

# DIAGRAMMES DE LEXIS ET COHORTES : LA TAILLE OU LE POIDS EN REMPLACEMENT DU TEMPS

Christophe VANDESCHRIK  
Institut de Démographie, UCL

---

## Introduction

Depuis longtemps, si pas depuis toujours, l'analyse démographique entretient une relation étroite avec le temps et d'autres notions dérivées, comme l'âge et la génération<sup>1</sup>. En effet, les phénomènes démographiques – natalité, mortalité et mobilité spatiale, pour ne citer que les plus couramment repris – semblent bien suivre de puissantes logiques temporelles, comme le montrent, par exemple, des graphiques illustrant la variation d'année en année ou de génération en génération des quotients de mortalité par âge.

La nécessité d'étudier les phénomènes démographiques par âge, période et cohorte s'est imposée à tout le moins comme un passage obligé, une première étape en quelque sorte exploratoire, sans devenir toutefois une fin en soi ; en effet, ce type d'analyse doit se compléter par d'autres pour approfondir l'explication et la compréhension des mécanismes en jeu. Cette première étape du processus classique d'analyse démographique sera dénommée dans la suite par le terme « cohortisation », tant cette notion de

---

1 Ce texte s'inspire assez largement du texte présenté le 24 février 2003 pour l'épreuve de confirmation d'une thèse portant sur ce sujet, ainsi que sur une présentation proposée lors des Midis de la recherche de l'Institut de Démographie en sa séance du 22 mai 2003. Par ailleurs, ce texte a été rédigé *a priori* pour un public de démographes, déjà au fait des questions et principes d'analyse d'application dans la discipline.

cohorte y tient un rôle important, voire central, dans l'élaboration de la mesure des phénomènes.

Au dix-neuvième siècle, des démographes – allemands pour la plupart – ont progressivement mis au point un outil très précieux dans cette perspective d'analyse par âge, période et cohorte : le diagramme dit – sans doute erronément – « de Lexis ». Ce diagramme vise simplement à représenter sur une figure en plan les événements démographiques en fonction du triple critère du moment, de l'âge et du moment de naissance. En fait, il s'agit d'un « diagramme en trajectoires »<sup>2</sup> : la vie de chaque individu s'y matérialise par une trajectoire, ligne oblique inclinée à 45° et représentant l'évolution de son âge en fonction du temps ; cette trajectoire peut servir de support pour localiser les événements affectant l'individu en cause.

Grâce à cette figure, il est aisé de montrer comment les données s'organisent et de définir, tout en les justifiant, des méthodes correctes pour le calcul de différents indices : par exemple, dans le cas de la mortalité, il devient aisé de rapporter un nombre de décès à l'effectif adéquat de survivants pour obtenir un quotient de mortalité.

C'est notamment sous l'impulsion donnée par les possibilités offertes par le diagramme de Lexis, que la démographie a pu progresser, par exemple, en matière de mesures des phénomènes débarrassées des effets de taille et de structure ou en matière de distinction des effets d'âge, de période et de cohorte, si tant est que ce soit possible.

Si cette procédure d'analyse en cohortisation porte la marque de fabrique « Démographie » et trouve tout naturellement à s'employer en cas de variables temporelles, est-elle pour autant à réserver à ce seul champ d'application ? Le *corpus* analytique de la démographie pourrait-il s'exporter avec fruit en cas de variables non temporelles ?

Ce document voudrait envisager les points suivants :

- la possibilité de constituer des figures présentant toutes les caractéristiques et qualités d'un diagramme de Lexis au départ de variables autres que celles qui sont habituellement utilisées et singulièrement au départ de variables non temporelles ;
- en conséquence, la possibilité de recourir aux méthodes de l'analyse démographique classique, à savoir la cohortisation, pour étudier certains phénomènes sous un angle non temporel ;
- une application de la cohortisation dans ce type de circonstances devrait permettre de produire des mesures potentiellement intéressantes, car débarrassées, à l'image de ce que réalise l'analyse démographique, de certains effets gênants, notamment en cas de comparaison.

---

2 On pourrait aussi dire « diagramme de trajectoires », mais comme la finalité de cette construction n'est pas la représentation des trajectoires (qui, le plus souvent, sont même retirées du dessin), nous préférons l'expression « diagramme en trajectoires ».

La réflexion que nous nous proposons de mener ici abordera les questions évoquées d'abord et essentiellement d'un point de vue théorique : sans chercher à en évaluer initialement l'éventuelle portée pratique, on cherchera prioritairement à établir la possibilité théorique de cohortiser un phénomène selon des variables non temporelles.

Cet essai d'application de la cohortisation n'ira pas sans problème lié aux caractéristiques des variables en jeu. En effet, le temps présente un déroulement immuable, irréversible, linéaire et continu, que rien ne peut arrêter, ni même simplement perturber. Vu ces qualités, le temps – et les variables dérivées, comme l'âge – est en fait une variable relevant d'une catégorie très particulière. Les variables qui pourraient éventuellement le remplacer ne présentent pas nécessairement – toutes – ces caractéristiques (que l'on pense simplement ici à la non-réversibilité ou à la continuité du déroulement du temps). Il s'agira donc aussi d'apprécier les conséquences de ces différences sur les mesures issues de la cohortisation des phénomènes selon des variables non temporelles.

## 1. Analyse démographique, cohorte et diagramme de Lexis

### 1.1. Objectifs et spécificités de l'analyse démographique

Dans cette partie, on dégagera les spécificités de l'analyse démographique<sup>3</sup> dans le but final de la transférer dans des circonstances qui ne lui sont pas habituelles. On peut considérer que l'objectif principal de l'analyse démographique classique consiste à caractériser l'activité d'un phénomène dans une population via une mesure à l'état pur<sup>4</sup> de son intensité et de son

---

3 Précisons de suite, qu'à côté de la démographie, peuvent se ranger d'autres utilisateurs des méthodes en cause, comme les sciences actuarielles ou l'épidémiologie. Il nous importe peu ici de savoir si ces dernières squattent les méthodes de la démographie ou si au contraire, la démographie a phagocyté des méthodes dont la mise au point aurait été initiée par d'autres.

4 Brièvement, qu'est-ce qu'une mesure à l'état pur ? Par exemple, en cas d'étude portant sur la mortalité, outre ce phénomène, l'émigration retire aussi des individus de la population concernée. L'émigration, phénomène perturbateur, entrave l'observation du phénomène étudié. Dans ces circonstances, les démographes utilisent des méthodes adaptées pour obtenir des indices de mortalité comme si ce phénomène était le seul à frapper la population observée, en éliminant donc le rôle joué par l'émigration dans la variation d'effectif. Ces procédures de calcul peuvent s'adapter à l'immigration qui, en ajoutant des individus, perturbe aussi l'observation ainsi qu'à tout autre phénomène perturbateur éventuel. Les résultats ainsi obtenus, libres de l'influence des perturbations, se désigne par l'expression « mesures à l'état pur ». À ce sujet, cf. notamment WUNSCH G. et TERMOTE M. (1978), p. 14-48 ou VANDESCHRIK Chr. (2000), pp. 103-106.

calendrier<sup>5</sup>. Même si cette proposition peut sembler fort restrictive, elle nous semble acceptable dans le contexte du présent travail et est surtout à même de nous guider vers notre objectif final.

Certains auteurs ajoutent que l'analyse se développe au sein d'une cohorte. Pour notre part, nous préférons ne pas ajouter cette spécification ; en effet, elle nous couperait dès l'abord des mesures transversales au bénéfice des seules mesures longitudinales, ce qui nous ferait perdre un volet important et méthodologiquement intéressant de la façon de faire des démographes.

D'une manière générale, à notre avis, quel que soit le phénomène d'intérêt, l'analyse démographique classique procède en deux étapes :

- une première où les événements observés sont « dépondérés » par les effectifs pour obtenir une mesure du comportement via les propensions (taux ou quotients) ; par exemple, dans l'étude de la fécondité, les taux s'obtiennent en divisant les naissances observées pour des femmes d'un âge donné par l'effectif des femmes de cet âge ;
- une seconde où l'on « repondère » par les comportements pour obtenir les mesures finales ; en reprenant l'exemple de la fécondité, l'âge moyen à la maternité s'obtient via une moyenne pondérée des âges des femmes à la naissance de leurs enfants, les poids étant les taux calculés lors de la première étape.

Bénéfice de l'opération : les mesures du comportement, du calendrier ou de l'intensité/niveau ainsi obtenues sont débarrassées des effets de taille et de structure par âge<sup>6</sup> de la population sous observation, ainsi d'ailleurs que des effets des éventuels phénomènes perturbateurs. Il devient possible de comparer le comportement dans des populations de taille et de structure par âge variables.

La première étape vise à mesurer le comportement par « compartiment d'âge », le plus souvent à l'intérieur de cohortes, mais pas obligatoirement (cas des mesures dans les carrés). Via l'application des comportements à une cohorte, éventuellement purement fictive, la deuxième étape reconstitue l'activité du phénomène comme si elle était mesurée dans une cohorte observée à l'état pur. *Mutatis mutandis*, l'analyse démographique procède toujours de la sorte, quel que soit le phénomène et quelle que soit la nature des données. Pour certains phénomènes, la seconde étape suppose le détour par la construction d'une table, mais l'objectif reste le même (et à la limite, il est encore plus apparent, d'ailleurs) : que se passerait-il si une population standard était confrontée au comportement mesuré dans la première étape ?

---

5 À ce sujet, cf. notamment HENRY L. (1952), relayé entre autre par BLAYO Ch. (1991).

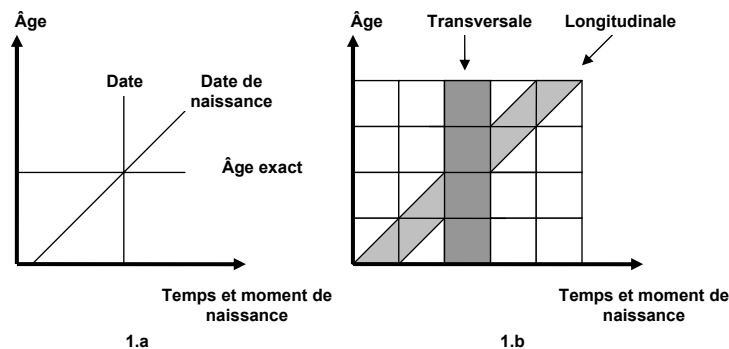
6 On pourrait encore améliorer la mesure en contrôlant en plus d'autres structures (matrimoniale ou par niveau d'instruction, par exemple), mais cela ne nous a pas semblé indispensable vu l'objectif du présent exposé.

Ce jeu de dépondération par les effectifs et de repondération par les comportements par âge combiné avec l'application à une cohorte pour mesurer le comportement global est, à notre avis, une des spécificités de l'analyse démographique classique. Pour la suite, nous dénommerons l'ensemble de cette procédure par le vocable de « cohortisation ».

## 1.2. Un outil pour visualiser la cohortisation

Pour illustrer graphiquement les analyses qu'elle pratique, la démographie dispose d'un outil qui lui est propre, à savoir le diagramme de Lexis (*cf.* figure 1).

**Figure 1. Coordonnées sur Diagramme de Lexis – Analyses transversale et longitudinale**



Il s'agit d'un diagramme en plan dont, le plus habituellement, les axes des abscisses et des ordonnées supportent respectivement le temps et l'âge<sup>7</sup>. Sur ce diagramme, il est évidemment possible de repérer des événements en fonction des deux coordonnées figurant sur les axes (comme c'est le cas dans n'importe quel type de diagramme cartésien), mais aussi, en plus, en fonction d'une troisième coordonnée, le moment de naissance (*cf.*

<sup>7</sup> La figure 1 propose le diagramme de Lexis dans sa version la plus classique. Le moment de naissance et l'âge peuvent laisser leur place respective à d'autres paires de variables comme le moment du mariage et la durée du mariage ou le moment d'une migration et la durée écoulée depuis cette migration. Ces deux exemples soulignent la relation qui unit les deux variables autres que le temps : celle qui remplace l'âge correspond à la durée écoulée depuis un événement particulier qui se place sur l'axe des abscisses (la naissance, le mariage, une migration...). Par ailleurs, le temps lui-même peut disparaître en tant que tel ; ainsi, un diagramme de Lexis pourra s'établir en substituant respectivement au temps, au moment de naissance et à l'âge, l'âge, l'âge au mariage et la durée du mariage. À ce sujet, *cf.* notamment VANDESCHRIK (1994).

figure 1.a). Un âge exact et une date se traduisent respectivement par une horizontale et une verticale et une date de naissance, par une oblique.

Via cette représentation, il sera également aisé d'opposer analyse longitudinale – dans le couloir oblique d'une génération – et analyse transversale – dans le couloir vertical d'une année (cf. figure 1.b). Cette opposition entre analyses longitudinale et transversale nous semble être aussi une spécificité de l'analyse démographique classique et, à tout le moins, un domaine sur lequel les techniques démographiques attirent de manière précise l'attention et pour lequel elles proposent des solutions intéressantes pour mesurer les phénomènes en pleine connaissance de cause.

### 1.3. Intérêt et limites de la cohortisation

La cohortisation est-elle toujours un passage obligé pour obtenir des mesures d'intensité et de calendrier ayant les qualités voulues ? Afin d'identifier les conditions où elle est indispensable, il faut prendre en compte les caractéristiques de l'observation, la constance ou non des phénomènes ainsi que la nature longitudinale ou transversale de l'analyse.

Prenons, à titre d'exemple, la primo-nuptialité étudiée parmi les célibataires, phénomène excluant de la population sous observation les individus qu'il touche. En cas d'observation longitudinale à l'état pur (en l'absence de perturbation, donc), le recours à la cohortisation, et singulièrement celui à la table d'éventualité, ne s'impose pas. En effet, la division des premiers mariages effectivement observés à chaque âge par l'effectif à la naissance donne directement accès aux événements réduits et donc, en les combinant, à la mesure de l'intensité et du calendrier. Au contraire, en cas d'observation transversale, même à l'état pur, il faudra passer par la cohortisation, sauf si l'effectif des cohortes à la naissance devait être constant dans le temps.

Pour systématiser la réflexion à ce sujet, il convient d'introduire le concept de population « intrinsèque »<sup>8</sup>. Âge par âge, elle désigne *l'ensemble des individus pour lesquels soit l'observation a pu saisir tous les événements matérialisant le phénomène étudié pour les individus l'ayant subi, soit aurait pu le faire, pour ceux qui ne l'ont pas encore connu*. Dans le cas de la primo-

---

8 Cette population dite « intrinsèque » est à distinguer de deux autres populations utilisées en analyse démographique. La population « soumise au risque » se définit comme suit : âge par âge, il s'agit du nombre d'individus qui, vu leur état ou statut, courent effectivement le risque de subir le phénomène étudié. La population « sous observation » quantifie le nombre d'individus observés, sans être nécessairement susceptibles de vivre le phénomène sous observation. Ces trois types de population interviennent de manière spécifique à différentes étapes du processus d'analyse des phénomènes démographiques.

nuptialité perturbée par l'émigration, un individu qui se marie une première fois (avant émigration éventuelle) restera définitivement dans cette population intrinsèque, et ce, quel que soit son devenir en matière de migration. En effet, l'observation de la primo-nuptialité à son égard est elle-même définitivement complète. Si un célibataire émigre, la population intrinsèque se modifie puisqu'à son endroit l'observation de la primo-nuptialité n'aura éventuellement pas été complète.

Si d'âge en âge, la population intrinsèque ne voit pas son effectif se modifier, l'observation du phénomène est complète et la cohortisation n'est pas nécessaire pour atteindre les mesures à l'état pur. Par contre, la présence de perturbations actives modifiant la population intrinsèque et/ou le caractère transversal de l'observation ou de la mesure conditionneront une modification de l'effectif de cette population intrinsèque d'un âge à l'autre et imposeront donc le passage obligé par la cohortisation pour obtenir les mesures adéquates du comportement, de l'intensité et du calendrier (*cf.* tableau 1).

**Tableau 1. Limites d'application de la cohortisation**

Observation	Longitudinale	Transversale
État pur	non	oui*
État perturbé	oui	oui, oui

\* si l'effectif des cohortes à la naissance est constant, la cohortisation n'est pas nécessaire, mais vu le caractère hautement improbable de cette spécification, nous avons préféré indiquer « oui » dans cette case.

#### 1.4. Un test simple avant cohortisation

Face à des données quelconques, on peut chercher à savoir si la cohortisation est simplement envisageable. Pour répondre à cette question, il suffit de voir s'il est possible de disposer les données en cause sur un diagramme ayant les caractéristiques d'un diagramme de Lexis classique. Ce test est pratique et, en définitive, facile à mettre en œuvre. Par ailleurs, autre avantage précieux, il permettra de bien visualiser ce que tente de réaliser la cohortisation. Dans le deuxième point, on envisagera donc les spécificités ainsi que la construction du diagramme de Lexis, ce qui nous amènera à formuler de façon précise les conditions mathématiques nécessaires pour opérer la cohortisation.

## 2. Diagramme de Lexis : spécificités, construction, historique

### 2.1. Le diagramme de Lexis, un diagramme particulier

#### 2.1.1. Deux axes, mais trois coordonnées dont une « parasite »

Une des spécificités marquantes du diagramme de Lexis, sans en être toutefois une stricte exclusivité (*cf.* point suivant), réside dans la possibilité qu'il offre d'utiliser non pas deux, mais bien trois coordonnées différentes pour y localiser un élément : outre l'âge et le temps occupant les deux axes, le moment de naissance peut aussi servir de coordonnée (*cf.* figure 1 et son commentaire). Cette troisième coordonnée parasite en fait l'axe des abscisses dont l'appellation doit donc être double, comme indiqué sur la figure 1 : l'axe horizontal supporte non seulement le temps, qui se traduit de manière classique par des verticales, mais aussi le moment de naissance, qui lui se traduit par une oblique croissante inclinée à 45 °.

#### 2.1.2. Des trajectoires « forcées »

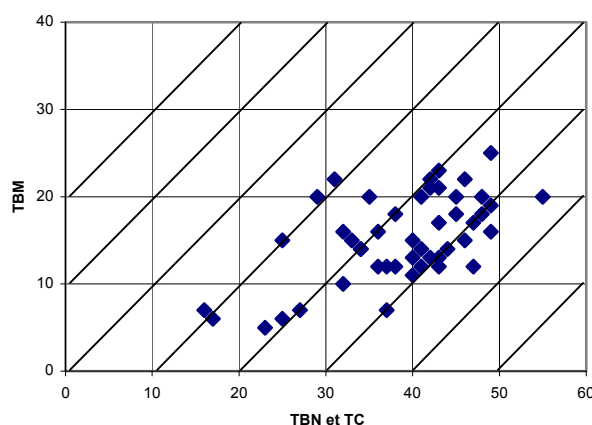
Dans le diagramme de Lexis, un individu se voit attribuer une trajectoire, sa ligne de vie. Il s'agit d'une oblique croissante inclinée à 45°, commençant à sa date de naissance et indiquant l'évolution de son âge en fonction du temps. Tous les faits concernant cet individu doivent se localiser sur cette trajectoire. À aucun moment, il ne pourra s'en écarter ; pour lui, cette trajectoire est « forcée ». La connaissance d'un seul point de la ligne de vie permet de la tracer entièrement. L'existence de ces trajectoires forcées représente un point essentiel pour différencier, parmi les diagrammes à deux axes et trois coordonnées, ceux qui méritent d'aller dans la catégorie des diagrammes de Lexis.

Ainsi, sur la figure 2, l'axe horizontal supporte à la fois le taux brut de natalité (TBN) et le taux de croissance (TC). Pour un pays, le taux brut de natalité se traduit par une verticale et un taux de croissance par une oblique. La valeur du taux de croissance se lit sur l'axe horizontal : l'oblique commençant à 10 ‰ relie tous les points du graphique où le taux de croissance vaut 10 ‰. Pour une unité d'observation s'agit-il d'une trajectoire forcée ? Nullement. Ainsi, lors de l'étude temporelle de la variation de ces trois variables pour un pays, la trajectoire obtenue en unissant les points le situant à différents moments risque bien de recouper les trois réseaux de droites verticales, horizontales ou obliques. Cette trajectoire n'aura rien de



forcé ; la connaissance d'un point de cette trajectoire ne permettra pas de la dessiner dans son ensemble.

**Figure 2. Relation entre le TBN et le TBM (en %) – Afrique 2002**



(Source : POPULATION REFERENCE BUREAU, 2002. *World Population Data Sheet*)

En conclusion, même si cette figure 2 comporte trois coordonnées distribuées sur seulement deux axes, elle ne peut rejoindre la catégorie des diagrammes de Lexis car elle ne comporte pas l'équivalent des trajectoires forcées que ce type de figure doit receler.

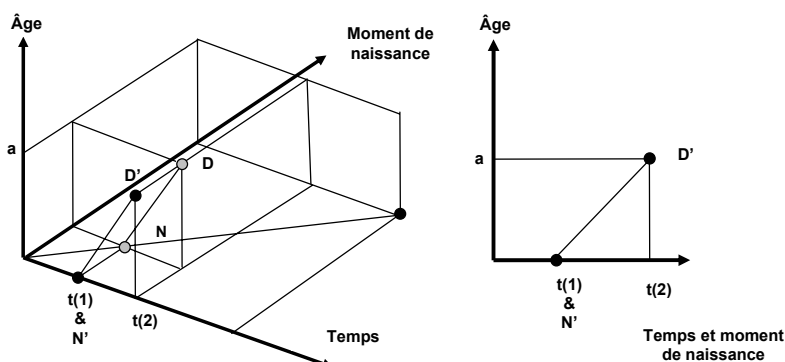
## 2.2. La construction du diagramme de Lexis

En partant d'un diagramme à trois axes (un pour l'âge, un autre pour le temps et un troisième pour le moment de naissance), il est possible de montrer que les lignes de vie se regroupent en fait sur un plan, négligeant le reste du volume décrit par les trois axes<sup>9</sup>. Il est dès lors possible, par projection des lignes de vie sur un des trois plans définis par deux des trois coordonnées, de réduire le diagramme en volume à un diagramme en plan sans perte d'information ou de précision dans la localisation. Lors de cette opération, une des coordonnées (celle dont l'axe a disparu) vient « parasiter » un des deux axes restant. Ainsi, dans la version utilisée actuellement, on a conservé les axes du temps et de l'âge ; la coordonnée moment de naissance

<sup>9</sup> À ce sujet et pour plus de détails, cf. VANDESCHRIK Chr. (1994) ou VANDESCHRIK Chr. (2001).

vient parasiter l'axe du temps. En conséquence, elle se traduit par une oblique et non par une verticale ou une horizontale (cf. figure 3).

**Figure 3. Réduction du nombre d'axes**



La projection peut s'effectuer également sur le plan défini par les axes de l'âge et du moment de naissance ou sur celui défini par les axes du temps et du moment de naissance. En cas de projection sur le plan « âge-moment de naissance », les lignes de vie sont en position verticale ; la coordonnée parasite est le temps, qui se traduit par des obliques décroissantes (cf. figure 4.a). En cas de projection sur le plan « moment de naissance-temps », les lignes de vie sont horizontales ; la coordonnée parasite est cette fois l'âge, qui se traduit par des obliques croissantes (cf. figure 4.b).

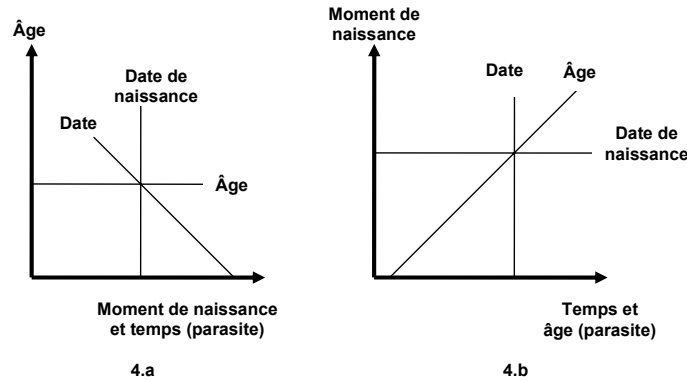
Les figures 1, 4.a et 4.b sont strictement équivalentes : chacune permet de localiser les données comme il se doit. Dans chacune, une coordonnée parasite un axe et se traduit par des obliques. La différence vient simplement du plan choisi pour projeter les lignes de vie au départ du graphique en trois dimensions. Chacune comporte aussi des trajectoires forcées, correspondant chaque fois à la traduction de la coordonnée du moment naissance dans l'espace défini par les deux axes, à savoir :

- une oblique croissante, dans le cas de la figure 1 ;
- une verticale dans le cas de la figure 4.a ;
- une horizontale dans celui de la figure 4.b.

Pour la suite, on ne prendra en compte, parmi les trois possibilités qui viennent d'être présentées, que celle de la figure 1. Cette version, actuellement la plus utilisée, a été inventée par Brasche en 1870 et reprise par Pressat dans les années 1950 ; celle de la figure 4.b a été proposée par Becker en 1874, soit approximativement un an avant que Lexis ne propose la sienne (figure 4.a), qui en fait était la réplique d'un graphique déjà proposé en 1869 par Zeuner sous une forme un peu différente et sous une forme semblable par

Verweij en 1874<sup>10</sup>... Comme on le voit, l'appellation finalement retenue pour ce diagramme ne semble pas en respecter la genèse<sup>11</sup> !

Figure 4. Deux autres versions du diagramme



### 2.3. Conditions mathématiques à l'élaboration d'un diagramme de Lexis

À ce stade, il convient de se demander quelles sont les conditions nécessaires pour pouvoir opérer la réduction du nombre d'axes au départ de la figure en trois dimensions et obtenir des trajectoires « forcées ». Pour la réduction du nombre d'axes tout d'abord, il faut que les trois coordonnées soient reprises dans une combinaison linéaire du type :

$$Z = aX + bY$$

Dans ce cas, si l'on décide de supprimer l'axe  $Z$ , la droite pour une valeur constante  $k$  de cette variable sur le plan  $XY$  aura pour équation :

$$k = aX + bY,$$

ce qui correspondra à une oblique de pente  $-a/b$ .

Si, comme dans le diagramme de Lexis classique,  $Z$  = le moment de naissance,  $X$  = le temps et  $Y$  = l'âge, la combinaison linéaire qui unit ces variables s'écrit comme suit :

$$Z = X - Y, \text{ avec } a = 1 \text{ et } b = -1.$$

Une valeur  $k$  de  $Z$  aura pour équation :

10 À ce sujet, cf. BRASCHE O. (1870), BECKER K. (1874), LEXIS W. (1875), VERWEIJ A. (1874), ZEUNER G. (1869).

11 À ce sujet VANDESCHRICK Chr. (1992, 1994 et 2001).

$$k = X - Y,$$

soit une oblique croissante de pente 1, ce qui correspond à une ligne de vie.

Si, comme dans le cas de la figure 2,  $Z$  = le taux de croissance,  $X$  = le taux brut de natalité et  $Y$  = celui de mortalité, la combinaison linéaire qui unit ces variables s'écrit comme suit :

$$Z = X - Y, \text{ avec } a = 1 \text{ et } b = -1.$$

Une valeur  $k$  de  $Z$  aura pour équation :

$$k = X - Y,$$

soit une oblique croissante de pente 1. À la différence des lignes de vie, ces obliques ne sont pas des trajectoires forcées. Quelle est la condition supplémentaire nécessaire pour obtenir des trajectoires de ce type ?

En fait, il faut que l'une des trois variables de la combinaison linéaire soit invariante unité d'observation par unité d'observation, comme l'est le moment de naissance. Si au contraire, comme dans le cas du diagramme avec les taux, aucune des coordonnées de la combinaison ne jouit de cette propriété, aucune trajectoire forcée ne pourra apparaître et, donc, en dépit de la réduction du nombre d'axes, une telle figure ne pourra rejoindre la catégorie des diagrammes de Lexis ; par ailleurs, pour souligner l'importance de cette dernière condition, il suffira d'ajouter que sans elle, point n'est possible de définir des cohortes et donc, de développer des analyses en cohortes !

En conclusion, pour obtenir un diagramme de Lexis, il suffit que les trois variables en jeu remplissent les deux conditions suivantes :

- être reliées par une combinaison linéaire, ce qui autorisera la réduction du nombre d'axes ;
- une des variables se doit d'être invariable unité d'observation par unité d'observation, ce qui conditionnera l'apparition de trajectoires forcées.

La détermination des conditions mathématiques précises permettant d'élaborer un diagramme de Lexis nous sera bien sûr très précieuse dans la phase de généralisation de l'emploi de ce type de figure aux cas de variables non temporelles.

En résumé de ce deuxième point et avant de passer à la généralisation, au-delà du possible recours à trois coordonnées pour se localiser sur une figure ne comportant que deux axes (caractéristique qu'il partage avec d'autres diagrammes cartésiens), la spécificité majeure du diagramme de Lexis vient de la possibilité d'y tracer pour chaque individu une ligne de vie qui représente sa trajectoire **forcée** dans le système défini par les axes du graphique. Pourquoi « forcée » ? La connaissance d'un seul point de cette trajectoire permet de la tracer dans son ensemble ; il n'est jamais possible pour un individu de quitter sa trajectoire pour en suivre une autre.

### 3. Diagramme de Lexis : généralisation

Dans les points qui suivent, nous essayerons de généraliser l'emploi du diagramme de Lexis et des méthodes d'analyse démographique en cas de variables non temporelles (et de variables temporelles, mais généralement non utilisées pour faire des diagrammes de ce type) et d'évaluer dans quelle mesure cela peut s'avérer profitable pour mieux comprendre l'observé. Le point 3 portera sur la seule généralisation du diagramme de Lexis ; en effet, si des données peuvent se distribuer sur une telle figure, la cohortisation de l'étude du phénomène en cause est envisageable, du moins sur le plan théorique, comme montré dans le point 4. Enfin, dans le cinquième point, nous listerons une série de problèmes spécifiques à dépasser lors de la cohortisation en cas des variables non traditionnelles.

#### 3.1. Taille, poids, etc...

Le diagramme de Lexis classique réduit le graphique initial en volume sur un plan, et ce, sans perte d'information. Par ailleurs, ce diagramme comporte des trajectoires forcées pour les unités d'observation. Les trois variables campant sur les axes du volume initial (avant réduction) se doivent de respecter des conditions très précises pour obtenir une construction plane ayant les caractéristiques voulues :

- la relation entre elles doit correspondre à une combinaison linéaire, ce qui autorise la réduction du volume en plan ;
- une des trois doit rester constante individu par individu, ce qui assure la présence de trajectoires forcées.

Les trois variables « taille », « taille à la naissance »<sup>12</sup> et « gain de taille depuis la naissance » satisfont pleinement les deux conditions :

- elles se regroupent dans une combinaison linéaire :

$$\text{gain de taille} = \text{taille} - \text{taille à la naissance} ;$$

- la taille à la naissance est constante pour un individu donné.

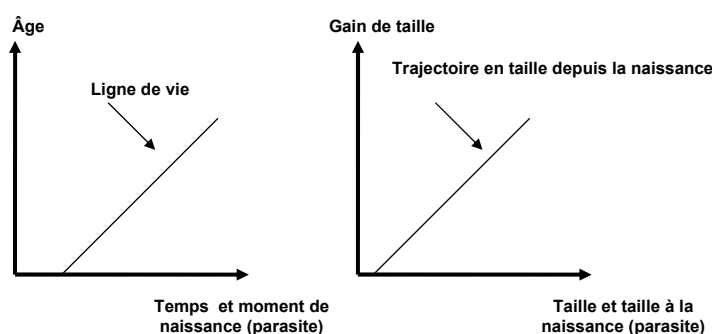
Des observations classées selon ces trois variables peuvent sans difficulté aucune se localiser sur un diagramme de Lexis (cf. figure 5). L'axe des abscisses y renseignera la taille (en remplacement du temps) ; celui des

---

12 L'on pourrait s'étonner de trouver une référence temporelle explicite dans le nom de cette variable. En fait, nos mesures s'opèrent très généralement dans un contexte temporel précis. Ainsi, le temps est d'une certaine façon bien présent dans le contexte des trois variables choisies : la taille et le gain de taille depuis la naissance se mesurent à un moment donné ; la taille à la naissance se réfère à un moment bien précis. Cette constatation ne transforme pas pour autant les trois variables en variables temporelles. Elles s'expriment bien en unités de longueurs (mètres, centimètres...) et nullement de temps (années, jours, secondes...) comme il sied aux variables temporelles.

ordonnées, le gain de taille (en remplacement de l'âge) et finalement, la taille à la naissance tiendra le rôle du moment de naissance, en parasitant l'axe de la taille<sup>13</sup>. Pour chaque individu, il est possible de tracer une trajectoire de taille, figurant l'évolution du gain de taille depuis la naissance en fonction de la taille en lieu et place de la classique évolution de l'âge en fonction du temps.

Figure 5. Diagrammes de Lexis en temps et en taille

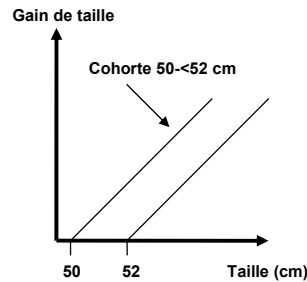


Vu les relations entre les trois variables de taille et l'identité des unités sur les deux axes, les trajectoires se présenteront toutes sous forme d'obliques croissantes inclinées à 45 degrés. La connaissance d'un seul point de la trajectoire est suffisante pour la déterminer dans son ensemble. À tous égards, ces trajectoires « forcées » jouissent des mêmes propriétés que les classiques lignes de vie. Et les cohortes dans tout cela ? Une cohorte regroupera tous les individus partageant une même gamme de taille à la naissance, par exemple entre 50 et 52 centimètres exacts<sup>14</sup> (50-<52 cm). Elle sera délimitée par les trajectoires forcées correspondant aux tailles à la naissance de 50 et 52 centimètres exacts (*cf.* figure 6).

13 Pour les graphiques, nous avons toujours opté pour les substitutions d'axes donnant le résultat le plus immédiatement comparable à un diagramme de Lexis classique. Un diagramme en taille peut se décliner sous trois – ou six – visages, exactement comme le diagramme en temps. Il ne nous a pas paru indispensable de développer les versions concurrentes, d'autant que pour ce faire, il suffirait de suivre la logique qui a été exposée au point 2.2 à propos de la construction du diagramme de Lexis.

14 La taille à la naissance est théoriquement une variable continue. Dans la pratique, elle s'exprime en centimètres sans plus de précisions. La cohorte en cause regroupera donc les individus dont la taille déclarée à la naissance était soit de 50, soit de 51 centimètres. La gamme de valeurs effectivement couverte par la cohorte dépendra de la façon dont la valeur est approximée, par arrondi ou troncature. N'oublions pas non plus le problème que pose la collecte d'une donnée de ce type au niveau de sa précision. Par ailleurs, nous supposons que les données ont été collectées parfaitement, ce qui pour le cas d'espèce est loin d'être assuré dans les faits !

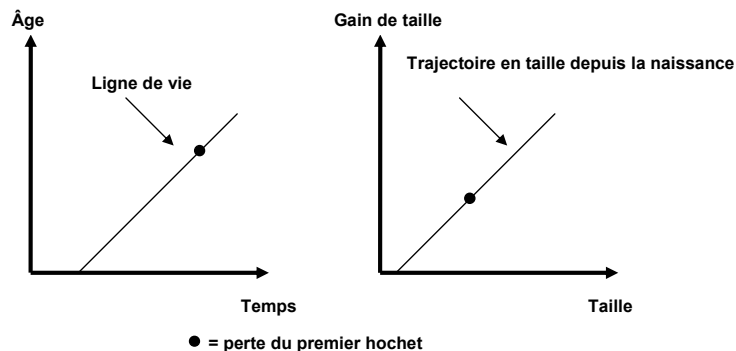
Figure 6. Diagrammes de Lexis en temps et en taille



Insistons : dans le cas classique, une cohorte regroupe les individus qui partagent la même année de naissance et dans le cas de la figure 6, ceux qui partagent une même gamme de poids à la naissance. Dans ces deux cas, l'évènement commun fondateur de la cohorte est la naissance ; la différence se marque au niveau du système de référence dans lequel il est repéré : système fondé soit sur le temps, soit sur la taille.

Autant il est possible de localiser des évènements sur les lignes de vie d'un diagramme de Lexis classique, autant il est possible de le faire sur un diagramme de Lexis en taille. Quels évènements ? Tous, aussi bien les évènements démographiques (le décès, l'émigration...) que non démographiques (le doublement du poids depuis la naissance, le premier anniversaire, le début de la marche, la perte du premier hochet...). Sur un plan purement théorique, les phénomènes correspondant à ces évènements pourront s'étudier via les outils de l'analyse démographique classique et singulièrement via la cohorte. Ainsi, sur la figure 7, la perte du premier hochet a été localisée sur les trajectoires d'âge et de taille d'un individu ; l'évènement localisé pourrait tout aussi bien être la contraction d'une maladie que le doublement de poids depuis la naissance ou une migration.

Figure 7. Localisation sur des diagrammes en temps et en taille



Il est à noter que la localisation sur le diagramme en taille se fait indépendamment du temps. Pour un individu, un événement peut correspondre à un âge et pour un autre individu, à un âge tout à fait différent. De même, sur le diagramme en temps, la localisation se fait en gommant complètement la notion de taille.

Le tableau 2 reprend différents exemples de variables qui satisfont aux deux conditions. La première colonne reprend l'exemple classique (moment de naissance, temps et âge) ; il est ainsi aisé de comparer les différents exemples. Le cas 2 reprend l'exemple du diagramme en taille. Pour les cas 3, 4 et 5, il est évidemment possible de produire un diagramme de Lexis et de constituer des cohortes respectivement de poids, de chiffre d'affaire et d'effectif de personnel.

**Tableau 2. Différents exemples de variables**

Cas 1	Cas 2	Cas 3	Cas 4	Cas 5
Temps	Taille	Poids	Chiffre d'affaire	Nombre d'employés
Moment de naissance	Taille à la naissance	Poids à la naissance	Chiffre d'affaire à un événement commun	Nombre d'employés à un événement commun
Age	Gain de taille depuis la naissance	Gain de poids depuis la naissance	Gain du chiffre d'affaire depuis l'évènement commun	Gain dans le nombre d'employés depuis l'évènement commun

En fait tous ces exemples sont semblables : même si ce n'est pas habituel dans le langage courant, le moment de naissance correspond au « temps à la naissance », la naissance étant l'évènement commun ; l'âge correspond quant à lui, à « un gain de temps depuis la naissance ». Chacun des cas se compose donc des mêmes trois éléments :

- une variable « absolue » (le temps, la taille...);
- la valeur de cette variable absolue à un évènement commun et fondant la cohorte ;
- une variation en terme de cette variable absolue depuis l'évènement commun.

Dans les trois premiers cas, l'évènement commun est le même ; il s'agit de la naissance. Par contre dans les deux autres, cet évènement commun n'est pas précisé. Il pourrait tout aussi bien s'agir d'une date précise (correspondant à l'entrée en vigueur d'une nouvelle loi économique ou sociale ou à un crac boursier, par exemple) ou alors d'un moment variable d'une firme à l'autre correspondant, par exemple, au moment où la firme commence à exporter ; dans ce dernier cas, il s'agirait donc d'une référence temporelle « flottante » par rapport au temps-calendrier. Par ailleurs, si l'on veut quitter le point de vue purement théorique pour atteindre éventuellement une certaine portée pratique, l'évènement commun devrait correspondre à un phénomène au moins potentiellement significatif pour les unités d'observation et leur « évolution » par rapport à la variable absolue, comme



le sont éventuellement la naissance ou l'entrée en vigueur d'une loi ou le début des exportations.

Si le diagramme de Lexis peut se généraliser, la notion de cohorte doit évidemment suivre le mouvement. Sans entrer dans les détails, une cohorte pourrait se définir comme suit : « ensemble des individus ayant connu l'évènement fondateur pour une même gamme de valeurs de la variable absolue ». En remplaçant « évènement fondateur » par « naissance » et « une même gamme de valeurs de la variable absolue » par « même année », le parallélisme avec la définition la plus classique de la cohorte s'établit sans peine (ensemble des individus nés la même année).

### 3.2. Des diagrammes de Lexis qui s'ignorent

Dans une partie précédente de ce texte, nous avons montré que, si le diagramme de Lexis est bien un diagramme cartésien, il n'en présente pas moins des particularités qui débouchent sur la possibilité de cohortiser les phénomènes pour les mesurer. Dans la littérature, on peut rencontrer des diagrammes cartésiens qui en fait mériteraient d'être qualifiés « de Lexis » sans que la cohortisation ne soit mise à profit.

Dans ce point, nous allons partir de figures extraites de la publication suivante : CASELLI Gr., MESLÉ Fr. et VALLIN J., *Les entorses au schéma de la transition épidémiologique*. Communication présentée à la séance 18 : « Nouvelles menaces sanitaires », du Congrès international de la population. Salvador. Brésil. Août 2002, Rome/Paris, 44 p. En fait, les figures reprises ci-après ont été légèrement modifiées par rapport aux figures originales sans que cela remette en cause le raisonnement qui sera suivi.

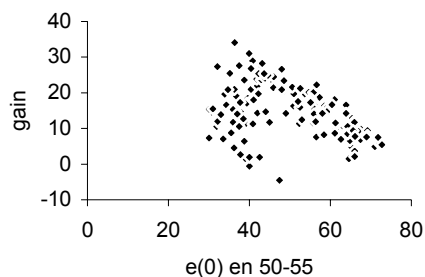
La première de ces figures se propose de localiser différents pays en fonction de leur espérance de vie en 1950-1955 et de leur gain d'espérance de vie entre 1950-1955 et 1995-2000<sup>15</sup> (cf. figure 8). La deuxième de ces figures localise les pays en fonction de leur espérance de vie en 1950-1955 et en 1995-2000<sup>16</sup> (cf. figure 9).

---

15 La figure originale (figure 3 en p. 7 de la publication de CASELLI et al. (2002)) diffère par le fait que le gain d'espérance de vie portait sur la période allant de 1950-1955 à 1970-1975.

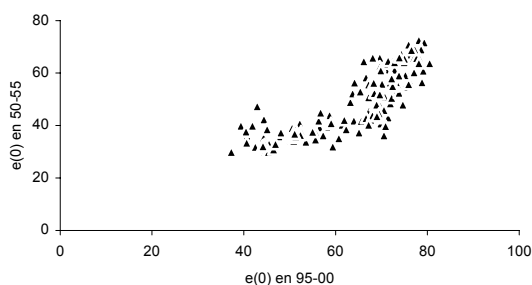
16 La figure originale (la figure 14 en p. 19 de la publication de CASELLI et al. (2002)), reprenait les espérances de vie en 1965 sur l'axe des abscisses et en 1995 sur l'axe des ordonnées. Nous avons mis en abscisses les données les plus récentes et non les plus anciennes ; on comprendra par la suite la raison de cette inversion.

**Figure 8. Localisation en fonction de l'espérance de vie initiale et du gain d'espérance de vie**



(Source : UNITED NATIONS (2001))

**Figure 9 Localisation en fonction des espérances de vie initiale et finale**

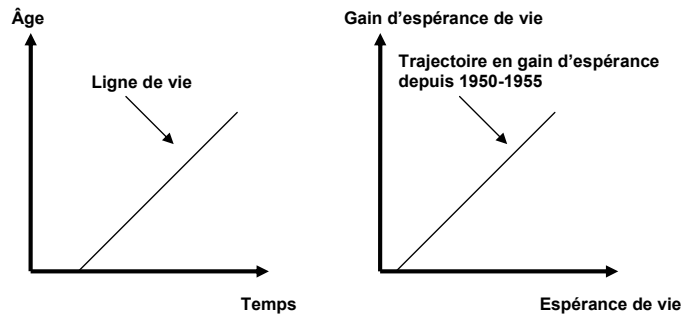


(Source : UNITED NATIONS (2001))

Pour montrer que ces deux dernières figures méritent le qualificatif « de Lexis », nous allons en proposer... une troisième reprenant le même ensemble de données. Dans cette figure, l'espérance de vie jouera le rôle dévolu classiquement au temps, l'espérance de vie en 1950-1955, celui du moment de naissance et le gain d'espérance depuis 1950-1955, celui de l'âge (*cf.* figure 10). Sur ce diagramme, une trajectoire a été ajoutée ; elle montre l'évolution du gain d'espérance de vie en fonction de l'espérance de vie pour une espérance de vie en 1950-1955 donnée (expression équivalente à celle d'application dans le diagramme de Lexis classique : évolution de l'âge en fonction du temps pour un moment de naissance donné). Un seul point de cette trajectoire suffit à la déterminer dans son ensemble. Autrement dit, ce type de trajectoire est comparable à une ligne de vie classique<sup>17</sup>.

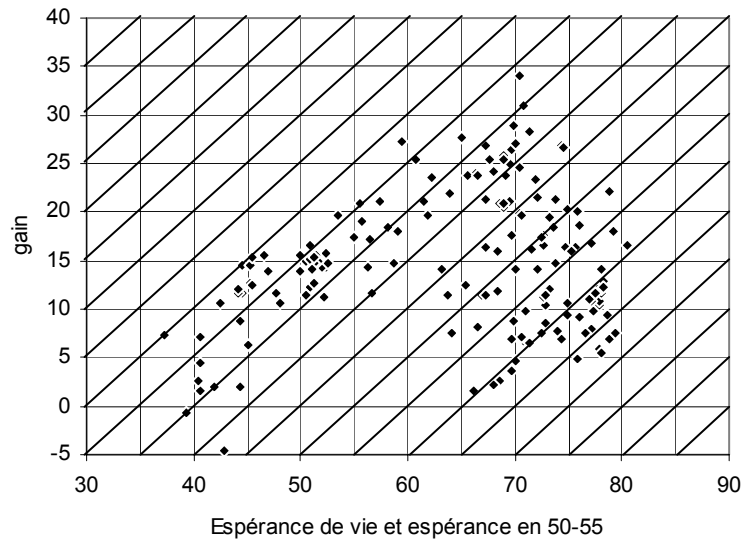
<sup>17</sup> Il est à noter que, d'un point de vue théorique, les trajectoires en cause peuvent se prolonger à l'infini.

Figure 10. Un diagramme de Lexis en espérance de vie



Comment dans ce contexte définir une cohorte ? Une cohorte regroupera des trajectoires présentant une gamme de valeurs voisines pour leur espérance de vie en 1950-1955, par exemple l'ensemble des trajectoires pour lesquelles cette espérance variait entre 45 et 50 ans. La figure 11 a été élaborée selon cette méthode de construction ; les diagonales y délimitent des cohortes quinquennales d'espérance de vie en 1950-1955.

Figure 11. Localisation en fonction de l'espérance de vie finale et du gain d'espérance

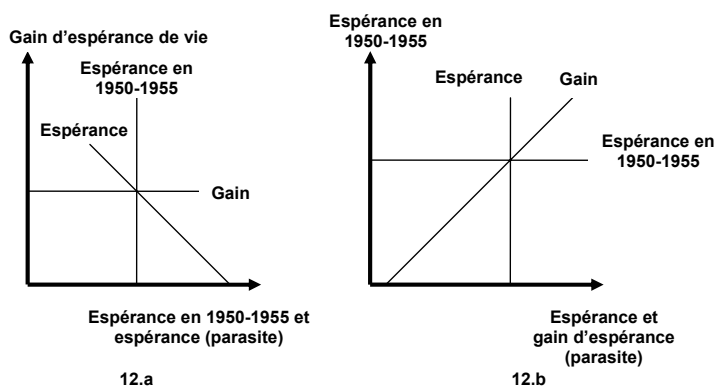


(Source : UNITED NATIONS (2001))

Quel est l'évènement matérialisé par les points de la figure 11 ? En fait, ces points permettent de connaître pour chaque pays à la fois l'espérance de vie en 1995-2000 en dessinant une verticale jusqu'à l'axe horizontal ; le gain d'espérance de vie entre 1950-1955 et 1995-2000 en traçant une horizontale jusqu'à l'axe vertical et finalement l'espérance de vie en 1950-1955 en traçant une diagonale jusqu'à l'horizontale en face d'un gain nul. Un point représente donc la période 1995-2000 pour une unité d'observation<sup>18</sup>.

Attention, l'axe horizontal est bien dévolu à l'espérance de vie (et non à l'espérance de vie en 1995-2000) ; l'espérance de vie joue le rôle de variable « absolue » telle que définie sous le tableau 2. L'on pourrait être tenté d'en rester à la dénomination « espérance de vie en 1995-2000 » pour cet axe, mais cela équivaudrait, dans un diagramme de Lexis classique, à baptiser l'axe horizontal « moment du mariage » si les points sur le diagramme devaient représenter cet évènement. Insistons, la situation décrite par la figure 11 est comparable à un diagramme de Lexis classique où les points représenteraient, par exemple, le décès des individus ou leur premier mariage : par une verticale, on connaîtra la date de l'évènement ; par une horizontale, leur âge à l'évènement et par une oblique, leur date de naissance. À notre avis, ceci suggère que le graphique 11 est un diagramme de Lexis.

**Figure 12. Diagrammes en espérance à la façon de Lexis et à celle de Becker**



Qu'en est-il des figures 8 et 9 ? Il s'agit aussi de diagrammes de Lexis. En fait, la figure 8, correspond à un diagramme de Lexis à la façon de Lexis, avec des trajectoires (et donc des cohortes) en position verticale et une coordonnée parasite (la valeur de l'espérance de vie) se matérialisant par des obliques décroissantes (*cf.* figure 12.a). Pour sa part, la figure 9 peut

<sup>18</sup> Pour l'interprétation des trajectoires forcées et des points, *cf.* annexe 1 et pour un décompte des observations par triangle, *cf.* annexe 2.

s'assimiler à un diagramme de Lexis à la façon de Becker avec des trajectoires (et donc des cohortes) en position horizontale et une coordonnée parasite (le gain d'espérance de vie) se matérialisant par des obliques croissantes (cf. figure 12.b).

En fait, tout graphique mettant en relation des données individuelles d'une même variable à deux moments différents constitue un diagramme de Lexis, qui éventuellement s'ignore. Pour s'en rendre compte, il suffit de faire émerger, moyennant quelques calculs intermédiaires, la variable absolue (qui remplacera le temps du diagramme de Lexis classique), la valeur de cette variable à l'évènement commun (qui remplacera le moment de naissance) et le gain en terme de cette variable par rapport à sa valeur à l'évènement commun (qui remplacera l'âge). Le système de repérage ainsi constitué au départ de deux coordonnées respectera les conditions à l'établissement d'un diagramme de Lexis. Finalement, en cas de données individuelles, et non plus collectives, l'exigence des trois coordonnées se réduit : il suffit que deux soient disponibles pour reconstituer l'ensemble des trois qui respectera les conditions précisées plus haut.

Le but de ce point était de généraliser l'usage du diagramme de Lexis au cas de variables inhabituelles pour ce faire. Nous le considérons comme atteint<sup>19</sup>. Nous allons montrer maintenant que les cas envisagés jusqu'ici peuvent déboucher sur des applications potentiellement intéressantes des méthodes de l'analyse démographique classique.

#### 4. Cohortisation : généralisation

Dans ce point, nous allons développer cinq exemples d'application de la cohortisation en situation inhabituelle, avec un souci avant tout théorique. Ceci n'exclut pas la volonté d'identifier des circonstances où l'application des principes de l'analyse démographique classique pourrait déboucher sur des résultats permettant d'appréhender différemment la situation telle que découlant de l'observation.

---

19 La généralisation pourrait être poussée plus avant encore. Il est ainsi possible de construire l'équivalent d'un diagramme de Lexis en substituant respectivement au temps, au moment de naissance et à l'âge, la température, la température au sol et la variation par rapport au sol de la température à différentes altitudes ; ce diagramme illustrerait une variation non temporelle d'une variable elle-même non temporelle. Un autre diagramme de Lexis pourrait s'élaborer au départ des trois variables suivantes mesurées au sein d'une entreprise : bénéfice, aide perçue et salaire ; dans ce derniers cas, les obliques pourraient prendre une inclinaison quelconque, se libérant de la contrainte des 45°. Il ne nous a pas semblé utile de développer ces exemples dans ce document étant donné qu'ils ne pourront se prolonger par des analyses intéressantes sur base de la cohortisation. Par contre, nous comptons les développer dans le cadre de notre thèse.

#### 4.1. Données en poids ou en taille sur diagramme de Lexis

Les données utilisées pour ce premier exemple d'application proviennent de la Banque de Données médico-sociales (BDMS), base de données gérée par l'Office de la Naissance et de l'Enfance de la Communauté française de Belgique (ONE). Cette base recueille notamment des données sur la grossesse, la naissance et les enfants en bas âge jusque l'âge de 1 an<sup>20</sup>. L'extrait que nous en avons reçu dans le cadre d'une convention avec l'ONE ne comporte aucune information permettant d'identifier les personnes en cause, ce qui assure le strict anonymat des données. Les deux exemples proposés ici porteront successivement sur la taille et le poids des nourrissons.

Pour les individus sélectionnés (en fonction de la disponibilité des données nécessaires pour les calculs et de leur plausibilité), le gain de taille durant la période d'observation a été calculé en retirant de la taille à la dernière visite celle à la naissance, ces deux variables étant intégrées dans la BDMS<sup>21</sup>. Par analogie avec la situation classique en démographie, les gains correspondent à l'équivalent d'âges en différence de millésimes obtenus en retirant l'année de survenance de l'évènement de celle de la naissance.

Avec le jeu de données ainsi constitué, un diagramme de Lexis a été élaboré comme explicité au point précédent : la variable *taille* se place sur l'axe des abscisses ; la variable *taille à la naissance* l'y rejoint en tant que coordonnée parasite et la variable *gain de taille depuis la naissance* se place sur l'axe des ordonnées (cf. figure 13) L'évènement à localiser sur cette figure correspond, pour chaque individu, à la fin de l'observation dont les coordonnées s'expriment en terme des trois variables précitées.

Assimilable à des évènements classés en différence de millésimes, ces évènements se localisent donc dans des parallélogrammes reposant sur un sommet. À titre d'exemple, les individus qui mesuraient 50 centimètres à la naissance et 53 centimètres à la fin de l'observation ont enregistré un gain de taille de 3 centimètres en différence de tailles, ce qui recouvre une gamme de gains exacts allant de 2 à 4 centimètres (4 exclu), comme le montre la figure 14.

Dans les figures 13 et 14, une cohorte est constituée par l'ensemble des individus de même taille « révolue » à la naissance. Seules les cohortes allant de 40 à 55 centimètres ont été conservées pour des questions d'effectifs

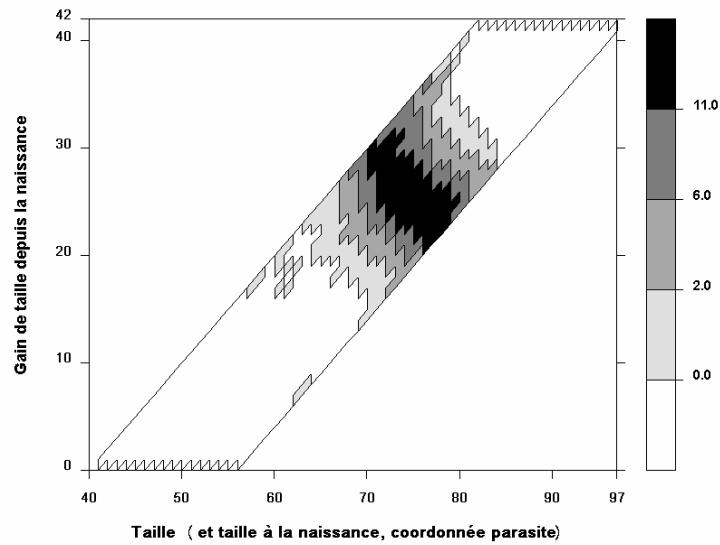
---

20 Pour une description de cette base de données et une analyse de son contenu, cf. OFFICE DE LA NAISSANCE ET DE L'ENFANCE (2000 et 2001). Nous tenons à remercier l'ONE pour l'accès à la BDMS.

21 Nous partirons de l'hypothèse hautement improbable que les mesures de la taille des nourrissons ne sont entachées d'aucune erreur ; vu l'objectif plutôt théorique poursuivi ici (montrer la possibilité d'établir un diagramme de Lexis en taille et poids), cette hypothèse n'est pas en soi trop gênante.

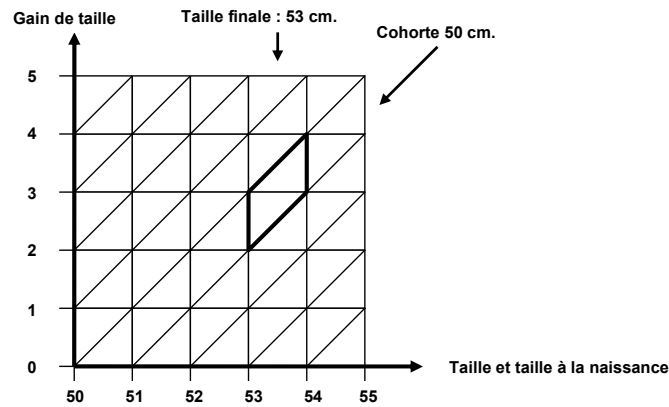
(cohorte de 40 centimètres : 144 individus ; cohorte de 55 centimètres : 308 individus ; effectif maximal : cohorte de 50 centimètres avec 14 111 individus ; au total, pour les 17 cohortes : 68 021 individus).

**Figure 13. Taille en fin d'observation : évènements réduits en % (taille en centimètres)**



(Source : Banque de données médico-sociales de l'ONE)

**Figure 14. Trame du diagramme de Lexis en taille (en centimètres)**



Le diagramme de Lexis de la figure 13 représente les événements réduits calculés par cohortes et gain en différence de tailles<sup>22</sup>. Par exemple, la cohorte 50 centimètres compte un effectif de 14 111, dont 2 049 mesureraient 74 centimètres en fin de l'observation et présentaient donc un gain en différence de tailles de 24 centimètres ; traduits en événements réduits, cela donne 14,52 % (2 049/14 111). Sur le diagramme de Lexis de la figure 13, le parallélogramme dévolu à ce résultat rentrera dans la catégorie la plus élevée (regroupant les valeurs supérieures à 11 %).

La figure 13 montre que les données s'organisent sous l'influence principale de deux effets :

- un premier effet de gain de taille (qui se dénommerait « effet d'âge » dans la configuration classique du diagramme de Lexis) : dans chacune des cohortes, la densité des événements varie en fonction du gain de taille depuis la naissance ; par ailleurs, la dispersion des événements épouse des schémas proches dans les différentes cohortes ;
- un deuxième effet de cohorte de taille à la naissance (dont la correspondance dans la configuration classique est immédiate) : à mesure que la taille à la naissance augmente, les événements se localisent pour des gains de taille de plus en plus faibles.

Par ailleurs, sur ce graphique, les données « aberrantes » apparaissent clairement, comme par exemple dans le cas de la cohorte la plus à droite pour des gains de taille légèrement inférieur à 10 centimètres.

Pour ce premier exemple, la durée de l'observation n'a pas été contrôlée. À l'inverse, dans l'exemple qui va suivre, seuls seront retenus les individus dont l'âge à la fin de l'observation était d'approximativement un an. Vu que la date à la fin de l'observation est souvent absente des données de base, les effectifs disponibles pour cet exercice s'effondrent. Il a donc fallu encore restreindre le nombre de cohortes conservées pour établir le diagramme de Lexis. Seules huit cohortes allant de 46 centimètres (avec un effectif de 160 individus) à 53 centimètres (avec un effectif de 136 individus) ont pu être conservées, pour un total de 2 981 observations (*cf.* figure 15).

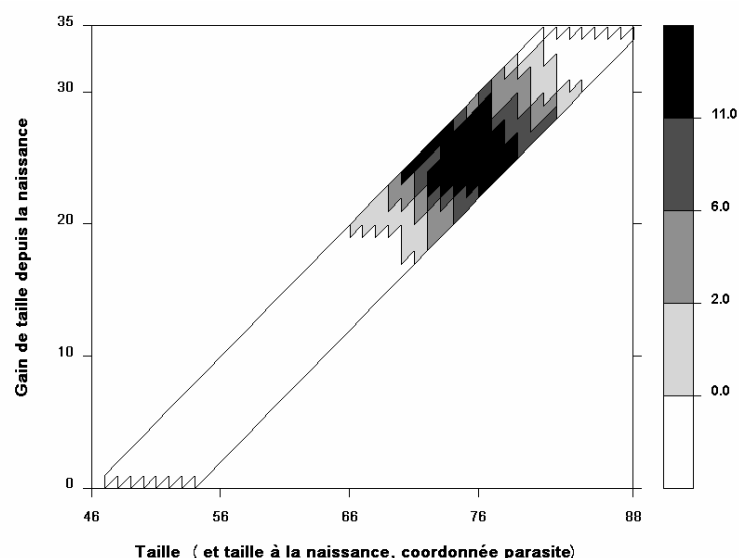
En gros, les observations suivent la même logique d'organisation que pour la figure 13, et ce, malgré que la tranche utilisée soit de moins grande ampleur : cohorte par cohorte, les dispersions se ressemblent ; le gain de taille a tendance à diminuer à mesure que la taille à la naissance augmente. Si la figure 15 semble un peu moins systématique que la précédente, cela provient sans doute, au moins en partie, d'effets liés aux petits effectifs en jeu ainsi qu'au faible nombre de cohortes utilisables.

---

22 Ce graphique a été élaboré sur base du programme « Lexis », mis au point par Kirill ANDREEV. Ce programme est d'accès libre (renseignements auprès de Kirill ANDREEV, adresse : [andreevk@post.queensu.ca](mailto:andreevk@post.queensu.ca) ou auprès du Max Planck Institute for Demographic Research, adresse : <http://www.demogr.mpg.de/>).



**Figure 15. Taille à un an : évènements réduits en % (taille en centimètres)**



(Source : Banque de données médico-sociales de l'ONE)

Une étude poussée de la variable *taille* nous amènerait à contrôler d'autres variables (comme le sexe de l'enfant, l'âge de la mère, sa parité, sa situation en terme de tabagisme, la durée de gestation...). Vu l'exiguïté de la population de référence, nous préférons arrêter ici le processus d'analyse, mené, pour rappel, avec un souci essentiellement à portée théorique.

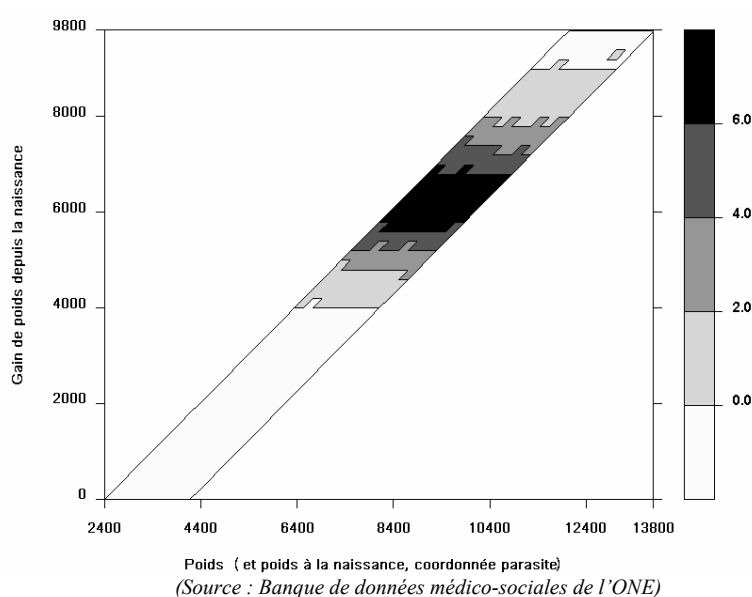
La figure 16, issue des mêmes données, propose cette fois un diagramme en poids. Ici, les cohortes sont constituées sur base de classes de poids à la naissance de 200 grammes, ainsi d'ailleurs que les gains de poids. Par ailleurs, les gains de poids ont été obtenus en retranchant du poids exact final le poids exact initial (à la naissance) puis regroupés par classe de 200 grammes. En conséquence, les données se localisent dans des parallélogrammes reposant sur une base tout comme des données qui croiseraient un âge révolu et une cohorte.

Contrairement aux figures en taille, ici, les cohortes adoptent toutes un comportement similaire en termes et de dispersion et de niveau de gains de poids<sup>23</sup>. Ce graphique est donc dominé exclusivement par un effet de gain

23 Cette constatation n'est valable que pour les gammes de cohortes de poids et de taille retenues, choix imposés par la disponibilité en nombre suffisant des données par cohorte. Elle ne résisterait pas nécessairement en cas de comparaison sur base de gammes de cohortes plus étendues ou de meilleur contrôle de l'âge.

de poids (qui se substitue à un effet d'âge sur le diagramme de Lexis classique), l'effet de cohorte qui était bien visible en terme de taille ayant complètement disparu ici. Il va de soi qu'une étude plus poussée du phénomène sur la base des graphiques entraînerait *de facto* le contrôle d'un certain nombre de variables, notamment celles citées précédemment. Nous ne le ferons pas ici.

**Figure 16. Poids en fin d'observation : évènements réduits en % (poids en grammes)**



Face à un diagramme de Lexis en taille ou en poids, on pourrait être tenté d'objecter qu'il s'agirait en fait de diagrammes temporels déguisés : le temps serait en quelque sorte sous-jacent aux variables utilisées. À notre avis, même s'il est vrai que les références au temps s'inscrivent bien dans les mesures en jeu (par exemple, la taille est toujours mesurée à un moment donné : la naissance, à la sortie de l'observation), il conviendrait au contraire d'inverser complètement la perspective : le temps ne représente en définitive qu'un proxy des variables explicatives qui restent sous-jacentes dans le diagramme de Lexis classique<sup>24</sup>.

Ainsi, en démographie, il est courant de parler d'effet d'âge, ce qui pourrait laisser accroire une valeur explicative substantielle pour cette variable. En fait, généralement, l'âge en lui-même n'a aucune vertu

<sup>24</sup> À ce sujet et pour la suite de la discussion, cf. notamment HOBBCRAFT J. et al. (1982).

explicative, si ce n'est médiatisé à travers des phénomènes physiologiques ou sociologiques, par exemple. En prenant le cas de la mortalité des personnes âgées, l'âge peut être, au moins en partie, considéré comme un proxy du processus physiologique du vieillissement. Insistons, ce dernier phénomène peut se développer à des vitesses différentes selon les individus ; même s'il y a certaines formes de régularité en fonction de l'âge, des personnes d'un même âge en terme de calendrier peuvent être à des étapes très différentes du vieillissement physiologique et connaître des risques de mourir bien différenciés. Dans le cas de la mortalité violente (par accident de la circulation) des jeunes adultes, et plus spécifiquement la mortalité des nuits du week-end, l'âge devient un proxy d'un phénomène « sociologique » en rapport notamment avec l'obtention du permis de conduire. Que l'âge légal pour ce permis se modifie et cette mortalité violente risque bien d'épouser une autre logique de répartition en fonction de l'âge.

Pourquoi dès lors un tel engouement en démographie pour les effets d'âge, mais aussi de période ou de cohorte ? Tout d'abord, il faut reconnaître que, malgré leur statut de proxy, ces variables n'en demeurent pas moins capables d'organiser efficacement l'observé<sup>25</sup>. En effet, elles se révèlent souvent plutôt bien corrélées avec les causes effectives et même avec leurs éventuels complexes enchevêtrements. Ensuite, si le repérage dans le temps des événements est largement disponible et aisé à organiser, il n'en va plus de même si l'objectif vise un classement en fonction d'autres variables : si un enfant en bas âge migre, la date de l'évènement sera connue sans grande difficulté ; pour ce qui est de sa taille à l'évènement...

Grâce aux deux exemples en taille et en poids proposés au début de ce point, nous pensons avoir montré la capacité du diagramme de Lexis à représenter de manière suggestive des données qui habituellement ne donnent pas lieu à l'établissement d'une telle représentation. Dans le point suivant, nous allons aborder de façon plus approfondie la cohortisation : au-delà de la simple représentation des données, nous allons essayer de montrer en quoi les méthodes de la démographie classique pourraient trouver à s'employer utilement dans des circonstances inhabituelles pour elles.

#### 4.2. Gain de taille et analyse en cohortes.

Le tableau 3 contient les données et résultats qui vont être utilisés. Les observations pour la cohorte de 50 centimètres à la naissance s'y classent selon le gain de taille exprimé en différence de tailles avec une gamme de valeurs allant de 12 à 33 centimètres (*cf.* figure 14 pour la disposition

---

25 À ce sujet, *cf.* notamment VAUPEL J. et al.(1997). Ce livre montre des diagrammes de Lexis avec des distributions de phénomènes en fonction des variables classiques de temps.

effective des données sur diagramme de Lexis). Par ailleurs, pour chaque classe de gain, une distinction s'établit selon que les individus concernés ont quitté l'observation avant un an ou à un an, catégorie regroupant en fait les sorties d'observation à 51 et 52 semaines de façon à obtenir des effectifs plus ou moins conséquents<sup>26</sup>.

**Tableau 3. Table d'extinction en gain de taille au 1<sup>er</sup> anniversaire pour la cohorte de 50 centimètres à la naissance**

Observations				Calculs				
Gain de taille	Durée de l'observation (semaines révolues)			Population sous observation	Diviseur	Quotient	Table d'extinction	
	50 et moins	51 & 52	Total				"lx"	"dx"
12	1	0	1	2291	2290,5	0,000	10000	0
13	2	0	2	2290	2289,0	0,000	10000	0
14	4	0	4	2288	2286,0	0,000	10000	0
15	6	0	6	2284	2281,0	0,000	10000	0
16	18	0	18	2278	2269,0	0,000	10000	0
17	23	0	23	2260	2248,5	0,000	10000	0
18	34	2	36	2237	2220,0	0,001	10000	9
19	76	1	77	2201	2163,0	0,000	9991	5
20	112	7	119	2124	2068,0	0,003	9986	34
21	146	22	168	2005	1932,0	0,011	9953	113
22	188	53	241	1837	1743,0	0,030	9839	299
23	279	61	340	1596	1456,5	0,042	9540	400
24	230	78	308	1256	1141,0	0,068	9141	625
25	226	109	335	948	835,0	0,131	8516	1112
26	159	78	237	613	533,5	0,146	7404	1083
27	113	69	182	376	319,5	0,216	6322	1365
28	64	42	106	194	162,0	0,259	4956	1285
29	19	33	52	88	78,5	0,420	3671	1543
30	14	8	22	36	29,0	0,276	2128	587
31	3	8	11	14	12,5	0,640	1541	986
32		2	2	3	3,0	0,667	555	370
33		1	1	1	1,0	1,000	185	185
Total	1717	574	2291	0				10000

(Source : Banque de données médico-sociales de l'ONE)

<sup>26</sup> Les sorties d'observation effectuées au-delà de la 52<sup>ème</sup> semaine ont été exclues, ce qui représentaient 148 cas.

Même si ce n'est pas immédiat (mais c'est l'une des rares possibilités offertes vu la structure des données disponibles), on peut, sur base de ces données, étudier le phénomène *1<sup>er</sup> anniversaire* en fonction de la variable *gain de taille durant la 1<sup>ère</sup> année de vie*. Cette situation d'analyse correspond en tous points à une autre bien plus habituelle pour les démographes : l'étude de la mortalité en fonction de l'âge. Dans les deux cas, le phénomène est non renouvelable et exclusif : la mort termine la vie (l'augmentation de l'âge) au même titre que le 1<sup>er</sup> anniversaire clôt le gain de taille durant la première année de vie. Dans les deux cas, le phénomène pourra s'étudier selon la même méthode, à savoir la table d'extinction.

Dans l'étude de la mortalité, il faut en général tenir compte de l'action des perturbations que représentent les sorties d'observation par émigration (et les entrées via les immigrations). Pour ces migrations, si l'âge est connu, le calcul des quotients pourra intégrer leur rôle perturbateur et initier l'élaboration d'une table sur base de quotients à l'état pur. Dans l'étude du 1<sup>er</sup> anniversaire en fonction du gain de taille, des sorties d'observation au pouvoir perturbateur se produisent également ; il s'agit des individus qui quittent l'observation avant leur 1<sup>er</sup> anniversaire. Pour ces sorties, le gain de taille est connu. Il est donc possible de tenir compte de la perturbation que ces sorties représentent pour calculer des quotients à l'état pur.

Avant d'explicitier ce calcul, insistons sur l'interprétation de quelques données du tableau 3. Au total, la cohorte de 50 centimètres comptait 2 291 individus à la naissance. Pour un gain de taille de 12 centimètres (première ligne de chiffres dans le tableau), on a constaté une sortie d'observation avant un an (soit un individu touché par la perturbation) et aucun 1<sup>er</sup> anniversaire (soit 0 individu touché par le phénomène étudié). Pour le gain de taille suivant (13 centimètres), la population sous observation n'est plus que de 2 290 individus, soit les 2 291 moins l'unique sortie à 12 centimètres. Autres exemples plus loin dans le tableau : pour le gain de 22 centimètres, parmi les 1 837 individus restant disponibles (soit les 2 291 du départ moins l'ensemble de ceux qui ont quitté l'observation avant un an ou ont fêté leur 1<sup>er</sup> anniversaire pour un gain inférieur ou égal à 21 centimètres), 188 ont quitté l'observation avant un an et 53 ont fêté leur 1<sup>er</sup> anniversaire ; au total, il y a donc eu 241 sorties d'observation, ce qui laisse une population sous observation de 1 596 individus pour le gain suivant. Etc.

Revenons-en maintenant au calcul des quotients à l'état pur. En prenant l'exemple de la mortalité perturbée par la seule émigration, une façon de procéder consiste à corriger le diviseur du quotient en retirant la moitié des sorties d'observation par perturbation<sup>27</sup>. Dans le tableau 3, ces diviseurs ont été obtenus en retirant de la population sous observation la moitié des sorties d'observation avant un an et s'étant produites pour la valeur de gain de

---

27 À ce sujet, cf. notamment WUNSCH G. et TERMOTE M. (1978), pp.37-38.

taille concernée ; par exemple, pour un gain de taille de 22 centimètres, la population sous observation vaut 1 857 ; les sorties d'observation, 188 et le diviseur, 1 743, soit  $1\,857 - (188/2)$ .

Ensuite, ce diviseur appliqué au nombre d'évènements observés (les 1<sup>ers</sup> anniversaires) donne le quotient ; ainsi pour le même gain de taille, le quotient vaut 0,030, soit  $53/1\,743$ . Comment interpréter ce quotient ? Il s'agit de la probabilité de fêter son 1<sup>er</sup> anniversaire à un gain de taille de 22 centimètres, probabilité calculée parmi ceux qui avaient atteint 21 centimètre de gain sans avoir fêté un an et sans avoir quitté l'observation par perturbation. En comparaison avec la mortalité étudiée en concurrence avec la seule perturbation de l'émigration, le quotient à un âge donné mesure la probabilité de mourir à cet âge parmi les survivants de cet âge et non émigrants.

Ce quotient correspondra à une mesure à l'état pur, dans le sens où elle est débarrassée de l'effet perturbateur lié aux sorties d'observation. Le reste du tableau contient les calculs partiels d'une classique table d'extinction. Le gain moyen de taille au 1<sup>er</sup> anniversaire (non présent dans le tableau) remplacera la classique espérance de vie. À la naissance, il vaut 27,37 ans alors que ce gain moyen calculé au départ des effectifs observés vaut 25,15 ans. Cette différence est comparable à celle qui prévaudrait entre des âges moyens au décès utilisant comme poids soit les nombres de décès de la table, soit les nombres absolus observés dans la population concernée.

Une procédure de constitution d'une table selon les mêmes principes pourrait s'appliquer à d'autres cohortes de taille à la naissance.

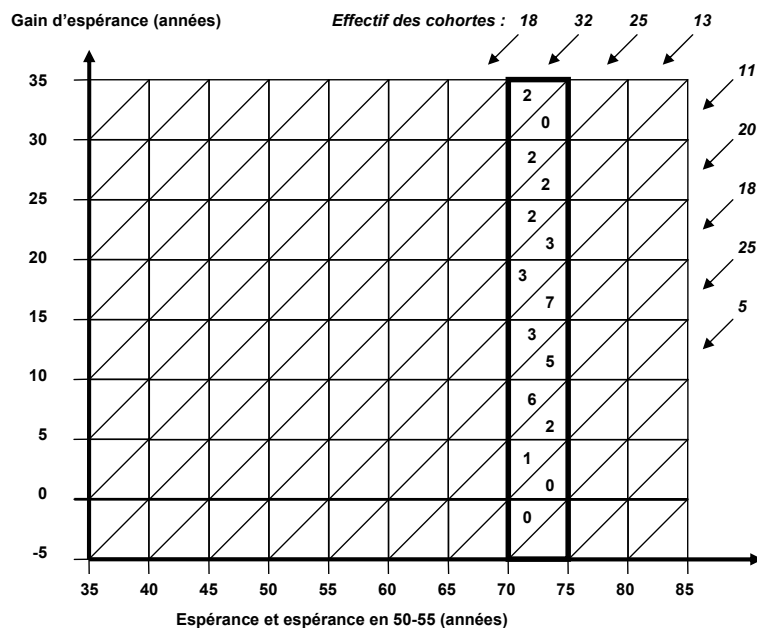
À notre sens, voici donc un exemple où les méthodes des démographes trouvent à s'employer utilement. Rappelons ici la portée essentiellement théorique de l'exercice en cours. Par ailleurs, il y aurait aussi lieu de s'interroger sur l'impact possible des classiques hypothèses sous-jacentes à ce type de calcul, mais nous ne voyons pas, en première analyse, en quoi l'indépendance et la répartition uniforme seraient fondamentalement plus – ou moins – gênantes ici qu'en analyse démographique classique.

#### **4.3. Espérance de vie, cohortisation et une première mesure transversale**

Nous repartirons ici de la figure 11 montrant la répartition des pays en fonction des trois coordonnées suivantes : l'espérance de vie, l'espérance de vie en 1950-1955 et le gain d'espérance depuis 1950-1955. L'évènement étudié est la période 1995-2000 via sa localisation sur les lignes d'espérance de vie des pays. Vu la nature des données et de l'observation (pas de

perturbation entravant l'observation<sup>28</sup>), l'exemple retenu portera sur une mesure transversale, à savoir : que vaut le gain moyen d'espérance de vie entre 1950-1955 et 1995-2000 parmi les pays ayant atteint une espérance de vie variant de 70 à 75 ans en 1995-2000 ? Dans la figure 17, la colonne entourée d'un trait gras indique les observations concernées par la question ; les chiffres dans les triangles rappellent le nombre d'observations qui y tombent.

Figure 17. Données brutes sur trame de Lexis



(Source : UNITED NATIONS (2001))

La mesure voulue peut être qualifiée de « transversale » vu qu'elle se fera dans une colonne recoupant toute une série de cohortes définies sur base de la valeur de l'espérance de vie en 1950-1955 ; le parallélisme avec une mesure transversale sur le diagramme de Lexis classique est immédiat en substituant la colonne mise en évidence sur la figure 17 par celle d'une année donnée. Par ailleurs, nous considérerons *a priori* que cette mesure transversale revêt un intérêt pratique dans le cadre d'une analyse donnée.

28 Le phénomène « gain d'espérance de vie entre 1950-1955 et 1995-2000 » pourrait aussi s'étudier longitudinalement. Toutefois, vu que l'observation est à l'état pur, le calcul avec les nombres absolus d'événements et celui avec des événements réduits donneraient des résultats absolument identiques, ce qui ne permettrait pas de souligner l'intérêt de la cohortisation du phénomène.

Enfin, même si, vu le faible nombre d'observations, l'hypothèse de répartition uniforme des événements dans les triangles peut sembler hautement irréaliste, nous la poserons quand même, étant donné la portée purement théorique de l'exercice en cours.

Le gain moyen d'espérance de vie entre 1950-1955 et 1995-2000 parmi les pays où l'espérance de vie en 1995-2000 valait entre 70 et 75 ans peut s'obtenir via une moyenne pondérée, les poids étant les nombres absolus d'observations des différents triangles. En calculant de la sorte, on obtient un gain moyen de 16,45 années (*cf.* annexe 2).

On pourrait reprocher à cette mesure de ne pas mettre sous contrôle l'effectif des différentes cohortes en cause<sup>29</sup>. Celui-ci varie fortement : de 11 pour la cohorte 50-<55 ans d'espérance de vie en 1950-1955 à 32 pour la cohorte 35-<40 ans. Pour mettre sous contrôle ces différences d'effectifs, il faut passer par l'intermédiaire des événements réduits qui remplaceront les effectifs observés comme poids dans le calcul de la moyenne (*cf.* figure 18). Ainsi, le calcul se fera comme si toutes les cohortes comptaient le même nombre d'individus et le résultat ne dépendra donc plus de l'importance des différentes cohortes. Cette procédure donnerait une moyenne de 16,83 années, en travaillant sur base des triangles (*cf.* annexe 2).

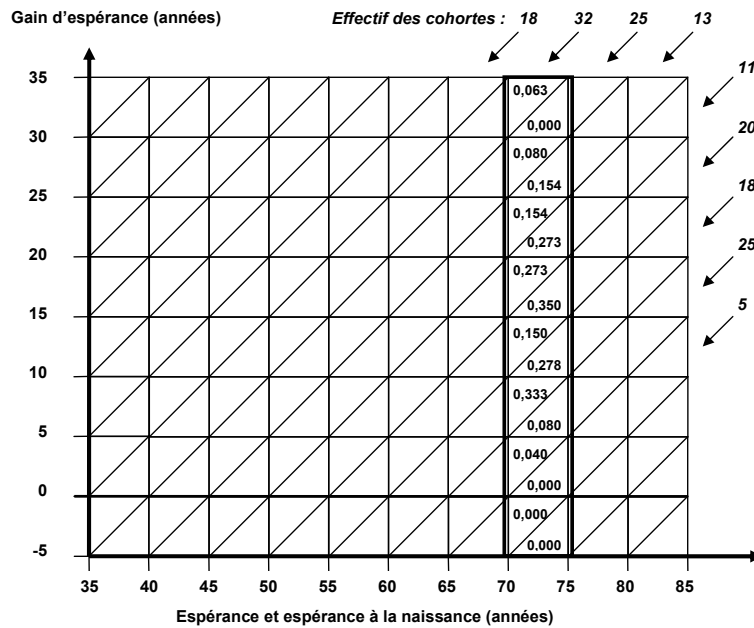
Par rapport à un calcul direct avec les effectifs observés, la mesure ainsi obtenue aura les mêmes qualités que celles de la mesure transversale du calendrier de la fécondité en remplaçant, comme poids dans la formule de l'âge moyen, les nombres absolus de naissances par âge par les taux de fécondité. En définitive, cette démarche suppose d'abord de « dépondérer » les événements par les effectifs des cohortes pour mesurer les comportements, puis de pondérer par les comportements ainsi mesurés les gains d'espérance pour obtenir une moyenne.

---

29 Pour établir un parallélisme précis avec les phénomènes démographiques et leur classification par rapport au choix des méthodes à leur appliquer, ajoutons que le phénomène « gain d'espérance de vie entre 1950-1955 et 1995-2000 », d'une part, est non renouvelable et, d'autre part, n'exclut pas les individus de l'observation : la trajectoire se poursuit au-delà (ou en deçà, en cas de régression de l'espérance de vie) du point y marquant l'évènement observé.



Figure 18. Évènements réduits sur trame de Lexis



(Source : UNITED NATIONS (2001))

La différence entre les deux mesures (16,45 contre 16,83) vient de ce que la première ne met pas sous contrôle l'importance relative des cohortes les unes par rapport aux autres, ce que réalise la deuxième mesure. Question importante à creuser, mais dont la réponse nous paraît affirmative : *est-il intéressant de calculer selon la deuxième voie ?* Premier élément de réponse : par exemple, le fait que le triangle ayant pour coordonnées une espérance de vie en 1995-2000 de 70-<75 ans, une espérance de vie en 1950-1955 de 55-<60 ans et un gain d'espérance de vie de 15-<20 ans compte 7 observations, dépend en partie du comportement de la cohorte en matière de gain d'espérance (comportement mesuré via les évènements réduits), mais aussi du fait que cette cohorte compte 20 individus.

Si on s'intéresse au comportement à des fins de comparaison, il faut mettre sous contrôle l'effet de structure (l'importance relative des différentes cohortes) et calculer selon la deuxième voie. C'est bien ce que font habituellement les démographes pour calculer en transversal l'âge moyen à la maternité : la fécondité d'une année dépend de l'âge et donc de l'année de naissance ; autrement dit, le comportement en matière de fécondité durant une année donnée varie en fonction de la cohorte de naissance et le passage par le calcul des taux permet de contrôler l'influence de l'effectif des cohortes. De même, dans notre exemple, on peut penser que le comportement

en matière de gain d'espérance de vie se module en fonction de l'espérance de vie initiale (en 1950-1955). Il faut donc, dans notre mesure, contrôler l'effectif des cohortes constituées sur cette base.

Notons enfin que ce type de procédure pourrait s'appliquer avec des références temporelles d'une autre nature. Pour ce faire, il faudrait imaginer, par exemple, que l'espérance de vie en 1950-1955 cède sa place à l'espérance de vie au moment de la perte d'indépendance pour le pays (en supposant que l'étude porte sur des pays ayant connu une phase de colonisation) et que l'évènement localisé ne soit plus la période 1995-2000, mais l'accès à l'indépendance (*cf.* annexe 1). Les références temporelles deviendraient ainsi « flottantes ».

En fait, ce qui a été exposé jusqu'ici en matière de cohortisation revient à un simple clonage de ce que réalise l'analyse démographique classique au départ des variables habituelles, à savoir le temps, l'âge et le moment de naissance. La suite du texte devrait montrer que, sur un plan théorique au moins, la cohortisation pourrait aussi trouver à s'employer utilement dans d'autres circonstances. Ainsi, à la transversalité classique qui vient d'être explicitée pourrait s'adjoindre une transversalité d'une autre nature, par recouplement des cohortes dans l'horizontalité (point suivant) ; le dernier point montrera une possible influence du raisonnement en cohorte sur les mesures de corrélation et régression.

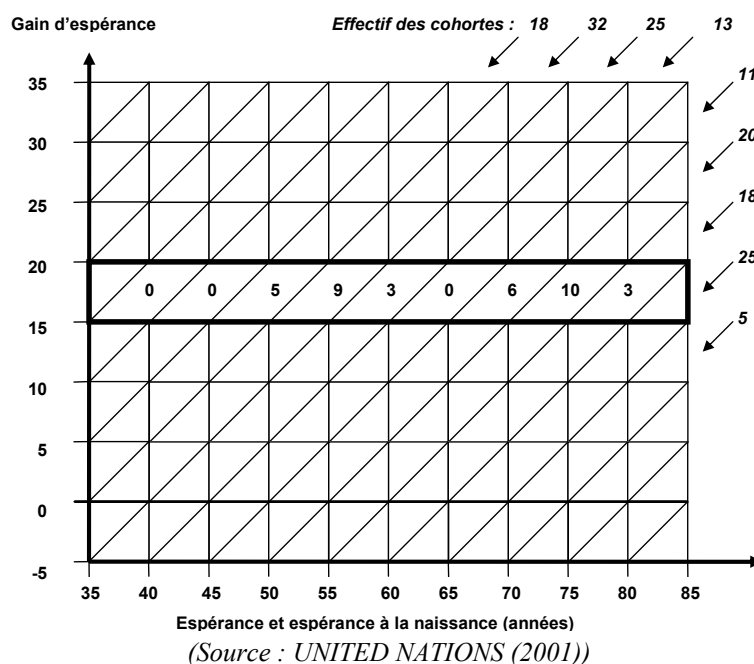
#### 4.4. Cohortisation et transversalité horizontale

Le calcul transversal présenté au point précédent ne se démarque en rien du classique calcul transversal des démographes. La transversalité pourrait aussi porter sur d'autres questions, comme par exemple, en reprenant l'exemple des gains d'espérance de vie entre 1950-1955 et 1995-2000 :

- que vaut l'espérance de vie moyenne en 1995-2000 parmi les pays présentant une certaine gamme de gain d'espérance entre 1950-1955 et 1995-2000 ?
- que vaut l'espérance de vie moyenne en 1950-1955 parmi les pays présentant une certaine gamme de gain d'espérance entre 1950-1955 et 1995-2000 ?

Les données nécessaires pour répondre à ces questions se localisent dans le couloir horizontal du diagramme de Lexis correspondant à la gamme de gain d'espérance choisie. Ainsi, sur la figure 19, le trait gras encadre la zone réservée aux gains d'espérance valant entre 15 et 20 ans. Par ailleurs, les chiffres dans les losanges renseignent le nombre d'observations qui y tombent. Cette tranche recoupe bien une série de cohortes ; la mesure lui correspondant mérite donc bien le qualificatif « transversal ».

Figure 19. Transversalité horizontale



L'espérance de vie en 1995-2000 pour les pays concernés (la 1<sup>ère</sup> question) peut se calculer de deux manières :

- tout d'abord, en pondérant les espérances en 1995-2000 (et plus précisément les centres de classe en supposant la répartition uniforme des observations) par les effectifs des losanges ;
- ensuite, en les pondérant par les événements réduits calculés en divisant les effectifs des losanges par l'effectif des cohortes.

La deuxième mesure offre l'avantage d'être débarrassée des effets liés aux différences numériques entre les effectifs des cohortes ; cette situation est en fait comparable à ce qui se passe en cas de transversalité verticale. Pour un gain d'espérance de 15 à 20 ans, la pondération selon les effectifs donne 64,86 ans contre 66,40 ans pour celle selon les événements réduits.

Vu la relation existant entre les trois variables en jeu, il n'est pas nécessaire de calculer selon les mêmes principes l'espérance de vie moyenne en 1950-1955 pour le même groupe de pays (2<sup>ème</sup> question). En effet, les résultats seront simplement décalés du gain moyen d'espérance de vie, soit 17,5 ans, toujours selon l'hypothèse de répartition uniforme des événements (à ce sujet, cf. annexe 3).

Dans la situation classique de l'analyse démographique, l'équivalent de cette transversalité horizontale existe au moins potentiellement. En prenant l'exemple de la mortalité à 3 ans révolus :

- quelle est la date moyenne de décès parmi les individus décédés à 3 ans révolus ?
- quelle est la date de naissance moyenne parmi ces mêmes individus ?

Vu la formulation de ces questions, on comprendra aisément que les démographes ne se soient pas rués sur la transversalité de nature horizontale ! Par contre, en l'appliquant à d'autres domaines, elle pourrait se révéler utile. Notons au passage que la transversalité verticale peut aussi se décliner sous deux formes distinctes :

- tout d'abord, classiquement, quel est l'âge moyen à l'évènement parmi les individus l'ayant vécu durant une année donnée ?
- ensuite, quelle est leur date de naissance moyenne ?

Une fois de plus, la deuxième question n'est guère affriolante en matière d'analyse démographique. On comprendra qu'elle ne retienne que peu ou prou l'attention<sup>30</sup>. Par ailleurs, cette date moyenne de naissance correspondra tout simplement au moment moyen de survenance de l'évènement (soit le milieu de la période en cause) diminué de l'âge moyen obtenu en répondant à la première question. Cette relation automatique<sup>31</sup> diminue d'autant l'intérêt de cette question.

Sur le plan de la méthode générale développée dans ce point, il est à signaler que nous avons préféré travailler sur base de losanges et non plus de triangles, comme dans le point précédent.

#### 4.5. Corrélation/régression et cohortes

En reprenant les données sur l'espérance de vie en 1950-1955 et 1995-2000 (cf. figures 8 et 9), on peut chercher à quantifier la corrélation entre ces deux mesures et à estimer la droite de régression (cf. figure 20). Cette figure est un diagramme de Lexis avec en abscisse la variable *espérance de vie en 1950-1955*, en ordonnée la variable *espérance de vie en 1995-2000* et en coordonnée parasite, la variable *gain d'espérance de vie entre 1950-1955 et 1995-2000*. Cette variable parasite se matérialisera par des obliques de pente unitaire. Les valeurs de cette variable parasite se liront sur l'axe vertical de manière directe (la diagonale qui intercepte l'axe vertical

---

30 À cela s'ajoute le fait que cette transversalité horizontale ne débouche pas sur la constitution d'une cohorte fictive, comme le fait la transversalité verticale.

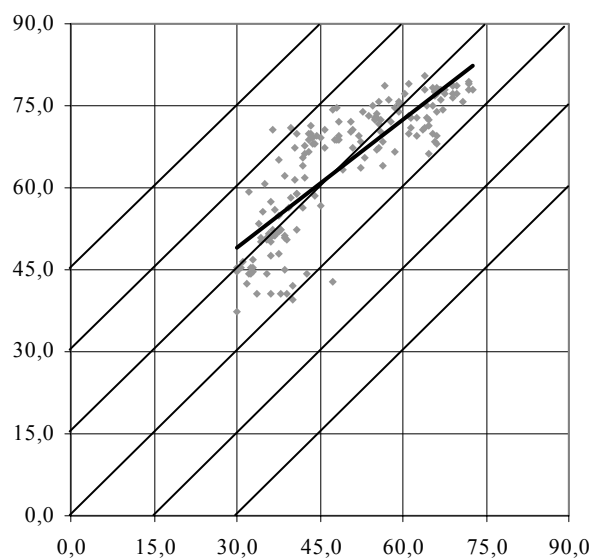
31 Pour la transversalité horizontale, nous avons montré en annexe 3 l'équivalence des résultats. La même démonstration pourrait se faire ici, moyennant le recours à des losanges reposant sur des sommets et non plus sur un côté.

à la valeur 15 indique un gain de 15) et de manière négative sur l'axe horizontal (la diagonale qui intercepte l'axe horizontal à la valeur 15 indique un gain de -15). Sur cette figure 20, les cohortes<sup>32</sup> se définissent selon la valeur de l'espérance en 1950-1955 ; elles se retrouvent en position verticale.

Voici les résultats des calculs de régression et corrélation :

- équation de la droite de régression :  $0,79x + 25,34$
- coefficient de détermination : 0,70.

**Figure 20. Corrélation et régression**



(Source : UNITED NATIONS (2001))

En fait, ces valeurs sont influencées par la densité des points-observations dans les différentes zones du graphique. Cette densité est elle-même fonction de deux éléments :

- le comportement des pays en matière d'évolution de l'espérance de vie ;
- le nombre de pays dans les différentes cohortes.

32 Problème récurrent : comment définir une cohorte ? Comment choisir leur amplitude ? Pour l'exemple, nous avons constitué des cohortes de 5 ans : les pays qui avaient une espérance de vie comprise entre 40 et 45 ans, 45 et 50 ans... en 1950-1955. Ceci peut sembler arbitraire, mais nos cohortes annuelles classiques ne le sont-elles pas tout autant, malgré le caractère « naturel » de la longueur d'une année ? Une façon d'éviter au moins partiellement la question est de diminuer l'amplitude des cohortes. Jusqu'où ? Jusqu'aux données individuelles ? On retomberait alors dans le cas initial où chaque observation a un poids unitaire...

Le deuxième point nous paraît gênant : les hasards de l'histoire auraient pu découper les pays autrement. Autrement dit, l'effectif initial des cohortes est ce qu'il est par hasard et donc, c'est par hasard que les nombres de points dans les zones du graphique sont ce qu'ils sont. Or ce que l'on cherche à estimer, c'est la relation et la corrélation mesurant un comportement. Il faudrait donc des mesures indépendantes du fait que les cohortes soient plus ou moins nombreuses.

Une solution pour dépasser ce problème : pondérer les observations non par 1, mais par  $1/n_c$  avec  $n_c$  désignant l'effectif de la cohorte  $c$ , ce qui revient, en langage démographique à pondérer par des événements réduits. De la sorte, on obtient une mesure portant sur le comportement et débarrassée des effets de structure liés à l'effectif des cohortes. Ce type de procédure débouche sur les résultats suivants :

- équation de la droite de régression :  $0,75x + 27,17$
- coefficient de détermination :  $0,72$ .

Est-ce que calculer la régression et la corrélation en contrôlant l'effectif des cohortes<sup>33</sup> a un sens ? Est-ce intéressant ? Nous pensons que oui, autant ou aussi peu que pour le classique âge moyen à la maternité par rapport à l'âge moyen des mères à la naissance de leur(s) enfant(s), par exemple. À notre sens, la réponse est d'autant plus affirmative qu'il existe une logique de comportement de cohorte sous-jacente au phénomène visé, ce qui revient à dire que, pays par pays, la valeur de 1995-2000 est influencée, au moins pour partie, par sa valeur en 1950-1955, ce qui ne paraît guère abusif dans le cas d'espèce. Finalement, la deuxième mesure nous paraît plus appropriée pour quantifier la relation entre les deux variables.

#### 4.6. Conclusions à propos de la généralisation de la cohortisation

Les différents exemples développés dans les pages qui précèdent montrent, à notre avis, une certaine efficacité, voire une efficacité certaine, de la cohortisation pour atteindre certains objectifs en matière d'analyse des données et ce, dans des circonstances débordant largement du champ traditionnellement dévolu à la démographie.

Les diagrammes de Lexis en poids et en taille permettent ainsi de représenter aisément des masses importantes de données et surtout d'en révéler des logiques d'organisation. Pourront éventuellement y apparaître les

---

33 Il est possible qu'une meilleure mesure serait obtenue si au lieu de se contenter de contrôler le nombre de pays, on contrôlait, par exemple, carrément la population (le nombre d'individus) que représente chaque point. Il est prévu d'étudier plus avant cette question si celle portant sur l'intérêt même de la procédure résiste suffisamment à la critique. Par ailleurs, ajoutons que, même en cas de corrélation entre deux variables de nature différente, il pourrait être utile de transposer la réflexion menée ici.

correspondants des effets d'âge, de période et de cohorte, si chers aux démographes.

Plus fondamentalement, les méthodes de l'analyse démographique classique pourront s'utiliser, même en cas de variables inhabituelles. Ainsi, une table d'extinction pourrait très bien s'établir non en fonction de l'âge, mais en fonction d'un gain – ou variation – portant sur d'autres variables, comme le poids ou la taille. De même, les ressorts de l'analyse transversale peuvent trouver à s'employer dans ces circonstances nouvelles pour elles. L'analyse transversale des démographes pourrait même s'enrichir : à côté de la transversalité verticale classique émerge une transversalité horizontale, dont la mise en évidence s'impose d'ailleurs de façon plus évidente en dehors du champ des variables temporelles classiques pour les démographes.

Finalement, le réflexe de lire et d'organiser les données en cohortes susceptibles d'adopter des comportements différenciés par rapport au phénomène étudié pourrait aussi influencer des méthodes d'analyse par ailleurs bien établies. Ainsi, la corrélation et la régression pourraient avoir intérêt à intégrer les notions de cohorte et de comportement de cohorte de façon à produire des mesures libérées de certains effets, ce qui pourrait en augmenter la qualité dans une perspective de comparaison, par exemple.

Comme on le voit à travers ces différents exemples, le champ d'application de la cohortisation s'élargit, au moins potentiellement, de façon importante ; par ailleurs, cette généralisation ne se limite pas à une simple transposition du savoir faire des démographes en changeant de variables puisqu'elle pourrait en plus se traduire par des ajouts à l'arsenal des méthodes démographiques et par des implications à propos de certaines méthodes de la statistique générale.

D'une manière générale, comment définir les circonstances où la cohortisation pourrait se révéler utile à l'analyse ? À notre avis, outre la condition préalable de la possibilité d'établir l'équivalent d'un diagramme de Lexis classique, deux conditions supplémentaires devraient intervenir :

- pour l'objectif ou la méthode de l'analyse en cause, la densité des points dans les différentes zones du graphique représentant les données est porteuse de sens ;
- même si cela pourra sembler tautologique, pour le phénomène étudié, on peut subodorer un comportement « grégaire » de cohorte ; en langage imagé, cela revient à dire qu'une idée du genre « les riches deviennent de plus en plus riches et les pauvres... » peut se défendre.

Moyennant ces deux conditions, la cohortisation devrait permettre de quantifier l'activité des phénomènes sur une base intéressante par contrôle de certains effets indésirables et par là, de faciliter certaines analyses, notamment en matière de comparaison.

Par ailleurs, rappelons qu'à la fin du premier point, il a été signalé que la cohortisation s'imposait en cas d'observation transversale et/ou perturbée. Au contraire, en cas d'observation à la fois longitudinale et à l'état pur, la cohortisation ne produira pas de résultats différents de ceux obtenus par calcul direct sur base des événements effectivement observés. Dans ces circonstances, seul le diagramme de Lexis apportera une éventuelle plus-value pour la représentation graphique condensée des données.

Les différents exemples développés ici montrent, à notre avis, la possibilité d'étendre l'application de la transversalité et de l'idée de la perturbation des observations à d'autres circonstances que celles auxquelles se confrontent habituellement les démographes. Densité des points, comportement de cohorte, état perturbé et transversalité-longitudinalité, voici des mots-clés qui ouvrent de larges perspectives d'utilisation pour les méthodes de l'analyse démographique classique.

## 5. Obstacles et problèmes à la généralisation

### 5.1. Un premier frein : la collecte des données

Pour recourir à la cohortisation, il faut que la collecte des données soit organisée en conséquence. En démographie, il a d'ailleurs fallu du temps et des colloques au dix-neuvième siècle pour que s'impose l'idée de la collecte des données non seulement en fonction de l'âge et du moment, mais aussi en fonction du moment de naissance<sup>34</sup>. Sans l'existence de ces données très spécifiques en matière de collecte, la démographie n'aurait pu développer son *corpus* de méthodes d'analyse basées sur la mesure des comportements au sein des cohortes.

Via la Banque de Données médico-sociales de l'ONE, nous avons eu accès à des données respectant l'exigence du triple classement en fonction d'une variable absolue (le temps, la taille...), de la valeur de cette variable absolue à un événement commun (le moment de naissance, la taille à la naissance...) et de la variation de cette variable absolue depuis l'événement commun (l'âge, la variation de taille depuis la naissance...). Toutefois, il n'en découle malheureusement que de faibles possibilités d'exploitation ; en effet, vu sa structure, un seul événement peut subir la cohortisation, à savoir la sortie du champ d'observation, puisque c'est le seul moment, avec la

---

34 À ce sujet, cf. notamment BUQUET L. (1974), p. 93 ou DUPÂQUIER J. et M. (1985), p. 310.



naissance, où les valeurs de certaines variables, comme le poids et la taille, s'enregistrent.

Pour étayer davantage les méthodes proposées ici tout en restant dans le même cadre général, il faudrait avoir accès à des données mettant en relation les événements de la vie des nourrissons, comme par exemple la contraction d'une maladie ou l'acquisition de certaines capacités, non seulement avec leur âge à ces événements, mais aussi avec leur taille ou leur poids. La cohortisation pourrait ainsi être menée en parallèle sur base de l'âge et le la variation de taille ou de poids<sup>35</sup>. Cet exercice serait d'autant plus indiqué que, pour les phénomènes à analyser, le développement « physiologique » de l'enfant serait un des facteurs explicatifs envisageables ; en effet, dans ces circonstances, les études se feraient donc non seulement en fonction du proxy de ce développement qu'est l'âge, mais aussi en fonction d'autres variables qui, à certains égards, pourraient en constituer de meilleurs révélateurs.

Pour assurer une cohortisation de qualité, il faudrait ne pas perdre de vue les éventuels phénomènes perturbateurs dont l'observation devrait, à l'instar du phénomène étudié, suivre la même exigence du triple classement selon le système de repérage utilisé (le temps, la taille ou le poids, pour reprendre les exemples déjà évoqués). Cette dernière condition est sans doute encore moins évidente à rencontrer que le triple classement pour un phénomène d'intérêt, sauf en ce qui concerne le temps.

Comme déjà signalé, chaque fois que l'on met en relation des données à propos d'une même variable à deux moments différents, les conditions nécessaires à la cohortisation sont rencontrées. Cette circonstance précise ouvre des perspectives intéressantes pour l'utilisation des méthodes proposées ici. Ainsi, sur base du couplage des revenus déclarés à deux dates différentes, il serait possible, via la cohortisation, de tester une idée du genre « les riches deviennent de plus en plus riches, les pauvres... » sur une base très fine, à l'image des analyses par âge des démographes et hors influence d'éventuelles perturbations, si les données le permettent.

## 5.2. Des observations limitées

En démographie, chaque année, des gens naissent, engendrant de nouvelles cohortes ; chaque année, des individus meurent ou migrent, modifiant ainsi l'effectif des cohortes dont ils ressortissent. La continuité du

---

35 Le « Dossier médical confidentiel », rempli par le médecin qui assure le suivi de l'enfant, représente dans ce domaine une source très intéressante. Toutefois, se pose le délicat problème du secret médical et donc de l'accès à ces données, qui, par ailleurs, ne jouissent pas d'une informatisation centralisée.

déroulement de ces processus rejailit sur le diagramme de Lexis, pour ne parler que de l'aspect graphique des choses : toutes les plages du graphique peuvent contenir des données. À l'inverse, la taille où le poids à la naissance ne peuvent prendre qu'une gamme finie de valeurs ; en conséquence, des parties des diagrammes de Lexis constitués sur cette base resteront définitivement vides, comme sur les figures 13, 15 et 16.

Cette différence ne nous semble en rien rédhitoire. Elle ne remet pas en cause, à notre avis, l'intérêt pour la cohortisation : ainsi, les analyses longitudinales et transversales gardent leur intérêt et surtout chacune conserve ses spécificités et sa capacité à répondre à des questions précises.

### 5.3. Types de trajectoire et problèmes spécifiques

Si sur le plan théorique, nous pensons avoir montré que des variables non temporelles pouvaient remplacer dans la cohortisation les variables classiques des démographes, il n'en faut pas moins souligner des différences. Dans le diagramme de Lexis classique, le temps, ainsi que les lignes de vie et l'âge, présentent un déroulement immuable, linéaire, et continu : ils progressent de manière régulière et inexorable. Qu'en est-il pour d'autres types de trajectoires ?

Au contraire du temps, la taille et le poids, par exemple, ne varient pas de manière continue et inexorable. Un individu peut voir stagner sa taille ou son poids pendant un certain temps, puis se modifier à des rythmes différents selon les périodes et les individus. Par ailleurs, si la taille est un phénomène irréversible, du moins durant une bonne part de la vie et en dehors de situations exceptionnelles, le poids, par contre, est réversible : un individu peut perdre du poids, comme c'est notamment le cas après la naissance ; il faut donc envisager des variations négatives, voire des variations négatives intermédiaires.

En reprenant l'exemple de la variation de l'espérance de vie entre 1950-1955 et 1995-2000 (*cf. supra*), il se peut que certains pays aient enregistré une valeur maximale avant 1995-2000 pour connaître ensuite une régression. Pour dépasser cette difficulté, au moins deux possibilités s'offrent à nous :

- plutôt que de faire une analyse sur l'ensemble de la période, la scinder en autant de séquences temporelles que les données permettent d'en distinguer ;
- introduire des sous-cohortes selon que les pays aient ou non connu des périodes de régression ; cette idée de sous-cohortes trouve son pendant immédiat en démographie classique : en cas d'analyse de la fécondité, par exemple, identification de sous-cohortes « qualitative » selon le statut matrimonial ou « quantitative » selon l'âge au mariage.

Une autre façon, certes moins élégante, consisterait à remplacer l'espérance de vie en 1950-1955 (base de la constitution des cohortes adoptée jusqu'ici) par cette valeur amputée de 5 ans, ce qui gommerait tous les chiffres négatifs observés et résoudrait ce point, mais pas celui du maximum temporellement intermédiaire. Dans la même ligne, pour le poids des enfants, on pourrait préférer le poids minimal de l'enfant à la fin de la période d'amaigrissement après la naissance à son poids à la naissance pour constituer les cohortes.

Si les trajectoires en tailles, en poids ou en espérance de vie présentent des spécificités par rapport aux lignes de vie, cela ne remet pas en cause les trajectoires en elles-mêmes ; de même, la localisation d'événements sur ces trajectoires reste licite sur le plan théorique et potentiellement significative sur le plan pratique. Les conséquences précises des spécificités des trajectoires sur la cohortisation et son cortège de calculs restent à évaluer dans le détail.

#### **5.4. Portée pratique de la généralisation de la cohortisation**

En analyse démographique classique, la pertinence du diagramme de Lexis et des calculs dérivés n'est plus à démontrer tant les phénomènes démographiques suivent de puissantes logiques temporelles. Par exemple, l'étude de la mortalité via ces méthodes permet de mettre de l'ordre dans les observations ; en effet, vont apparaître des logiques d'âge, des logiques de temps et des logiques générationnelles qui organisent les données en mettant en évidence des généralités ou des « lois » de distribution des événements.

Via le diagramme de Lexis en taille et les calculs dérivés, par exemple, on pourra voir si des logiques comparables se manifestent : les données s'organisent-elles en fonction de logiques de taille, de gain de taille depuis la naissance et de taille à la naissance ? Si ces logiques ne se dégagent pas de l'analyse, on pourra en conclure que le critère de la taille n'est pas significatif par rapport au phénomène étudié et l'abandonner. Si, au contraire, l'analyse permet d'isoler des logiques de taille, c'est que ce critère est potentiellement pertinent pour organiser les données et éventuellement comprendre le phénomène étudié. Si les démographes n'avaient pas constaté toute la force des variables temporelles pour organiser leurs observations en matière de mortalité, fécondité, mobilité spatiale..., ils auraient abandonné, dans leurs applications temporelles du moins, les outils qui leur sont devenus maintenant incontournables.

En analyse démographique, la cohortisation se justifie pleinement – car produisant des mesures de qualité des phénomènes – en cas d'observation transversale ou perturbée. Si ces deux circonstances n'ont plus à démontrer leur pertinence dans le champ de la démographie, qu'en est-il

ailleurs ? Le fait d'observer un phénomène d'intérêt en choisissant un cadre de référence non temporel n'élimine nullement le pouvoir perturbateur de certains phénomènes : par exemple, ce n'est pas parce que la contraction d'une maladie chez les nourrissons serait observée en fonction du poids, et non en fonction de l'âge, qu'une émigration perdrait son rôle perturbateur. À notre sens, il s'agit d'une circonstance où le contrôle des perturbations se révélera aussi – peu – pertinent indépendamment du cadre de référence choisi.

Par ailleurs, à notre avis, la transversalité des démographes peut également se justifier en cas de cadre de référence non temporel : est-il plus ou moins sensé de s'interroger sur le comportement des individus si l'observation porte sur une année en recoupant différentes cohortes de naissance que si elle devait porter sur des individus d'un poids donné en recoupant différentes cohorte de poids à la naissance ? Sur un plan théorique, nous ne voyons pas de raison d'établir une différence fondamentale entre ces deux types d'analyse.

En résumé, loin de se limiter aux seules variables temporelles classiques des démographes, la cohortisation<sup>36</sup> pourrait s'appliquer dans le cas d'autres types de variables.

## 6. Conclusions

La cohortisation peut être *a priori* considérée comme un outil spécifique à l'analyse démographique et, par association de situations, à toutes les circonstances d'analyse où le moment de survenance des événements, le moment de naissance des individus (ou celui de tout autre événement origine) et l'âge des individus (ou une autre durée si l'évènement origine n'est pas la naissance) permettent de dégager des logiques temporelles qui organisent l'observé.

Nous pensons avoir montré que la cohortisation, loin d'être une technique à réserver aux seules situations classiques de l'analyse démographique, pourrait trouver à s'employer dans d'autres domaines. Quelles sont en définitive les conditions nécessaires pour voir émerger depuis les données des cohortes auxquelles les techniques de l'analyse démographique pourraient s'appliquer ? Le système de repérage utilisé pour localiser les données doit se constituer de trois coordonnées présentant les caractéristiques suivantes :

---

36 Par ailleurs, si nos données le permettent en qualité et quantité, nous comptons procéder à des applications de modèles « APC » (« Âge, période, cohorte ») et identifier, autant que faire se peut, les équivalents des effets d'âge, période et cohorte en cas de variables non classiques.

- s'agencer dans une combinaison linéaire ;
- compter une coordonnée demeurant constante individu par individu.

Outre le système de repérage basé sur le temps et utilisé classiquement par les démographes (le temps, le moment de naissance et l'âge) ou ses formes dérivées (par exemple, l'âge, l'âge à un événement particulier et la durée écoulée depuis cet événement), nous en avons identifié d'autres respectant les conditions énoncées et laissant, par ailleurs, imaginer d'autres nombreux exemples. Ils se basent soit sur la taille, soit sur le poids soit sur l'espérance de vie. Le diagramme de Lexis et les méthodes de l'analyse démographique pourront donc s'appliquer dans ces circonstances, pour autant, bien sûr, que la collecte à leur endroit ait respecté l'exigence du triple classement. Il est aussi à noter que la présence éventuelle de phénomènes perturbateurs étend *de facto* l'exigence du triple classement aux événements les matérialisant ; sans cela, il ne sera pas envisageable d'inclure l'information les concernant dans l'établissement des mesures du phénomène étudié.

L'application de trois techniques ou méthodes plutôt démographiques a été testée sur ces systèmes de repérage inhabituels : la représentation du diagramme de Lexis d'événements réduits ; le calcul d'une table d'extinction avec présence d'un phénomène perturbateur en analyse longitudinale ; les analyses transversales de type vertical ou horizontal. Par ailleurs, il a aussi été montré que la corrélation et la régression pourraient éventuellement gagner à intégrer la notion de cohorte et de comportement de cohorte. Malgré certains obstacles, nous pensons avoir montré, au-delà de la simple possibilité technique, l'intérêt à un recours aux méthodes des démographes en dehors de leur champ habituel.

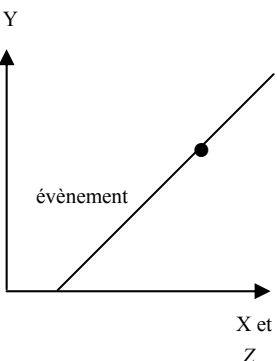
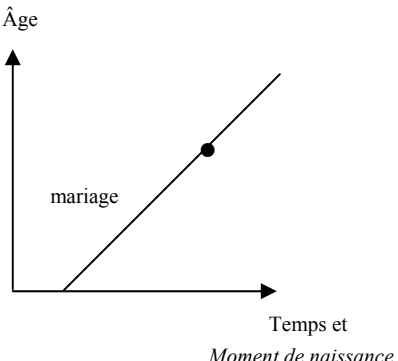
Cette généralisation et la plus value qu'elle pourrait apporter en matière de mesure et de compréhension de certains phénomènes mettraient la démographie dans une position de force, notamment dans le monde des sciences humaines, et, plus spécifiquement encore, en matière de pédagogie, tant sont évidents l'intérêt et la nécessité de la cohortisation pour analyser les observations de son champ d'activité et tant est relativement simple la compréhension de cette cohortisation quand les variables temporelles habituelles servent à définir les trajectoires sur lesquelles les événements se localisent.

### Bibliographie

- BECKER K. (1874), *Zur Berechnung von Sterbetafeln an die Bevölkerungsstatistik zu Stellende Anforderungen*, Berlin, Verlag des Königlichen statistischen Bureaus.
- BLAYO Ch. (1991), « Choix des cohortes et des sous-cohortes : règles générales et application à l'avortement », *Population*, 6, pp. 1379-1404.
- BRASCHE O. (1870), *Beitrag zur Methode der Sterblichkeitsberechnung*, Würzburg, A. Struber's Buchhandlung.
- BUQUET L. (1974), *Démographie*, Paris, Masson et Cie.
- CASELLI Gr., MESLÈ Fr. et VALLIN J., *Les entorses au schéma de la transition épidémiologique. Communication présentée à la séance 18 : « Nouvelles menaces sanitaires », du Congrès international de la population*. Salvador. Brésil. Août 2002, Rome/Paris, 44 p.
- DUPÂQUIER J. et M. (1985), *Histoire de la démographie*, Paris, Librairie Académique Perrin.
- HENRY L. (1952), *Fécondité des mariages, nouvelle méthode de mesure*, Travaux et Documents. Cahier n° 16, Paris, INED.
- HOBcraft J. MENKEN J. et PRESTON S. (1982), Age, Period, and Cohort Effects in Demography: A Review, *Population Index*, Volume 48, N° 1, pp. 4-43.
- LEXIS W. (1875), *Einleitung in die Theorie der Bevölkerungsstatistik*, Strasbourg, Karl J. Trübner.
- OFFICE DE LA NAISSANCE ET DE L'ENFANCE (2000), *Banque de données médico-sociales. Rapport 2000*, Bruxelles, Office de la Naissance et de l'Enfance.
- OFFICE DE LA NAISSANCE ET DE L'ENFANCE (2001), *Banque de données médico-sociales. Rapport 2001*, Bruxelles, Office de la Naissance et de l'Enfance.
- UNITED NATIONS (2001), *World Population Prospects. The 2000 Revision. Volume I : Comprehensive Tables*, New York, United Nations.
- VANDESCHRICK Chr. (1992), Le diagramme de Lexis revisité, *Population*, 5, 1992, pp. 1241-1262.

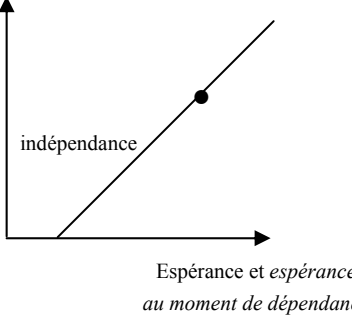
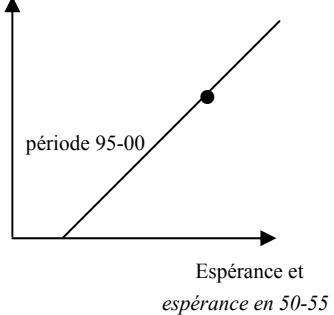
- VANDECHRICK Chr. (1994), Le temps dans le temps en démographie. Le diagramme de Lexis : bilan et perspectives, in VILQUIN E. (éd.), *Le temps et la démographie. Chaire Quetelet 1993*, Louvain-la-Neuve/Paris, Academia/L'Harmattan, pp. 271-307.
- VANDECHRICK Chr. (2000), *Analyse démographique*, Louvain-la-Neuve/Paris, Academia/L'Harmattan.
- VANDECHRICK Chr. (2001), The Lexis diagram, a misnomer, *Demographic Research*, Volume 4, Article 3, pp. 95-124 (adresse web : <http://www.demographic-research.org>).
- VAUPEL J., WANG Z., ANDREEV K. et YASHIN A. (1997), *Population data at a glance*, Odense; Odense University Press (adresse web : <http://www.demographic-research.org>).
- VERWEIJ Abr. (1874). *De waarnemingen der bevolkingsstatistiek. Academisch Proefschrift, ter verkrijging van den graad van doctor in de wis- en natuurkunde, aan de Hoogschool te Utrecht, te verdedigen op vrijdag den 18 december 1874. Des namiddags te 3 uren*, Deventer, Rustering en Vermandel.
- WUNSCH G. et TERMOTE M. (1978), *Introduction to demographic Analysis. Principles and Methods*, New York and London, Plenum Press.
- ZEUNER G. (1869), *Abhandlungen aus der Mathematischen Statistik*, Leipzig, Verlag von Arthur Felix.

## Annexe 1. Interprétation des points sur les trajectoires

	Cas 1
Phénomène étudié :	primo-nuptialité
Population observation :	femmes
Axe X :	temps, date
Axe Y :	âge (ou gain de temps depuis la naissance)
Axe Z (supprimé) :	moment de naissance
	
Évènement :	premier mariage
Un individu sous observation :	une femme
Trajectoire :	ligne de vie
<p>● =</p> <ul style="list-style-type: none"> <li>- un évènement</li> <li>- affectant un individu sous observation</li> <li>- localisé sur la trajectoire de l'individu</li> </ul> <p><b>ou</b></p> <p>situation d'un individu, au moment de l'évènement (sur sa trajectoire)</p> <p><i>ce qui se fait en fonction de</i></p> <ul style="list-style-type: none"> <li>- la valeur de X à l'évènement</li> <li>- la valeur de Y à l'évènement</li> <li>- la valeur de Z (k pour i)</li> </ul>	<p>● =</p> <ul style="list-style-type: none"> <li>le premier mariage</li> <li>d'une femme</li> <li>localisé sur la ligne de vie de la femme</li> </ul> <p><b>ou</b></p> <p>situation d'une femme au moment de son premier mariage (sur sa ligne de vie)</p> <p><i>ce qui se fait en fonction de</i></p> <ul style="list-style-type: none"> <li>- de la date <u>du mariage</u></li> <li>- de l'âge au mariage</li> <li>- la date de naissance de la femme</li> </ul>
Un individu dans DL =	une ligne de vie



## Annexe 1 (suite)

Cas 2	Cas 3
accès à l'indépendance	la période 95-00 (bizarre, n'est-il pas ?)
pays ayant été colonisés	pays
espérance de vie (ou « espérance »)	espérance de vie
gain d'espérance depuis la date de dépendance	gain d'espérance depuis 50-55
espérance au moment de la dépendance	espérance en 50-55
<p>Gain d'espérance depuis la dépendance</p>  <p>indépendance</p> <p>Espérance et <i>espérance</i> au moment de dépendance</p>	<p>Gain d'espérance depuis 50-55</p>  <p>période 95-00</p> <p>Espérance et <i>espérance en 50-55</i></p>
indépendance	la période 95-00
<i>un pays</i>	<i>un pays</i>
ligne d'espérance de vie	ligne d'espérance de vie
<p>● = l'indépendance d'un pays localisée sur la ligne d'espérance du pays</p> <p><b>ou</b></p> <p>situation d'un pays au moment de son indépendance (sur sa ligne d'espérance de vie)</p> <p><i>ce qui se fait en fonction de</i></p> <ul style="list-style-type: none"> <li>- de l'espérance <u>moment de l'indépendance</u></li> <li>- du gain d'espérance depuis la dépendance</li> <li>- de l'espérance à la dépendance du pays</li> </ul>	<p>● = la période 95-00 pour un pays localisée sur la ligne d'espérance du pays</p> <p><b>ou</b></p> <p>situation d'un pays en 95-00 (sur sa ligne d'espérance de vie)</p> <p><i>ce qui se fait en fonction de</i></p> <ul style="list-style-type: none"> <li>- de l'espérance <u>en 95-00</u></li> <li>- du gain d'espérance depuis 50-55</li> <li>- de l'espérance en 50-55 du pays</li> </ul>
une ligne d'espérance de vie	une ligne d'espérance de vie

**Annexe 2. Calcul des gains d'espérance sans et avec cohortisation**

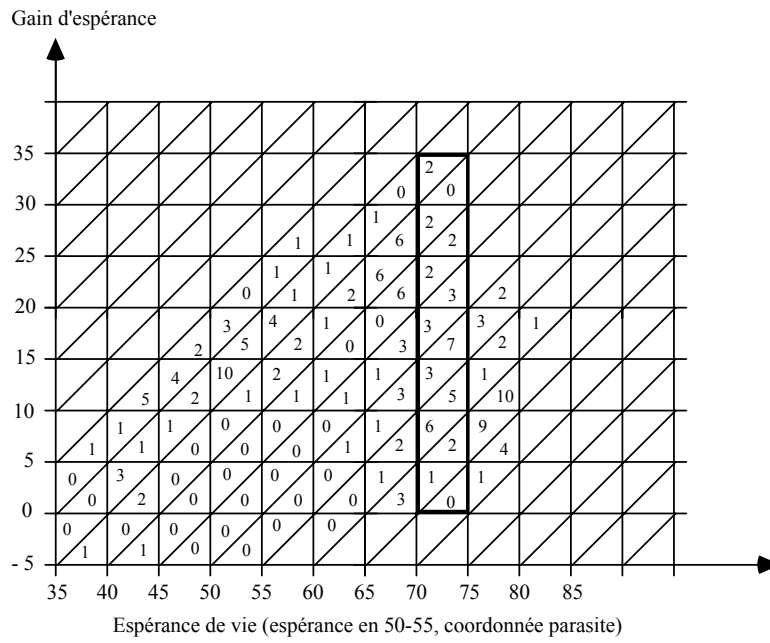
1 Cohorte	2 Effectif de la cohorte	3 Triangle	4 Gain moyen	5 Evène- ments	6 4*5	7 Evènements réduits (5/2)	8 7*4
65-<70	25	1	3,333	1	3,333	0,040	0,133
65-<70	25	2	6,667	2	13,333	0,080	0,533
60-<65	18	1	8,333	6	50,000	0,333	2,778
60-<65	18	2	11,667	5	58,333	0,278	3,241
55-<60	20	1	13,333	3	40,000	0,150	2,000
55-<60	20	2	16,667	7	116,667	0,350	5,833
50-<55	11	1	18,333	3	55,000	0,273	5,000
50-<55	11	2	21,667	3	65,000	0,273	5,909
45-<50	13	1	23,333	2	46,667	0,154	3,590
45-<50	13	2	26,667	2	53,333	0,154	4,103
40-<45	25	1	28,333	2	56,667	0,080	2,267
40-<45	25	2	31,667	0	0,000	0,000	0,000
35-<40	32	1	33,333	2	66,667	0,063	2,083
Somme				38	625,000	2,227	37,470
Moyenne					16,447		16,827

**Contenu des colonnes :**

- 1 : espérance de vie en 1950-1955 ;
- 2 : nombre de pays composant la cohorte ;
- 3 : triangle 1 = triangle entre un gain révolu et un gain exact ; triangle 2 = « triangle inverse ».
- 4 : gain moyen dans le triangle sous hypothèse de répartition uniforme ;
- 5 : nombre d'évènements de chaque triangle ;
- 7 : nombre d'évènements d'un triangle divisé par l'effectif de la cohorte.

**Localisation des données sur diagramme de Lexis : cf. page suivante.**

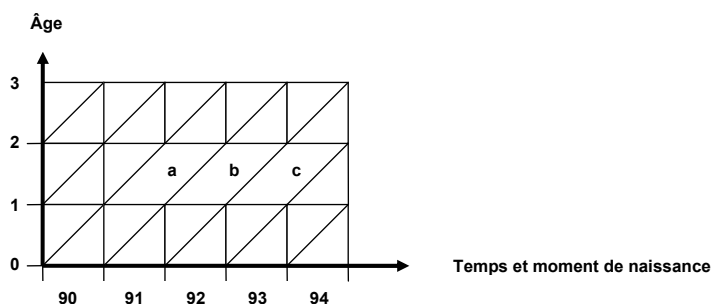
**Figure 21. Disposition des évènements sur diagramme de Lexis**



### Annexe 3. Équivalence des transversalités horizontales des deux types

Soit à calculer parmi les individus décédés à 1 an révolu la date moyenne de leur décès (transversalité horizontale de 1<sup>er</sup> type) et la date moyenne de leur naissance (transversalité horizontale de 2<sup>ème</sup> type). Comme suggéré par la figure 22, nous supposons que seules trois cohortes ont subi l'observation ; les lettres dans les losanges renseignent indifféremment les effectifs absolus de décédés ou les événements réduits y correspondant. Par ailleurs, à chaque fois que nécessaire, nous poserons l'hypothèse de répartition uniforme des événements.

Figure 22. Transversalité horizontale sur diagramme de Lexis



Quelle est la date moyenne au décès ? Pour les  $a$  événements du premier losange, il faut compter 92 (soit le 1/1/92) ; pour les  $b$  du deuxième, 93 et pour les  $c$  du troisième, 94. La date moyenne vaudra donc :

$$\bar{d} = \frac{(92 * a) + (93 * b) + (94 * c)}{a + b + c}$$

La date moyenne de naissance des décédés s'obtiendra en comptant respectivement 90,5 ; 91,5 et 92,5 pour  $a$ ,  $b$  et  $c$  :

$$\bar{dn} = \frac{(90,5 * a) + (91,5 * b) + (92,5 * c)}{a + b + c}$$

Ces deux formules diffèrent au niveau des coefficients affectant les nombres d'événements ou événements réduits des losanges. De façon précise, en passant de la 1<sup>ère</sup> à la 2<sup>ème</sup> formule, les coefficients diminuent de 1,5 ans, soit précisément l'âge moyen au décès des décédés à 2 ans révolus. Ce résultat est logique vu la relation existant entre les 3 variables en jeu :

$$\text{âge au décès} = \text{date du décès} - \text{date de naissance.}$$

Les deux types de calculs transversaux horizontaux donnent bien des résultats équivalents.