisième édition revue et corrigée*, Louvain-la-Neuve et Paris, Bruylant-Academia/L'Harmattan, Collection Population et développement, 1, 215 p.

Vandeschrick C. (2005), *Diagramme de Lexis et cohortes : du temporel au non-temporel*, Thèse présentée en vue de l'obtention du titre de Docteur en démographie, Louvain-la-Neuve, Université catholique de Louvain. Département des sciences de la population et du développement, 316 p. (http://edoc.bib.ucl.ac.be:81/ETD-db/collection/available/BelnUcetd-04152005-223721/).

Wilmoth J. R., Andreev K., Jdanov D., and Glei D.A. with the assistance of Boe C., Bubenheim M., Philipov D., Shkolnikov V., Vachon P. (Last Revised: May 31, 2007 (Version 5)), Methods Protocol for the Human Mortality Database. (http://www.mortality.org/Public/Docs/MethodsProtocol.pdf).

Wunsch G. (2002), « 1. The Life Table: a Demographic Overview », *in* Wunsch G., Mouchart M. et Duchêne J., *The Life Table. Modelling Survival and Death*, Dordrecht, Kluwer Academic Publishers, European Studies of Population, 11, pp. 13-31.

Wunsch G. (1999), *Introduction to Demographic Analysis*, Louvain-la-Neuve, DUC Diffusion universitaire/CIACO.

Wunsch G. and Termote M. (1978), *Introduction to Demographic Analysis*, New York and London, Plenum Press.

Annexes

Annexe 1. Calcul de l'âge moyen par sommation quand n tend vers l'infini

Le calcul de l'âge moyen au décès dans le 1er triangle peut s'exprimer comme suit :

$$\bar{y} = \frac{\sum_{x=1}^{n}[(2n+1)-(2x)]^2}{(4n)*\sum_{x=1}^{n}[(2n+1)-(2x)]}$$, avec n = le nombre de sous-cohortes.

Vu que le calcul porte sur la limite de cette expression pour n tendant vers l'infini, n+1 est assimilable à n. Dès lors, l'expression devient : $\bar{y} = \frac{\sum_{x=1}^{n}[(2n)-(2x)]^2}{(4n)*\sum_{x=1}^{n}[(2n)-(2x)]}$.

α. Dénominateur de l'âge moyen au décès dans le 1er triangle

Point de départ : $(4n)*\sum_{x=1}^{n}[(2n)-(2x)]$

- La sommation :

  $\sum_{x=1}^{n}[(2n)-(2x)] = (n(2n)) - \sum_{x=1}^{n}[(2x)] = (2n^2) - 2*\left(\frac{n^2}{2}\right) = n^2$,

  où $\sum_{x=1}^{n}[(2x)] = 2*\sum_{x=1}^{n}x = 2*\left[\frac{(n+1)*n}{2}\right] = 2*\left[\frac{n^2}{2}\right] = n^2$

  vu que n tend vers l'infini, n+1 peut être assimilé à n.

  Finalement : $\sum_{x=1}^{n}[(2n)-(2x)] = (2n^2) - n^2 = n^2$.
- Si on réinjecte 4n, le dénominateur vaut : $4n^3$.

β. Numérateur de l'âge moyen au décès dans le 1er triangle :

Point de départ : $\sum_{x=1}^{n}[(2n)-(2x)]^2 = \sum_{x=1}^{n}\left[(2n)^2 - (2*2x*(2n)) + (4x^2)\right]$

Sachant que :

- $\sum_{x=1}^{n}x^2 = \frac{n(2n+1)(n+1)}{6}$ et vu que n tend vers l'infini, n+1 peut être assimilé à n et donc :

  $\sum_{x=1}^{n}x^2 = \frac{n(2n+1)(n+1)}{6} = \frac{n(2n)(n)}{6} = \frac{2n^3}{6} = \frac{n^3}{3}$.
- $\sum_{x=1}^{n}x = \frac{n^2}{2}$ (*cf. supra*).

Dès lors :

$$\sum_{x=1}^{n}[((2n))-(2x)]^2 = \sum_{x=1}^{n}\left[(2n)^2 - (2*2x*(2n)) + (4x^2)\right] = \sum_{x=1}^{n}\left[(2n)^2 - (8nx) + (4x^2)\right]$$

$$= \left[\sum_{x=1}^{n}(2n)^2\right] - \left[8n*\sum_{x=1}^{n}x\right] + \left[4*\sum_{x=1}^{n}x^2\right] = 4n^3 - 8\frac{n^3}{2} + 4\frac{n^3}{3} = \frac{4n^3}{3}.$$

γ. Calcul de l'âge moyen au décès dans le 1er triangle

Finalement[26] : $\bar{y} = \frac{\sum_{x=1}^{n}[(2n)-(2x)]^2}{(4n)*\sum_{x=1}^{n}[(2n)-(2x)]} = \frac{4n^3/3}{4n^3} = \frac{1}{3}$, cqfd.

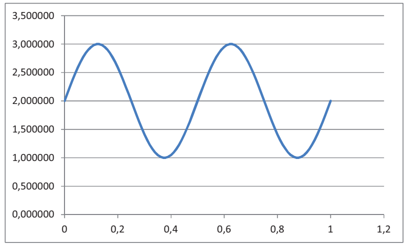
Quand n tend vers l'infini, l'âge moyen tend donc bien vers 1/3.

Annexe 2. Saisonnalités cycliques

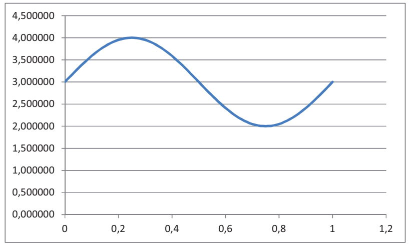
L'objectif de cette annexe est de montrer un exemple de calcul avec des fonctions cycliques pour la saisonnalité des naissances et des décès combinées avec une répartition par âge proche de l'observé. Les fonctions cycliques pour les naissances et les décès permettent certes de représenter sans doute déjà mieux (même si très imparfaitement) une variation saisonnière, avec notamment une continuité d'une année à l'autre. Toutefois, la complexité des calculs est telle que, dans la pratique, ils en deviennent ingérables, comme nous l'allons montrer ci-dessous. Notons que cette annexe va nettement au-delà de l'objectif de ce texte : donner à l'apprenti(e) démographe des outils simples pour mieux comprendre la table de mortalité, ses qualités et défauts, ses limites et ses forces.

Figure A.2. Fonctions cycliques pour les saisonnalités

a. Saisonnalité des naissances



b. Saisonnalité des décès



Supposons :

- la variation en fonction de l'âge modélisé par la fonction suivante :

  $Dens_y = \left(\left(\frac{((2-2y)^{40})*(10-10y)}{2^{37}}\right)*1{,}368750\right) + 0{,}35.$
- la saisonnalité des naissances représentée par une fonction avec deux cycles sur l'année :

  $SaisN_z = 2 + (\sin(4*\pi*z)) = 2 + (\sin(4*\pi*(x-y)))$,

[26] Merci à Monsieur J.-Ph. Conard pour son aide précieuse pour cette démonstration et aussi pour d'autres calculs présentés dans ce texte.

avec, pour rappel, dans le 1er triangle, le moment de naissance (z) = le moment (x) – l'âge (y). (*Cf.* figure A.2.a.)

- la saisonnalité des décès représentée par un seul cycle sur l'année : $\text{SaisD}_x = 3+(\sin(2*\pi*x))$. (*Cf.* figure A.2.b.)

Dans ces conditions, la fonction à intégrer pour obtenir le nombre de décès (FD) s'exprime :

$$FD = (2+(\sin(4*\pi*(x-y))))*(3+(\sin(2*\pi*x)))*\left(\left(\frac{((2-2y)^{40})*(10-10y)}{2^{37}}\right)*1{,}368750\right)+0{,}35.$$

Le nombre de décès (ND) s'obtient en intégrant cette fonction sur x et y :

$$NDT_1 = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}\left((2+(\sin(4*\pi*(x-y))))*(3+(\sin(2*\pi*x)))*\left(\left(\frac{((2-2y)^{40})*(10-10y)}{2^{37}}\right)*1{,}368750\right)+0{,}35\right)*dy\right]*dx$$

À ce jour, les calculateurs utilisés n'ont pas donné de réponse à ce calcul. Par contre, des calculs ont pu être réalisés en simplifiant la situation. Ainsi, en cas de saisonnalité des naissances et des décès et avec une variation en fonction de l'âge nulle (soit 1 à tous les âges), l'âge moyen au décès vaut :

- $NDT_1 = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}((2+(\sin(4*\pi*(x-y))))*(3+(\sin(2*\pi*x))))*dy\right]*dx = 3-\frac{1}{4\pi} = 2{,}92042..$
- $ACDT_1 = \int_{x=0}^{x=1}\left[\int_{y=0}^{y=x}((2+(\sin(4*\pi*(x-y))))*(3+(\sin(2*\pi*x))))*y*dy\right]*dx = -\frac{1+\pi-8\pi^2}{8\pi^2} = 0{,}947546.$
- $\bar{y} = \frac{ACDT_1}{NDT_1} = -\frac{0{,}947546}{2{,}92042} = 0{,}32446.$

Cet exercice sur l'influence des saisonnalités n'a pas été poussé plus avant ici.

Annexe 3. Âge médian via la surface sous l'horizontale de l'âge médian

La répartition de la surface du triangle de la figure 8 reste d'application. Dès lors, en cas de répartition uniforme et unitaire, le calcul de l'âge médian en prenant en compte la surface sous la ligne à l'âge médian s'écrit comme suit :

$$(1)+(2) = 0{,}25 \Rightarrow \left(\int_{x=0}^{x=m}\int_{y=0}^{y=x} dy*dx\right)+\left(\int_{x=m}^{x=1}\int_{y=0}^{y=m} dy*dx\right) = 0{,}25 \Rightarrow \left(\int_{x=0}^{x=m}\left[y\right]_0^x*dx\right)+\left(\int_{x=m}^{x=1}\left[y\right]_0^m*dx\right) = 0{,}25$$

$$\Rightarrow \left(\int_{x=0}^{x=m} x*dx\right) + \left(\int_{x=m}^{x=1} m*dx\right) = 0{,}25 \qquad \Rightarrow \left(\left[\frac{x^2}{2}\right]_0^m\right) + \left(m*\left[x\right]_m^1\right) = 0{,}25$$

$$\Rightarrow \left(\frac{m^2}{2}\right) + \left(m-m^2\right) = 0{,}25 \qquad \Rightarrow m-\left(\frac{m^2}{2}\right)-0{,}25 = 0.$$

Sans surprise, on retombe sur le résultat obtenu en suivant la méthode de Boyarski ou celui obtenu en faisant porter le calcul sur le triangle au-dessus de l'horizontale de l'âge médian, qui vaut, pour rappel, 0,29289. Pour suivre le calcul en cas de répartition non uniforme (densité des décès deux fois plus forte à 0 an qu'à 1 an) :

$$(1)+(2) = \frac{5}{12} \Rightarrow \left(\int_{x=0}^{x=m}\int_{y=0}^{y=x}(2-y)*dy*dx\right) + \left(\int_{x=m}^{x=1}\int_{y=0}^{y=m}(2-y)*dy*dx\right) = \frac{5}{12}$$

$$\Rightarrow \left(\int_{x=0}^{x=m}\left(2x-\frac{x^2}{2}\right)*dx\right) + \left(\int_{x=m}^{x=1}\left(2m-\frac{m^2}{2}\right)*dx\right) = \frac{5}{12}$$

$$\Rightarrow \left(2\frac{m^2}{2}-\frac{m^3}{6}\right) + \left(\left(2m-\frac{m^2}{2}\right)-2m^2+\frac{m^3}{2}\right) = \frac{5}{12}$$

$$\Rightarrow m^2-\frac{m^3}{6}+2m-\frac{m^2}{2}-2m^2+\frac{m^3}{2}-\frac{5}{12} = 0$$

$$\Rightarrow \frac{1}{12}\left(24m-18m^2+4m^3-5\right) = 0.$$

Par itération, on obtient la valeur suivante : 0,25398 (ou avec plus de décimales : 0,25398334941409), soit le résultat obtenu au point 1.7.

Annexe 4. Âge moyen de l'effectif à 0 an révolu au départ du quotient entre 0 exact et 0 révolu

Supposons connu le quotient entre 0 an exact et 0 an révolu, soit q_b (avec « b » = « birth »). Il s agit du quotient couvrant le 1er triangle délimité par 0 an exact et 0 an révolu. Par ailleurs, supposons la répartition uniforme des décès dans ce 1er triangle.

Pour les besoins spécifiques du calcul, supposons que la répartition uniforme des décès s'applique aussi au 2e triangle et que la densité soit identique à celle du 1er triangle. Dans ces conditions, la valeur du quotient entre 0 et 1 an exact (soit le risque q) vaut le double du q_b : $q = 2*q_b$.

Dès lors, la méthode suivie dans le corps du texte peut être reprise en remplaçant q par $2*q_b$.

$$l_{\underline{0}} = \int_{y=0}^{y=1}(1-(2*q_b*y))*dy = \left[y-\frac{2*q_b*y^2}{2}\right]_0^1 = 1-q_b.$$

Ce résultat est logique. En effet, l'effectif à 0 an révolu est, par définition, le nombre de survivants à 0 an révolu. La proportion de ceux qui meurent dans le 1er triangle est le risque q_b et, donc, la proportion de ceux qui survivent est le complément à l'unité de ce risque de mourir. Étant donné, pour rappel, que nous avons choisi une densité uniforme et unitaire des naissances, ce complément à l'unité du risque de mourir entre 0 exact et 0 révolu quantifie le nombre de survivants à 0 an révolu. Par ailleurs, vu qu'en utilisant le quotient entre 0 et 1 an exact, la réponse était de 1 diminué de la moitié du quotient et vu que ce quotient entre 0 et 1 an exact vaut deux fois le quotient du 1er triangle, il est logique de retrouver le résultat mentionné au-dessus de ce paragraphe.

L'âge total des individus de l'effectif à 0 an révolu se calcule comme suit :

$$\text{Âge tot} = \int_{y=0}^{y=1}(1-(2*q_b*y))*y*dy = \int_{y=0}^{y=1}\left(y-\left(2*q_b*y^2\right)\right)*dy = \left[\frac{y^2}{2}-\frac{2*q_b*y3}{3}\right]_0^1 = \frac{1}{2}-\frac{2}{3}q_b.$$

Finalement, l'âge moyen de l'effectif à 0 an révolu ($\bar{y}$) s'obtient par le rapport entre l'âge cumulé pour l'ensemble de cet effectif et cet effectif lui-même :

$$\bar{y} = \frac{\text{Âge tot}}{l_0} = \frac{\int_{y=0}^{y=1} \left(1-\left(2*q_b*y\right)\right)*y*dy}{\int_{y=0}^{y=1} \left(1-\left(2*q_b*y\right)\right)*dy} = \frac{\frac{1}{2}-\frac{2}{3}q_b}{1-q_b}.$$

Supposons : $q_b = 0{,}25$

Selon la formule de cette annexe, l'âge moyen vaut : $\bar{y} = \frac{\frac{1}{2}-\left(\frac{2}{3}*\frac{1}{4}\right)}{1-\frac{1}{4}} = \frac{\frac{1}{2}-\frac{2}{12}}{\frac{3}{4}} = \frac{4/12}{3/4} = \frac{1/3}{3/4} = \frac{4}{9}.$

Vérification en reprenant la formule du corps du texte avec $q_{0\to1} = 0{,}5$:

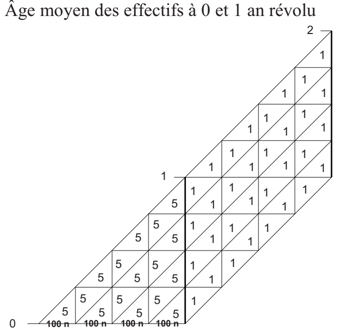
$$\bar{y} = \frac{\int_{y=0}^{y=1} \left(y-q*y^2\right)*dy}{\int_{y=0}^{y=1} \left(1-q*y\right)*dy} = \frac{\frac{1}{2}-\frac{1}{3}q}{1-\frac{1}{2}q} = \frac{\frac{1}{2}-\left(\frac{1}{3}*0{,}5\right)}{1-\left(\frac{1}{2}*0{,}5\right)} = \frac{2/6}{3/4} = \frac{8}{18} = \frac{4}{9}.$$

Notons enfin que si $q_b = 1$, le nombre de survivants à 0 an révolu est nul, ce qui rend le calcul de l'âge moyen sans intérêt (car il n'y a aucun survivant à 0 an révolu) et surtout impossible (car le dénominateur est nul). Par ailleurs, comment assurer la répartition uniforme alors qu'il faut que l'ensemble des sous-cohortes décède avant d'atteindre l'effectif à 0 an révolu, c'est-à-dire dans un délai soit d'un année pour celle née au temps 0, soit d'une fraction de seconde pour celle née juste avant le temps 1 ? En fait cette situation prévaut dès que le quotient du 1er triangle atteint ou dépasse 0,5.

Annexe 5. Calcul discret pour l'évolution de l'écart par rapport à x + 0,5

Dans cette annexe, nous allons étudier l'âge des effectifs révolus successifs et, plus spécifiquement, la différence entre l'hypothèse habituelle (l'âge révolu + 0,5) et ce que donnent des calculs effectifs à ce sujet.

Figure A.5.
Âge moyen des effectifs à 0 et 1 an révolu



Le raisonnement sera fait sur un exemple simplifié, partiellement représenté sur la figure A.5. Soit une cohorte subdivisée en 4 sous-cohortes. Chacune de celles-ci compte 100 naissances. Entre 0 an exact et 0 an révolu, chaque petit triangle compte 5 décès. Entre 0 an révolu et 1 an révolu, il s'agira d'un seul décès (malgré le caractère peu réaliste de cette hypothèse par comparaison à la situation dans le 1er triangle). Le tableau A.1 permet d'entrevoir la suite de l'histoire : entre 1 et 2 ans révolus, il y a un demi-décès par petit triangle (les sous-cohortes perdent 4 unités entre 1 et 2 ans révolus) et 2, entre 2 et 3 ans révolus (les sous-cohortes perdent 16 unités entre 2 et 3 ans révolus).

Le tableau A.5.A montre l'évolution des 4 sous-cohortes à différents âges. Ainsi, à 0 an exact, chacune compte 100 individus. En fin de 1er triangle soit à 0 an révolu), la 1re sous-cohorte compte 65 individus (soit les 100 initiaux moins les 7*5 décès) ; la 4e en compte encore 95 individus, ne déplorant que 5 décès (un seul petit triangle). Entre 0 an révolu et 1 an révolu, l'effectif de toutes les sous-cohortes a perdu 8 individus supplémentaires. Entre 1 et 2 ans révolus (2 et 3), ce sont 4 (16) individus qui disparaissent par sous-cohorte.

Le tableau A.5.C montre le calcul de l'âge moyen des effectifs révolu. Par exemple le 46,875 dans la colonne « 0 an révolu » pour la 2e sous-cohorte a été obtenu en multipliant 75 par 0,625, soit le nombre d'individus dans cette sous-cohorte dans l'effectif à 0 an révolu multiplié par la moyenne entre 0,75 et 0,5 an, les âges extrêmes des individus de cette sous-cohorte (tout en étant bien conscient que cette moyenne n'est qu'une approximation). Autre exemple : le 192,375 dans la colonne « 3 ans révolus » et sur la ligne de la 3e sous-cohorte a été obtenu en multipliant 57 par 3,375, soit le nombre d'individus dans cette sous-cohorte dans l'effectif à 3 ans révolus multiplié par la moyenne entre 3,25 et 3,5 an. Dans une colonne, la division du total du sous-tableau C par l'effectif de l'âge révolu correspondant (total du sous-tableau A) permet d'obtenir l'âge moyen de l'effectif révolu. Ainsi, à 2 ans révolus, l'âge moyen est de 2,454, soit 667,5/272.

Tableau A.5. Âge moyen aux effectifs révolus

A. Effectif de survivants à l'âge révolu par sous-cohorte					
Sous-cohorte	0 exact	0 an révolu	1 an révolu	2 ans révolus	3 ans révolus
1	100	65	57	53	37
2	100	75	67	63	47
3	100	85	77	73	57
4	100	95	87	83	67
Total	**400**	**320**	**288**	**272**	**208**
B. Part de la sous-cohorte dans l'effectif à l'âge révolu					
1	25,00%	20,31%	19,79%	19,49%	17,79%
2	25,00%	23,44%	23,26%	23,16%	22,60%
3	25,00%	26,56%	26,74%	26,84%	27,40%
4	25,00%	29,69%	30,21%	30,51%	32,21%
Total	**100,00%**	**100,00%**	**100,00%**	**100,00%**	**100,00%**
C. Effectif multiplié par l'âge central pour la sous-cohorte					
1	0	56,875	106,875	152,375	143,375
2	0	46,875	108,875	165,375	170,375
3	0	31,875	105,875	173,375	192,375
4	0	11,875	97,875	176,375	209,375
Total	**0,000**	**147,500**	**419,500**	**667,500**	**715,500**
Âge moyen	**0,000**	**0,461**	**1,457**	**2,454**	**3,440**

L'âge moyen des effectifs révolus s'écarte de plus en plus de l'hypothèse habituelle (l'âge révolu + 0,5 an). Cela provient du fait que la 4e sous-cohorte voit son effectif relatif se renforcer d'un âge au suivant (sous-tableau B). Ainsi, elle représente 25 % de l'effectif à 0 an exact, mais 32,21 % à 3 ans révolus. Autrement dit, dans le total, la sous-cohorte qui présente l'âge moyen le plus bas voit augmenter sa part ; il en va de même, mais dans une mesure moindre pour la 3e sous-cohorte. Pour les 1re et 2e sous-cohortes, la situation est inversée. Dès lors, tout naturellement, l'écart avec l'hypothèse habituelle augmente avec l'âge.

Comité d'édition
Philippe Bocquier, Sandra Brée, Thierry Eggerickx, Paul Majérus, Ester Rizzi, Jean-Paul Sanderson, Isabelle Theys, Christophe Vandeschrick.

Responsable
Thierry Eggerickx

Référence de ce document
Vandeschrick C. (2015), *Répartition des décès et table de mortalité : à boire et à manger !, Démographie et sociétés, Document de Travail 5*, Centre de recherche en démographie, Louvain-la-Neuve, 43 p.

Contact
Isabelle Theys
Centre de recherche en démographie
Université catholique de Louvain
1 Place Montesquieu bte L2.08.03
1348 Louvain-la-Neuve, Belgique
Tél. 32 10 47 29 51 Fax 32 10 47 29 52
isabelle.theys@uclouvain.be

juin 2015
www.uclouvain.be/demo