

Population et conditions sociales 3/2003/E/n° 26

Méthodologie relative au calcul des indicateurs démographiques d'Eurostat

Rapport détaillé de l'Observatoire
Démographique Européen

G. Calot, J.-P. Sardon - ODE



COMMISSION
EUROPÉENNE



THEME 3
Population
et conditions
sociales

3

Europe Direct est un service destiné à vous aider à trouver des réponses aux questions que vous vous posez sur l'Union européenne.

**Un nouveau numéro unique gratuit:
00 800 6 7 8 9 10 11**

De nombreuses autres informations sur l'Union européenne sont disponibles sur l'internet via le serveur Europa (<http://europa.eu.int>).

Luxembourg: Office des publications officielles des Communautés européennes, 2004

ISBN 92-894-8007-6
ISSN 1725-0668

© Communautés européennes, 2004

Population et conditions sociales 3/2003/E/n° 26

**Méthodologie relative au calcul des
indicateurs démographiques d'Eurostat -**
Rapport détaillé de l' Observatoire Démographique
Européen
G. Calot, J.-P. Sardon - ODE

Les idées exprimées dans ce document sont celles de l'auteur et ne reflètent pas nécessairement le point de vue de la Commission européenne.

Copyright: Commission européenne 2003

Table des matières

Introduction	p. 1
L'âge à la survenance d'un événement	p. 5
L'âge moyen à la survenance d'un événement	p. 6
La conversion des distributions par âge de la population résidente à une date autre que le 1 ^{er} janvier en distributions au 1 ^{er} janvier	p. 8
La calcul des taux bruts annuels	p. 9
Le rang des naissances vivantes	p. 9
L'analyse des données journalières	p. 10
L'analyse des divorces	p. 11
Chapitre 1 : Opérations préalables au traitement des données dans la chaîne d'exploitation	p. 12
<u>A) Eclatement par année d'âge des données pluri-annuelles</u>	p. 13
<i>I – Problèmes affectant la répartition par âge des effectifs de la population</i>	p. 13
1- La répartition des effectifs de population est connue dans un découpage non annuel de l'âge (le plus souvent quinquennal)	p. 13
a) Multiplicateurs de Sprague	p. 13
<i>Séquence des opérations</i>	p. 14
b) Amélioration des estimations obtenues à l'aide des multiplicateurs de Sprague	p. 15
<i>Séquence des opérations</i>	p. 15
2- Pyramides d'âge disponibles par année d'âge pour les années encadrant celle pour laquelle on ne dispose que de la répartition par groupe d'âge	p. 15
<i>Séquence des opérations</i>	p. 16
3- Pyramides d'âge disponibles par année d'âge pour l'année précédant immédiatement celles pour lesquelles on ne dispose que de la répartition par groupe d'âge	p. 16
<i>Séquence des opérations</i>	p. 17
4- Eclatement par année d'âge des pyramides dont le dernier groupe est antérieur à « 99 ans ou plus »	p. 17
<i>Séquence des opérations</i>	p. 17
<i>II – Problèmes affectant la répartition par âge des événements démographiques</i>	p. 18
1- Les nombres d'événements démographiques selon l'âge sont connus dans un découpage non annuel de l'âge (le plus souvent quinquennal)	p. 18
a) Pyramides d'âge disponibles au 1 ^{er} janvier par année d'âge	p. 19
<i>Séquence des opérations</i>	p. 19
b) Pyramides d'âge disponibles au 1 ^{er} janvier, seulement par groupe d'âge	p. 19
<i>Séquence des opérations</i>	p. 20
<i>L'« ajustement » à une loi de Gompertz</i>	p. 20

<i>L'« ajustement » d'une distribution empirique à une loi de Gompertz</i>	p. 21
B) <u>Autres corrections à apporter aux données de base</u>	p. 23
1- Redressement des non-déclarés	p. 23
<i>Séquence des opérations</i>	p. 23
2- Conversion des pyramides à une date autre que le 1 ^{er} janvier	p. 23
a) Naissances mensuelles disponibles	p. 23
<i>Séquence des opérations</i>	p. 24
b) Naissances mensuelles non disponibles	p. 24
3- Les estimations intercensitaires de population	p. 24
a) Aucune révision intercensitaire	p. 25
<i>Séquence des opérations</i>	p. 26
b) Révision du seul effectif total de la population	p. 26
<i>Séquence des opérations</i>	p. 27
4- Mise en accord des marges des différents tableaux concernant une même variable	p. 27
C) <u>Estimation des données manquantes</u>	p. 28
1- Trou dans une série chronologique	p. 28
a) La distribution des événements par âge (ou par durée) n'est pas disponible, mais le total annuel est disponible	p. 28
<i>Séquence des opérations</i>	p. 28
b) La distribution des événements par âge (ou par durée) n'est pas disponible, pas plus que le total annuel	p. 29
<i>Séquence des opérations</i>	p. 29
c) L'effectif de la population par âge (ou par durée) n'est pas disponible, mais le total annuel est disponible	p. 29
<i>Séquence des opérations</i>	p. 30
d) L'effectif de la population par âge (ou par durée) n'est pas disponible, pas plus que le total annuel	p. 30
<i>Séquence des opérations</i>	p. 30
2- Données manquantes en fin de série (en vue d'une agrégation géographique)	p. 30
a) L'effectif de la population n'est pas disponible	p. 30
<i>Séquence des opérations</i>	p. 31
b) Les événements démographiques ne sont pas disponibles	p. 31
<i>Séquence des opérations</i>	p. 32
Chapitre 2 : La construction des tables de survenance d'événements	p. 33
A) <u>Les différents types d'événements</u>	p. 33
1- Événements renouvelables	p. 33
2- Événements non renouvelables	p. 34
3- L'observation des nombres absolus d'événements et des effectifs soumis au risque	p. 35
B) <u>Événements renouvelables</u>	p. 36
1- Nombre de personnes-années d'exposition au risque	p. 36
<i>Séquence des opérations</i>	p. 36

2- Estimation des taux et construction des tables de survénance pour les événements renouvelables (comme la fécondité)	p. 37
3- Calcul des taux (de seconde catégorie)	p. 42
a) Age en années révolues	p. 42
Méthode fruste	p. 42
Méthode élaborée	p. 42
• Constance du risque à l'intérieur du carré	p. 42
• Prise en compte de la variabilité du risque selon l'âge à l'intérieur du carré	p. 48
b) Age atteint dans l'année	p. 50
Méthode fruste	p. 50
Méthode élaborée	p. 51
4- Conversion des événements et des taux dans une autre figure de Lexis	p. 51
<i>Séquence des opérations</i>	p. 51
C) Événements non renouvelables	p. 53
1- Estimation des quotients et construction des tables de survénance pour les événements non renouvelables (comme la mortalité)	p. 53
a) Age en années révolues	p. 53
Méthode fruste	p. 53
Méthode élaborée	p. 54
b) Age atteint dans l'année	p. 55
Méthode fruste	p. 55
Méthode élaborée	p. 55
c) Spécificité de la mortalité à 0 et 1 an	p. 56
<i>Séquence des opérations</i>	p. 57
D) Tables de fécondité par rang	p. 60
E) Table de divortialité	p. 62
<i>Séquence des opérations</i>	p. 65
Chapitre 3 : Les indicateurs déduits des tables de survénance	p. 67
1- Indicateurs dérivés déduits des tables	p. 67
2- Précautions à prendre pour le calcul de ces indicateurs	p. 70
3- Autres indicateurs dérivés	p. 70
a) Les taux bruts	p. 70
b) L'effectif moyen des générations soumises au risque	p. 71
c) Recombinaison longitudinale des taux et quotients : les indicateurs longitudinaux	p. 71
d) La construction d'indicateurs dérivés mensuels	p. 72
e) L'extrapolation à court terme des indicateurs transversaux annuels	p. 72
<i>Formulation des indicateurs dérivés</i>	p. 73
Conclusion	p. 74
Annexe 1 : The construction of life tables	p. 75
Annexe 2 : L'analyse démographique conjoncturelle	p.111
Index des indicateurs	p.136

Méthodologie relative au calcul des indicateurs démographiques d'Eurostat

Rapport détaillé de l'ODE

par Gérard Calot et Jean-Paul Sardon
Observatoire Démographique Européen

Juin 2002

Introduction

L'organisation de la collecte d'informations démographiques, auprès des Offices Nationaux de Statistique, en vue de la gestion permanente d'une base de données *internationale*, contenant à la fois des informations *de base* et des informations *dérivées* (informations résultant de la combinaison d'informations de base), est conditionnée par les réponses données à une série de questions préalables.

- Quels *domaines* d'analyse sont couverts par la base de données ?

Les domaines d'analyse couverts par la base de données, pour ce qui est de chacun des *Etats*¹ concernés, et mis à part les migrations internationales dont l'approche très particulière justifie qu'il n'en soit pas traité dans la suite de ce document, sont : l'effectif et les structures générales de la population résidente, la primo-nuptialité, le divorce, la fécondité et la mortalité.

Le degré de détail démographique avec lequel chacun de ces domaines est étudié est précisé lors de l'énoncé des informations de base. Il est clair cependant qu'un organisme international ne saurait entrer dans un très grand détail, que l'hétérogénéité des définitions nationales rendrait d'ailleurs rapidement illusoire.

La *période* couverte par la base de données variera d'un pays à l'autre et d'une information à l'autre, selon l'étendue des données disponibles. Il est souhaitable de réunir le maximum d'informations de base sur la période écoulée depuis 1960. Toutefois, des données plus anciennes remontant à la Seconde guerre mondiale, voire à la Première ou même au début du XX^{ème} siècle, peuvent être incluses dans la base lorsqu'elles existent. Certaines analyses, telles celle du vieillissement de la population ou l'établissement de données longitudinales, nécessitent en effet la mobilisation d'informations observées sur longue période.

- Quelles *définitions* adopter en matière d'*unités statistiques* (définition, par exemple, de la personne *résidente*, de la *naissance vivante*, du *mort-né*, ...) et en matière de *variables statistiques* (définition de l'*âge* d'une personne à la survenance d'un événement, du *rang* d'une naissance vivante, ...) ?

Si, concernant les *unités* statistiques, il existe des recommandations internationales élaborées par l'ONU, qui sont dans l'ensemble assez bien suivies par les pays européens², il n'en

¹ Nous ne traiterons pas ici des bases de données à l'échelon infra-national.

² Encore que certaines difficultés subsistent, notamment en ce qui concerne la notion de population *résidente*, en particulier dans le cas des pays qui ont été longtemps des pays d'*émigration* et où existe un registre de population :

existe guère en matière de *variables* statistiques et les définitions adoptées dans les divers pays peuvent différer notablement. Généralement, une seule définition existe pour un pays donné (pour certains, elle a changé à quelque moment du passé) et ce sont les informations de base répondant à cette définition qu'il convient de collecter, d'insérer dans la base de données et, le cas échéant, de publier avec une note précisant la définition retenue. Mais, au niveau des informations dérivées, il est souhaitable, chaque fois que cela est possible, de présenter les indicateurs résumés de *tous* les pays de telle sorte qu'ils correspondent à la *même* définition et soient ainsi *directement comparables*.

- Selon quelle *périodicité* les informations de base sont-elles collectées ?

Il semble raisonnable de collecter les informations de base selon une périodicité *annuelle*, à une date de l'année telle que la plupart des pays disposent d'un lot de données annuelles *nouvelles* par rapport à leur envoi antérieur et si possible de données *définitives* (l'idéal est qu'ils disposent des données *définitives* de l'année *précédente*).

Mais le rythme d'élaboration des données annuelles diffère selon les pays. Il faudra donc, s'agissant notamment d'établir des informations de base ou dérivées par *agrégation géographique*, par exemple pour l'*ensemble* de l'Union Européenne³, gérer au mieux la *diversité*, à l'issue d'une collecte annuelle, des millésimes les plus récents pour lesquels une information donnée est disponible.

- Collecte-t-on des données exclusivement *annuelles* ?

Il paraît judicieux, dans une publication *annuelle*, de faire figurer certaines données *mensuelles*, de façon à permettre de *dater* avec un détail plus précis qu'à l'échelle annuelle les retournements ou les variations brusques des indicateurs dérivés les plus usuels, suite notamment à des changements de législation ou à des événements politiques ou sociaux particuliers.

En outre, comme on le verra plus loin, les données mensuelles sur longue période permettent d'améliorer le calcul des tables (de primo-nuptialité, fécondité et mortalité) et de faciliter la conversion des distributions annuelles par sexe et âge de la population résidente à une date *autre* que le 1^{er} janvier en des distributions au 1^{er} janvier.

D'autre part, le suivi des indicateurs conjoncturels *mensuels* permet, à tout moment, de disposer d'une *projection tendancielle* du phénomène étudié, ce qui fournit un moyen commode et immédiat pour estimer des données *manquantes* de façon, lorsqu'il en est besoin, à estimer notamment le nombre *annuel* d'événements et l'indicateur conjoncturel correspondant dans le pays considéré, puis de procéder à d'éventuelles agrégations géographiques.

Enfin, la publication régulière d'informations *conjoncturelles*, en particulier sous la forme de *graphiques* d'indicateurs *mensuels*, permet d'intéresser la presse et les médias en général aux évolutions démographiques à moyen terme et de promouvoir ainsi la base de données auprès d'un public plus large que le cercle des démographes.

la population résidente est alors souvent la population *de droit*, celle figurant dans le registre et donc non radiée du registre après émigration, alors qu'il serait souhaitable qu'elle fût la population de *résidence habituelle*. En effet, le point important n'est pas tant la *définition* de la population de référence que sa *cohérence* avec les flux observés par le système statistique. Or ces flux ne peuvent que difficilement porter sur des événements qui se produisent *hors du territoire national*, même si des efforts sont faits pour collecter des données à l'étranger par le canal des consulats.

³ Nous envisageons ici l'*agrégation géographique* des données au niveau de l'Union Européenne, mais le problème est le même pour n'importe quel ensemble d'Etats : pays de l'Union économique et monétaire, pays de l'Espace économique européen, etc.

- Quels outils de *validation* des données utilise-t-on *avant* insertion dans la base de données ?

Un examen critique des informations nouvellement reçues est indispensable *avant* leur introduction dans la base de données : contrôle de la *cohérence*, logique ou comptable, d'un lot d'informations annuelles, contrôle de la *vraisemblance* de chacune des données. Le jugement qu'il convient d'exercer en matière de vraisemblance des données se fonde le plus souvent sur la *régularité chronologique* de certains indicateurs dérivés et sur la nécessité d'une explication satisfaisante des éventuelles discontinuités (qui peuvent tenir à des changements de définition, à la prise en compte des résultats d'un nouveau recensement sans qu'on ait révisé les effectifs intercensitaires par sexe et âge, mais aussi ... à des *erreurs matérielles*, en particulier de saisie ou de confection de tableaux électroniques).

Le contrôle *a posteriori* de la qualité de la base doit aussi être organisé et donner lieu épisodiquement à des confrontations avec les publications nationales et à un examen périodique de la régularité de différentes séries d'indicateurs dérivés.

- Comment s'assurer que les gestionnaires de la base de données sont informés des *révisions ultérieures* touchant les informations de base ?

Lorsqu'un office national est amené à réviser ses données de base, par exemple lorsque des données *définitives* sont substituées à des données jusque-là *provisoires*, ou encore lorsque, suite aux résultats d'un nouveau recensement, de nouvelles évaluations annuelles de la population résidente par sexe et âge sont établies pour toute la période intercensitaire, il est indispensable que l'organisme international gérant la base de données soit informé de cette révision et que les nouvelles données soient insérées dans la base.

S'il est tout à fait exclu de demander aux offices nationaux de vérifier, ne serait-ce qu'une seule fois, le contenu intégral de la base de données pour leur pays, la vérification de quelques grandeurs clés annuelles (population totale, nombres annuels d'événements, ...) peut se révéler utile pour détecter les révisions inconnues de la base de données.

- Selon quelle *méthodologie* convient-il d'établir les informations dérivées ?

Les choix méthodologiques concernant le calcul des indicateurs dérivés nous semblent devoir être guidés par quatre ordres de considérations :

Comparabilité internationale des résultats

Qualité statistique des méthodes employées

Simplicité de ces méthodes

Disponibilité et fourniture, sur demande, d'informations présentant ces méthodes

Les différences, même de faible importance, entre les indicateurs établis et publiés par les offices nationaux et ceux résultant de la méthodologie adoptée à l'échelon international posent un problème délicat. A notre sens, le principe de *comparabilité* des résultats, tant d'une année à l'autre pour un même pays que, *surtout*, d'un pays à l'autre, doit, en dernier ressort, l'emporter sur d'autres considérations et conduire les offices nationaux à comprendre et à accepter qu'un organisme international élabore et publie des informations dérivées pouvant différer *un peu*⁴ de celles figurant dans les publications nationales. Une note indiquant l'éventualité de telles différences peut cependant être insérée dans les publications.

Les objectifs de comparabilité des résultats et de qualité statistique se recouvrent en grande partie : la comparabilité des résultats est une qualité essentielle pour une bonne méthode statistique (absence de biais, correction des effets perturbateurs, stabilité et régularité des résultats), surtout dans un contexte international. En revanche, qualité statistique et

⁴ Elles ne sauraient différer *beaucoup*, sauf erreur dans la mise en œuvre de l'une ou l'autre des méthodes.

simplicité sont loin d'aller automatiquement de pair. S'il est toujours possible d'exposer clairement les buts d'une méthode, d'expliquer les raisons qui conduisent à prendre en compte tels ou tels éléments, de présenter les propriétés des résultats à attendre, il ne serait pas réaliste d'espérer obtenir en toutes circonstances des résultats élaborés par application d'une formule simple. Des compromis peuvent certes être recherchés entre qualité statistique et simplicité, mais ce ne devrait pas être au prix d'une détérioration de la qualité statistique.

Les problèmes méthodologiques que pose la constitution d'une base de données internationale ne concernent pas que les règles de calcul des indicateurs dérivés. En voici quelques-uns qui portent sur les informations de base elles-mêmes :

Comment estimer la répartition par *année d'âge* de la population résidente appartenant au groupe *terminal* d'âge fourni par l'office national (par exemple *75 ans ou plus*) ? Cette répartition est nécessaire à une construction *esthétique* de la pyramide des âges du pays, mais aussi souhaitable pour construire les tables de mortalité par âge et indispensable pour agréger géographiquement les effectifs par année d'âge.

Comment estimer une répartition par *simple année d'âge* lorsqu'on ne dispose, pour une année donnée, que de totaux par *groupe quinquennal* d'âge (par exemple en matière de naissances selon l'âge de la mère : cas notamment de l'Espagne avant 1971) ? Cette répartition est indispensable au calcul des indicateurs *longitudinaux*, c'est-à-dire par année de naissance. De façon analogue, comment désagréger par *simple année d'âge* un effectif se rapportant à un groupe d'âge, par exemple un effectif *quinquennal*, de population résidente ?

Comment redresser les cas de *non déclaration* dans les distributions statistiques faisant intervenir une ou plusieurs variables ?

Comment résoudre la *non cohérence* entre informations de base (par exemple : total *annuel* des nombres d'événements *mensuels* différent du total *tous âges* des nombres d'événements par *âge*, ou par *âge et rang de naissance*) ?

Comment, d'une manière plus générale, pallier diverses sortes de *lacunes* dans l'information disponible pour un pays donné à un moment donné et, en particulier :

Comment estimer telle ou telle information de base pour un pays qui ne l'a pas encore établie, de façon à permettre une agrégation géographique ?

Comment estimer le nombre *annuel* d'événements lorsqu'on ne dispose pas encore des données mensuelles se rapportant aux *derniers* mois de l'année ?

Comment estimer l'indicateur conjoncturel de primo-nuptialité ou de fécondité quand on ne dispose que du *nombre total* (éventuellement provisoire) d'événements de l'année et pas encore de la *répartition par âge* de ces événements ?

Comment estimer les espérances de vie masculine et féminine à la naissance quand on ne dispose que du *nombre total de décès* de l'année et pas encore de sa *répartition par sexe et âge* ?

Comment estimer la descendance finale d'une génération qui n'a pas encore atteint 50 ans l'année où ont été observées les données les plus récentes ?

Comment estimer la proportion de non-célibataires à 50 ans pour une génération qui n'a pas encore atteint cet âge dans les dernières statistiques disponibles ?

La nature des informations de base disponibles pour chaque pays et chaque année pouvant varier d'un pays à l'autre, voire d'une année à l'autre pour un même pays, il est nécessaire que le système de traitement des informations de base permettant d'obtenir les informations dérivées soit assez souple pour fournir *en toutes circonstances* les meilleurs résultats

possibles compte tenu des données qui sont disponibles au moment où on réalise ce traitement.

Il ne faut ainsi pas perdre de vue que la gestion d'une base de données forme un *tout* et que les choix effectués en amont du calcul des indicateurs dérivés conditionnent ce calcul.

L'âge à la survenance d'un événement

L'un des tout premiers problèmes que l'on rencontre en matière de *comparabilité* internationale dans le domaine de la démographie est celui de la définition de l'*âge* d'une personne à la survenance d'un événement : âge au mariage (d'un nouveau marié), âge de la mère à la naissance d'un enfant, âge au décès d'un décédé, ancienneté d'un mariage à sa dissolution par divorce, ...

L'*âge du langage courant* est l'*âge en années révolues* au moment de l'événement, c'est-à-dire l'*âge au dernier anniversaire*. Cette définition de l'âge est celle que retiennent certains pays, tel le Royaume-Uni, pour établir leurs statistiques de mariages, de naissances ou de divorces. La définition de l'âge adoptée par d'autres pays est celle qui déduit l'âge de l'*année de naissance* : différence, par exemple en matière de fécondité, entre le millésime de l'année de naissance de l'enfant et le millésime de l'année de naissance de la mère. Ce dernier âge, encore appelé âge *atteint* (sous-entendu : durant l'année civile de l'événement), est en moyenne *inférieur* d'une demi-année à l'âge révolu.

Il arrive qu'il soit spécialement important de préciser à quelle définition de l'âge on se réfère. C'est ainsi que la fécondité des *adolescentes* diffère fréquemment de plus d'un *quart*, selon qu'on la mesure à moins de 20 ans en âge *révolu* ou à moins de 20 ans en âge *atteint* : les comparaisons internationales sont entièrement faussées si on ne prend pas garde à cette différence.

Certains pays fournissent des données *à la fois* par âge révolu et par âge atteint (*triangles* du diagramme de Lexis). Tel est très souvent le cas des pays européens en matière de *décès*, mais moins fréquemment en matière de mariages, de naissances ou de divorces, notamment pour les années les moins récentes.

En conséquence, si on veut assurer la *comparabilité* internationale des résultats, il faut adopter l'une des deux conventions concernant l'âge. Celle de l'âge *atteint* est préférable : d'une part l'addition longitudinale des taux par âge atteint fournit des indicateurs qui se rapportent à une année de naissance *unique*, d'autre part les *biais* dont il est fait mention ci-dessous sont moindres s'agissant des taux par âge atteint que des taux par âge révolu.

Mais alors la question se pose, pour un pays qui, une année donnée, dispose *seulement* d'informations par âge *révolu*, de *convertir* ces informations en âge *atteint*. Cette conversion est nécessaire si on veut le comparer aux autres pays (ou le comparer à lui-même à travers le temps, dans le cas où, à quelque moment du passé, il a changé sa définition de l'âge), mais aussi pour agréger géographiquement les données d'une même année.

L'Observatoire Démographique Européen (ODE) a élaboré une méthodologie de la construction des tables de primo-nuptialité, fécondité et mortalité, dans les deux définitions de l'âge. De cette façon, les *deux* systèmes de taux (et de quotients en matière de mortalité), par âge révolu et par âge atteint, sont établis, et les nombres absolus d'événements sont estimés par *triangle* du diagramme de Lexis quand les données de base sont observées *uniquement* par âge révolu ou *uniquement* par âge atteint, ce qui permet les agrégations géographiques dans les deux définitions.

Mais, ce qui est au moins aussi important, cette méthodologie permet d'une part de *lisser* les courbes de taux et, d'autre part et *surtout*, de *corriger* certains taux affectés de *biais* statistiques notables, pouvant dépasser 10%, dus à des accidents historiques (les deux

guerres mondiales notamment, surtout la première). Ces biais tiennent au régime saisonnier exceptionnel de la natalité durant les années marquées par ces accidents historiques. Les taux, à *tous* les âges, des générations nées durant ces années-là sont biaisés, en valeur relative d'une quantité qui, pour une même génération, varie peu avec l'âge et avec le phénomène étudié (primo-nuptialité, fécondité, mortalité). Les indicateurs *longitudinaux* de ces générations sont en conséquence entachés de biais que la méthode adoptée élimine de manière satisfaisante. En revanche, les indicateurs *transversaux* sont assez peu affectés : la correction apportée à un indicateur conjoncturel ou à un âge moyen transversal est généralement faible.

Pour appliquer la méthode, la connaissance de la répartition *mensuelle* des naissances sur longue période est souhaitable, mais il s'agit là d'une information dont disposent presque tous les pays européens. A défaut, la répartition, observée dans un recensement, de la distribution de la population par *année et mois* de naissance fournit les éléments requis. Si on ne dispose même pas de cette informations, on admet l'*uniformité* du régime saisonnier, ce qui permet d'améliorer l'estimation des taux par rapport aux résultats du calcul fruste (taux non corrigés et non lissés).

La méthodologie du calcul des taux est assez souple pour permettre d'obtenir, quelles que soient les informations disponibles au moment où est effectué ce calcul, la meilleure estimation possible de chaque taux. Sur option, elle fournit aussi les taux non corrigés et non lissés.

Il convient de faire observer que l'estimation des nombres de naissances en triangles du diagramme de Lexis, obtenue en divisant par 2 le nombre observé dans le carré, conduit à des nombres absolus par parallélogrammes, puis à des taux par âge atteints établis par la méthode classique, dont l'irrégularité est manifeste lorsqu'on les compare à ceux fournis par la méthode de l'ODE (cf. graphique 1 page suivante). Or l'estimation des nombres absolus d'événements en parallélogrammes est indispensable pour permettre de réaliser les agrégations géographiques.

En matière de mortalité, le document, placé en annexe, intitulé *The construction of life tables*, présente les méthodes établies par l'ODE pour construire les tables de mortalité par sexe et âge.

L'âge moyen à la survenance d'un événement

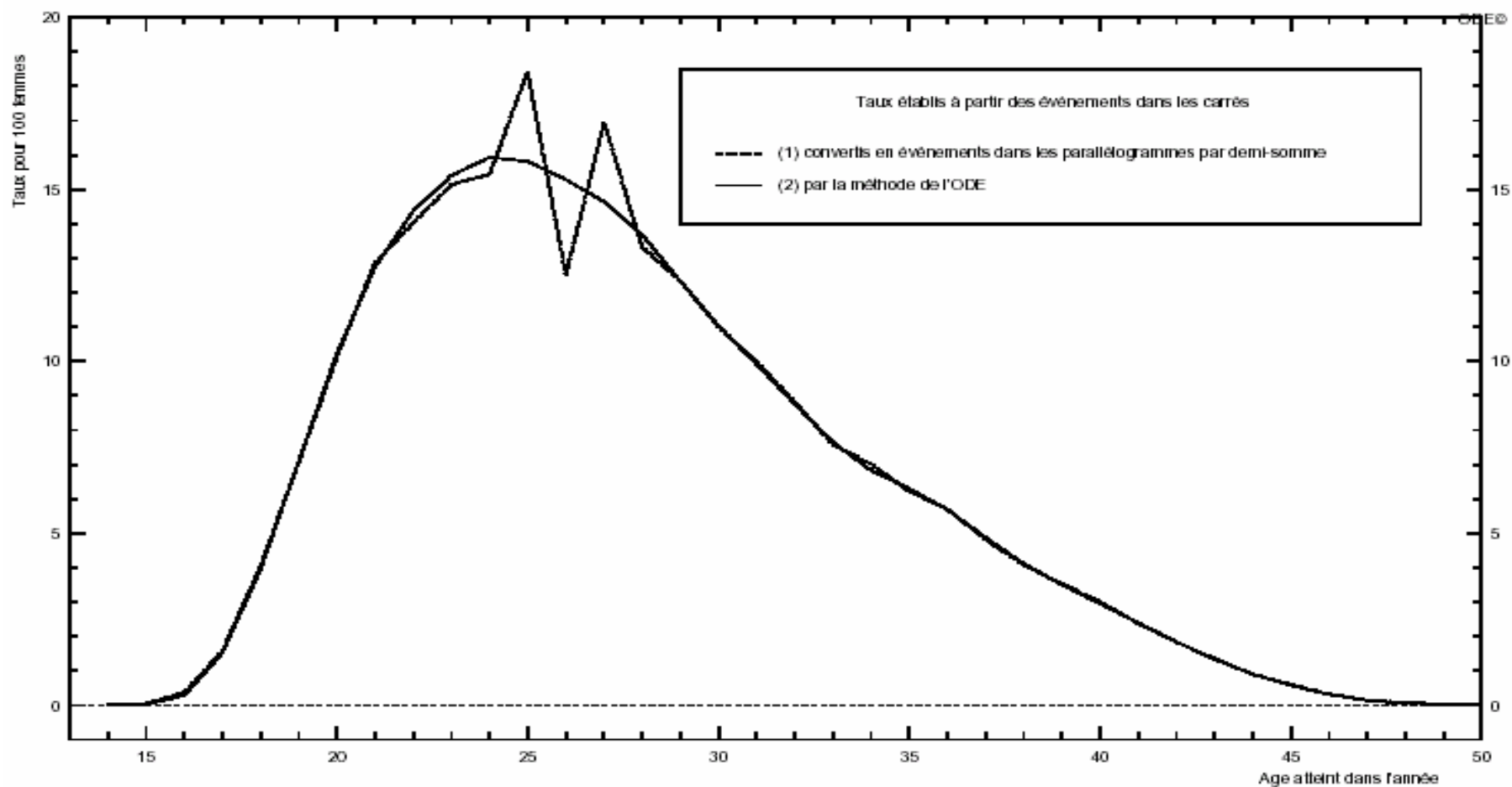
L'âge moyen à la survenance d'un événement peut être calculé de deux manières différentes : en pondérant les âges par les *nombres absolus* d'événements ou par les *taux*.

Dans le premier cas, on obtient, par exemple en matière de fécondité, l'âge moyen des femmes qui ont eu un enfant dans l'année. Cet indicateur, qui répond à la définition classique d'une moyenne statistique (âge moyen, à l'accouchement, des *parturientes* de l'année), correspond généralement à l'information dont souhaitent disposer les autorités sanitaires du pays. Mais ce nombre moyen est directement affecté par les *irrégularités* de la pyramide des âges féminine : si les femmes de 20-24 ans se trouvent être particulièrement *nombreuses* dans la population résidente cette année-là (par exemple parce que les générations du baby-boom se trouvent alors appartenir à ce groupe d'âge), l'âge moyen sera exceptionnellement *abaissé*.

Figure 1. FINLANDE, année d'observation 1967

Taux de fécondité, par âge de la mère atteint durant l'année civile de la naissance,
déduits des nombres d'événements observés dans les carrés

- (1) convertis en nombres d'événements dans les parallélogrammes par demi-somme dans les carrés adjacents
- (2) par la méthode de l'ODE



Il s'ensuit notamment que les variations temporelles de l'âge moyen calculé en utilisant les nombres absolus comme coefficients de pondération reflètent *simultanément* deux sortes de variations : celles touchant la composition par âge de la population féminine et celles concernant le calendrier transversal de la fécondité.

Au contraire, l'âge moyen obtenu en utilisant les *taux de fécondité* par âge comme coefficients de pondération n'est pas affecté par le phénomène perturbateur qu'est la composition par âge de la population féminine et permet des comparaisons aussi bien à travers le temps qu'à travers l'espace.

Si les écarts entre les deux âges moyens étaient toujours faibles, le *distinguo* n'aurait pas grande portée, mais tel n'est pas le cas : ils peuvent différer de presque un an (alors qu'on exprime généralement un âge moyen en années avec *une* décimale), faussant ainsi les comparaisons.

Dans une publication internationale, il convient, pour des raisons de comparabilité, de ne faire figurer que des âges moyens fondés sur les taux.

En matière de primo-nuptialité, il convient, en outre, de faire choix d'un âge limite en deçà duquel le *premier* mariage est étudié. Cet âge limite est conventionnellement fixé au 50^{ème} anniversaire : la fréquence du *célibat définitif* est mesurée à cet âge. En dépit de l'élévation contemporaine de l'âge au mariage, nous ne pensons pas qu'il y ait lieu de remettre en cause le choix de cette limite pour la construction des tables de primo-nuptialité. L'âge moyen au premier mariage, établi sur la base des taux, doit naturellement porter sur les seuls mariages de *célibataires* en *deçà* de cette limite.

La conversion des distributions par âge de la population résidente à une date autre que le 1^{er} janvier en distributions au 1^{er} janvier

La plupart des pays européens évaluent les effectifs de la population résidente par sexe et année d'âge au 1^{er} janvier de chaque année. Mais certains pays adoptent une autre date : le Royaume-Uni qui retient le 1^{er} juillet (*mid-year*), l'Irlande qui retient le 15 avril. Dans ce cas, en se fondant sur les statistiques *mensuelles* de naissances sur longue période (jusqu'à l'année la plus récente), l'Observatoire Démographique Européen a établi une méthodologie permettant d'évaluer les effectifs au 1^{er} janvier.

La conversion systématique de toutes les évaluations de population par sexe et âge en évaluations au 1^{er} janvier permet d'utiliser les *mêmes* logiciels de construction des tables pour tous les pays et toutes les années et aussi d'agréger géographiquement les effectifs par sexe et âge.

Finalement, *quelle que soit* la manière dont ont été établies les données de base concernant tant les effectifs de *résidents* par sexe et âge (évaluations au 1^{er} janvier ou à une autre date de l'année) que les *flux* annuels d'événements par âge (premiers mariages, naissances, décès observés par âge atteint, par âge révolu ou les deux à la fois), les tables de primo-nuptialité, fécondité et mortalité peuvent être construites de façon à être *directement comparables* d'un pays à l'autre et d'une année à l'autre.

En outre, les divers taux sont également établis par année de naissance et âge révolu (à *cheval* sur deux années d'observation), ce qui fournit des *cumuls* par génération permettant de caractériser les générations successives par leur situation au moment d'un *anniversaire* (par exemple : proportion des célibataires qui se marient avant leur 25^{ème} anniversaire ou entre leur 25^{ème} et leur 30^{ème} anniversaire, proportion des enfants qui naissent entre le 25^{ème} et le 30^{ème} anniversaire de leur mère).

La calcul des taux bruts annuels

Les indicateurs démographiques les plus simples (mais aussi ... les plus grossiers) sont les *taux bruts annuels*, obtenus en rapportant le nombre annuel total d'événements observés à la population *moyenne* de l'année. La plupart des pays européens retiennent pour population *moyenne* la demi-somme des effectifs de résidents au 1^{er} janvier de l'année et au 1^{er} janvier de l'année suivante.

Mais, d'une part il peut arriver que la population au 1^{er} janvier de l'année suivante diffère de celle au 31 décembre de l'année, parce que l'office national a décidé d'opérer un *ajustement statistique* pour pallier une incohérence comptable entre la variation de la population totale, le solde du mouvement naturel (balance naissances – décès) et le solde des migrations internationales⁵. Dans ce cas, il semble souhaitable, dans une publication internationale où figurent les éléments du mouvement de la population en nombres absolus et sous la forme des taux bruts (taux brut de variation de la population totale, taux brut d'accroissement naturel, solde migratoire relatif), de retenir pour population moyenne lors du calcul de ces derniers la demi-somme des effectifs au 1^{er} janvier et au 31 décembre et d'indiquer dans une note la valeur de l'ajustement statistique.

D'autre part, certains pays qui disposent d'un registre de population établissent des séries *mensuelles* ou *trimestrielles* de la population résidente et retiennent pour population moyenne de l'année la moyenne arithmétique des valeurs mensuelles ou trimestrielles. Les taux bruts qui en résultent, exprimés habituellement en *pour mille* avec une décimale, ne diffèrent généralement guère plus que par le jeu des arrondis de ceux obtenus par référence à la demi-somme des effectifs aux deux extrémités de l'année. C'est pourquoi, il apparaît souhaitable que les taux bruts déduits d'une base de données internationale soient calculés *uniformément* pour tous les pays de cette dernière manière.

Le rang des naissances vivantes

Le *rang* d'une naissance vivante est une variable de grand intérêt pour l'étude de la fécondité. Mais deux définitions différentes du rang coexistent en Europe : le rang *biologique*, défini par rapport à l'histoire génésique de la mère, et qui s'applique à *toutes* les naissances vivantes, et le rang *dans le mariage en cours* qui ne s'applique qu'aux naissances *dans le mariage*⁶.

Aux époques où à la fois les naissances *hors mariage* et les divorces et veuvages, donc les *remariages*, sont rares, les deux définitions sont sensiblement équivalentes. Mais quand les unes ou les autres sont fréquents, notamment dans la période contemporaine, le rang dans le mariage en cours perd une grande partie de sa *pertinence*. Or d'assez nombreux pays européens, notamment l'Allemagne, la France et le Royaume-Uni, continuent à ne produire de statistiques d'état civil que selon la définition du rang dans le mariage en cours. D'autres ne produisent de données annuelles par rang biologique que depuis une époque relativement récente (ainsi la Suède depuis 1974, l'Autriche depuis 1984).

Aussi la question se pose-t-elle de savoir si, pour les pays qui ne disposent pas de données annuelles par rang biologique, il ne serait pas souhaitable de faire figurer dans les

⁵ Cette situation se produit lorsque, par suite d'une différence dans le degré de complétude de deux recensements successifs, l'estimation de la variation intercensitaire de la population totale n'est pas égale à la somme du solde du mouvement naturel et du solde des migrations internationales.

⁶ Indiquons pour mémoire que certains pays déterminent le rang d'une naissance en ajoutant, comme il convient, une unité au nombre de naissances *antérieures*, mais en incluant dans ce nombre les enfants mort-nés ou encore les enfants adoptés. Par ailleurs, il arrive que la statistique des naissances par rang ne se limite pas aux seules naissances vivantes, mais porte aussi sur les enfants mort-nés, ou encore sur l'ensemble des accouchements.

publications internationales, *en plus* – et même éventuellement *à la place* – des séries annuelles traditionnelles fondées sur les statistiques d'état civil, quelques données (qui ne sauraient guère être annuelles) obtenues selon le rang biologique par voie d'*enquêtes* ou grâce à quelque méthode fiable d'estimation. Cette remarque vaut à la fois pour les indicateurs *transversaux* (par année ou période d'observation) et pour les indicateurs *longitudinaux* (par année ou période de naissance des générations), en particulier pour la proportion des femmes *sans enfant* (mesure de l'infécondité) dans les générations. Il en va de même plus généralement de la décomposition de la descendance finale selon le nombre final d'enfants.

Cette remarque est à rapprocher de la suivante, qui tient à une autre faiblesse des statistiques d'état civil. Les événements comptabilisés par ces statistiques se limitent généralement à ceux qui se produisent sur le *territoire* du pays considéré. Or les migrations internationales peuvent être à l'origine de discordances plus ou moins marquées entre les données de *flux* enregistrées par l'état civil et les données de *stock* obtenues lors des recensements. Par exemple, s'agissant de primo-nuptialité, certains immigrés, entrés dans le pays comme célibataires, peuvent aller se marier dans leur pays d'origine ; dans ce cas, leur mariage échappe aux statistiques du pays d'accueil mais ils apparaissent comme mariés si un recensement vient à y être effectué. C'est pourquoi il peut être jugé souhaitable, au moins pour la dernière période intercensitaire, de confronter systématiquement l'augmentation de la proportion de non célibataires par génération obtenue par cumul des statistiques d'état civil au cours de la période intercensitaire et celle obtenue par différence entre les statistiques des deux recensements. Cette confrontation permet aussi de déceler d'éventuelles erreurs dans la base de données.

Ce même type de confrontation entre flux et stocks peut être effectué en ce qui concerne la répartition des femmes selon le nombre d'enfants déjà nés⁷ ou encore la proportion de personnes nées hors du pays selon le sexe (test de la cohérence entre la variation intercensitaire de cette proportion selon la génération et les soldes migratoires par sexe et âge de la période intercensitaire).

Pour les pays qui disposent, durant une période assez longue d'observation, d'informations annuelles sur les naissances vivantes selon le rang *biologique*, il est suggéré d'établir les *probabilités d'agrandissement*, c'est-à-dire les fréquences du passage de r à $r + 1$ enfants selon l'âge de la mère, ainsi que leurs agrégats *tous âges réunis*, tant transversaux que longitudinaux. De la même façon en matière de primo-nuptialité masculine et féminine, il est suggéré d'établir non seulement les taux analogues aux taux de fécondité (taux appelés encore taux de *deuxième* catégorie dans la littérature), mais également les *quotients* de primo-nuptialité (ou taux de *première* catégorie) ainsi que leurs synthèses transversales (indicateurs conjoncturels et âges moyens fondés sur les quotients).

L'analyse des données journalières

Sans être indispensable, le recueil de données journalières, en matière de naissances et surtout de mariages, est souhaitable pour affiner la correction des variations saisonnières. En effet, les mariages d'une semaine se concentrant, dans la plupart des pays européens, de plus en plus sur un jour donné (le samedi dans certains pays, le vendredi dans d'autres), un nombre brut de mariages se rapportant au même mois de deux années différentes n'aura pas la même signification selon que le mois correspondant comprend quatre ou cinq samedis (ou vendredis). Ces données journalières sont fréquemment établies par les office nationaux

⁷ A la condition que les statistiques d'état civil soient établies selon le rang *biologique*.

à des fins de contrôle statistique, et donc souvent disponibles. Ajoutons que les données journalières en matière de naissances fournissent des indications intéressantes sur les pratiques médicales.

L'analyse des divorces

La divortialité s'analyse, comme la fécondité selon l'âge de la mère, au moyen des *taux* selon l'ancienneté du mariage. Mais, si les offices nationaux estiment les effectifs féminins par âge au 1^{er} janvier de chaque année, compte tenu à la fois de la mortalité et des migrations, ils ne tiennent pas à jour une comptabilité analogue des mariages subsistants à chaque 1^{er} janvier, compte tenu à la fois des dissolutions d'unions (par veuvage ou divorce) et des migrations. Aussi les seuls taux de divortialité qu'on puisse établir rapportent les nombres de divorces observés à une ancienneté donnée au nombre *initial* de mariages célébrés dans le pays. Ces taux sont entachés de l'effet perturbateur des veuvages et surtout des migrations.

Par ailleurs, la définition du divorce varie selon les pays. Certains établissent leurs statistiques annuelles sur la base du nombre de jugements de divorce *prononcés* par les tribunaux. D'autres établissent leurs statistiques sur la base des divorces *transcrits* à l'état civil. Or certains divorces ne sont jamais transcrits (par exemple parce qu'aucun des ex-conjoints ne se remarie), tandis qu'un divorce prononcé une année donnée peut, s'il vient à être transcrit, l'être plusieurs années seulement après le jugement.

La conversion des taux de divorce par ancienneté du mariage atteinte ou révolue ne saurait être opérée avec le même souci de qualité statistique qu'en matière de fécondité, de primo-nuptialité ou de mortalité, d'autant plus que les informations disponibles sont parfois seulement par groupes quinquennaux d'ancienneté. Toutefois les taux transversaux de divortialité peuvent être additionnés longitudinalement pour évaluer l'intensité du divorce dans les promotions de mariages successives.

Chapitre 1

Opérations préalables au traitement des données dans la chaîne d'exploitation

Le traitement de l'information reçue des instituts nationaux de statistique consiste, si aucune anomalie n'est décelée dans cette information, en l'application de méthodes statistiques diverses, à l'issue desquelles sont créés des fichiers de résultats, fichiers que nous appellerons *fichiers dérivés*. Si aucune anomalie majeure n'est décelée dans l'information reçue, la création des fichiers dérivés doit pouvoir être obtenue de façon entièrement automatique. Mais diverses sortes d'anomalies sont susceptibles d'être rencontrées.

- Certaines seront résolues par la chaîne de traitement elle-même (cas des anomalies fréquentes ou dont la correction ne donne pas lieu à un calcul complexe). Dans cette première catégorie, on trouve celles qui tiennent à :
 1. la définition de l'âge (âge atteint et âge révolu),
 2. la présence de non-déclarés,
 3. des écarts dans les nombres totaux d'événements selon le critère de classement (par exemple : naissances vivantes selon l'âge de la mère, le rang de naissance ou le mois d'occurrence).
- D'autres anomalies en revanche donneront lieu à rejet par la chaîne de traitement automatique, nécessitant une correction *préalable* des données reçues de l'office national. Cette correction préalable peut elle-même être réalisée :
 1. par voie manuelle
 2. ou par voie automatique, à l'aide d'un programme *ad hoc*, après consultation éventuelle de l'office national. En effet, avant de soumettre les données officielles d'un pays à un traitement qui va nécessairement les modifier par voie automatique, il est généralement souhaitable de prendre contact avec les responsables de la statistique de ce pays afin de leur expliquer les modifications envisagées et d'obtenir, autant que possible, leur accord. Il conviendra, notamment, de s'entendre avec le pays lorsqu'il y a plusieurs choix possibles.

Les anomalies de cette seconde catégorie, qui se rapporteront le plus souvent à des périodes *anciennes*, seront résolues compte tenu des informations disponibles au moment où est effectué le traitement. Généralement, elles consisteront en un traitement particulier réalisé une fois pour toutes à partir des données reçues de l'office national.

Dans ce premier chapitre nous traiterons des anomalies du second type, qui imposent l'application de traitements préalables aux données reçues des offices nationaux, de façon qu'elles satisfassent les contraintes de la chaîne automatisée.

Nous aborderons, successivement, trois cas particuliers :

- La décomposition par année d'âge des données regroupées,
- Les autres corrections à apporter aux données de base,
- L'estimation des données manquantes.

A) Eclatement par année d'âge des données pluriannuelles

I - Problèmes affectant la répartition par âge des effectifs de la population

1- La répartition des effectifs de population est connue dans un découpage non annuel de l'âge (le plus souvent quinquennal)

Les estimations annuelles de *population par sexe et âge* peuvent être disponibles, non pas dans un *découpage annuel* couvrant la *totalité* des âges, mais dans un découpage *moins détaillé*. Dans la suite de ce chapitre, nous emploierons le terme *quinquennal* pour désigner un découpage *non annuel* sur l'étendue des âges étudiés (il peut être annuel dans certaines zones d'âge et pluriannuel, avec éventuellement des amplitudes de classe *variables* d'une classe à l'autre, aux autres âges).

Cette non disponibilité de la répartition par sexe et année d'âge de la population empêche le calcul direct des taux démographiques se rapportant à une seule cohorte, ce qui complique singulièrement la recombinaison longitudinale de ces taux et donc la mise à jour des indices se rapportant à des cohortes ou générations particulières.

Elle empêche également l'agrégation de ces données avec celles correspondant à un découpage plus fin, sauf à opérer sur ces dernières un regroupement similaire.

Plusieurs options sont ici disponibles :

1. On peut décider de pas prendre en considération ces années pour lesquelles le découpage n'est pas annuel,
2. on utilise le découpage existant et l'on adapte le découpage des événements pour le rendre compatible avec celui de la population, ce qui permet le calcul des taux et des indicateurs transversaux, mais interdit la transposition longitudinale, sauf à considérer que les taux sont constants sur l'ensemble du groupe d'âge,
3. on utilise le découpage existant, on calcule des taux par groupe d'âge et on annualise ensuite la courbe des taux cumulés, à partir d'un ajustement à la loi de Gompertz,
4. on éclate les effectifs par groupes quinquennaux d'âge en effectif par année d'âge.

Si l'on choisit cette dernière option, plusieurs méthodes existent, qui permettent de procéder à l'éclatement des groupes quinquennaux d'âge d'une *pyramide des âges* quelconque. Elles se fondent toutes sur des régularités observées dans l'évolution d'un âge au suivant et peuvent s'appliquer en toutes circonstances puisqu'elles ne nécessitent la connaissance que de la seule pyramide à annualiser. Parmi ces méthodes, la plus connue est celle des *multiplicateurs de Sprague*⁸.

a) Multiplicateurs de Sprague

La mise en œuvre de cette méthode est la suivante :

Soit un groupe quinquennal d'effectif N_0 que l'on désire ventiler en cinq effectifs par année d'âge, tels que :

$$n_1 + n_2 + n_3 + n_4 + n_5 = N_0.$$

⁸ Cf. Thomas Bond Sprague, « Explanation of a New Formula for Interpolation », *Journal of the Institute of Actuaries*, 22 :270, 1880-1881.

Il faut d'abord distinguer les deux premiers ainsi que les deux derniers groupes quinquennaux qui nécessitent un traitement spécifique, sur lequel nous reviendrons ci-après.

Pour les autres groupes, on utilise la *table intermédiaire* des multiplicateurs de Sprague (dont les totaux en colonnes sont égaux à 0 ou à 1 et ceux en lignes à 0,2) :

	N_{-2}		N_{-1}		N_0		N_{+1}		N_{+2}		Total
n_1	-	0,0128	+	0,0848	+	0,1504	-	0,0240	+	0,0016	0,2000
n_2	-	0,0016	+	0,0144	+	0,2224	-	0,0416	+	0,0064	0,2000
n_3	+	0,0064	-	0,0336	+	0,2544	-	0,0336	+	0,0064	0,2000
n_4	+	0,0064	-	0,0416	+	0,2224	+	0,0144	-	0,0016	0,2000
n_5	+	0,0016	-	0,0240	+	0,1504	+	0,0848	-	0,0128	0,2000
Total		0,0000		0,0000		1,0000		0,0000		0,0000	1,0000

Séquence des opérations

Pour estimer, par exemple, l'effectif à 27 ans, qui correspond à n_3 dans le groupe quinquennal d'âge 25-29, on a :

$$N_{-2} = N_{-15-19}$$

$$N_{-1} = N_{-20-24}$$

$$N_0 = N_{-25-29}$$

$$N_{+1} = N_{-30-34}$$

$$N_{+2} = N_{-35-39}$$

En utilisant les coefficients de la ligne n_3 , on écrit :

$$n_3 = +0,0064 N_{-15-19} - 0,0336 N_{-20-24} + 0,2544 N_{-25-29} - 0,0336 N_{-25-29} + 0,0064 N_{-35-39}$$

ce qui fournit l'effectif des individus âgés de 27 ans.

Pour les groupes extrêmes, on utilise des coefficients qui ne font intervenir que quatre groupes quinquennaux.

Pour les groupes extrêmes, on utilise des coefficients qui ne font intervenir que quatre groupes quinquennaux.

Table pour le *premier* groupe quinquennal

	N_0		N_{+1}		N_{+2}		N_{+3}		Total
n_1	+	0,3616	-	0,2768	+	0,1488	-	0,0336	0,2000
n_2	+	0,2640	-	0,0960	+	0,0400	-	0,0080	0,2000
n_3	+	0,1840	+	0,0400	-	0,0320	+	0,0080	0,2000
n_4	+	0,1200	+	0,1360	-	0,0720	+	0,0160	0,2000
n_5	+	0,0704	+	0,1968	-	0,0848	+	0,0176	0,2000
Total		1,0000		0,0000		0,0000		0,0000	1,0000

Table pour le *deuxième* groupe quinquennal

	N_{-1}		N_0		N_{+1}		N_{+2}		Total
n_1	+	0,0336	+	0,2272	-	0,0752	+	0,0144	0,2000
n_2	+	0,0080	+	0,2320	-	0,0480	+	0,0080	0,2000
n_3	-	0,0080	+	0,2160	-	0,0080	+	0,0000	0,2000
n_4	-	0,0160	+	0,1840	+	0,0400	-	0,0080	0,2000
n_5	-	0,0176	+	0,1408	+	0,0912	-	0,0144	0,2000
Total		0,0000		1,0000		0,0000		0,0000	1,0000

Table pour l'avant-dernier groupe quinquennal

	N_{-2}		N_{-1}		N_0		N_{+1}		Total
n_1	-	0,0144	+	0,0912	+	0,1408	-	0,0176	0,2000
n_2	-	0,0080	+	0,0400	+	0,1840	-	0,0160	0,2000
n_3	+	0,0000	-	0,0080	+	0,2160	-	0,0080	0,2000
n_4	+	0,0080	-	0,0480	+	0,2320	+	0,0080	0,2000
n_5	+	0,0144	-	0,0752	+	0,2272	+	0,0336	0,2000
Total		0,0000		0,0000		0,0000		0,0000	1,0000

Table pour le dernier groupe quinquennal

	N_{-3}		N_{-2}		N_{-1}		N_0		Total
n_1	+	0,0176	-	0,0848	+	0,1968	+	0,0704	0,2000
n_2	+	0,0160	-	0,0720	+	0,1360	+	0,1200	0,2000
n_3	+	0,0080	-	0,0320	+	0,0400	+	0,1840	0,2000
n_4	-	0,0080	+	0,0400	-	0,0960	+	0,2640	0,2000
n_5	-	0,0336	+	0,1488	-	0,2768	+	0,3616	0,2000
Total		0,0000		0,0000		0,0000		0,0000	1,0000

b) Amélioration des estimations obtenues à l'aide des multiplicateurs de Sprague

La limite de cette méthode est qu'elle ne mobilise pas la totalité de l'information disponible. Or, pour aboutir à la meilleure estimation possible de la répartition de la population par année d'âge, il est souhaitable de prendre en compte le maximum d'informations. A cet effet, il peut s'avérer nécessaire de distinguer diverses situations en fonction du type de données sur lesquelles ces estimations pourront s'appuyer. Ainsi, dans une seconde étape, si les nombres de naissances vivantes annuelles sont disponibles sur longue période ou si on dispose d'une répartition par sexe et âge détaillé à une date quelconque (antérieure et/ou postérieure à celle de l'estimation), on pourra, *si on le souhaite*, calculer les coefficients apparents de survie, puis les lisser pour déterminer, à partir de ces coefficients lissés, une répartition par âge, probablement plus proche de la répartition réelle.

Séquence des opérations

Si SP_x est l'effectif estimé à l'âge x à partir des multiplicateurs de Sprague,

et $(1-x)qa_0 = \frac{SP_x}{N_{n-x}}$, le coefficient apparent de survie de la naissance à l'âge x ,

Il faut :

- calculer la suite des rapports $\frac{SP_x}{N_{n-x}}$,
- lisser la série des rapports obtenus par moyenne mobile pour obtenir $(1-x)qa_0''$,
- parvenir aux estimations améliorées $SP_x'' = N_{n-x} * (1-x)qa_0''$.

2 - Pyramides d'âge disponibles par année d'âge pour les années encadrant celle pour laquelle on ne dispose que de la répartition par groupe d'âge.

Lorsque l'on dispose de la répartition par année d'âge de la population pour les années qui encadrent celle(s) pour laquelle seule la répartition par *groupe* d'âge est disponible, la procédure à mettre en œuvre est simple et se déroule en deux étapes :

- On procède à une interpolation linéaire entre les effectifs d'une même génération entre $n-1$ et $n+1$,

- Puis on amène, par correction proportionnelle, les sommes quinquennales des effectifs interpolés à coïncider avec l'effectif quinquennal observé correspondant.

Avec cette procédure d'estimation des effectifs par génération, la seule difficulté réside dans la détermination des effectifs du premier et du dernier âge : zéro et 99 ans et plus.

L'effectif à 0 an sera estimé en appliquant au nombre de naissances de l'année $n-1$ un coefficient *apparent*⁹ de survie à 0 an révolu au 1^{er} janvier de l'année n . Celui du groupe terminal ouvert, 99 ans et plus, sera obtenu en appliquant à chacun des âges, 98 et 99 et plus, une probabilité de survivre 1 an de plus, puis en faisant la somme des deux nombres ainsi obtenus.

Cette procédure est identique à celle utilisée, lorsque l'on ne dispose pas de l'estimation par âge et sexe de la population pour une année particulière¹⁰.

Séquence des opérations

1- Interpolation linéaire : $P_n^g = \frac{P_{n-1}^g + P_{n+1}^g}{2}$,

2- Correction proportionnelle : comme $\sum_g^{g+a} P_n^g = P_n^{g,g+a}$

$$P_n^{*g} = P_n^g * \frac{D_n^{g,g+a}}{\sum_g^{g+a} P_n^g}$$

3- Effectif à 0 an : $P_n^0 = N_{n-1} * (1 - ka_0)$, avec ka_0 = coefficient apparent *partiel* de survie

4- Effectif à 99 ans et plus : $P_n^{99+} = \sum_{98}^w (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99} * (1 - q_{99})) + (P_{n-1}^{100} * (1 - q_{100})) + \dots$

$$P_n^{99+} \approx \sum_{98}^w (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99+} * (1 - q_{99}))$$

3 - Pyramides d'âge disponibles par année d'âge pour l'année précédant immédiatement celles pour lesquelles on ne dispose que de la répartition par groupe d'âge.

Lorsque *seule* est disponible la répartition par âge de la population de l'année précédant immédiatement celles pour lesquelles on ne dispose que de la répartition par groupe d'âge, la méthode par interpolation décrite ci-dessus ne peut être utilisée. Il convient, alors, pour chacune des années successives de faire choix de coefficients *perspectifs* apparents de survie entre le 1^{er} janvier et le 31 décembre de l'année, de les appliquer à chaque âge, puis d'utiliser, pour estimer les effectifs d'âge 0 et 99 et plus au 1^{er} janvier de l'année n , la procédure décrite au paragraphe précédent.

On amène ensuite, par correction proportionnelle, chaque somme quinquennale des effectifs estimés à coïncider avec l'effectif quinquennal observé.

Puis, par itération successive de cette procédure, on estime la répartition, par année d'âge, de la population résidente de chacune des années pour lesquelles seule la répartition quinquennale est disponible.

⁹ Puisqu'il fait intervenir le solde migratoire.

¹⁰ Nous la présenterons plus en détail lorsque nous traiterons des données manquantes au chapitre I C, page

Séquence des opérations

1- choix des coefficients apparents de survie : $(1 - qa_x)$,

2- calcul des effectifs de chaque âge l'année n : $P_n^x = P_{n-1}^{x-1} * (1 - qa_{x-1})$,

3- Effectif à 0 an : $P_n^0 = N_{n-1} * (1 - ka_0)$,

4- Effectif à 99 ans et plus : $P_n^{99+} = \sum_{98}^w (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99} * (1 - q_{99})) + (P_{n-1}^{100} * (1 - q_{100})) + \dots$

$$P_n^{99+} \approx \sum_{98}^w (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99+} * (1 - q_{99})),$$

5- Correction proportionnelle : comme $\sum_x^{x+a} P_n^x = P_n^{x,x+a}$

$$P_n^{*x} = P_n^x * \frac{P_n^{x,x+a}}{\sum_x^{x+a} P_n^x}$$

4 - Eclatement par année d'âge des pyramides dont le dernier groupe est antérieur à « 99 ans ou plus »

Lorsque les pyramides par *année d'âge* se terminent par un groupe ouvert dont la borne inférieure est *antérieure* à 99 ans (par exemple, comme ce fut longtemps le cas dans de nombreux pays européens, groupe d'âge terminal : 85 ans ou plus), il est souhaitable d'éclater par année d'âge le groupe terminal pour permettre une représentation graphique esthétique de la pyramide des âges, mais surtout pour construire la table de mortalité.

Dans une telle situation, l'idéal est d'obtenir de l'office national de statistique la répartition selon le découpage souhaité si celle-ci existe ou si elle peut être estimée par ses soins. Dans le cas contraire, il incombe à Eurostat de procéder à l'éclatement de ce groupe ouvert.

Pour procéder à l'extension de la pyramide, on utilisera une table de mortalité adaptée aux conditions du pays traité, généralement la table d'une année récente de ce pays ou d'un pays voisin.

L'utilisation de cette table de mortalité peut se faire de deux manières différentes, selon le type de population associée à cette table de mortalité que l'on retient :

- La première suppose que la population traitée est la population *stationnaire* associée à la table de mortalité, mais avec une dérive variant linéairement avec l'âge.
- La seconde suppose que la population est une population *stable* associée à la table utilisée, le taux de croissance de cette population étant à déterminer.

On retient la méthode qui conduit aux effectifs estimés variant le plus régulièrement selon l'âge.

Séquence des opérations

1- Population stationnaire

- estimation de chacun des effectifs des 10 derniers âges observés et 99 ans, en appliquant aux effectifs de la population stationnaire la règle de proportionnalité entre le total de ces effectifs dans la population stationnaire et dans la population observée ;

- calcul du rapport entre les effectifs observés et ceux, de même total, estimés ci-dessus à partir de la population stationnaire ;
- estimation de la dérive linéaire ;
- calcul du rapport précédent ajusté de la dérive linéaire ;
- estimation ajustée des effectifs par application du rapport ajusté ci-dessus ;
- estimation finale en appliquant à l'estimation ajustée la règle de proportionnalité entre le total du groupe ouvert observé et celui de la population stationnaire.

2- Population stable

- estimation du taux de croissance à partir du rapport des effectifs de chacun des 10 derniers âges, observés et déduits de la population stationnaire ;
- détermination des effectifs de la population stable à partir du taux de croissance déterminé précédemment ;
- estimation de chacun des effectifs entre le dernier âge observé et 99 ans, en appliquant à chacun de ces effectifs la règle de proportionnalité entre le total du groupe ouvert observé et celui de la population stable.

II) Problèmes affectant la répartition par âge des événements démographiques

1 - Les nombres d'événements démographiques selon l'âge sont connus dans un découpage non annuel de l'âge (le plus souvent quinquennal)

Les statistiques transversales de *flux d'événements par âge* (naissances tous rangs ou d'un rang spécifié, premiers mariages d'un sexe spécifié, décès d'un sexe spécifié, divorces) peuvent être disponibles, non pas dans un *découpage annuel* couvrant la *totalité* des âges, mais dans un découpage *moins détaillé*. Comme plus haut, nous emploierons le terme *quinquennal* pour désigner un découpage *non annuel* sur l'étendue des âges étudiés (il peut être annuel dans certaines zones d'âge et pluriannuel, avec éventuellement des amplitudes de classe *variables* d'une classe à l'autre, aux autres âges).

Deux cas sont à distinguer :

- les *pyramides d'âge* sont disponibles au 1^{er} janvier de chaque année civile *par année d'âge* sur *toute l'étendue* des âges concernés par le flux d'événements considéré
- les *pyramides d'âge* ne sont disponibles au 1^{er} janvier de chaque année civile que *par groupes d'âge*

Le redressement des nombres absolus *non déclarés* doit naturellement être effectué *préalablement* à tout autre calcul.

a) Pyramides d'âge disponibles au 1^{er} janvier par année d'âge

Les *nombre absolus* d'événements *quinquennaux*, ainsi que les *taux quinquennaux*, doivent être rangés à des places spécifiques définies par le domaine (par exemple : naissances vivantes, naissances vivantes de rang biologique égal à 2, naissances hors mariage, mariages d'hommes célibataires, décès de femmes, ou divorces) ainsi que par la figure du diagramme de Lexis auquel se réfèrent ces agrégats (CARRE, PV ou PH).

La conversion de données *quinquennales* en données *annuelles* est réalisée par « ajustement » des *taux quinquennaux* à une loi de Gompertz¹¹ dont la fonction cumulative coïncide avec celle des taux quinquennaux observés.

Les *taux annuels estimés* sont ceux de la loi de Gompertz « ajustée ». De ces *taux annuels estimés*, on déduit, par l'intermédiaire des effectifs de population par âge, les *nombre absolus annuels estimés* d'événements.

Ces *nombre absolus estimés* par année d'âge sont ensuite traités par la chaîne automatique comme s'il s'agissait de *nombre observés*.

Séquence des opérations

1- Saisie des effectifs observés, comme s'ils étaient par année d'âge sur la *totalité* de l'échelle des âges, en attribuant l'effectif d'un groupe quinquennal à l'âge *central* de ce groupe (l'effectif à 35-39 ans est attribué à 37 ans). Les non déclarés sont, comme d'habitude, provisoirement attribués à l'âge maximal augmenté d'une unité.

2- Redressement automatique des non déclarés, en gardant copie du fichier initial qui comportait les effectifs de non déclarés.

3- Création du fichier des *nombre absolus par groupe rigoureusement quinquennal d'âge*.

4- Création du fichier des *taux frustes par groupe rigoureusement quinquennal d'âge* :

$$t_n^{x,x+5} = \frac{E_n^{x,x+5}}{\sum_x P_n^x}$$

5- « Ajustement » de Gompertz, et création du fichier des *taux « ajustés » par année d'âge* t_n^{*x} .

6- Création de *deux* fichiers : d'une part celui qui contient les *nombre absolus « ajustés » par année d'âge* et, d'autre part, celui qui contient les *nombre absolus par année d'âge finalement retenus* : $E_n^{*x} = P_n^x * t_n^{*x}$

b) Pyramides d'âge disponibles au 1^{er} janvier, seulement par groupe d'âge

Lorsque les pyramides d'âge ne sont pas disponibles par *année d'âge* mais seulement par *groupe d'âge*, dans un découpage de l'âge *cohérent* avec celui des *nombre absolus* d'événements par âge, on commence par établir, en utilisant un programme *ad hoc*, les fichiers de *taux quinquennaux*. On réalise leur annualisation, comme précédemment, à l'aide d'un « ajustement » de Gompertz. Mais on ne peut déduire des *taux annuels estimés* les *nombre absolus annuels estimés*.

¹¹ La procédure d'ajustement est décrite un peu plus loin.

Dans le cas où les pyramides d'âge disponibles sont *rigoureusement* quinquennales et où les nombres absolus d'événements sont disponibles en *carrés* par groupes *rigoureusement* quinquennaux d'âge, un programme *ad hoc* est alors nécessaire.

Les taux annuels ainsi estimés pourront être utilisés dans le cadre de la recombinaison longitudinale pour compléter les informations disponibles pour chacune des cohortes et suppléer l'absence, partielle ou totale, de données par année de naissance ou année d'âge.

Séquence des opérations

1 et 2- Idem page 19

3- Création du fichier des *nombres absolus par groupe rigoureusement quinquennal d'âge*.

4- Création du fichier des *taux frustes par groupe rigoureusement quinquennal d'âge* :

$$t_n^{x,x+5} = \frac{E_n^{x,x+5}}{\sum_x P_n^{x,x+5}}$$

5- « Ajustement » de Gompertz, et création du fichier des *taux « ajustés » par année d'âge* $t_n''^{x,x+5}$.

L'« ajustement » à une loi de Gompertz

La loi de Gompertz s'énonce en termes de *fonction cumulative* d'une distribution statistique de la façon suivante :

La variable statistique *continue* X suit la loi de Gompertz si sa fonction cumulative $F(x)$, proportion des individus dont le caractère X est inférieur à x , répond à la formule :

$$F(x) = \exp\left\{-\exp\left[P_k\left(\frac{x-b}{a}\right)\right]\right\}, \text{ soit } \text{Log}[-\text{Log}[F(x)]] = P_k\left(\frac{x-b}{a}\right),$$

expressions où a et b sont des *constantes* et où $P_k(x)$ est un polynôme de degré k en x .

Pour qu'il s'agisse bien d'une fonction cumulative, c'est-à-dire d'une fonction *monotone croissante* variant de 0 à 1 quand x varie de $-\infty$ à $+\infty$, il est notamment nécessaire que le degré k du polynôme soit *impair* et que le coefficient a_k du terme de degré k soit *négatif*. Toutefois, la propriété de fonction monotone croissante n'est vraiment nécessaire que dans l'intervalle de variation de X : en dehors de cet intervalle, il est indifférent que les variations de F soient monotones ou non et que les valeurs de F soient comprises entre 0 et 1 ou non.

Le *quantile* d'ordre a de cette distribution est égal à :

$$X_a = ax + b$$

où x est la racine, qui doit être *unique*, de l'équation :

$$P_k(x) = \text{Log}[-\text{Log}(a)]$$

Observons que les paramètres a et b sont de *faux* paramètres : si on se donne une loi de Gompertz, les valeurs de a , b , a_s , $s = 0, 1, \dots, k$ étant fixées, cette même loi peut être énoncée avec d'autres valeurs *arbitraires* des coefficients a et b (auxquelles sont associées des valeurs correspondantes des coefficients a_s). En effet, si $\text{Log}[-\text{Log}[F(x)]]$ est un polynôme de degré k en $\frac{x-b}{a}$, c'est aussi un polynôme de degré k

¹² Nous employons le terme *ajusté* bien qu'il ne s'agisse pas d'un ajustement à proprement parler de la courbe cumulative, mais de la détermination de la courbe cumulative de Gompertz qui passe *exactement* par un *sous-ensemble* choisi donné de points empiriques. C'est pourquoi nous écrivons « ajustement » et non ajustement.

en $\frac{x-\hat{b}}{\hat{a}}$, quels que soient \hat{a} et \hat{b} . La *multiplicité* des lois de Gompertz dont le polynôme P est de degré k est ainsi d'ordre $k+1$.

L' « ajustement » d'une distribution empirique à une loi de Gompertz

Soit une distribution empirique dont la fonction cumulative F est donnée sous la forme des valeurs de F correspondant à $r+1$ limites supérieures de classes : à la limite supérieure de classe x_j se rapporte la valeur F_j ($0 < F_j < 1$) de la fonction cumulative, $j = 1, 2, \dots, r+1$. Si on se fixe à la fois les valeurs de a et b , le degré k du polynôme de Gompertz et les $k+1$ couples (x_i, F_i) , $i = 1, 2, \dots, k+1$ pour lesquels on astreint la courbe ajustée à passer, le polynôme P de degré k est entièrement déterminé.

Ses coefficients a_s , $s = 0, 1, \dots, k$ sont les racines du système de $k+1$ équations linéaires à $k+1$ inconnues :

$$\sum_{s=0}^{s=k} a_s \left(\frac{x_i - b}{a} \right)^s = \text{Log}[-\text{Log}(F_i)]$$

La courbe cumulative *ajustée*¹² passe dans ces conditions exactement par les $k+1$ points (x_i, F_i) sélectionnés pour déterminer les coefficients a_s , mais il n'est aucunement garanti que les valeurs qui s'en déduisent pour ces coefficients fassent de la fonction $\exp\left\{-\exp\left[P_k\left(\frac{x-b}{a}\right)\right]\right\}$ une fonction monotone croissante à valeurs comprises entre 0 et 1. En particulier, il peut fort bien se produire que la fréquence ajustée d'une classe donnée, disons de la classe $(x_u \leq x \leq x_{u+1})$, c'est-à-dire la quantité :

$$\exp\left\{-\exp\left[P_k\left(\frac{x_{u+1}-b}{a}\right)\right]\right\} - \exp\left\{-\exp\left[P_k\left(\frac{x_u-b}{a}\right)\right]\right\},$$

ne soit pas comprise entre 0 et 1, bien que, pour *toutes* les classes résultant de la sélection opérée pour déterminer les coefficients a_s , les fréquences ajustées soient *exactement* égales aux fréquences empiriques.

Deux cas sont à distinguer :

a) on se fixe le degré k du polynôme à ajuster, mais on ne fixe pas a priori les $k+1$ valeurs x_i par lesquelles on astreint la courbe des fréquences cumulative à passer

b) on se fixe les $k+1$ valeurs x_i par lesquelles on astreint la courbe des fréquences cumulatives à passer.

Dans le cas b), les fréquences cumulées F_i associées aux valeurs x_i sont les fréquences cumulatives empiriques déduites des fréquences quinquennales (compte non tenu éventuellement de classes à fréquence très petite, telles que Moins de 15 ans ou 45 ans ou plus) et le degré du polynôme k est par exemple pris égal à 5 si les $k + 1 = 6$ classes retenues vont de Moins de 20 ans à 40 ans ou plus

Dans le cas a), pour limiter le risque que certaines fréquences cumulatives « ajustées » soient négatives ou supérieures à 1, il est recommandé de sélectionner les $k+1$ points servant à l'ajustement de la façon suivante :

- Première valeur sélectionnée telle que F soit assez proche de 0 mais pas trop proche, disons telle que F soit de l'ordre de 1%
- Dernière valeur sélectionnée telle que F soit assez proche de 1 mais pas trop proche, disons telle que F soit de l'ordre de 99%
- Autres valeurs x_i sélectionnées telles que les différences $F_i - F_{i-1}$ soient assez proches les unes des autres, mais plus petites pour $i = 2$ et $i = k$

En pratique, si on ajuste un polynôme de degré k , on s'efforcera de sélectionner les $k+1$ valeurs *distinctes* x_i sur la base des fréquences cumulées empiriques F_j de telle sorte que l'on se trouve le plus près possible de la situation où :

- pour $i = 1$: $F_1 = 1\%$

- pour $i = 2$: $F_2 = (1 + \frac{49}{k-1})\%$
- pour i quelconque de 2 à k : $F_i = (1 + \frac{49(2i-3)}{k-1})\%$
- pour $i = k$: $F_k = (1 + \frac{49(2k-3)}{k-1})\%$
- pour $i = k+1$: $F_{k+1} = 99\%$

Ainsi, lorsqu'on convient d'ajuster un polynôme de degré 6, on sélectionne 7 limites de classes x_i avec approximativement :

$$F_1 = 1\%, F_2 = 11\%, F_3 = 30\%, F_4 = 50\%, F_5 = 70\%, F_6 = 89\%, F_7 = 99\%$$

B) Autres corrections à apporter aux données de base

1- Redressement des non-déclarés

L'une des difficultés rencontrées dans les comparaisons internationales provient du traitement des non-déclarés dans les tableaux statistiques. En effet, certains offices nationaux publient des tableaux contenant des lignes et colonnes spécifiques pour les non-déclarés, alors que d'autres procèdent à ce redressement avant publication, sans indication sur l'existence éventuelle de non-déclarés ou sur les procédures utilisées.

Dans une base de données alimentant un système de calcul d'indicateurs, on est conduit à redresser *tous* les tableaux contenant des non-déclarés. Mais il convient de conserver une copie de chacun des tableaux sous leur forme originelle, de façon à permettre de retrouver ultérieurement d'éventuelles erreurs.

La procédure de redressement généralement retenue consiste à répartir les non-déclarés *au prorata* des éléments déclarés. Il est, en effet, difficile de définir, pour l'ensemble des pays et des phénomènes étudiés, des procédures simples fondées sur des hypothèses de sélection des populations pour lesquelles l'information est manquante.

Pour que cette méthode, indubitablement fruste mais qui a l'avantage de la simplicité, soit efficace et robuste, il convient qu'elle soit faite sur les sous-ensembles les plus petits. Ainsi, en ce qui concerne les naissances vivantes, il conviendra, chaque fois qu'il est possible, de procéder au redressement de l'âge non-déclaré de la mère, non pas à partir des naissances vivantes classées selon l'âge de la mère, mais à partir des naissances vivantes classées selon la qualité juridique de la naissance. En effet, dans nombre de pays l'âge non-déclaré est surtout le fait de femmes *non-mariées*.

Séquence des opérations

$$E_n^{n,x} = E_n^x * \frac{\sum_{x=a,x}^w E_n^x}{\left(\sum_{x=a,x}^w E_n^x \right) - ND_n}$$

2- La conversion des pyramides d'âge établies à une date autre que le 1^{er} janvier en pyramides d'âge au 1^{er} janvier

Certains pays européens (Royaume-Uni, Irlande) et non-européens (Etats-Unis, Canada, Japon, ...) établissent chaque année la répartition estimée de la population résidente par sexe et année d'âge, mais la date de référence à laquelle se rapporte chaque pyramide n'est pas le 1^{er} janvier mais une autre date, identique d'une année à l'autre : 1^{er} juillet pour le Royaume-Uni, 15 avril pour l'Irlande.

De façon à traiter les données de tous les pays avec les *mêmes* logiciels, il est nécessaire de convertir *au préalable* les pyramides d'âge du Royaume-Uni et de l'Irlande en pyramides au 1^{er} janvier. Deux possibilités s'offrent alors selon que la série des naissances par mois d'occurrence est disponible ou non.

a) Naissances mensuelles disponibles

La conversion que réalise le système se fonde sur la série *mensuelle* des naissances vivantes. Considérons une année n : le nombre de résidents à la date a de l'année n (a étant mesuré en fraction d'année à partir du 1^{er} janvier n) de personnes d'âge révolu i est noté $P_i^{n,a}$. L'effectif à la naissance des membres de cette génération (née entre la date a de l'année $n-i-1$ et la date a de l'année $n-i$) est estimé à $v = \int_{x=0}^{x=1} \frac{d(x)}{1-S(w)} e^{-\int_0^x d(x)} dx - m^i$ sur la base de la série mensuelle des

naissances vivantes de l'époque. Le rapport apparent de survie de cette génération entre la naissance et l'âge i révolu est $P_i^{n,a} / N_{n-i-1,a} \rightarrow n-i,a$. L'effectif au 1^{er} janvier n de la génération $n-i-1$ est noté P_i^n tandis qu'à la naissance il était de N_{n-i-1} . Le rapport apparent de survie, a_s , peut être estimé par interpolation linéaire entre deux rapports apparents de survie observés¹³, ce qui fournit l'effectif recherché P_i^n dès lors qu'on connaît N_{n-i-1} .

Par exemple, pour l'Irlande, on établit de cette manière la pyramide d'âge au 1^{er} janvier 1990 connaissant les pyramides au 15 avril 1989 et au 15 avril 1990 (interpolation linéaire). On peut également estimer *provisoirement* la pyramide au 1^{er} janvier 1991 par extrapolation linéaire sur la base des pyramides précédentes, cette estimation devant être soumise à révision lorsqu'on connaîtra la pyramide au 15 avril 1991, ce qui permettra d'opérer une *interpolation* et non une *extrapolation* linéaire).

Séquence des opérations

- 1- Nombre de résidents à la date a de l'année n (a étant mesuré en fraction d'année à partir du 1^{er} janvier n) : $P_i^{n,a}$
- 2- Effectif à la naissance des membres de cette génération : $v = \int_{-\infty}^{+\infty} x^2 \frac{d(x)}{1-S(x)} e^{-\int_0^x d(x)} dx - m^2$
- 3- Rapport apparent de survie : $P_i^{n,a} / N_{n-i-1,a} \rightarrow n-i,a$
- 4- Effectif au 1^{er} janvier n de la génération $n-i-1$: P_i^n (à la naissance il était de N_{n-i-1})
- 5- Rapport apparent de survie, a_s , estimé par interpolation linéaire entre deux rapports apparents de survie observés.
- 6- Fournit l'effectif recherché P_i^n dès lors qu'on connaît N_{n-i-1}

b) Naissances mensuelles non disponibles

Si la répartition mensuelle des naissances n'est pas disponible pour certaines générations, on peut néanmoins procéder aux interpolations et extrapolations linéaires sur la base, faute de mieux, d'un mouvement saisonnier *uniforme* des naissances. Mais il est clair qu'il est bien meilleur de se référer à la répartition des naissances mensuelles.

3- Les estimations intercensitaires de population

Le bilan de la population résidente entre deux dates, qu'il s'agisse de deux 1^{er} janvier consécutifs ou des dates de deux recensements successifs, traduit l'égalité comptable :

Variation de l'effectif de la population = Naissances – Décès + Solde Migratoire

Il n'est pas rare que les données agrégées officielles publiées par l'office national ne soient pas cohérentes entre elles ou soient incohérentes avec les nombres totaux d'événements tous âges correspondants.

Dans le premier cas, on se résout généralement à recourir à la notion d'*ajustement statistique*, sorte de pis-aller qui reconnaît une *rupture* irréductible censée se produire le 31 décembre à 24 heures :

Population au 1^{er} janvier $n+1$ = Population au 31 décembre n + Ajustement statistique

¹³ Le rapport $N_{i+1} / N_{n-i-1,a} \rightarrow n-i,a$ est à référer à la date $n-i-1+a$, le rapport $P_{i-1}^{n,a} / N_{n-i,a} \rightarrow n-i+1,a$ à la date $n-i+a$ et le rapport P_{i-1}^n / N_{n-i} à la date $n-i$.

Cette solution, pour pratique qu'elle soit, nécessite une adaptation de la définition des dénominateurs des taux d'occurrence, bruts ou par âge : on retient pour population moyenne soumise au risque durant l'année n la moyenne arithmétique des populations au 1^{er} janvier n et au 31 décembre n et non, comme en l'absence d'ajustement statistique, entre deux 1^{er} janvier successifs.

Dans le second cas, se pose le problème délicat de la non cohérence entre les données globales officielles et les totaux des répartitions de nombres d'événements selon l'âge. Une telle situation se produit lorsque, suite à la réalisation d'un nouveau recensement, la révision intercensitaire des effectifs de population opérée par l'office national se limite à l'*effectif total* de la population et laisse inchangées les répartitions par âge.

Avant d'analyser ce cas particulier, examinons les difficultés liées à l'absence totale de révision des estimations de population.

Tous les pays européens produisent des estimations annuelles de population, mais le nombre de ceux qui, une fois le recensement suivant effectué, *révisent* les estimations annuelles produites antérieurement est assez réduit. Cela ne pose généralement guère de problèmes majeurs ; cependant, lorsque les mouvements migratoires sont relativement forts et mal connus du fait de l'absence de leur enregistrement, les écarts entre population résidente *réellement* présente et population estimée peuvent être importants¹⁴. Une détection implacable des effets de cette non révision consiste à suivre, au fil des années, l'effectif d'une même génération.

Dans le cas où la dérive entre la population estimée et la population réelle croît d'une année à l'autre, c'est-à-dire lorsqu'un nouveau recensement met en évidence un décalage non négligeable entre la population estimée à la date du recensement et la population recensée, sans que l'on procède à une révision intercensitaire, l'évolution de tous les indices faisant intervenir l'effectif de la population risque de présenter des dents de scie.

La liste des pays procédant à la révision des effectifs de population intercensitaires reste à établir, mais elle est, semble-t-il, relativement courte. En effet, si la plupart des pays européens, ceux du moins qui procèdent encore à un recensement général de la population¹⁵, révisent l'effectif *total* de leur population après un nouveau recensement¹⁶, seule la France semble procéder à une révision de la répartition par âge et sexe.

Plusieurs attitudes peuvent être adoptées par Eurostat selon la politique suivie par les offices statistiques dans ce domaine.

a) Aucune révision intercensitaire

Dans le cas d'un pays qui ne procède à aucune révision intercensitaire, deux solutions peuvent être retenues :

- garder les seuls effectifs par sexe et âge disponibles, ce qui risque d'introduire des discontinuités, plus ou moins fâcheuses, dans les séries calculées ;
- procéder à la révision des effectifs de population.

La révision consiste par exemple en l'interpolation linéaire sur toute la période intercensitaire des effectifs par âge et sexe entre la pyramide fondée sur le premier recensement et celle fondée sur le second.

¹⁴ Ces écarts peuvent également tenir, pour une part plus ou moins grande, à un défaut de complétude différentiel du recensement par rapport au précédent. C'est sans doute la raison de l'écart de 480 000 personnes entre la population recensée en France en 1999 et l'estimation de la population héritée du recensement précédent.

¹⁵ Les Pays-Bas et le Danemark ont renoncé au recensement général de la population.

¹⁶ C'est le cas de tous les pays de l'Union européenne à l'exception de la Belgique et de l'Allemagne.

Séquence des opérations

$$\text{Interpolation linéaire : } P_{n+x}^g = P_n^g * (1 + \sqrt[\frac{r+a}{r}]{\frac{P_{r+a}^g}{P_r^g} - 1})^x$$

b) Révision du seul effectif total de la population

Lorsque le pays limite la révision des données à l'effectif *total* de la population au 1^{er} janvier de chaque année de la période intercensitaire, se pose alors le problème délicat de la non cohérence, chaque année, des chiffres officiels globaux et des nombres totaux d'événements par âge. Pour faire face à cette difficulté, **plusieurs solutions peuvent être adoptées**, dont la pertinence dépend, en fait, de l'écart entre les populations estimées sur la base du recensement précédent et les populations révisées à la suite du nouveau recensement.

- Si cet écart est réduit et ne se trouve pas concentré sur les sous-populations les plus concernées par le phénomène étudié, il n'y aura pas d'incidence majeure sur les indicateurs démographiques fondés sur l'ancienne estimation, non révisée, par âge de la population.
- On peut également, si cet écart n'est pas tout à fait négligeable et peut conduire à une sous-estimation ou une surestimation des indicateurs, prendre en compte (par simple proportionnalité) l'écart entre les deux estimations successives¹⁷.

Toutefois, cette solution conduit à la présence dans la base de données de *deux* effectifs annuels de population pour une même date, l'un fondé sur l'effectif officiel révisé, l'autre sur la somme des effectifs classés par âge (et sexe) hérités du recensement précédent. Cela peut nécessiter des notes de bas de page explicatives, conduire à des confusions ultérieures et se révéler particulièrement gênant.

Pour pallier cette difficulté, deux possibilités s'offrent :

1. Une première solution consiste à ne retenir l'effectif officiel révisé que dans les opérations de calcul du bilan démographique, le total du tableau répartissant la population résidente par âge et sexe étant utilisé dans tous les autres cas.
2. La seconde solution, *qui a notre préférence*, consiste en la correction de la répartition par sexe et âge fournie par le pays, de façon à en aligner le total sur la donnée annuelle officielle.

Dans ce cas deux procédures sont possibles :

- répartir la nouvelle estimation de l'effectif de la population selon la répartition par âge et sexe de l'ancienne ;
- adapter la procédure utilisée dans le cas où il n'y a pas de révision intercensitaire. C'est-à-dire interpoler annuellement les effectifs par âge et sexe entre les premières pyramides disponibles après deux recensements successifs, puis les aligner sur les effectifs annuels révisés par correction proportionnelle.

Cette seconde méthode est un peu plus complexe à mettre en œuvre, mais *est sans doute préférable*. Elle prend, en effet, en compte les dérives toujours possibles sur la répartition par âge.

¹⁷ Un problème semblable se pose dans les pays, comme l'Italie, qui utilisent une définition de population pour établir l'effectif de la population, ainsi que sa répartition, qui diffère de celle utilisée pour enregistrer les événements classés selon l'âge de survenance.

Séquence des opérations

1 - Interpolation linéaire : $P_{n+x}^g = P_n^g * (1 + \sqrt[\alpha]{\frac{P_{r+a}^g}{P_r^g} - 1})^x$

avec P_r^g = Effectif de la génération g au 1er recensement et P_{r+a}^g l'effectif de cette même génération au 2nd recensement

2 - Correction proportionnelle : $P_n^{*g} = P_n^g * \frac{PR_n}{\sum_{g=a}^w P_n^g}$

avec PR = effectif total révisé de la population

4- Mise en accord des marges des différents tableaux concernant une même variable

Nous avons évoqué ci-dessus les difficultés que peut entraîner la disponibilité d'estimations différentes de la population totale. Mais on peut se trouver dans une situation de même nature avec les événements démographiques. En effet, le nombre *annuel total* de naissances vivantes, par exemple, peut être obtenu comme marge de plusieurs tableaux :

- répartition des naissances selon le mois d'occurrence,
- répartition des naissances selon le sexe,
- répartition des naissances selon l'âge de la mère,
- répartition des naissances selon leur rang,

Avec l'utilisation des ordinateurs pour traiter les statistiques d'état civil (à partir des années 1960 dans la plupart des pays), les incohérences entre marges ont généralement disparu, mais pour les périodes anciennes on enregistre des écarts.

Dans une telle situation, afin que soit assurée la cohérence interne des données de la base, il convient de privilégier l'une des marges, généralement celle qui sera indiquée par le pays comme étant le nombre officiel d'événements, utilisé dans le bilan démographique définitif de l'année considérée. Ce nombre sera ensuite repris comme total à atteindre dans chacun des tableaux consacrés à cette variable, l'écart entre cette valeur et la marge du tableau étant introduit en *non déclarés* et traité par la suite comme tel, c'est-à-dire réparti au prorata des effectifs de chacune des cases du tableau considéré.

Il faut toutefois être conscient que cette façon de procéder conduit à des écarts entre les tableaux redressés et ceux qui ont été initialement envoyés par l'office statistique.

C) Estimation des données manquantes

Se trouver face à des données manquantes est situation relativement fréquente, qu'il s'agisse d'un trou dans une série chronologique d'un pays particulier ou de l'absence de données pour un élément d'un ensemble géographique. Ceci peut être dommageable car la non disponibilité de cet élément peut interdire le calcul de toute une série d'indicateurs. Il convient pour résoudre ce problème de prévoir des méthodes d'estimation visant moins à donner une estimation précise de la donnée manquante que de permettre le calcul de tous les indices dont cette donnée n'est qu'un élément, quelconque mais indispensable, de l'ensemble.

Les procédures à utiliser sont nombreuses et le choix à faire dépendra avant tout du type de donnée et d'usage que l'on veut faire de cette estimation.

1- Trou dans une série chronologique

En présence d'un trou dans une série chronologique, l'attitude à adopter dépendra des informations disponibles. Quatre cas peuvent être distingués.

a) La distribution des événements par âge (ou par durée) n'est pas disponible, mais le total annuel est disponible

Dans ce cas, qui est le plus favorable, on a le choix entre deux procédures d'estimation :

1. soit procéder à une estimation de la distribution des événements en deux phases :
 - faire la moyenne des données des années encadrant le trou,
 - répartir, au prorata du poids de chacun des âges (ou des durées), l'écart entre le nombre total d'événements observés l'année considérée et celui obtenu par moyenne des événements des années encadrantes.
2. Une autre manière de procéder, *un peu plus satisfaisante* mais plus complexe à mettre en œuvre, consiste à travailler sur les taux des années encadrant le trou, par une procédure voisine de celle décrite ci-dessus :
 - estimer le taux de l'année considérée par la moyenne des taux des années encadrant cette dernière,
 - déterminer les événements par âge en multipliant chacun des taux par âge, ainsi estimés, par la population moyenne de chacun de ces âges l'année considérée,
 - répartir, au prorata du poids de chacun des âges (ou des durées), l'écart entre le nombre total d'événements observés l'année considérée et celui obtenu précédemment par la somme des événements par âge.

Cette situation est, par exemple, celle du Danemark pour l'année 1970 en matière de primo-nuptialité tant masculine que féminine.

Séquence des opérations

1^{ère} méthode

1a- Interpolation linéaire :
$$E_n'^g = \frac{E_{n-1}^g + E_{n+1}^g}{2},$$

avec $E_n'^g$ = événements estimés l'année n dans la génération g

1b- Correction proportionnelle :
$$E_n''^g = E_n'^g * \frac{EO_n}{\sum_{g=a}^w E_n'^g}$$

avec EO_n = nombre total d'événements observés

2^{ème} méthode

2a- Interpolation linéaire : $t'_n{}^g = \frac{t_{n-1}^g + t_{n+1}^g}{2}$,

avec t_n^g = Taux de la génération g observé l'année n

2b- Estimation des événements : $E_n'^g = \overline{P}_n^g * t_n'^g$

2c- Correction proportionnelle : $E_n''^x = E_n'^x * \frac{EO_n}{\sum_{g=a}^w E_n'^g}$

b) La distribution des événements par âge (ou par durée) n'est pas disponible, pas plus que le total annuel

Si le nombre total d'événements n'est pas disponible, on utilisera l'une des deux procédures précédentes, en excluant la dernière phase.

Séquence des opérations

1^{ère} méthode

1a- Interpolation linéaire : $E_n'^g = \frac{E_{n-1}^g + E_{n+1}^g}{2}$,

avec $E_n'^g$ = événements estimés l'année n dans la génération g

2^{ème} méthode

2a- Interpolation linéaire : $t'_n{}^g = \frac{t_{n-1}^g + t_{n+1}^g}{2}$,

avec t_n^g = Taux de la génération g observé l'année n

2b- Estimation des événements : $E_n'^g = \overline{P}_n^g * t_n'^g$

c) L'effectif de la population par âge (ou par durée) n'est pas disponible, mais le total annuel est disponible

Pour estimer la répartition par âge de la population pour une année manquante il convient de :

- faire la moyenne arithmétique des effectifs appartenant à la même génération le 1^{er} janvier de chacune des deux années encadrant le trou.
- répartir la différence entre le total annuel observé et la somme annuelle des effectifs estimés au prorata des effectifs annuels estimés.

Avec cette procédure d'estimation des effectifs par génération, comme pour celle décrite plus haut (paragraphe AI1, page 13) lorsque l'on ne dispose, pour l'année considérée, que de la répartition par groupe d'âge, la seule difficulté réside dans la détermination des effectifs du premier et du dernier âge : zéro et 99 et plus.

L'effectif à 0 ans sera estimé en appliquant au nombre de naissance de l'année $n-1$ un coefficient *apparent* de survie à 0 an révolu au 1^{er} janvier de l'année n . L'effectif du groupe terminal ouvert, 99 et plus, sera obtenu en appliquant à chacun des âges, 98 et 99 et plus, une probabilité de survivre 1 an, puis en faisant la somme des deux nombres ainsi obtenus.

Séquence des opérations

1- Interpolation linéaire : $P'_n{}^g = \frac{P_{n-1}^g + P_{n+1}^g}{2}$,

2- Correction proportionnelle : $P_n{}^{g'} = P_n{}^g * \frac{PO_n}{\sum_{g=a}^w P_n{}^g}$

avec PO_n = effectif total observé de la population

3- Effectif à 0 an : $P_n^0 = N_{n-1} * (1 - ka_0)$, avec ka_0 = coefficient apparent *partiel* de survie

4- Effectif à 99 ans et plus : $P_n^{99+} = (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99} * (1 - q_{99})) + (P_{n-1}^{100} * (1 - q_{100})) + \dots$

$$P_n^{99+} \approx (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99+} * (1 - q_{99}))$$

d) L'effectif de la population par âge (ou par durée) n'est pas disponible, pas plus que le total annuel

Si l'effectif total n'est pas disponible, on utilisera la première phase de la procédure précédente.

Séquence des opérations

1- Interpolation linéaire : $P'_n{}^g = \frac{P_{n-1}^g + P_{n+1}^g}{2}$,

2- Effectif à 0 an : $P_n^0 = N_{n-1} * (1 - ka_0)$, avec ka_0 = coefficient apparent *partiel* de survie

3- Effectif à 99 ans et plus : $P_n^{99+} = (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99} * (1 - q_{99})) + (P_{n-1}^{100} * (1 - q_{100})) + \dots$

$$P_n^{99+} \approx (P_{n-1}^{98} * (1 - q_{98})) + (P_{n-1}^{99+} * (1 - q_{99}))$$

2- Données manquantes en fin de série (en vue d'une agrégation géographique)

A défaut d'estimations provisoires, ou extrapolées (dans le cadre du système de traitement des données), de la grandeur considérée et si le nombre d'entités géographiques, ou plus exactement leur poids dans l'ensemble constitué, n'est pas trop grand, il conviendra de procéder, autant que possible, à une estimation.

Sont naturellement concernés ici tant les effectifs de population que les événements démographiques.

a) L'effectif de la population par âge n'est pas disponible

En ce qui concerne l'estimation de la répartition par âge d'une population, **on pourra recourir à diverses méthodes.**

1. *La plus satisfaisante*, mais la plus longue à mettre en œuvre, consiste à reconduire les coefficients apparents de survie déterminés à partir des deux dernières pyramides disponibles. Cela conduit à reprendre la première partie de la méthodologie décrite, au chapitre I-A-1-b, pour faire face à une situation dans laquelle on veut transformer une pyramide par groupe d'âge en une pyramide annuelle.
2. Une solution plus rapide consiste à procéder à une extrapolation des effectifs de chacune des générations sur la base des deux dernières pyramides disponibles, les premiers et derniers âges étant estimés par une extrapolation fondée, non plus sur les effectifs classés selon l'année de naissance mais sur l'âge.

3. La méthode la plus simple à réaliser sera fondée sur le simple vieillissement d'un an d'âge (si l'écart entre la dernière pyramide observée et celle à estimer n'est que d'un an) de la dernière pyramide disponible. Le dernier groupe ouvert sera maintenu constant et l'effectif à 0 an sera estimé par reconduction du dernier effectif observé à cet âge.

Si le seul effectif total de la population est à estimer, on procédera à une extrapolation, ou à un simple gel, de cet effectif.

Séquence des opérations

1^{ère} méthode

1.-Calculer les coefficients apparents perspectifs de survie d'un 1^{er} janvier au suivant pour les deux dernières années pour lesquelles ce calcul est possible :

$$(1 - qa_x^{n-2}) = \frac{P_{x+1}^{n-1}}{P_x^{n-2}} \text{ et } (1 - qa_x^{n-1}) = \frac{P_{x+1}^n}{P_x^{n-1}}$$

2.-Calculer les coefficients moyens de survie :

$$(1 - qa_x^n) = \frac{(1 - qa_x^{n-2}) + (1 - qa_x^{n-1})}{2}$$

3.-Calculer les effectifs au 1^{er} janvier de l'année n :

$$P_x^n = P_{x-1}^{n-1} * (1 - qa_{x-1}^n)$$

4.-Calculer les coefficients apparents de survie de la naissance au 1^{er} janvier suivant :

$$(1 - qa_0^{n-2}) = \frac{P_0^{n-2}}{N_{n-3}} \text{ et } (1 - qa_0^{n-1}) = \frac{P_0^{n-1}}{N_{n-2}}$$

5.-En calculer la moyenne :

$$(1 - qa_0^n) = \frac{(1 - qa_0^{n-2}) + (1 - qa_0^{n-1})}{2}$$

6.-Calculer l'effectif à 0 an au 1^{er} janvier de l'année n :

$$P_0^n = N^{n-1} * (1 - qa_0^n)$$

2^{ème} méthode

1.-Extrapolation : $P_n^g = P_{n-1}^g * (1 + \sqrt[3]{\frac{P_{n-1}^g}{P_{n-3}^g} - 1})$

3^{ème} méthode

1.-Reconduction : $P_n^g = P_{n-1}^g$

b) Les événements démographiques ne sont pas disponibles

Lorsque ce sont les événements démographiques qui ne sont pas disponibles pour l'ensemble des pays concernés par l'agrégation géographique, **plusieurs procédés sont possibles.**

On peut :

- procéder à une extrapolation linéaire des dernières valeurs observées,
- extrapoler les taux d'occurrence de l'événement considéré par âge, puis en déduire les événements en appliquant ces taux extrapolés à la population dont les effectifs par âge sont supposés connus,
- extrapoler l'indicateur conjoncturel puis multiplier cet indicateur extrapolé par la génération moyenne en âge de l'événement considéré, observée ou elle même extrapolée.

Les mêmes procédures pourront être facilement adaptées si ce sont les indicateurs démographiques et non les événements qui font défaut.

Séquence des opérations

1^{ère} méthode

1-.Extrapolation linéaire des événements : $E'_n = E_{n-1} * (1 + \sqrt[2]{\frac{E_{n-1}^x}{E_{n-3}^x} - 1})$

2^{ème} méthode

1-.Extrapolation linéaire des taux : $t'_n = t_{n-1}^x * (1 + \sqrt[2]{\frac{t_{n-1}^x}{t_{n-3}^x} - 1})$

2-.Estimation des événements : $E'_n = P_n^x * t'_n$

3^{ème} méthode

1-.Extrapolation linéaire des indicateurs : $IC'_n = IC_{n-1} * (1 + \sqrt[2]{\frac{IC_{n-1}}{IC_{n-3}} - 1})$

2-.Extrapolation linéaire de la génération moyenne : $GM'_n = GM_{n-1} * (1 + \sqrt[2]{\frac{GM_{n-1}}{GM_{n-3}} - 1})$

3-.Estimation des événements : $E'_n = GM'_n * IC'_n$

Chapitre 2

La construction des tables de survénance d'événements

L'un des objets essentiels de l'analyse démographique est l'étude de la *survenance* d'événements divers au sein d'univers statistiques : ainsi le premier mariage chez les célibataires, la naissance d'un enfant (de rang spécifié ou non) chez les femmes ou les couples, le décès chez les personnes de tel sexe. L'*intensité* de la survénance est généralement mesurée de manière différentielle selon l'âge (ou la durée écoulée depuis une date prise pour origine).

Parmi les événements susceptibles de survenir, on distingue les événements *renouvelables*, qui peuvent se produire *plusieurs* fois durant la vie d'un même individu, comme la naissance d'un enfant (de rang *non spécifié*), et les événements *non renouvelables*, qui ne peuvent se produire qu'*une seule fois au maximum* durant la vie d'un même individu, comme le premier mariage, la naissance d'un enfant de rang *biologique spécifié* ou le décès. Un événement non renouvelable qui se produit *nécessairement* une fois durant la vie de tout individu est qualifié de *fatal*, comme le décès.

A) Les différents types d'événements

1- Evénements renouvelables

Lorsque l'événement étudié est *renouvelable* (naissances vivantes sans distinction de son rang) ou traité comme tel (premier mariage et divorce), on mesure l'intensité de sa survénance par un *taux* : rapport entre le nombre d'occurrences de l'événement et le nombre de personnes-années d'exposition au risque, parmi un groupe homogène d'individus *ayant ou n'ayant pas* déjà subi l'événement, exposés *indépendamment* les uns des autres et avec la *même intensité* au risque de survénance de l'événement. Désignons par $f(x)dx$ le taux de survénance entre les âges x et $x + dx$: la probabilité qu'un individu donné subisse l'événement entre les âges x et $x + dx$ est égale à $f(x)dx$. Le nombre d'occurrences de l'événement au cours de la vie de l'individu, c'est à dire l'intensité du phénomène, a pour espérance mathématique $I = \int_{x=a}^{x=W} f(x)dx$, en désignant par a et W les âges extrêmes où l'événement est susceptible de se produire.

L'âge à la survénance de l'événement est la variable aléatoire continue dont la densité de répartition est $\frac{f(x)}{\int_{x=a}^{x=W} f(x)dx}$ et la fonction cumulative $F(x) = \frac{\int_{x=a}^{x=x} f(x)dx}{\int_{x=a}^{x=W} f(x)dx}$. Sa moyenne m et

sa variance V sont respectivement $m = \frac{\int_{x=a}^{x=W} x f(x)dx}{\int_{x=a}^{x=W} f(x)dx} = W - \int_{x=a}^{x=W} F(x)dx$ et

$V = \frac{\int_{x=a}^{x=W} x^2 f(x)dx}{\int_{x=a}^{x=W} f(x)dx} - m^2$. Si l'événement est par exemple la *naissance d'un enfant* pour une

femme, I est, au sein d'un groupe constitué d'un grand nombre de femmes, le *nombre moyen d'enfants par femme* et m l'*âge moyen à la maternité*. Si l'événement est le *premier mariage* pour un célibataire, la primo-nuptialité étant traitée comme produisant des événements renouvelables, W est conventionnellement fixé à 50 ans (en âge *exact*), I est la *proportion de non-célibataires à 50 ans* (complément à l'unité de la fréquence du *célibat définitif*) et m l'*âge moyen au premier mariage* (fondé sur les *taux* de primo-nuptialité).

2- Evénements non renouvelables

Lorsque l'événement étudié est *non renouvelable* (décès, mariages de célibataires), on mesure l'intensité de sa survenance par un *quotient* : rapport entre le nombre d'occurrences de l'événement considéré et le nombre de personnes-années d'exposition au risque, parmi un groupe homogène d'individus *n'ayant pas subi l'événement* et exposés *indépendamment* les uns des autres et avec la *même intensité* à la survenance de l'événement. Désignons par $q(x)dx$ le quotient de survenance de l'événement entre les âges x et $x+dx$: la probabilité qu'un individu donné subisse l'événement entre les âges x et $x+dx$ sachant que cet événement ne s'est pas produit avant l'âge x est égale à $q(x)dx$. La probabilité $S(x)$ que l'événement se produise *après* l'âge x sachant qu'il s'est produit *après* l'âge x_0 est

$S(x) = S(x_0) e^{-\int_{x_0}^x q(x)dx}$. En particulier, pour $x_0 = a$, âge minimal à la survenance de l'événement, $S(x) = e^{-\int_a^x q(x)dx}$. La probabilité $S(x)$ que l'événement se produise entre les âges x et $x+dx$ sachant qu'il s'est produit *après* l'âge x_0 est $S(x) - S(x+dx) = S(x_0) q(x) e^{-\int_{x_0}^x q(x)dx} dx$. La probabilité pour un individu d'échapper à l'événement est $S(w) = e^{-\int_a^w q(x)dx}$.

L'âge à la survenance de l'événement, pour ceux qui le subissent, est la variable aléatoire continue dont la densité de répartition est $\frac{q(x) e^{-\int_a^x q(x)dx}}{1 - S(w)}$ et la fonction cumulative $\frac{1 - S(x)}{1 - S(w)}$. Sa moyenne m et sa variance V sont respectivement égales à

$$m = \int_{x=a}^{x=w} x \frac{q(x) e^{-\int_a^x q(x)dx}}{1 - S(w)} dx = a + \int_{x=a}^{x=w} \frac{S(x)}{1 - S(w)} dx, \quad V = \int_{x=a}^{x=w} x^2 \frac{q(x) e^{-\int_a^x q(x)dx}}{1 - S(w)} dx - m^2.$$

Si l'événement est par exemple le *décès* pour un individu, w est la durée maximale de la vie humaine, $S(w)$ est *nul* (événement *fatal*) et, au sein d'un groupe constitué d'un grand nombre d'individus, m est l'*espérance de vie à la naissance*, égale à l'*âge moyen au décès*. L'*espérance de vie à l'âge x* est l'espérance mathématique $e(x)$ du nombre d'années restant à

vivre à un individu *encore en vie* à l'âge x : $e(x) = \frac{\int_{x=x}^{x=w} S(x) dx}{S(x)} - x$.

Si l'événement est le *premier mariage* (événement traité comme non renouvelable) pour un célibataire, w est l'âge fixé conventionnellement à 50 ans pour mesurer le *célibat définitif*, $S(x)$ est la proportion de non célibataires à l'âge x , $S(w)$ le complément à l'unité de la fréquence du célibat définitif et, au sein d'un groupe constitué d'un grand nombre d'individus, m est l'*âge moyen au premier mariage* fondé sur les *quotients*, établi sur les non célibataires à 50 ans. L'âge moyen au premier mariage établi sur les personnes qui se marient entre les âges

x et $w = 50$ ans est $e(x) = \frac{\int_{x=x}^{x=w} S(x) dx}{S(x) - S(w)}$.

3- L'observation des nombres absolus d'événements et des effectifs soumis au risque

L'expression mathématique du risque de survenance d'un événement peut être vue comme mesurant une *probabilité* d'occurrence pour un individu donné ou une *fréquence* d'occurrence au sein d'un groupe constitué d'un grand nombre d'individus *homogènes* (en particulier nés quasi simultanément).

Or les données d'observation dont on dispose se rapportent à des groupes ayant une certaine *amplitude* : individus appartenant à un groupe d'âge au minimum *annuel*, observés sur une *durée* d'au moins une *année*. Il convient donc d'adapter les définitions précédentes au matériau statistique disponible.

Les estimations courantes de l'état de la population résidente portent sur les effectifs de population par sexe et année d'âge au 1^{er} janvier de chaque année. Les statistiques annuelles d'état civil fournissent les nombres d'événements observés durant une année d'observation selon l'âge de la personne concernée¹⁸. Cet âge, qui est toujours un *entier*, peut être (Figure 2) l'âge en années *révolues* au moment de l'occurrence de l'événement (nombres d'événements situés dans un *carré* du diagramme de Lexis) ou l'âge que la personne concernée a *atteint* ou *atteindra* durant l'année civile de l'événement (nombres d'événements situés dans un *parallélogramme* à côtés *verticaux* du diagramme de Lexis).

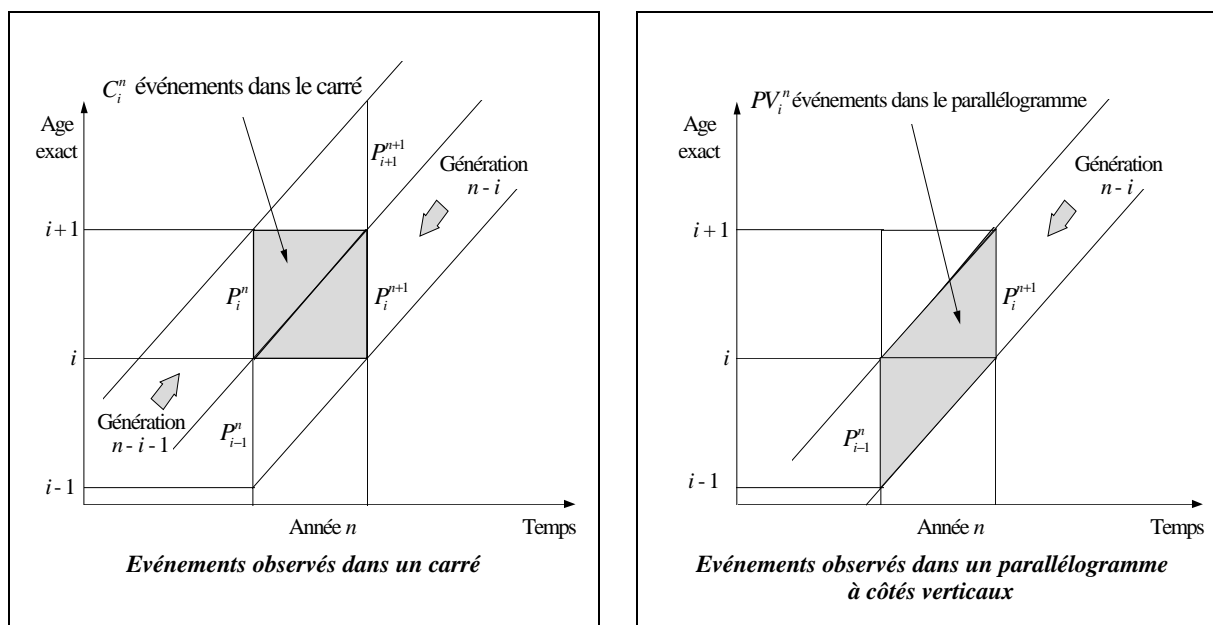


Figure 2. Nombres absolus d'événements et effectifs soumis au risque

¹⁸ Ou de l'unité statistique concernée (par exemple : couple dans le cas d'un divorce) si l'événement étudié est subi par des unités statistiques autres que des personnes.

B) Evénements renouvelables

1- Nombre de personnes-années d'exposition au risque

C'est le nombre de personnes-années d'exposition au risque qui, en toute rigueur, constitue le dénominateur du calcul des taux. La règle précise est de faire figurer chacun des individus qui constituent la population, à laquelle se rapportent les événements étudiés, pour le temps durant lequel il a été présent au cours de l'année. C'est donc la somme de ces temps de présence qui doit figurer au dénominateur des taux. Il s'agit donc d'une moyenne pondérée des individus, les poids étant les fractions d'année de présence.

Cette comptabilité devenant impossible pour des populations, on adopte comme population de référence, la population observée au cours de l'année. L'effectif de la population variant constamment au cours de l'année, on retient comme population de référence la *population moyenne*, moyenne arithmétique des effectifs de la population en début et en fin d'année. On parle ici de la population à la fin de l'année et non au début de l'année suivante, car les pays peuvent être amenés, pour faire face à des incohérences comptables, à introduire des ajustements statistiques qui créent une discontinuité entre 31 décembre et le 1^{er} janvier suivant.

Cette population moyenne doit être adaptée à la définition de l'âge utilisée pour classer les événements :

- Si les événements figurent dans des carrés du diagramme de Lexis (classement selon l'âge révolu), on fera la moyenne arithmétique des effectifs de même âge à deux 1^{er} janvier consécutifs,
- S'ils sont classés selon l'âge atteint dans l'année (ou en différence de millésime) dans des parallélogrammes à base verticale du diagramme de Lexis, on fera la moyenne arithmétique de l'effectif d'une même génération aux deux dates.

Certains pays utilisent comme population moyenne l'estimation de population faite au 30 juin de l'année, d'autres font la moyenne des estimations réalisées pour chacun des mois (comme l'Allemagne ou l'Autriche). Si c'est sans doute cette dernière méthode qui est la plus rigoureuse, à la condition toutefois que l'ensemble des événements soient observés de manière précise (en particulier les mouvements migratoires), l'écart obtenu sur le calcul des taux démographiques par âge est tellement faible qu'il convient de retenir la méthode la plus simple, ce qui, de plus, permet de disposer d'une méthode de calcul unique pour l'ensemble des pays étudiés.

Cette population moyenne n'est qu'un premier intermédiaire dans la chaîne de calcul des taux, et n'a nul besoin d'être mis en fichier puisqu'à l'issue de cette chaîne seront disponibles les taux par âge calculés, dans les deux définitions de l'âge, âge révolu et âge atteint dans l'année, en les corrigeant des conséquences de l'inégalité de l'alimentation des cohortes, ainsi que des variations aléatoires.

Séquence des opérations

1 - Taux selon l'âge atteint dans l'année

$$\text{moyenne arithmétique : } \bar{P}_i^n = \frac{P_{i-1}^n + P_i^{n+1}}{2},$$

2- Taux selon l'âge révolu

$$\text{moyenne arithmétique : } \bar{P}_i^n = \frac{P_i^n + P_i^{n+1}}{2},$$

2- Estimation des taux et construction des tables de survenance pour les événements renouvelables (comme la fécondité)

Tableau 1. FINLANDE, 1966. Données sur la fécondité par âge, en années révolues

Age révolu (années)	Pop, féminine au 1 ^{er} janvier		Naissances 1966	Taux de fécondité 1966		
	1966	1967		Fruste	Elaboré	Différence
14	43269	44180	7	0.0002	0.0002	0
15	45448	42987	28	0.0006	0.0006	0
16	46568	45066	220	0.0048	0.0062	-0.0014
17	48485	46231	1155	0.0244	0.0242	0.0002
18	48191	48133	2608	0.0542	0.0554	-0.0012
19	47154	47220	4306	0.0913	0.0898	0.0014
20	42481	46724	5711	0.1280	0.1196	0.0085
21	34458	41692	4919	0.1292	0.1407	-0.0115
22	32804	34402	5221	0.1554	0.1544	0.0010
23	26224	32271	4682	0.1601	0.1608	-0.0007
24	37820	25854	5092	0.1599	0.1618	-0.0018
25	26916	37519	4968	0.1542	0.1600	-0.0058
26	32002	26905	4843	0.1644	0.1554	0.0091
27	30138	31881	4526	0.1460	0.1457	0.0002
28	28446	29690	3844	0.1322	0.1315	0.0008
29	27432	28580	3258	0.1163	0.1205	-0.0042
30	27397	27273	3139	0.1148	0.1109	0.0040
31	27106	27737	2708	0.0988	0.1003	-0.0015
32	25528	26938	2312	0.0881	0.0881	0
33	27781	25619	2118	0.0793	0.0802	-0.0008
34	28684	27846	2041	0.0722	0.0720	0.0002
35	30334	28646	1841	0.0624	0.0626	-0.0002
36	30678	30412	1630	0.0534	0.0539	-0.0005
37	30264	30407	1453	0.0479	0.0473	0.0006
38	29448	30145	1239	0.0416	0.0409	0.0007
39	29976	29449	989	0.0333	0.0340	-0.0007
40	30409	29820	824	0.0274	0.0270	0.0003
41	29954	30397	633	0.0210	0.0216	-0.0006
42	30967	29790	521	0.0172	0.0166	0.0006
43	29886	30856	363	0.0120	0.0123	-0.0003
44	30702	29853	251	0.0083	0.0082	0.0001
45	31341	30561	149	0.0048	0.0048	0.0001
46	23154	31192	59	0.0022	0.0023	-0.0001
47	27659	22943	26	0.0010	0.0009	0.0001
48	27199	27633	7	0.0003	0.0003	0
49	26474	27106	6	0.0002	0.0002	0
TOTAL	1172777	1183958	77697	2.407	2.411	-0,004
AGE MOYEN	30.38	30.35	27.04	27.78	27.77	-0,01

Prenons l'exemple de la fécondité : la naissance sans spécification de rang est un événement renouvelable et considérons le cas de la Finlande en 1966 à l'âge de 21 ans révolus, année pour laquelle on dispose des nombres d'événements en *carrés*. Une première estimation du taux de fécondité à 21 ans, que nous qualifierons de *fruste*¹⁹, est obtenue en rapportant le nombre des naissances observées à 21 ans révolus (4 919) à la demi-somme²⁰ des effectifs féminins âgés de 21 ans révolus respectivement au 1^{er} janvier (34 458) et au 31 décembre (41 692) de l'année 1966, soit $f = \frac{4919}{(34458 + 41692)/2} = 0.1292$ (Tableau 1).

Sur le graphique représentant selon l'âge les taux frustes ainsi obtenus pour l'année 1966 (Figure 3), on constate trois *couples* d'anomalies (à 20-21 ans, 25-26 ans et 29-30 ans) : dans chaque couple, un premier taux exceptionnel est suivi immédiatement d'un second, également exceptionnel, mais dont le caractère exceptionnel est inverse du précédent : ici, par exemple, à 20 ans le taux observé est trop élevé, à 21 ans trop faible.

Ces anomalies peuvent tenir à un *régime saisonnier* des naissances exceptionnel à l'époque où sont nées les générations concernées. Cela est visible sur les effectifs à 21 ans révolus : 34 458 au 1^{er} janvier 1966, puis 41 692 au 1^{er} janvier 1967 : un tel écart de 21% en seulement un an témoigne d'une variation rapide du nombre des naissances entre 1944 et 1945. On le vérifie plus en détail en considérant l'évolution temporelle des caractéristiques déduites de la statistique des naissances mensuelles (Figure 4) : du fait du conflit qui a opposé la Finlande à l'Union Soviétique en 1940, puis en 1945, le régime saisonnier des naissances ces deux années-là a été fortement perturbé et la formule de calcul fruste s'avère imprécise parce qu'elle repose (implicitement) sur l'hypothèse d'un régime saisonnier *uniforme* des naissances au sein des couples de générations dont on établit les taux. Lorsqu'une telle perturbation se produit, la méthode fruste conduit à des taux biaisés à deux âges consécutifs, successivement dans un sens puis dans l'autre. Tel est le cas ici des taux à 20, puis à 21 ans (couples de générations 1944-45 et 1945-46, comprenant la génération perturbée 1945) et à 25-26 ans (couples de générations 1939-40 et 1940-41, comprenant la génération perturbée 1940).

En revanche, tel n'est pas le cas des anomalies observées à 29-30 ans (couples de générations 1935-36 et 1936-37, comprenant la génération 1936). Mais, alors que les anomalies liées aux perturbations des générations 1940 et 1945 persistent en 1967 et les années suivantes, celles concernant les deux couples de générations contenant la génération 1936 ont disparu (Figure 5). On n'a affaire vraisemblablement qu'à une variation de nature *accidentelle*, due à une erreur d'observation portant à ces âges soit sur les effectifs féminins, soit sur le nombre des naissances.

C'est pour éliminer les biais provoqués par les régimes saisonniers exceptionnels au sein de certaines générations, mais aussi pour *lisser* les résultats de façon à limiter l'effet des erreurs accidentelles, que l'Observatoire Démographique Européen a mis au point une méthodologie de construction des tables de survenance. Cette méthodologie se fonde sur des nombres absolus d'événements pouvant être observés indifféremment par âge révolu (*carrés* du diagramme de Lexis), par âge atteint (*parallélogrammes* à côtés verticaux) ou les deux à la fois (*triangles*). Elle fournit les estimations élaborées des *deux* sortes de taux selon l'année

¹⁹ C'est pourtant la façon dont la plupart des offices nationaux de statistique procèdent pour établir les tables de survenance d'événements.

²⁰ C'est cette estimation par demi-somme du nombre de personnes-années d'exposition au risque qui fait problème lorsque le régime saisonnier des naissances au sein du couple de générations concernées s'écarte fortement de l'uniformité, alors qu'en même temps l'intensité de survenance varie rapidement avec l'âge.

Figure 3. FINLANDE, année d'observation 1966

Taux de FECONDITE par AGE REVOLU obtenus par la méthode classique et par la méthode de l'ODE

Les taux obtenus par la méthode classique pour les biennums contenant les années 1936, 1940 et 1945 sont indiqués par un cercle

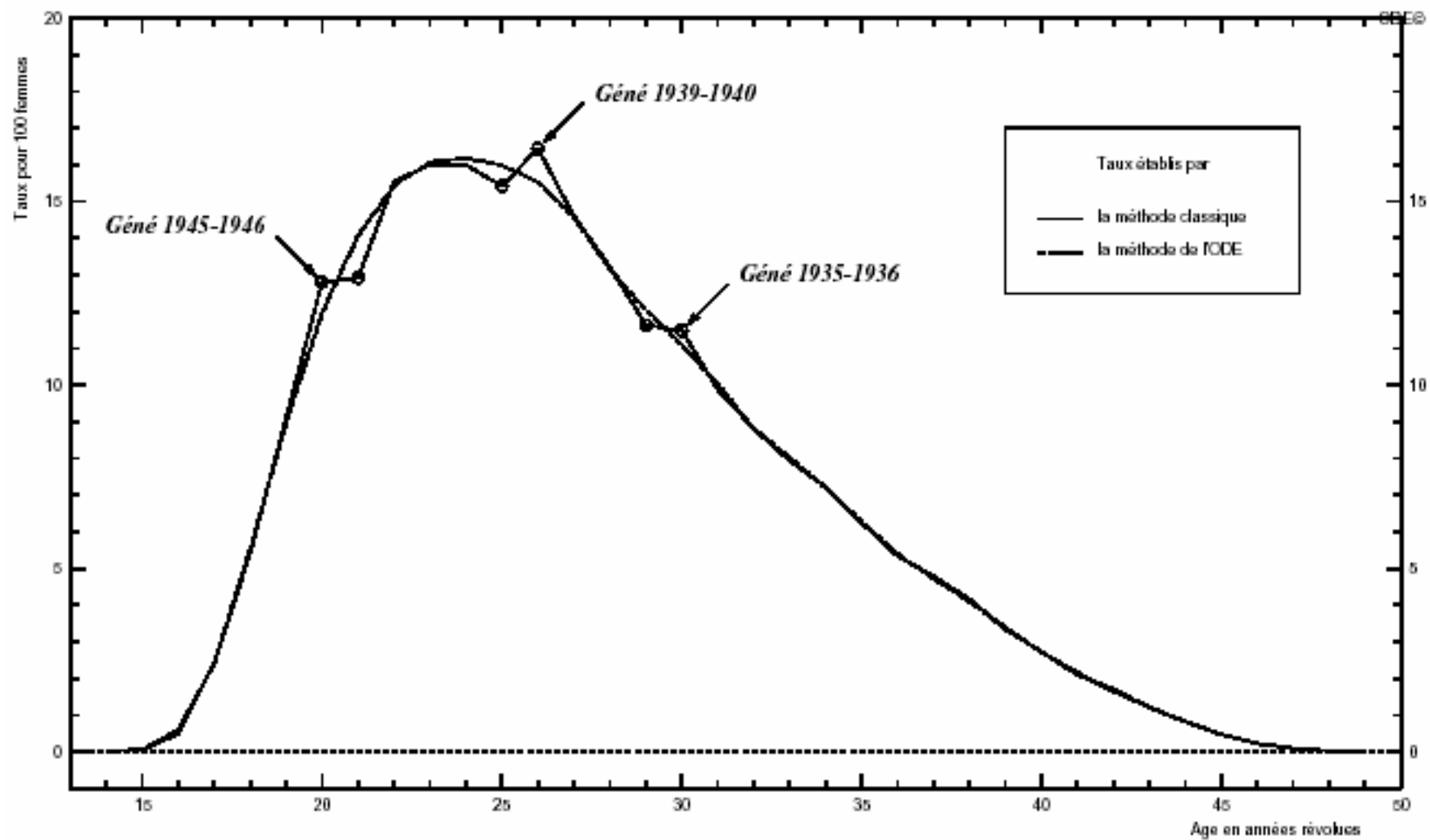


Figure 4. FINLANDE. CARACTERISTIQUES, à la NAISSANCE, des GENERATIONS nées de 1900 à 1998

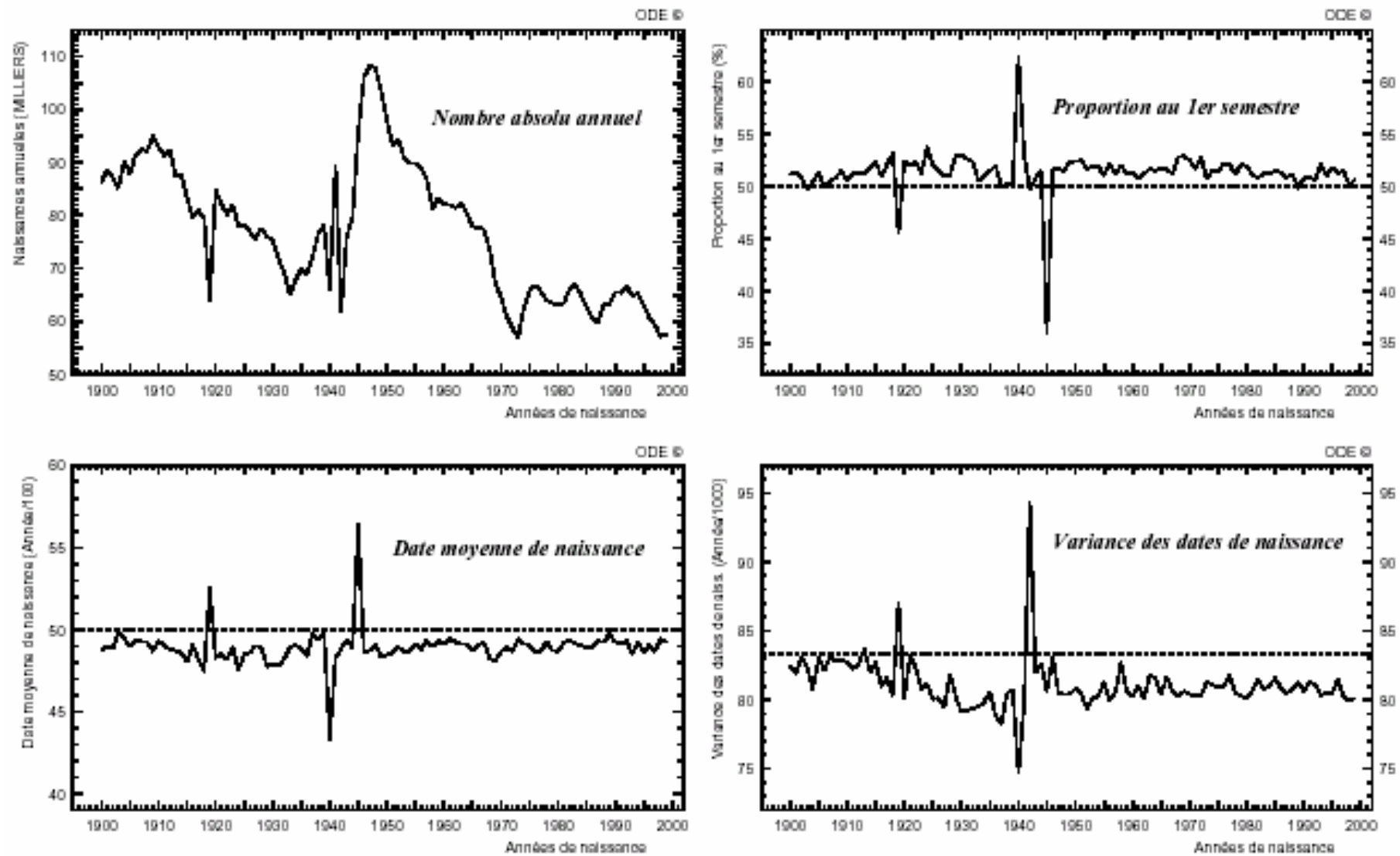
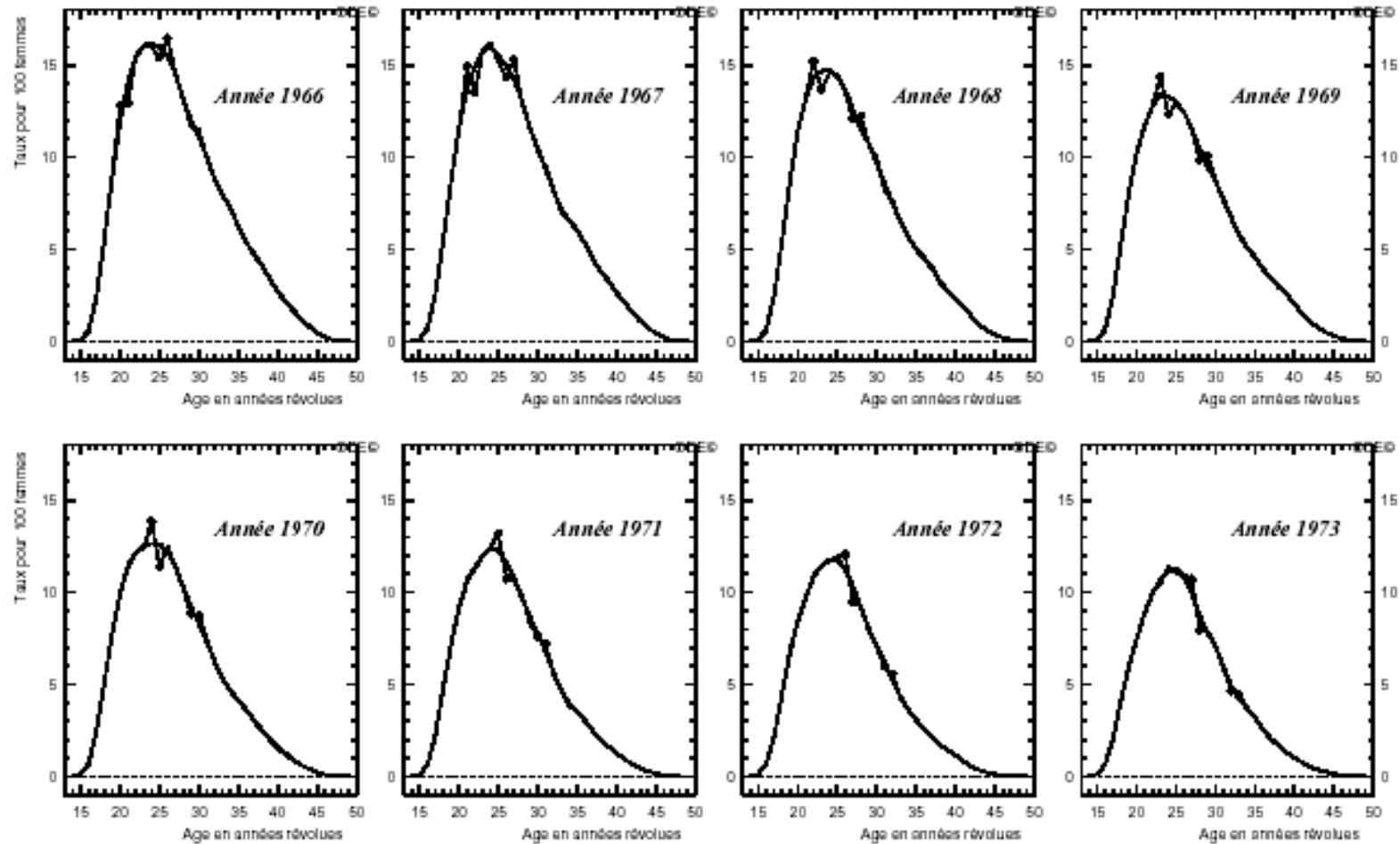


Figure 5. FINLANDE, années d'observation 1966-1973
 Taux de FECONDITE par AGE REVOLU obtenus par
 la méthode classique (trait continu) et par la méthode de l'ODE (pointillé)
 Les taux obtenus par la méthode classique pour les biennums contenant les années 1936, 1940 et 1945 sont indiqués par un cercle



d'observation : par âge révolu (estimation de $\int_x^{x+1} f(x)dx$) et par âge atteint (estimation de $\int_{x-1/2}^{x+1/2} f(x)dx$) ; elle fournit aussi l'estimation des nombres absolus d'événements par triangle. Finalement, elle permet toutes les agrégations géographiques possibles dans le diagramme de Lexis. Par ailleurs, elle fournit les taux par âge révolu pour une même génération, à cheval sur deux années de calendrier, ce qui permet d'établir les cumuls aux divers anniversaires par génération.

L'estimation élaborée du taux de fécondité à 21 ans en 1966 est ainsi de 0,1407 au lieu de 0,1292, soit un écart *relatif* de 8,9%. On montre que ce biais relatif affectant les couples de générations 1944-45 à 21 ans ne varie guère avec l'âge, ni même avec le phénomène étudié : il est sensiblement de 9% à tous les âges pour la primo-nuptialité (Figure 6), la mortalité ou la fécondité au sein du couple de générations 1944-45. Les perturbations liées au régime saisonnier des naissances concernent donc certains couples de générations et ceux-là seulement, mais lorsqu'un couple de générations est concerné, il l'est à *tous* les âges. Par ailleurs lorsqu'une année donnée un taux est biaisé de ce fait, un des deux taux adjacents est affecté d'un biais de sens contraire, sensiblement égal en valeur absolue. Il en résulte que :

- les indicateurs conjoncturels et les âges moyens transversaux sont très peu affectés par ces biais (Figures 7 et 8)
- en revanche, les indicateurs longitudinaux sont nettement régularisés lorsqu'on les établit à l'aide des taux élaborés (Figures 9 et 10)

3- Calcul des taux (de seconde catégorie)

Sont donc concernés par ce chapitre les taux de fécondité générale par âge, les taux de fécondité par âge et rang, ainsi que les taux de primo-nuptialité, masculine et féminine par âge.

Pour chacune des définitions de l'âge, **trois méthodes de calcul des taux par âge sont possibles**, la dernière étant préférable car c'est la seule qui permette de supprimer tous les biais de mesure

a) Age en années révolues

Méthode fruste

Les taux démographiques par âge *révolu*, fondés sur les nombres absolus d'événements observés durant une année n donnée dans les carrés du diagramme de Lexis, sont obtenus habituellement en rapportant le nombre d'événements E_i^n observés dans le carré d'âge révolu i à la demi-somme des effectifs P_i^n et P_i^{n+1} d'âge révolu i au 1^{er} janvier et au 31 décembre de l'année n considérée. Ces taux, que nous appelons *frustes*, sont affectés, comme nous l'avons vu au paragraphe ci-dessus (B2), de *biais* dus à la non uniformité de la distribution des anniversaires au sein de chaque couple de générations consécutives.

Méthode élaborée

- Constance du risque selon l'âge à l'intérieur du carré

Si le risque est *constant* à l'intérieur du carré, ne dépendant ni de l'âge exact x entre i et $i+1$, ni du moment t durant l'année n , l'estimation correcte de ce risque est, comme toujours, le rapport du *nombre absolu d'événements* à la *somme des durées d'exposition au risque*. Or, même si on fait abstraction de la mortalité et des migrations internationales, la durée d'exposition au risque *varie* d'un individu à l'autre selon sa date de naissance à l'intérieur du couple ($n-i-1$, $n-i$) : ceux nés vers le début de l'année $n-i-1$ ou la fin de l'année $n-i$ sont exposés au risque pendant des durées voisines de *zéro*, tandis que ceux nés vers la fin de l'année $n-i-1$ ou le début de l'année $n-i$

Figure 6. FINLANDE, années d'observation 1964-1971
 Taux de PRIMO-NUPTIALITE MASCULINE par AGE REVOLU obtenus par
 la méthode classique (trait continu) et par la méthode de l'ODE (pointillé)
 Les taux obtenus par la méthode classique pour les biennums contenant les années 1940 et 1945 sont indiqués par un cercle

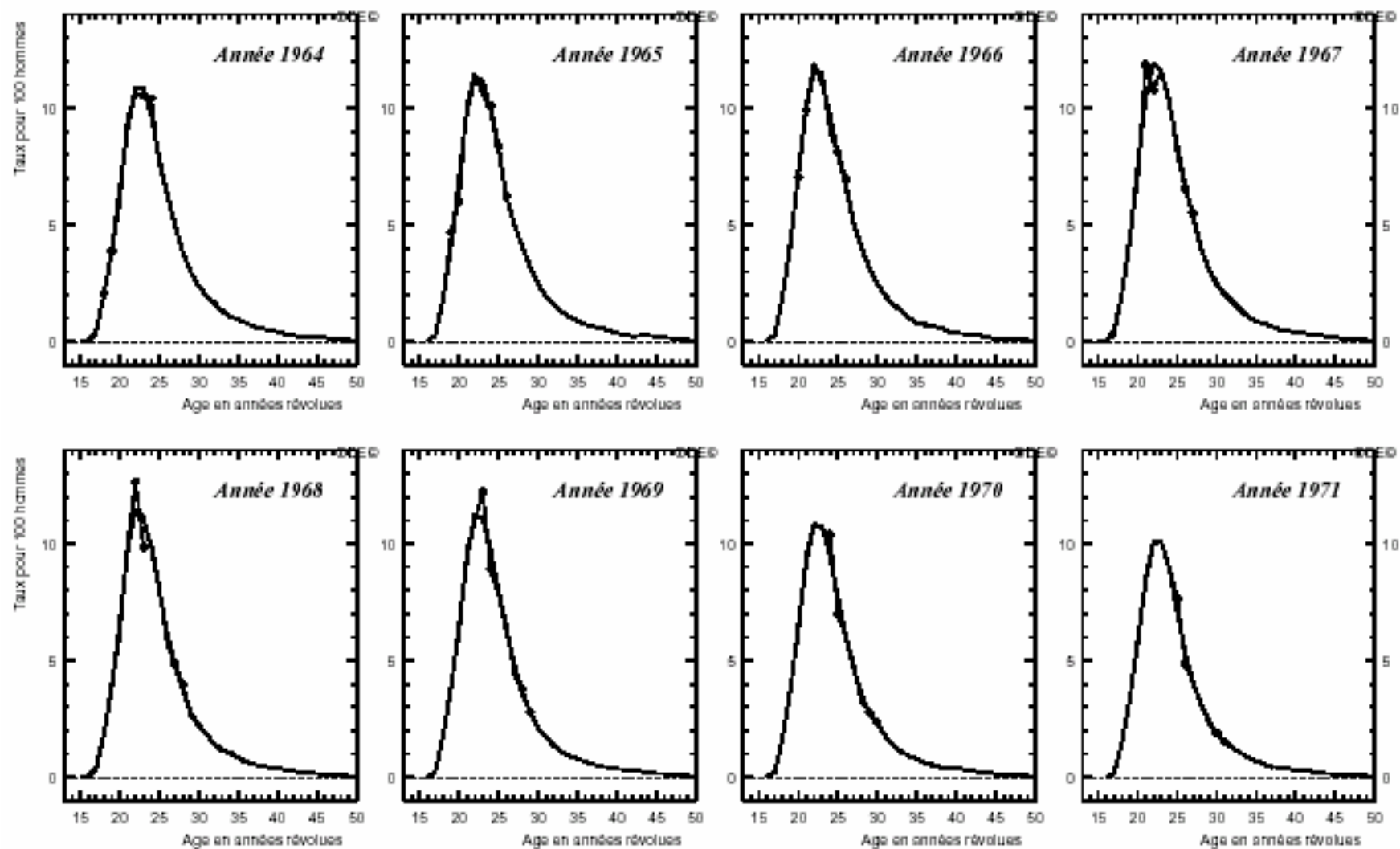


Figure 7. FINLANDE, Années d'observation 1941-1995. INDICATEUR CONJONCTUREL de FECONDITE
Comparaison des indicateurs déduits des taux obtenus par la méthode classique et par la méthode de l'ODE

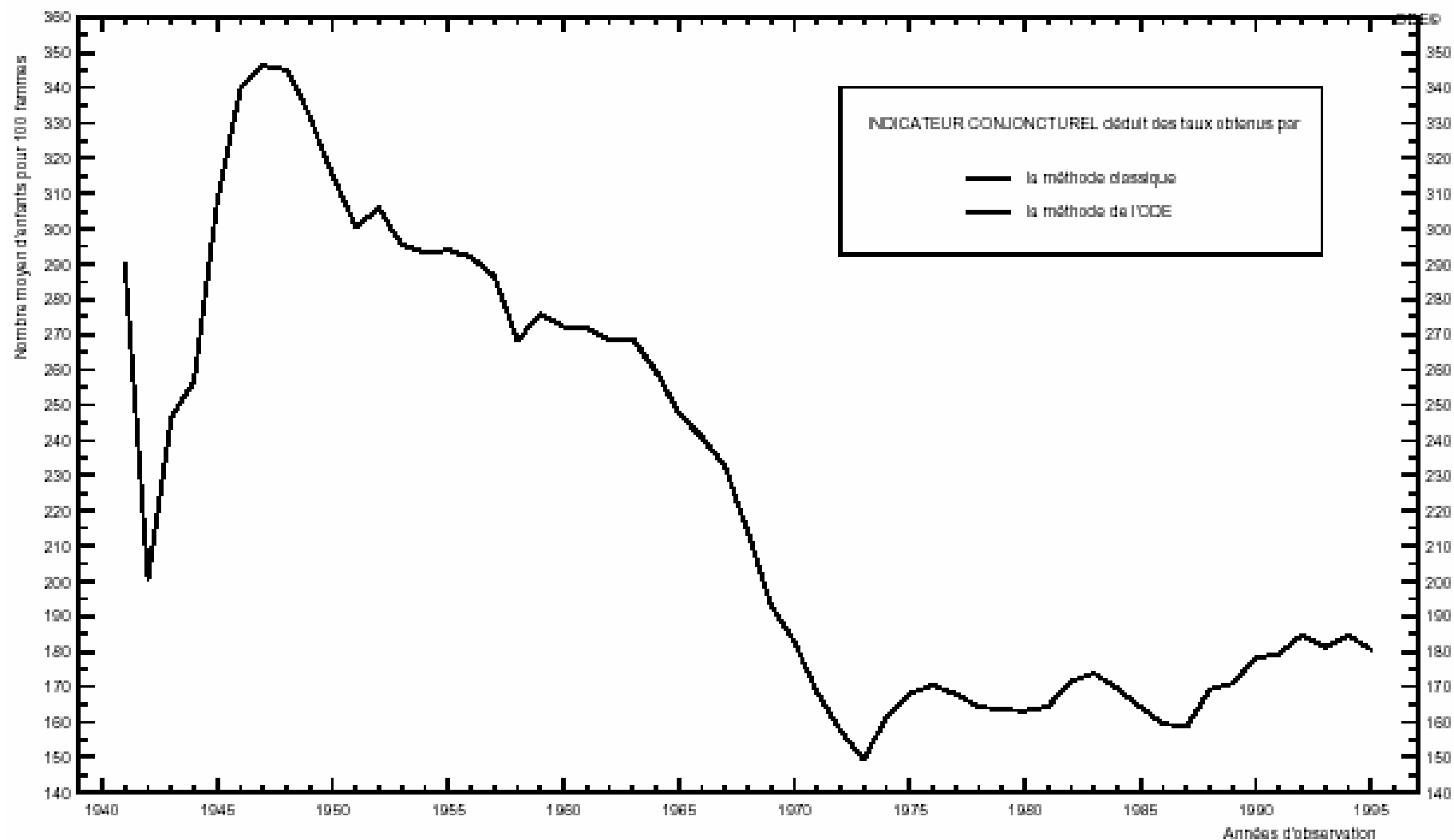


Figure 8. FINLANDE, Années d'observation 1941-1995. AGE MOYEN TRANSVERSAL à la MATERNITE
Comparaison des âges moyens obtenus par la méthode classique et par la méthode de l'ODE
ainsi que sur la base du nombre absolu d'événements

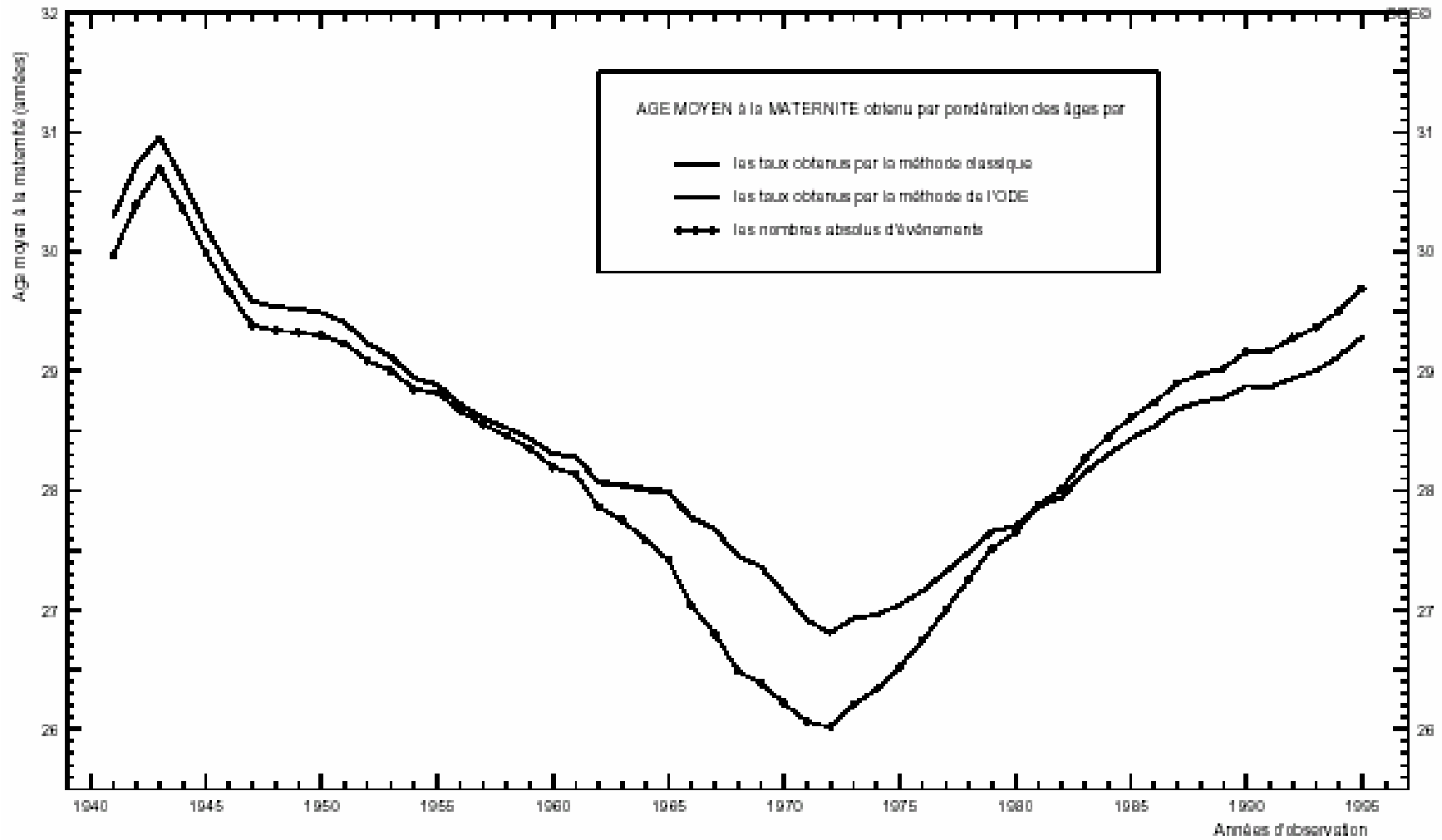


Figure 9. FINLANDE, Générations 1917-1960. DESCENDANCE FINALE
Comparaison des descendance finale déduites des taux obtenus par la méthode classique et par la méthode de l'ODE

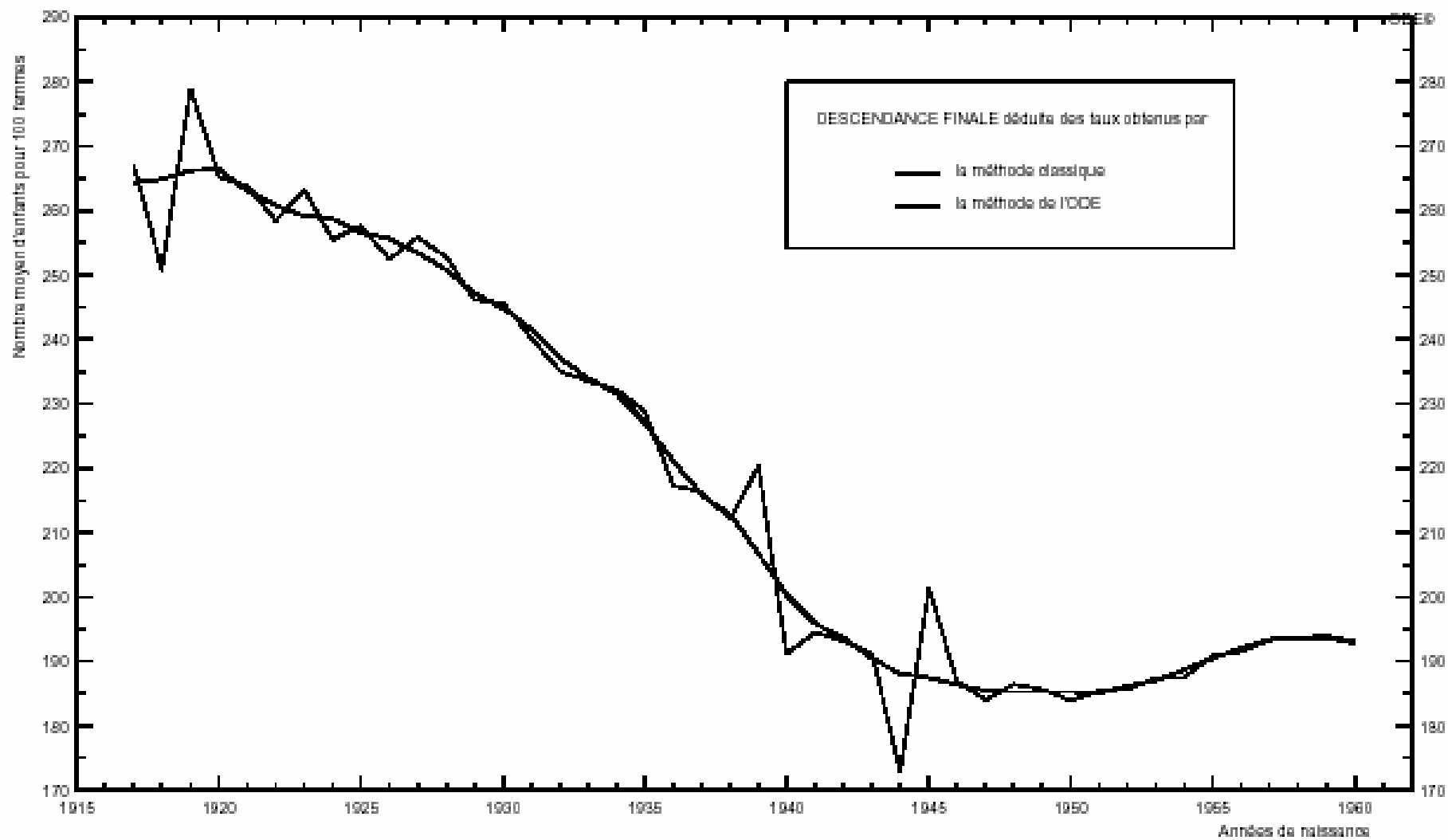
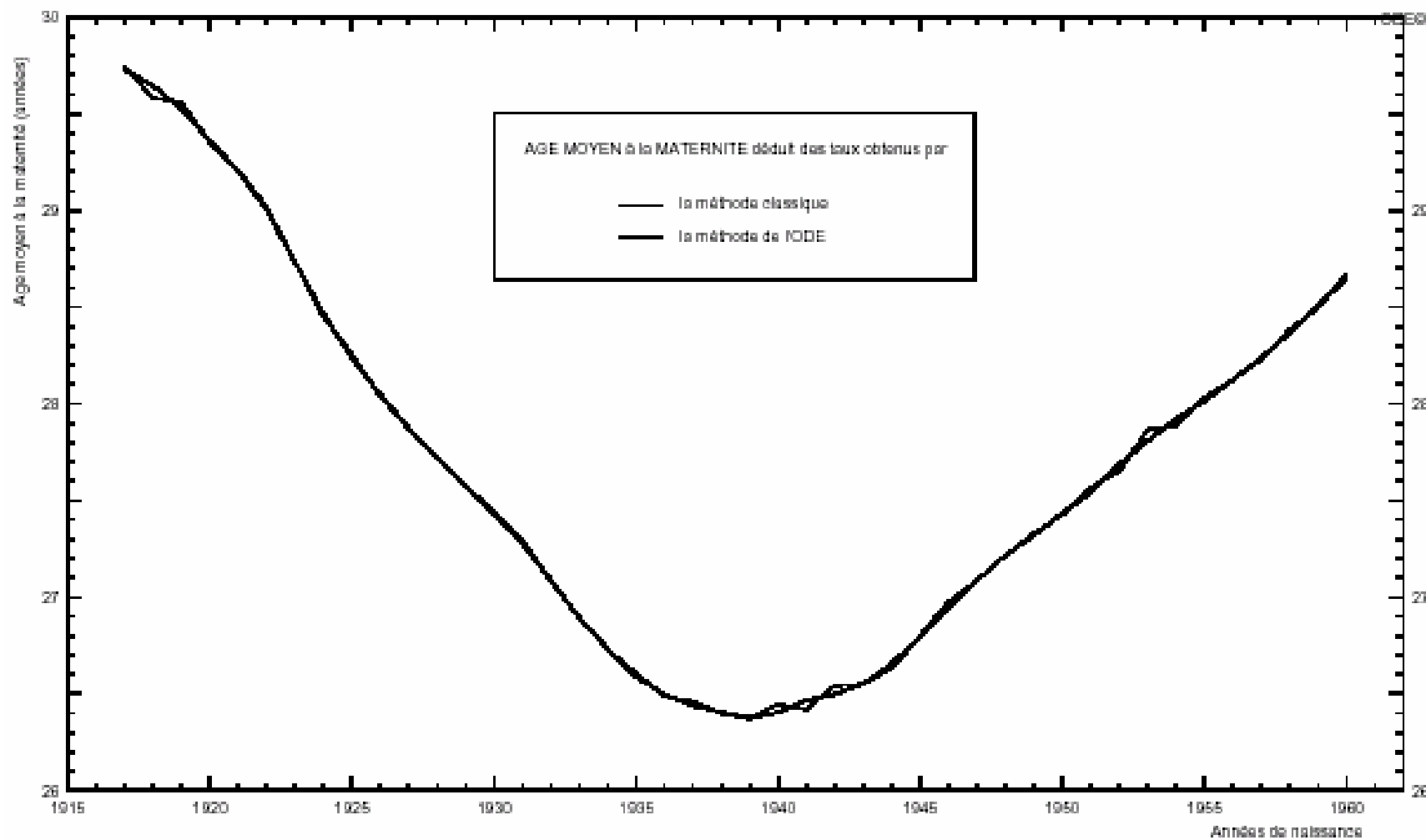


Figure 10. FINLANDE, Générations 1917-1960. AGE MOYEN LONGITUDINAL à la MATERNITE
Comparaison des âges moyens déduits des taux obtenus par la méthode classique et par la méthode de l'ODE



sont exposés au risque durant près d'une année entière. Pour considérer que les P_i^n individus présents le 1^{er} janvier n ont été exposés *en moyenne* pendant une *demi-année*, comme l'ont été aussi les P_i^{n+1} individus présents le 31 décembre de l'année n , il faut alors admettre l'hypothèse que la distribution des anniversaires à l'intérieur de chacune des générations nées en $n-i-1$ et $n-i$ est *uniforme*.

Sous cette hypothèse, on peut montrer que l'estimation *sans biais* du taux f_i^n à l'âge révolu i l'année n est *approximativement* :

$$\hat{f}_i^n = \frac{E_i^n}{P_i^n m_1 + P_i^{n+1} (1 - m_2)}$$

expression où m_1 et m_2 sont les dates *moyennes* de i ème anniversaire (comptées à partir du début de l'année et mesurées en année) pour les individus, *présents* lors de leur i ème anniversaire, qui sont nés respectivement l'année $n-i-1$ et l'année $n-i$.

Dans la mesure où la mortalité et les migrations internationales n'altèrent pas sensiblement l'allure de la distribution des anniversaires au sein d'une même génération, on peut se référer à l'époque où ces générations sont nées et déterminer les dates moyennes de naissance m_1 et m_2 sur la base de la distribution mensuelle des naissances vivantes au cours des années $n-i-1$ et $n-i$. On peut aussi se référer à un recensement récent qui fournirait la répartition de la population résidente par année *et mois* de naissance (voire *jour* de naissance).

Le biais qui affecte tous les taux d'un même couple de générations est ainsi un biais *purelement statistique*, sensiblement *constant* en valeur *relative*, *indépendant* à la fois de l'âge et du phénomène étudié, qu'il s'agisse de fécondité ou de primo-nuptialité. En matière de mortalité, le biais qui affecte les *quotients* est sensiblement le même, en valeur relative, que celui affectant les taux de fécondité ou de primo-nuptialité.

- Prise en compte de la variabilité du risque selon l'âge à l'intérieur du carré

Si on tient compte du fait que le risque $f(x)$ *varie* à l'intérieur du carré ($i \cdot x \cdot i+1$) ainsi que de la mortalité et des migrations internationales, on peut montrer que, moyennant les hypothèses précisées ci-dessous et en désignant par f_i , f_i' et f_i'' , les valeurs de $f(x)$ et de ses deux premières dérivées en $x = i + \frac{1}{2}$, le taux f_i^n à estimer à l'âge révolu i est :

$$f_i^n = f_i + \frac{f_i''}{24}$$

tandis que le nombre absolu d'événements E_i^n observés dans le carré²¹ est égal à :

²¹ L'équation suivante résulte de l'addition des équations correspondant à chacun des triangles :

$$E_1 = N_1 \left[f m_1 + \frac{f'}{2} (m_1 - m_1^2 - V_1) \right] + \frac{f''}{2} \left(\frac{m_1}{4} - \frac{m_1^2 + V_1}{2} + \frac{m_1^3}{3} + m_1 V_1 + \frac{m_1}{3} \right) - \frac{s_1}{6} \left(f + \frac{f''}{40} \right)$$

$$E_2 = N_2 \left\{ f(1 - m_2) - \frac{f'}{2} [m_2 - m_2^2 - V_2] + \frac{f''}{2} \left[\frac{1}{12} - \frac{m_2}{4} + \frac{m_2^2 + V_2}{2} - \frac{m_2^3}{3} - m_2 V_2 - \frac{m_2}{3} \right] \right\} + \frac{s_2}{6} \left[f + \frac{f''}{40} \right]$$

C'est à l'aide de ces formules que l'on estime le nombre d'événements dans chacun des triangles pour en déduire le nombre d'événements dans le parallélogramme à base verticale qui permettra de calculer les taux de survénances selon l'âge atteint dans l'année.

$$\begin{aligned}
E_i^n = & \left(P_i^n + \frac{s_1}{2} \right) \left\{ f m_1 + \frac{f'}{2} (m_1 - m_1^2 - V_1) + \frac{f''}{2} \left(\frac{m_1 - m_1^2 + V_1}{4} + \frac{m_1^3}{2} + m_1 V_1 + \frac{m_1}{3} \right) \right\} \\
& + \left(P_i^{n+1} - \frac{s_2}{2} \right) \left\{ f (1 - m_2) - \frac{f'}{2} [m_2 - m_2^2 - V_2] + \frac{f''}{2} \left[\frac{1}{12} - \frac{m_2}{4} + \frac{m_2^2 + V_2}{2} - \frac{m_2^3}{3} - m_2 V_2 - \frac{m_2}{3} \right] \right\} \quad (1) \\
& + \frac{s_2 - s_1}{6} \left(f + \frac{f''}{40} \right)
\end{aligned}$$

expression dans laquelle on a omis, pour alléger l'écriture, l'indice i de f , f' et f'' et où s_1 et s_2 désignent les densités de solde migratoire apparent dans les triangles respectivement supérieur et inférieur du carré :

$$s_1 = P_{i+1}^{n+1} - P_i^n, \quad s_2 = P_i^{n+1} - P_{i-1}^n$$

tandis que m_1 , V_1 et m_1 désignent respectivement la moyenne, la variance et le moment centré d'ordre 3 de la distribution des anniversaires au sein de la génération née en $n-i-1$, m_2 , V_2 et m_2 les quantités analogues pour la génération née en $n-i$. Lorsqu'au sein d'une génération la distribution des anniversaires est *uniforme*, c'est-à-dire lorsque la *densité* des lignes de vie est *constante*, les valeurs de m , V et m sont égales respectivement à $\frac{1}{2}$, $\frac{1}{12}$ et 0.

Ces résultats reposent sur les hypothèses ci-après :

- les migrants, avant leur émigration ou après leur immigration, de même que les personnes décédées, avant leur décès, sont exposées à un risque *identique* à celui des individus constamment présents ;
- dans le diagramme de Lexis, les points représentatifs des entrées (par immigration) et des sorties (par émigration ou décès) se répartissent *uniformément* à l'intérieur d'un même parallélogramme à côtés verticaux, spécifique d'une même génération durant l'année considérée ;
- la fonction de risque dépend de l'âge mais non du *moment* dans l'année : le phénomène étudié est supposé dépourvu de *saisonnalité* ;
- la fonction de risque $f(x)$ varie *lentement* sur un intervalle d'âge de plusieurs années consécutives, de sorte que les développements polynomiaux de $f(x)$ à l'ordre 2 sont satisfaisants ;
- les quantités m_1 , V_1 et m_1 ainsi que m_2 , V_2 et m_2 sont *connues* (évaluées sur la base de la distribution mensuelle de naissances au cours de l'année $n-i-1$ et de l'année $n-i$ ou sur la base d'un recensement récent).

L'estimation des taux f_i^n est réalisée par *itérations* successives. A la première itération, on suppose f_i' et f_i'' nuls pour tout i et on déduit de E_i^n , P_i^n , P_i^{n+1} , P_{i-1}^n et P_{i+1}^{n+1} , au moyen de la relation (1), f_i , puis $f_i^n = f_i$. A la chaque itération ultérieure, on estime f' et f'' par ajustement parabolique des valeurs de f obtenues à l'itération *précédente* sur, par exemple, 5 points consécutifs :

$$\begin{aligned}
f_i' &= \frac{1}{5} (f_{i+2} - f_{i-2}) + \frac{1}{10} (f_{i+1} - f_{i-1}) \\
f_i'' &= \frac{2}{7} (f_{i+2} + f_{i-2} - f_i) - \frac{1}{7} (f_{i+1} + f_{i-1})
\end{aligned}$$

et on déduit de E_i^n , P_i^n , P_i^{n+1} , P_{i-1}^n et P_{i+1}^{n+1} , au moyen à nouveau de la relation (1), f_i , puis $f_i^n = f_i + \frac{f_i''}{24}$. On arrête les itérations lorsque les estimations f_i^n sont stabilisées pour tout i . Cette procédure offre l'avantage d'opérer des *lissages* sur les données, qui atténuent les effets des variations accidentelles.

Si on néglige les quantités s_1 et s_2 ainsi que les dérivées d'ordre 1 et 2 de $f(x)$, une seule itération est nécessaire et l'estimation du taux f_i^n est celle indiquée plus haut :

$$f = \frac{E_i^n}{P_i^n m_1 + P_i^{n+1} (1 - m_2)}$$

alors que l'estimation classique est :

$$f = \frac{E_i^n}{(P_i^n + P_i^{n+1}) / 2}$$

Le biais dont est affectée l'estimation classique résulte ainsi :

- quand P_i^n et P_i^{n+1} sont voisins, de ce que m_1 et m_2 sont différents (situation extrêmement peu fréquente, car, lorsque m_1 et m_2 sont différents, P_i^n et P_i^{n+1} le sont généralement aussi) ;
- quand P_i^n et P_i^{n+1} sont différents, de ce que m_1 et m_2 diffèrent de $1/2$.

Lorsque la répartition des naissances durant le couple d'années n'est guère influencée que par les variations *saisonnnières* de la natalité, l'expérience montre que le biais est *négligeable*. Mais, pour les couples marqués par des variations *brutales* de la natalité, par exemple à la fin et, surtout, au *début* d'une guerre, le biais peut atteindre des valeurs appréciables et conduire à des erreurs d'un ordre de grandeur important par rapport à celui de la variation, d'un âge à l'autre, des taux observés la même année et, surtout, par rapport à la variation, d'une année à l'autre, des taux à âge égal. Alors le biais affecte le couple à *tous* les âges, c'est-à-dire tout au long de son existence, en gardant sensiblement la même valeur *relative*. Il s'ensuit notamment que l'estimation des caractéristiques *longitudinales* du couple de génération est affectée du biais relatif commun aux taux à chacun des âges.

Il est donc souhaitable que les logiciels soient établis à partir de la méthodologie décrite ci dessus et estiment, de manière unique pour tous les pays, les taux à tous les âges et pour toutes les années. Ils doivent également permettre d'établir, sur option, les taux selon la méthode classique (taux frustes). Lorsque les caractéristiques m , V et m d'un couple de générations ne sont pas disponibles, il faut retenir, *faute de mieux*, des valeurs égales respectivement à $\frac{1}{2}$, $\frac{1}{12}$ et 0.

b) Age atteint dans l'année

Méthode fruste

Les taux démographiques par âge *atteint dans l'année*, fondés sur les nombres absolus d'événements observés durant une année n donnée dans les *parallélogrammes à base verticale* du diagramme de Lexis, sont obtenus habituellement en rapportant le nombre d'événements E_i^n observés dans le parallélogramme d'âge atteint i à la demi-somme des effectifs P_{i-1}^n et P_i^{n+1} d'âge révolu i au 1^{er} janvier et $i+1$ au 31 décembre de l'année n considérée. Ces taux, *frustes*, sont affectés, eux aussi, de *biais*, mais d'une ampleur notablement plus faible que ceux des taux par âge révolu.

Méthode élaborée

Quand les événements sont classés dans un parallélogramme à côtés verticaux, les formules sont beaucoup plus simples car une seule génération intervient.

Le taux f_i^n à estimer à l'âge atteint i est toujours :

$$f_i^n = f_i + \frac{f_i''}{24}$$

tandis que le nombre absolu d'événements E_i^n observés dans le parallélogramme²² devient égal à :

$$E_i^n = \left(P_{i-1}^n + \frac{s}{2} \right) \left\{ f + f' \left(m - \frac{1}{2} \right) + \frac{f''}{2} \left(\frac{1}{3} - m + m^2 + V \right) \right\} + s \frac{f'}{6} \quad (2)$$

expression dans laquelle s désigne la densité de solde migratoire apparent dans le parallélogramme d'âge atteint i :

$$s = P_i^{n+1} - P_{i-1}^n$$

tandis que m et V désignent respectivement la moyenne et la variance de la distribution des anniversaires au sein de la génération née en $n-i$.

L'estimation des taux f_i^n étant toujours réalisée par *itérations* successives.

4- Conversions des événements et des taux dans une autre figure de Lexis

La procédure utilisée consiste à estimer les nombres d'événements dans chacun des triangles constitutifs de la figure originale à partir de la fonction de survéance, calculée dans cette dernière figure, appliquée dans chacune des équations spécifiques des divers triangles.

La phase suivante consiste donc dans l'agrégation des triangles dans les figures du schéma de Lexis adéquates, puis le calcul des taux dans ces nouvelles figures à l'aide de l'équation spécifique.

Séquence des opérations

1 - Taux frustes

- calcul de la population moyenne par moyenne arithmétique :

$$\bar{P}_i^n = \frac{P_i^n + P_{i+1}^n}{2} \quad \text{ou} \quad \bar{P}_{n-i}^n = \frac{P_{i-1}^n + P_i^{n+1}}{2}, \quad \text{selon le cas}$$

- calcul du taux :

$$f_i^n = \frac{E_i^n}{(P_i^n + P_{i+1}^n)/2} \quad \text{ou} \quad f_{n-g}^n = \frac{E_i^n}{(P_{i-1}^n + P_i^{n+1})/2}, \quad \text{selon le cas}$$

2- Taux élaborés (âge révolu)

- constance du risque

²² Sur la base de l'addition des deux triangles :

$$E_1 = N \left[f m - \frac{f'}{2} (m^2 + V) + \frac{f''}{6} (m^3 + 3mV + m) \right] - s \left(\frac{f}{6} + \frac{f'}{12} + \frac{f''}{40} \right)$$

$$E_2 = N \left[f(1-m) + \frac{f'}{2} [(1-m)^2 + V] + \frac{f''}{6} [(1-m)^3 + 3(1-m)V - m] \right] + s \left(\frac{f}{6} + \frac{f'}{12} + \frac{f''}{40} \right)$$

avec $N = \left(P_{i-1}^n + \frac{s}{2} \right)$

- calcul des caractéristiques à la naissance : m_1 et m_2 ,
- calcul du taux :

$$\hat{f}_i^n = \frac{E_i^n}{P_i^n m_1 + P_i^{n+1} (1 - m_2)}$$

b) risque variable

- calcul des caractéristiques à la naissance : m , V et m , en leur absence retenir des valeurs égales respectivement à $\frac{1}{2}$, $\frac{1}{12}$ et 0.

- estimation du taux, $f_i^n = f_i + \frac{f_i''}{24}$, réalisée par *itérations* successives

- à la 1^{ère} itération : - on suppose f_i' et f_i'' (dérivées première et seconde) *nuls* pour tout i

- on déduit de E_i^n , P_i^n , P_i^{n+1} , P_{i-1}^n et P_{i+1}^{n+1} , au moyen de la relation (1), f_i , puis $f_i^n = f_i$

- à chaque itération ultérieure : - on estime f' et f'' par ajustement parabolique des valeurs de f obtenues à l'itération *précédente*

- on déduit de E_i^n , P_i^n , P_i^{n+1} , P_{i-1}^n et P_{i+1}^{n+1} , au moyen à nouveau de la relation (1), f_i , puis $f_i^n = f_i + \frac{f_i''}{24}$.

- estimation des événements dans les triangles constitutifs des carrés pour en déduire ceux dans les parallélogrammes à base verticale.

3- Taux élaborés (âge atteint)

- calcul des caractéristiques à la naissance : m et V (en leur absence retenir $\frac{1}{2}$ et $\frac{1}{12}$).

- estimation du taux, $f_i^n = f_i + \frac{f_i''}{24}$, réalisée par *itérations* successives

- à la 1^{ère} itération : - on suppose f_i' et f_i'' (dérivées première et seconde) *nuls* pour tout i

- on déduit de E_i^n , P_i^{n+1} et P_{i-1}^n , au moyen de la relation (2), f_i , puis $f_i^n = f_i$.

- à chaque itération ultérieure : - on estime f' et f'' par ajustement parabolique des valeurs de f obtenues à l'itération *précédente*

- on déduit de E_i^n , P_i^{n+1} et P_{i-1}^n , toujours au moyen de la relation (2), f_i , puis $f_i^n = f_i + \frac{f_i''}{24}$;

- estimation des événements dans les triangles constitutifs des parallélogrammes pour en déduire ceux dans les carrés.

C) Événements non renouvelables

1- Estimation des quotients et construction des tables de survie pour les événements non renouvelables (comme la mortalité)

Les données statistiques disponibles pour analyser la survie des événements non renouvelables se présentent de la même façon que pour les événements renouvelables : nombres absolus d'événements en parallélogrammes à côtés verticaux, en carrés ou les deux à la fois, c'est-à-dire en triangles. En revanche, la méthode de construction des tables de survie est différente. En effet, dans le cas des événements renouvelables l'agrégation des taux se fait par addition, alors que pour les événements non renouvelables les compléments à 1 des probabilités d'occurrence s'enchaînent par multiplication.

De la même façon que pour le calcul des taux, il convient de distinguer deux cas, selon la définition de l'âge retenue pour classer les événements.

a) Age en années révolues

Méthode fruste

S'agissant des décès, la plupart des pays européens disposent de données en *triangles* depuis déjà de nombreuses années (par exemple la France depuis 1907).

Désignons par P_i^n l'effectif au 1^{er} janvier n de la population d'un sexe donné à l'âge révolu i et par C_i^n le nombre de décès situés dans le carré correspondant à l'âge révolu i l'année n .

La méthode *fruste*, appliquée aux données en triangles ou en carrés, consiste à estimer le quotient de mortalité ${}_1Q_i^n$ ²³ à l'âge révolu i de la façon suivante : on estime d'abord le *taux* de mortalité t_i^n comme s'il s'agissait d'un événement renouvelable par

$$t_i^n = \frac{C_i^n}{(P_i^n + P_{i+1}^n)/2}, \text{ puis le quotient lui-même par } {}_1Q_i^n = 1 - e^{-t_i^n} \approx \frac{t_i^n}{1 + t_i^n/2} \text{ (l'expression}$$

approchée ne vaut que quand le taux t_i^n est petit). La méthode fruste repose implicitement sur deux hypothèses : hypothèse d'*uniformité* du régime saisonnier des naissances durant le couple d'années $n-i-1$ et $n-i$ comme dans le cas des événements renouvelables, hypothèse de variations *faibles* du quotient instantané de mortalité entre les âges i et $i+1$.

C'est pourquoi une méthodologie plus élaborée doit être utilisée, qui ne suppose pas ces hypothèses satisfaites, et qui opère en outre des *lissages*, plus souhaitables encore en matière de mortalité qu'en matière de fécondité car les « risques » — et donc les nombres absolus d'événements — sont souvent nettement plus petits. En particulier, comme les données observées sont souvent de médiocre qualité aux grands âges et que, parallèlement, les effectifs se réduisent, les quotients prennent une allure erratique, la méthodologie de l'ODE retient, sur option, l'hypothèse supplémentaire qu'au-delà d'un certain âge (par exemple : 85 ans) le *logarithme* du quotient de mortalité varie *linéairement* avec l'âge : on ajuste alors une droite par exemple aux *dix* données précédentes (ici : de 70 à 89 ans) et on lit les données suivantes (ici : à partir de 90 ans) sur la droite ajustée²⁴. Cette méthodologie

²³ L'indice inférieur gauche 1 exprime le fait qu'il s'agit d'un quotient sur un intervalle d'*un* an : entre les âges exacts i et $i+1$.

²⁴ Le calcul de l'espérance de vie à tous les âges tient compte de ce qu'à 99 ans l'espérance de vie restante n'est pas nulle, ni fixée a priori : elle est égale à l'espérance de vie restante fondée sur les quotients extrapolés indéfiniment (extrapolation linéaire des logarithmes des quotients).

doit fournir, pour chaque âge i de 0²⁵ à 99 ans, le quotient de mortalité ${}_1Q_i^n$ à l'âge révolu i , le nombre de survivants S_i^n à l'âge exact i et l'espérance de vie restante à l'âge exact i .

Méthode élaborée

Cette méthodologie utilisée pour le calcul des quotients est tout à fait semblable à celle qui a été décrite précédemment pour le calcul des taux de seconde catégorie. La procédure est identique, seules les formules changent.

Dans le cas d'événements classés selon l'âge en années révolues (carrés de Lexis) on peut, comme pour le calcul des taux, faire le choix de deux hypothèses concernant l'évolution du risque à l'intérieur du carré : risque constant ou variable.

Nous n'entrerons dans le détail du calcul qui conduirait à répéter ce qui a été écrit au paragraphe précédent ; nous nous bornerons à fournir les formules à utiliser et à indiquer les quelques particularités²⁶.

- Constance du risque selon l'âge à l'intérieur du carré

Après avoir calculé les caractéristiques à la naissance des générations concernées, il convient de calculer le taux par la formule suivante, qui montre que l'effet approximatif des migrations sur le nombre de décès d'un carré du diagramme de Lexis est lié à la différence des balances migratoires des deux triangles constitutifs de ce carré :

$$\hat{q}_i^n = \frac{E_i^n}{P_i^n m_1 + P_i^{n+1}(1 - m_2) + \frac{s_2 - s_1}{6}}$$

et de convertir ce taux en quotient :

$${}_1Q_i^n = 1 - e^{-\hat{q}_i^n} \approx \frac{\hat{q}_i^n}{1 + \hat{q}_i^n / 2}$$

- Prise en compte de la variabilité du risque selon l'âge à l'intérieur du carré

Dans ce cas le nombre d'événements dans le carré prend la forme :

$$SQ_i^n = \left(P_i^n + \frac{s_1}{2} \right) \left[e^{qm_1 + \frac{q'}{2}[m_1(1-m_1)-V_1] + \frac{V_1}{2}q^2} - 1 \right] + \left(P_i^{n+1} - \frac{s_2}{2} \right) \left[1 - e^{-q(1-m_2) + \frac{q'}{2}[m_2(1-m_2)-V_2] + \frac{q^2}{2}V_2} \right] + \frac{s_1}{2} \left[1 - e^{-\frac{q}{3} + \frac{q^2}{36}} \right] + \frac{s_2}{2} \left[1 - e^{-\frac{q}{3} + \frac{q^2}{36}} \right]$$

(3)

Et l'on utilise pour estimer le quotient instantané q , ou taux de mortalité, la même procédure itérative que pour l'estimation du taux de seconde catégorie des événements renouvelables. Mais il existe deux possibilités :

- soit on considère q constant :
 - à la première itération $q'_i = 0$,
 - aux itérations ultérieures $q'_i = \frac{q_{i+1} - q_{i-1}}{2}$

²⁵ Aux âges 0 et 1 an, la méthodologie diffère de celle appliquée aux autres âges, en raison de la *rapidité* des variations du quotient instantané de mortalité avec l'âge. Indiquons d'autre part que le quotient de mortalité à 0 an est habituellement dénommé *taux de mortalité infantile*.

²⁶ On trouvera une description précise de cette méthode, utilisée par l'ODE, dans l'article de Gérard Calot et Ana Franco "The construction of life tables" figurant dans l'ouvrage publié par Guillaume Wunsch, Michel Mouchart et Josianne Duchêne (éditeurs), *Life Tables : Data, Methods, Models*, Kluwer, 2001, pp. 31-75, et reproduit en annexe de ce rapport.

- soit c'est la dérivée logarithmique $\frac{q'_i}{q_i}$ qui est constante :

- à la première itération la dérivée est égale à 0,
- aux itérations ultérieures à $\frac{\text{Log}(q_{i+1}) - \text{Log}(q_{i-1}))}{2} = 0,5 * \text{Log}\left(\frac{q_{i+1}}{q_{i-1}}\right)$

Ensuite, on calcule le quotient fini par :

$${}_1Q_i^n = 1 - e^{-q} \approx \frac{q}{1 + q/2}$$

Si le nombre de décès n'est connu qu'à l'intérieur du carré et non dans chacun des triangles la procédure itérative ci-dessus doit être précédée d'une étape, elle aussi itérative, au cours de laquelle sera estimé le nombre de décès dans chacun des triangles à partir des formules suivantes :

Triangle inférieur :

$$D_2 = N_2 - P_2 + \frac{S_2}{2}$$

$$\approx N_2 \left\{ 1 - e^{-q(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{v_2}{2}(q^2 - q') \right] \right\} + \frac{S_2}{2} \left[1 - e^{-\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) \right]$$

Triangle supérieur :

$$D_1 = P_1 - N_1 + \frac{S_1}{2}$$

$$\approx N_1 \left\{ e^{qm_1 + \frac{q'}{2}m_1(1-m_1)} \left[1 + \frac{v_1}{2}(q^2 - q') \right] - 1 \right\} - \frac{S_1}{2} \left[e^{\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) - 1 \right]$$

b) Age atteint dans l'année

Méthode fruste

Dans le cas du parallélogramme à côtés verticaux, la formule de calcul du taux est voisine de celle utilisée pour les décès classés selon l'âge en années révolues :

$$t_i^n = \frac{C_i^n}{(P_i^n + P_{i+1}^{n+1})/2}$$

Et la transformation du taux en quotient se fait de la même manière, à ceci près que le quotient est défini entre les âges exacts $i-0,5$ et $i+0,5$:

$${}_1Q_{i-0,5}^n = 1 - e^{-t_i^n} \approx \frac{t_i^n}{1 + t_i^n/2}$$

Méthode élaborée

Encore une fois la procédure est identique, seule l'équation donnant les événements dans le parallélogramme à côtés verticaux²⁷ diffère :

²⁷ Dans ce cas les équations correspondant à chacun des triangles sont :

$$E_1 = N \left[e^{qm - \frac{q'}{2}(m^2 + v) + \frac{q^2}{2}v} - 1 \right] + \frac{S}{2} \left[1 - e^{-\frac{q}{3} - \frac{q'}{6} + \frac{q^2}{36}} \right]$$

$$E_2 = N \left[1 - e^{-q(1-m) - \frac{q'}{2}[(1-m)^2 + v] + \frac{q^2}{2}v} \right] + \frac{S}{2} \left[1 - e^{-\frac{q}{3} - \frac{q'}{6} + \frac{q^2}{36}} \right]$$

avec $N = \left[\frac{P_{i-1}^n + P_i^{n+1}}{2} + \frac{D_{i,i}^n + D_{2,i}^n}{2} \right]$

$$VSP_i^n = \left[\frac{P_{i-1}^n + P_i^{n+1}}{2} + \frac{D_{1,i}^n + D_{2,i}^n}{2} \right] \left[1 - e^{-q - \frac{q'}{2}(1-2m)} \right] e^{qm - \frac{q'}{2}(m^2 + V) + \frac{q^2}{2}V} + \frac{s}{2} \left[2 - e^{\frac{q}{3} - \frac{q'}{6} + \frac{q^2}{36}} - e^{-\frac{q}{3} - \frac{q'}{6} + \frac{q^2}{36}} \right] \quad (4)$$

On estime le taux ou quotient instantané à partir de l'équation ci-dessus, en choisissant de considérer comme constant le quotient instantané ou sa dérivée logarithmique, puis on transforme le taux en quotient.

L'inconvénient de ces quotients, *perspectifs*, est qu'ils ne sont pas calculés entre anniversaires mais entre deux premier janvier successifs ; ils sont donc décalés d'une demi-année d'âge. On peut passer, pour la plupart des âges sauf les très jeunes, des quotients perspectifs aux quotients entre anniversaires par interpolation linéaire ou, de préférence, logarithmique. Mais il est préférable d'utiliser la procédure, décrite ci-dessus, pour estimer les décès dans chacun des triangles et d'en déduire une table entre anniversaires.

c) Spécificité de la mortalité à 0 et 1 an

Tout ce que vient d'être décrit dans ce chapitre consacré à l'estimation des quotients s'applique quel que soit le phénomène étudié (fécondité, nuptialité et mortalité) et l'âge considéré à l'exception de la mortalité à 0 an et, dans une moindre mesure, 1 an. En effet, la décroissance très rapide du quotient instantané, et donc de la fonction de risque, depuis la naissance met en échec l'approximation polynomiale utilisée.

On peut alors estimer le quotient de mortalité à l'âge 0, sous les deux hypothèses de migrations nulles dans le carré et de distributions des anniversaires des générations nées en $n-1$ et n identiques, par :

$$\hat{Q} = 1 - \left(1 - \frac{D_1}{P_1} \right) \left(1 - \frac{D_2}{N_2} \right)$$

avec D_1 et D_2 = décès des triangles supérieur et inférieur,

P_1 = population de 0 an révolu au 1^{er} janvier de l'année considérée,

et N_2 = naissances de l'année considérée

En présence de migrations dans ce premier carré la formule précédente peut se réécrire ainsi :

$$\hat{Q} = 1 - (1 - a)(1 - b)$$

Pour estimer a (qui se réfère au triangle supérieur) et b (triangle inférieur) deux hypothèses peuvent être faites quant à la place de ces migrations :

- pour le triangle supérieur, la première hypothèse suppose que les migrations de l'année n de la génération née en $n-1$ interviennent le 1^{er} janvier de l'année n (a_1) ; la seconde, que celles de la génération n interviennent le 31 décembre (a_2) :

$$a_1 = \frac{D_1^0}{P_2^1 + D_1^0 + D_2^1} \quad \text{et} \quad a_2 = \frac{D_1^0}{P_1^0}$$

- Pour le triangle inférieur, la première hypothèse suppose que les migrations de l'année n de la génération née en $n-1$ interviennent immédiatement après la naissance (b_1) ; la seconde, que celles de la génération n interviennent exclusivement le 31 décembre (b_2) :

$$b_1 = \frac{D_2^0}{P_2^0 + D_2^0} \quad \text{et} \quad b_2 = \frac{D_2^0}{N_2}$$

On peut alors calculer les bornes inférieure et supérieure de \hat{Q} par :

$$1 - [1 - \min(a_1, a_2)] [1 - \min(b_1, b_2)] \quad \text{et} \quad 1 - [1 - \max(a_1, a_2)] [1 - \max(b_1, b_2)]$$

Ces limites sont naturellement proches lorsque les balances migratoires sont faibles tant dans le parallélogramme à base verticale correspondant à la génération n-1 que dans le triangle inférieur. On retiendra la moyenne de ces bornes.

La même procédure peut être utilisée pour estimer la probabilité de décéder entre le 1^{er} et le 2^{ème} anniversaire. Les bornes inférieure et supérieure sont alors :

$$1 - [1 - \min(a_1, a_2)] [1 - \min(b_1, b_2)] \quad \text{et} \quad 1 - [1 - \max(a_1, a_2)] [1 - \max(b_1, b_2)]$$

avec

$$a_1 = \frac{D_1^1}{P_2^2 + D_1^1 + D_2^2}, \quad a_2 = \frac{D_1^1}{P_1^1},$$

$$b_1 = \frac{D_2^1}{P_2^1 + D_2^1} \quad \text{et} \quad b_2 = \frac{D_2^1}{P_1^0 - D_1^0}$$

Rappelons que la primo-nuptialité peut être traitée comme produisant des événements renouvelables ou non renouvelables. En matière d'indicateurs *longitudinaux*, les deux méthodes de traitement conduisent aux *mêmes* résultats²⁸, mais il en va différemment des indicateurs *transversaux*. En particulier, l'indicateur conjoncturel fondé sur les *taux* peut fort bien dépasser l'unité, ce qui n'est jamais le cas de l'indicateur conjoncturel fondé sur les *quotients*, ni de l'indicateur *longitudinal* (proportion de non célibataires à 50 ans). Par ailleurs, compte dûment tenu du décalage égal à l'âge moyen au premier mariage, les indicateurs longitudinaux et transversaux sont généralement beaucoup plus proches les uns des autres lorsque ces derniers sont fondés sur les quotients que lorsqu'ils sont fondés sur les taux. Malgré tout, la tradition observée dans les publications des offices statistiques européens consiste à traiter la primo-nuptialité de la même façon que la fécondité, c'est-à-dire comme produisant des événements *renouvelables*. Toutefois, les outils dont disposera Eurostat permettront de traiter *successivement* la primo-nuptialité comme produisant des événements renouvelables, puis non renouvelables.

Séquence des opérations

1 - *Méthode fruste* (données en triangles ou en carrés)

- estimation du taux : $t_i^n = \frac{C_i^n}{(P_i^n + P_i^{n+1})/2}$
- transformation du taux en quotient : ${}_1Q_i^n = 1 - e^{-t_i^n} \approx \frac{t_i^n}{1 + t_i^n/2}$

2- *Méthode élaborée (âge révolu)*

²⁸ C'est la raison qui justifie qu'on traite souvent des phénomènes à événements *non renouvelables* (ainsi la primo-nuptialité ou le divorce) comme s'ils produisaient des événements *renouvelables*.

a) constance du risque

- calcul des caractéristiques à la naissance : m_1 et m_2 ,

- calcul du taux : $\hat{q}_i^n = \frac{E_i^n}{P_i^n m_1 + P_i^{n+1}(1 - m_2) + \frac{s_2 - s_1}{6}}$

- transformation du taux en quotient : ${}_1Q_i^n = 1 - e^{-\hat{q}_i^n} \approx \frac{\hat{q}_i^n}{1 + \hat{q}_i^n / 2}$

b) risque variable

- calcul des caractéristiques à la naissance : m et V , en leur absence retenir des valeurs égales respectivement à $\frac{1}{2}$ et $\frac{1}{12}$.

- estimation du quotient instantané \hat{q}_i^n , réalisée par *itérations* successives au moyen de la relation (3), en supposant :

- soit q constant :

à la 1^{ère} itération : - on suppose $q'_i = 0$

à chaque itération ultérieure : $q'_i = \frac{q_{i+1} - q_{i-1}}{2}$

- soit $\frac{q'_i}{q_i}$ constant

à la 1^{ère} itération : - on suppose $\frac{q'_i}{q_i} = 0$

à chaque itération ultérieure : $\frac{q'_i}{q_i} = 0,5 * \text{Log} \left(\frac{q_{i+1}}{q_{i-1}} \right)$

- estimation du quotient par :

$${}_1Q_i^n = 1 - e^{-q} \approx \frac{q}{1 + q/2}$$

3- *Méthode fruste (âge atteint)*

- estimation du taux : $t_i^n = \frac{C_i^n}{(P_i^n + P_{i+1}^{n+1}) / 2}$

- transformation du taux en quotient : ${}_1Q_{i-0,5}^n = 1 - e^{-t_i^n} \approx \frac{t_i^n}{1 + t_i^n / 2}$

3- *Méthode élaborée (âge atteint)*

a) risque variable

- calcul des caractéristiques à la naissance : m et V , en leur absence retenir des valeurs égales respectivement à $\frac{1}{2}$ et $\frac{1}{12}$.

- estimation du quotient instantané \hat{q}_i^n , réalisée par *itérations* successives au moyen de la relation (4), en supposant :

- soit q constant :

à la 1^{ère} itération : - on suppose $q'_i = 0$

à chaque itération ultérieure : $q'_i = \frac{q_{i+1} - q_{i-1}}{2}$

- soit $\frac{q'_i}{q_i}$ constant

à la 1^{ère} itération : - on suppose $\frac{q'_i}{q_i} = 0$

à chaque itération ultérieure : $\frac{q'_i}{q_i} = 0,5 * \text{Log} \left(\frac{q_{i+1}}{q_{i-1}} \right)$

• estimation du quotient par :

$${}_1Q_{i-0,5}^n = 1 - e^{-q} \approx \frac{q}{1 + q/2}$$

D) Tables de fécondité par rang

Jusqu'à une époque relativement récente cohabitation sans mariage, ruptures d'union par divorce et remariages étaient peu fréquents. Ainsi, les naissances survenaient presque toutes chez des couples mariés, le plus souvent pour la première fois. Dans ces conditions, la définition retenue pour définir le rang d'une naissance n'avait qu'une importance relativement faible sur le niveau des indicateurs de fécondité par rang. En effet, le rang défini dans le mariage actuel différait peu de celui défini parmi toutes les naissances de la mère, le rang biologique²⁹. Aujourd'hui le développement des nouvelles formes d'union, avec pour corollaire l'élévation considérable des naissances hors mariage, ne permet plus d'analyser avec autant de pertinence la fécondité selon le rang de la naissance si l'on dispose du seul classement des naissances selon le rang dans le mariage en cours. C'est pourquoi certains pays, comme la France, change de définition pour retenir le classement selon le rang biologique, même si ce changement n'est pas aisé ; ou utilisent des données d'enquêtes pour produire des indices prenant en compte l'ensemble des naissances précédentes, comme la France avec l'enquête famille ; ou estiment, plus ou moins périodiquement, à partir de clefs de conversion fondées sur la comparaison des naissances, enregistrées au cours d'une enquête, classées selon les deux définitions, le rang biologique à partir du rang dans le mariage en cours, comme en Allemagne.

Il convient donc de bien distinguer les pays selon la définition du rang utilisée et de publier, dans des tableaux séparés, les indicateurs se rapportant à l'une ou l'autre des deux définitions. De même, pour les pays ayant changé de définition au cours de la période récente, on veillera à ne pas mélanger les informations fondées sur des définitions différentes, et l'on présentera les indicateurs issus de chacun de ces définitions dans des tableaux séparés. On privilégiera toujours, si l'on a le choix, le rang biologique.

Le traitement des données se fera de la même façon que le rang soit défini dans le mariage en cours ou parmi toutes les naissances de la mère. La seule distinction que l'on puisse faire, lorsque le rang est celui défini dans le mariage et ne concerne donc que les seules naissances légitimes, est de recommander l'ajout d'une colonne de rang 0 correspondant aux naissances hors mariage. De cette manière le total du tableau, ainsi complété, sera égal au nombre total de naissances vivantes et sa marge fournira la répartition des naissances par âge sans distinction du rang de la naissance.

De même on évitera de procéder au calcul des indices synthétiques transversaux et, plus encore, longitudinaux, lorsque le rang est défini dans le mariage en cours car leur interprétation est des plus délicates.

La fécondité par rang, qu'elle soit défini *dans le mariage en cours* (distinction, parmi les naissances vivantes issues d'une même union célébrée par un mariage, du premier-né, du deuxième, ...) ou *parmi toutes les naissances de la mère* (distinction, parmi les naissances vivantes issues d'une même femme, du premier-né, du deuxième, ...) produit, par nature, des événements *non renouvelables*, mais peut être traitée comme produisant des événements *renouvelables*. C'est notamment le cas dans le cadre des indices calculés par Eurostat.

²⁹ Le rang *biologique* d'une naissance vivante est le rang de l'enfant parmi les enfants nés-vivants issus de la *même mère*. Certains pays européens n'observent pas le rang biologique dans leurs statistiques courantes d'état civil mais, pour les seules naissances *dans le mariage* (traditionnellement dénommées naissances *légitimes*), le rang dans le *mariage en cours* : rang de l'enfant parmi les enfants nés-vivants issus du *même couple*. Définie par le rang dans le mariage en cours, la fécondité par rang est traitée comme produisant des événements renouvelables. Par ailleurs, elle ne saurait conduire à une notion analogue à la probabilité d'agrandissement, du moins si la fréquence des naissances *hors* du *premier* mariage (couples non mariés ou remariés) n'est pas nulle.

Les données de bases disponibles dans la base de données d'Eurostat ne permettent, en effet, pas de rapporter les naissances de chacun des rangs aux seules femmes soumises au risque de cet événement que constitue la naissance d'un enfant de rang déterminé.

En préalable au calcul des taux, il conviendra assez souvent de répartir les naissances d'âge et/ou de rang non déclaré, en utilisant la procédure au *prorata* décrite au chapitre IB1, page 23. Il convient, lorsque les deux tableaux, par âge et par âge et rang, sont disponibles de redresser les non-déclarés à partir de l'information plus détaillée contenue dans le tableau par âge et rang et de prendre, pour répartition par âge redressée des non-déclarés, la marge du tableau par âge et rang redressé.

Les taux de fécondité pour chacun des rangs, biologiques ou dans le mariage en cours, seront calculés selon une procédure strictement identique à celle utilisée pour le calcul des taux de fécondité générale par âge. Seuls les événements pris en considération changent : l'ensemble des naissances d'un âge donné, qui figure au numérateur, étant remplacé par les seules naissances de cet âge donné d'un rang donné.

Les diverses procédures de calcul possibles restent les mêmes, mais il convient, naturellement de retenir la même procédure dans les deux cas : fécondité générale par âge et fécondité par âge et rang de naissance.

En toute rigueur, en procédant aux mêmes calculs que pour la fécondité générale on aboutit pas aux mêmes indices, même si la distinction est assez rarement faite par les experts. En effet, on n'obtient pas des taux, de seconde catégorie, de fécondité par rang, comme l'on l'entend souvent dire, mais des *indices-composantes*. Ces divers « taux » par rang, pour un âge donné, sont, en effet, le résultat de la décomposition du taux de fécondité par âge en ses diverses composantes par rang, la somme des indices composantes par rang, pour un âge donné, redonnant le taux de fécondité générale de cet âge. Toutefois, nous utiliserons, dans la suite, les termes de *composante par âge et rang* et *taux par âge et rang* de manière équivalente.

La somme des composantes par âge, de rang biologique n , pour une année donnée donnera la composante de rang n de l'indicateur conjoncturel, parfois dénommée indicateur conjoncturel de rang n . On parlera de composante de rang n de la descendance finale, ou de descendance finale de rang n lorsque la somme concerne une même génération. On obtiendra l'âge moyen, transversal ou longitudinal, à la maternité de rang n en appliquant la même procédure de calcul que pour l'âge moyen à la maternité, tous rangs confondus.

La fécondité par *rang biologique* de naissance peut également être traitée comme produisant des événements *non renouvelables*. Ainsi, définit-on, en considérant les générations parvenues au *terme* de leur vie féconde, la notion de *probabilité d'agrandissement*³⁰ : proportion, parmi les femmes qui ont atteint à un certain moment de leur vie une *parité*³¹ donnée, de celles qui ont atteint ultérieurement la *parité* immédiatement supérieure. Cet indicateur sera défini au chapitre 3.

³⁰ notion utile à l'analyse de l'évolution de la fécondité par rang.

³¹ Le terme *parité* désigne en démographie le nombre d'enfants nés vivants.

E) Tables de divortialité

Le divorce est généralement traité comme un événement *renouvelable* (ce qu'il n'est pas réellement) survenant chez des *couples mariés* et étudié différenciellement selon la *durée écoulée depuis la célébration du mariage* (durée dénommée encore *durée de mariage*). Faute de disposer d'une comptabilité, par promotion de mariage³², du nombre de couples résidents, qui évolue au fil des années sous l'effet de la mortalité, du divorce et des migrations internationales, on calcule des taux de divortialité en rapportant les nombres d'événements observés aux diverses durées à l'effectif *initial* de chaque promotion de mariage. Les données dont on dispose généralement correspondent à des nombres de divorces selon l'année du divorce et la durée écoulée depuis le mariage, *exprimée en différence de millésime*. Les nombres de divorces observés chaque année sont donc classés par *promotion* de mariage, c'est-à-dire en durée de mariage *atteinte* durant l'année de calendrier du divorce (*parallélogrammes* à côtés verticaux du diagramme de Lexis, à l'exception des divorces prononcés l'année même du mariage qui correspondent à un *triangle*).

Les taux de divortialité, contrairement aux autres taux présentés plus haut (taux de fécondité et taux de primo-nuptialité), sont traités comme des taux frustes qui ne font pas intervenir l'alimentation plus ou moins irrégulière des promotions de mariage et ne prennent donc pas en considération l'évolution mensuelle du nombre de mariages. Ce choix avait été fait pour deux raisons :

- La première est que le nombre de mariages subsistants est le plus souvent inconnu, sauf, peut-être, dans quelques pays à registre de population ou lorsque l'on a, au préalable, procédé à une estimation de ce nombre en faisant intervenir la mortalité de chacun des sexes et les migrations,³³
- La seconde c'est que l'ensemble des pays fournissent, dans le cadre de la collecte commune, les divorces classés selon la durée atteinte du mariage dans l'année du divorce, configuration dans laquelle les biais de mesure sont toujours limités.

Mais, il s'agit, sans doute, d'une fiction. Il est, en effet, difficile de croire que tous les pays enregistrent l'année du mariage, alors même qu'ils n'enregistrent pas l'année de naissance de la mère et que le tableau à remplir, fourni par la collecte commune, n'offre pas, à l'office statistique, le choix entre les deux définitions possibles de la durée et impose de considérer les divorces comme étant en durée atteinte. C'est pourquoi, par précaution et pour se prémunir contre les erreurs d'interprétation de la définition de la durée utilisée, il convient de :

- Demander à chacun des pays la définition précise de la durée utilisée dans le classement des divorces,
- Prévoir la possibilité de traiter, de la manière la plus adéquate, les divorces selon chacune des définitions possibles, c'est-à-dire en prenant en considération les variations de l'alimentation des promotions de mariages.

Dans ce cas on traite le divorce comme les autres phénomènes démographiques et l'on calcule des taux *élaborés* selon une méthodologie transposée du calcul des taux de fécondité. Pour ce faire, il conviendra, au préalable, de calculer les caractéristiques, l'année du mariage, des promotions de mariages annuelles, à partir de la répartition mensuelle du nombre total de mariages, comme on détermine les caractéristiques, à la naissance, des générations annuelles à

³² Seule la répartition des personnes résidentes par sexe et année d'âge fait l'objet d'une comptabilité régulière par les offices nationaux européens de statistique.

³³ Si l'on fait intervenir les ruptures d'union par divorce dans le calcul des mariages subsistants, on se retrouve alors dans la situation de calcul de taux de première catégorie (ou de quotients) et non plus dans le cas de taux de seconde catégorie (ou événements réduits) additifs.

partir de la répartition mensuelle des naissances. Ensuite on calcule les taux élaborés de divortialité, comme on le fait pour la fécondité ou la primo-nuptialité, en substituant toutefois au dénominateur le nombre initial de mariages à la population moyenne.

Les taux *frustes* peuvent être en durée de mariage atteinte ou en durée révolue, selon le cas, mais ils utilisent tous deux le nombre *initial* de mariages comme dénominateur. Alors que le taux à la durée atteinte i pour l'année de divorce nn est le rapport entre le nombre de mariages célébrés l'année $nn-i$ et rompus l'année nn , au nombre total de mariages célébrés l'année $nn-i$; celui en durée révolue est le rapport entre le nombre de mariages rompus l'année nn à la durée i et la moyenne des mariages célébrés les années $nn-i$ et $nn-i-1$.

Comme pour chacun des phénomènes démographiques, il convient de fixer un intervalle d'âges fixes à la survenance des événements, et donc de définir un âge (ou une durée) minimal et un âge maximal, faute de quoi, la recombinaison longitudinale ultérieure des taux risque d'être fautive. En effet, si la borne supérieure³⁴, par exemple, varie au fil du temps, on risque, dans la mesure où généralement il s'agit d'un intervalle ouvert, de surestimer l'intensité du phénomène des cohortes pour lesquelles on aurait été amené à additionner des taux correspondants à plusieurs intervalles ouverts consécutifs.

L'intensité de la divortialité et l'ancienneté moyenne du mariage au moment du divorce, qu'elles soient transversales ou dans les promotions de mariage, se calculent par les mêmes procédures que dans le cas des autres événements renouvelables.

Lorsque toute l'information nécessaire au calcul des taux de divortialité par durée de mariage annuelle n'est pas disponible, on va voir ci-après qu'il est cependant souvent possible d'estimer l'indicateur conjoncturel de divortialité. La méthode d'estimation diffère selon le type d'information manquantes.

a) L'effectif initial des promotions de mariages correspondant aux durées de mariage les plus élevées n'est pas disponible

Il convient dans ce cas d'admettre, faute de mieux, l'hypothèse selon laquelle le nombre annuel de mariages est, pour ces années-là, identique à celui de l'année la plus ancienne dont l'effectif initial de mariages est connu.

b) Le classement des divorces par durée de mariage porte sur des groupes pluriannuels de durée

La procédure la plus simple consiste à calculer, pour chaque groupe pluriannuel, le nombre moyen annuel de divorces en divisant le nombre de divorces du groupe par son étendue (exprimée en années), puis de traiter la série annuelle obtenue en remplaçant les données de chaque durée par le nombre moyen annuel approprié.

Cette situation se rencontre assez fréquemment : la durée *terminale* correspond alors à un groupe *ouvert* de durées (par exemple 30 ans ou plus). Dans ce cas la procédure à suivre dépend de la borne inférieure de durée à laquelle se rapporte ce groupe terminal :

- Si cette borne inférieure est suffisamment élevée (35 ans par exemple) pour qu'on puisse admettre que le nombre de divorces est très faible au-delà de cette durée, on fera comme si tous les divorces du groupe ouvert avaient pour durée commune 35 ans et on rapportera les divorces à la durée 35 ans *ou plus* au seul effectif initial des mariages de l'année $n-35$.
- Si cette borne inférieure est relativement élevée, mais pas suffisamment pour que l'on puisse admettre l'approximation précédente (groupe terminal : 25 ans ou plus, par exemple), on pourra rapporter les divorces à la durée 25 ans ou plus au nombre initial de

³⁴ Si dans le cas précis du divorce, c'est la borne supérieure qu'il convient de limiter, puisque la première durée est toujours la durée 0, en matière de fécondité et de primo-nuptialité un âge minimal doit également être fixé.

mariages des promotions que l'on considère comme les plus affectées par le divorce à ces durées.

On peut également rapporter les divorces à 25 ans ou plus à la somme des nombres de mariages célébrés de $n-25$ à $n-30$, si l'on considère qu'au-delà de 30 ans de mariage il n'y a plus de divorces. Les taux ainsi calculés ayant une dimension annuelle, il ne faut pas oublier de les multiplier par le nombre de promotions prises en considération, lorsqu'on établit l'indicateur conjoncturel de divortialité.

On peut encore diviser le nombre de divorces du groupe ouvert par son amplitude *efficace* supposée, puis rapporter chacun des nombres obtenus à l'effectif initial de chacune des promotions participant à cette efficacité.

On pressent bien que plus la borne inférieure du groupe terminal ouvert de durée est basse³⁵, plus il serait souhaitable de pondérer chacune des promotions par sa part dans le risque de survenance du divorce plutôt que de considérer, comme dans les solutions proposées ci-dessus, que les risques sont plus ou moins également répartis.

- Si la durée est relativement basse, on procède en trois temps :
 1. on calcule de manière habituelle les taux de divortialité par durée jusqu'à la dernière durée d'amplitude annuelle,
 2. on rapporte les divorces du groupe ouvert à une moyenne pondérée des nombres de mariages donnant lieu à des divorces aux durées correspondant au groupe ouvert. Les coefficients de pondération sont empruntés à un calendrier de divorce estimé ou observé par ailleurs (autre année ou autre pays).
 3. on obtient l'indicateur conjoncturel en sommant les taux annuels et en ajoutant le taux pour le groupe ouvert.

c) On ne dispose que du nombre annuel de divorces, mais non de la répartition par durée de mariage.

Il est possible d'estimer l'indicateur conjoncturel de divortialité lorsqu'on ne dispose pas de la répartition du nombre des divorces selon la durée de mariage. En effet, la *méthode de la moyenne pondérée*, dénommée également *méthode du calendrier-type*, fournit une telle estimation si on peut se référer à un calendrier plausible de la survenance du divorce au fil des durées de mariage.

Ce calendrier, dont la somme des éléments vaut 1, peut correspondre à une observation faite, à une autre date dans le pays considéré ou dans un autre pays où les conditions de mise en union et de rupture d'union sont proches.

L'effet du choix du calendrier retenu sur le niveau de l'indicateur estimé est faible si le nombre de mariages n'est pas sujet à de brusques variations d'une année à l'autre.

Exprimons que le rapport entre le nombre de divorces D_x^n observés l'année n à la durée x et le nombre initial de mariages $M_g = M_{n-x}$ dans la promotion correspondante³⁶ est égal au produit de l'élément de calendrier effectif $a'_{g,x}$ de la promotion g par l'intensité effective p'_g de la divortialité dans cette promotion avec :

$$a'_{g,x} = \frac{D_x^{g+x}}{\sum_{x=a}^{x=W} D_x^{g+x}} \quad \text{et} \quad p'_g = \frac{\sum_{x=a}^{x=W} D_x^{g+x}}{M^g}$$

³⁵ Le recours à une moyenne pondérée, quoique de mise en œuvre plus longue, est toujours préférable à la simple répartition linéaire. Dans le cas de groupes quinquennaux de durée il est, en effet, toujours possible de répartir le nombre d'événements du groupe entre les promotions concernées en se référant à la part, observée une autre année ou dans un autre pays, de chacun des taux par durée annuelle dans la somme des taux par durée de ce groupe.

³⁶ Nous supposons, pour alléger les écritures, que les données sont observées en parallélogrammes à côtés verticaux : on a alors $g = n - x$.

Le nombre de divorces D^n d'une année donnée peut donc s'écrire comme la somme des produits des éléments de calendrier *effectifs*³⁷ $a'_{g,x}$ par les intensités *effectives* p'_g de la divortialité dans la promotion de mariage g , et par l'effectif des mariages initiaux M_g dans ces promotions :

$$D^n = \sum_{x=a}^{x=W} a'_{g,x} p'_g M_g$$

L'indicateur de divortialité recherché est :

$$ICD = \sum_{x=a}^{x=W} a_{g,x} p_g$$

que l'on peut assimiler à :

$$\sum_{x=a}^{x=W} a'_{g,x} p'_g \cdot$$

Sous l'hypothèse de stationnarité de la divortialité dans le temps, les deux formules précédentes peuvent s'écrire :

$$D^n = p'_g \sum_{x=a}^{x=W} a'_{g,x} M_g$$

et :

$$S = p'$$

Il en découle :

$$S = \frac{D^n}{\sum_{x=a}^{x=W} a'_{g,x} M_g}$$

Cette formule donne des résultats tout à fait acceptables lorsque, d'une part, les variations du calendrier et de l'intensité sont régulières et modérées et, d'autre part, les événements perturbateurs (migrations et veuvage) ne sont pas trop marqués. En effet cette formule, outre qu'elle consiste en la substitution au calendrier et à l'intensité *effectifs* d'un calendrier et d'une intensité *corrigés* des effets des phénomènes perturbateurs, suppose que la divortialité est stationnaire dans le temps.

Par ailleurs, dans cette méthode comme dans celle de la somme des taux de divortialité par durée de mariage, la substitution de l'effectif *initial* des promotions de mariage à l'effectif des mariages subsistants, a pour effet de *sous-estimer* l'indicateur de divortialité si le veuvage et l'émigration font sortir d'observation plus de mariages que l'immigration n'en fait entrer, et inversement dans le cas contraire.

Séquence des opérations

1 – Taux frustes (durée révolue)

- calcul du taux par rapport à l'effectif initial de mariages :

$$f_i^n = \frac{E_i^n}{(M^{n-i-1} + M^{n-i})/2}$$

2 – Taux frustes (durée atteinte)

- calcul du taux par rapport à l'effectif initial de mariages :

$$f_i^n = \frac{E_i^n}{M^{n-i}}$$

³⁷C'est-à-dire compte tenu de la manifestation des événements perturbateurs.

3- Taux élaborés (durée révolue)

a) constance du risque

- calcul des caractéristiques des promotions de mariage : m_1 et m_2 ,
- calcul du taux :

$$\hat{f}_i^n = \frac{E_i^n}{M^{n-i-1}m_1 + M^{n-i}(1-m_2)}$$

b) risque variable

- calcul des caractéristiques des promotions de mariage : m , V et m , en leur absence retenir des valeurs égales respectivement à $\frac{1}{2}$, $\frac{1}{12}$ et 0.

- estimation du taux, $f_i^n = f_i + \frac{f_i''}{24}$, réalisée par *itérations* successives

- à la 1^{ère} itération : - on suppose f_i' et f_i'' (dérivées première et seconde) *nuls* pour tout i

- on déduit, au moyen de la relation (1), f_i , puis $f_i^n = f_i$, de E_i^n , M^{n-i-1} et M^{n-i}

- à chaque itération ultérieure : - on estime f' et f'' par ajustement parabolique des valeurs de f obtenues à l'itération *précédente*

- on déduit, au moyen à nouveau de la relation (1), f_i , puis $f_i^n = f_i + \frac{f_i''}{24}$, de E_i^n , M^{n-i-1} et M^{n-i} .

4- Taux élaborés (durée atteinte)

- calcul des caractéristiques des promotions de mariage : m et V (en leur absence retenir $\frac{1}{2}$ et $\frac{1}{12}$).

- estimation du taux, $f_i^n = f_i + \frac{f_i''}{24}$, réalisée par *itérations* successives

- à la 1^{ère} itération : - on suppose f_i' et f_i'' (dérivées première et seconde) *nuls* pour tout i

- on déduit, au moyen de la relation (2), f_i , puis $f_i^n = f_i$, de E_i^n , M^{n-i-1} et M^{n-i} .

- à chaque itération ultérieure : - on estime f' et f'' par ajustement parabolique des valeurs de f obtenues à l'itération *précédente*

- on déduit, toujours au moyen de la relation (2), f_i , puis $f_i^n = f_i + \frac{f_i''}{24}$, de E_i^n , M^{n-i-1} et M^{n-i} .

Chapitre 3

Les indicateurs déduits des tables de survenance

Une table démographique est une collection des *taux* ou de *quotients*, selon l'âge (ou la durée écoulée depuis une date prise pour origine), se rapportant soit à la même *année d'observation* (table *transversale*), soit à la même *année de naissance* (table *longitudinale*). De cette table, on déduit divers indicateurs synthétiques destinés à la *résumer*.

1- Indicateurs dérivés déduits des tables

Les divers indicateurs destinés à résumer les informations contenues dans une table sont de deux types : indicateur d'intensité ou indicateur de tendance centrale qui synthétise le calendrier du phénomène. Ainsi distingue-t-on :

- Indicateur d'*intensité* d'une table de *taux* : c'est la *somme* de tous les taux de la table. L'indicateur ainsi obtenu porte le nom général d'*indicateur conjoncturel* du phénomène considéré si la table est *transversale*. Si la table est *longitudinale*, il porte un nom différent selon le phénomène : *descendance finale* dans le cas de la *fécondité*, *proportion de non célibataires à 50 ans* dans le cas de la *primo-nuptialité*, *proportion de mariages rompus par divorce* dans le cas de la *divortialité*.

Comme nous l'avons vu, au chapitre précédent à propos de la table de divortialité, en l'absence des éléments permettant de calculer une table transversale et donc d'en déduire l'indicateur synthétique d'intensité, il est, très souvent, possible d'en produire une estimation. Ces procédures d'estimation sont toutes fondées sur la méthode de la moyenne pondérée mise en œuvre par Corrado Gini. C'est à cette méthode que nous avons eu recours pour proposer une estimation de l'indicateur conjoncturel de divortialité en l'absence de la répartition des divorces selon l'ancienneté du mariage. C'est sur ce même principe que Gérard Calot a fondé son concept de *génération moyenne* (sur lequel nous reviendrons) qui permet d'estimer, plusieurs mois avant que ne soient disponibles les éléments nécessaires à son calcul, une estimation de l'indicateur conjoncturel à partir de la seule disponibilité du nombre absolu d'événements³⁸.

- Indicateur d'*intensité* d'une table de *quotients* : c'est la proportion des unités statistiques qui subissent l'événement considéré entre les âges extrêmes a et w . L'indicateur ainsi obtenu porte le *même* nom si la table est transversale ou longitudinale, mais alors on précise dans quel cas on se trouve. Ainsi la *proportion de non célibataires à 50 ans* fondée sur les *quotients* transversaux ou longitudinaux dans le cas de la primo-nuptialité. En matière de mortalité (événement fatal), l'indicateur d'intensité tant transversal que longitudinal vaut 1 et est sans objet³⁹.

³⁸ Cf. G. Calot, « Une notion intéressante : l'effectif moyen des générations soumises au risque. I. Présentation méthodologique », *Population*, 6, 1984, 947-976.

³⁹ Il faut toutefois signaler que, comme on le fait couramment pour l'analyse de la primo-nuptialité, la mortalité, *transversale*, pourrait également être traitée comme un événement renouvelable. Dans ce cas on calcule des taux de seconde catégorie, assimilables aux événements d'une table, qui débouchent sur le calcul, par sommation de ces taux, d'un indicateur conjoncturel de mortalité (dont la valeur s'écarte de 1) et d'un âge moyen au décès. Pour plus de détails voir « Un indicateur conjoncturel de mortalité : l'exemple de la France », *Population*, n° 2, 1993, pp. 347-368.

- Indicateur de *tendance centrale* d'une table de *taux* : c'est, lorsque l'événement étudié se produit, l'âge moyen des personnes concernées lors de sa survenance. L'indicateur ainsi obtenu porte le nom d'*âge moyen*, que la table soit *transversale* ou *longitudinale* : *âge moyen à la maternité* dans le cas de la *fécondité*, *âge moyen au premier mariage* fondé sur les taux (célibataires se mariant avant l'âge de 50 ans) dans le cas de la *primo-nuptialité*, *ancienneté moyenne des mariages rompus par divorce* dans le cas de la *divortialité*. Les âges moyens qui constituent les indicateurs de tendance centrale des événements renouvelables (fécondité) ou traités comme tels (primo-nuptialité) doivent donc être fondés sur les taux et non sur les événements.

La formule de calcul de cet âge moyen varie selon la figure du schéma de Lexis dans laquelle sont calculés les taux utilisés. Ainsi lorsque les taux sont par âge révolu ou entre anniversaire (carré de Lexis) l'âge moyen se calcule ainsi :

$$\bar{x} = 0,5 + \frac{\sum_{x=a}^{x=W} x f_x^n}{\sum_{x=a}^{x=W} f_x^n}$$

Si les taux sont définis selon l'âge atteint dans l'année ou en différence de millésime (parallélogramme à base verticale du schéma de Lexis) la formule devient :

$$\bar{x} = \frac{\sum_{x=a}^{x=W} x f_x^n}{\sum_{x=a}^{x=W} f_x^n}$$

n représentera l'année de calendrier lorsque c'est à l'âge moyen transversal que l'on s'intéresse, et la cohorte (année de naissance ou de mariage) dans le cadre de l'analyse longitudinale.

- Indicateur de *tendance centrale* d'une table de *quotients* : c'est encore, lorsque l'événement étudié se produit, l'âge moyen des personnes concernées lors de sa survenance. L'indicateur ainsi obtenu porte le nom d'*âge moyen au décès* ou *espérance de vie à la naissance* dans le cas de la *mortalité*, d'*âge moyen au premier mariage* fondé sur les *quotients* (célibataires se mariant avant l'âge de 50 ans) dans le cas de la *primo-nuptialité*.

La formule du calcul de l'âge moyen au décès, ou espérance de vie à la naissance, est très voisine de celle de l'âge moyen dans le cas d'un événement renouvelable repéré selon l'âge en années révolues, si ce n'est que l'on utilise les événements de la table $d_{x,x+1}$ ⁴⁰ :

$$e_0 = 0,5 + \frac{\sum_{x=a}^{x=W-1} x * d_{x,x+1}}{\sum_{x=a}^{x=W-1} d_{x,x+1}}$$

qui se simplifie en :

⁴⁰ Auxquels sont assimilables les taux de seconde catégorie.

$$e_0 = 0,5 + \frac{\sum_{x=1}^{x=W} S_x}{S_0}$$

où S_x est le nombre de survivants, à l'âge x , de la table de mortalité.

A ces indicateurs classiques on peut ajouter :

- L'*infécondité définitive*⁴¹ : c'est le complémentaire à 1 de l'indicateur d'intensité de la table des taux de fécondité de rang biologique 1, dans laquelle la naissance de rang 1, événement non renouvelable, est traitée comme un événement renouvelable par le calcul de taux de seconde catégorie qui, en toute rigueur, sont des indices-composantes et fournissent la composante de rang 1 des divers taux de fécondité générale par âge. Il est calculé, en transversal, à partir de l'indicateur conjoncturel de rang 1 et à partir de la descendance finale de rang en longitudinal. Ce complémentaire à l'indicateur d'intensité est également fréquemment utilisé dans l'analyse de la primo-nuptialité, où est connu sous la dénomination de *fréquence du célibat définitif*.
- La *proportion de femmes avec x enfants* : c'est la proportion de femme ayant mis au monde x enfants parmi l'ensemble des femmes. Elle peut se calculer en transversal comme dans les cohortes. La proportion de femmes *sans* enfant, mesure de l'infécondité, est, en transversal, le complément à 1 de l'indicateur conjoncturel de rang 1 ; dans les générations c'est le complément à 1 de la descendance finale. Pour les autres dimensions de la famille, c'est la différence entre l'indicateur d'intensité de la fécondité de rang biologique n et celui de rang $n+1$. Pour la dernière dimension retenue, la proportion de femmes ayant eu *4 enfant ou plus* par exemple, est égale à l'indicateur d'intensité de la fécondité de rang retenu, indicateur conjoncturel ou descendance finale de rang *4 ou +*, selon le cas.
- Les probabilités d'agrandissement a_n sont des indicateurs dérivés particuliers d'une collection de tables de fécondité par rang biologique. Ainsi la probabilité d'agrandissement du rang 1 au rang 2 (notée encore a_1) est la proportion, parmi les femmes ayant eu *au moins un enfant*, de celles en ayant eu *au moins deux*. De même, la probabilité d'agrandissement du rang 0 au rang 1 (notée encore a_0) est la proportion, parmi l'ensemble des femmes, de celles ayant eu *au moins un enfant* (son complément à l'unité est la *proportion des femmes sans enfant*, appelée encore *fréquence de l'infécondité définitive*). Les probabilités d'agrandissement se définissent aussi en termes *transversaux* : les proportions de femmes ayant atteint les diverses parités sont obtenues par rapport des cumuls tous âges des taux de fécondité transversaux par âge, comme le sont les proportions analogues longitudinales. Eurostat envisage d'établir *successivement*, dans son système, les probabilités d'agrandissement longitudinales et transversales.

Les probabilités d'agrandissement longitudinales peuvent être établies sur des parcours encore *incomplets* de la vie féconde. Ainsi, à l'âge i , on calculera la proportion⁴², parmi les femmes d'âge i qui ont déjà eu *au moins r* enfants, de

⁴¹ Cet indicateur, pour être pertinent, doit être calculé à partir du seul classement des naissances vivantes selon le rang *biologique* de la naissance, c'est-à-dire parmi l'ensemble des naissances vivantes de la mère.

⁴² La proportion, parmi les femmes de 50 ans qui ont eu *au moins r* enfants, de celles en ayant eu *au moins r+1* est précisément la probabilité d'agrandissement longitudinale $a_{r \rightarrow r+1}$.

celles en ayant déjà eu *au moins* $r+1$. Chacune est obtenue par rapport de l'indicateur d'intensité de rang $n+1$ à l'indicateur d'intensité de rang n .
La combinaison de ces probabilités, selon la formule :

$$a_0 + a_0 a_1 + a_0 a_1 a_2 + \dots$$

redonne l'indicateur d'intensité de la fécondité générale, indicateur conjoncturel ou descendance finale selon le cas.

2- Précautions à prendre pour le calcul de ces indicateurs

Il est indiqué au paragraphe précédent que l'indicateur d'intensité d'une table de taux est la somme de tous les taux de la table. Ainsi l'indicateur conjoncturel de fécondité, par exemple est la somme de tous les taux de fécondité par âge disponibles pour une année donnée.

Mais plusieurs difficultés peuvent surgir lorsque l'intervalle des âges disponibles varie au fil du temps :

- Tout d'abord, l'indicateur *transversal* de valeur centrale, l'âge moyen à la maternité par exemple, qui est, par construction, assez sensible aux valeurs extrêmes pourrait se ressentir de ces variations des âges limites des tables ;
- D'autre part, et c'est sans doute l'élément le plus important, la transposition longitudinale des taux pourrait aboutir à des mesures de l'intensité longitudinale dont la procédure de calcul varierait d'une cohorte à l'autre.

Cette variabilité des limites n'aurait aucune incidence sur la qualité des mesures si elle ne tenait qu'aux variations aléatoires de l'expression du phénomène étudiée, mais le plus souvent les offices statistiques opèrent des regroupements aux deux extrémités des âges. Si ces intervalles ouverts ne sont pas toujours identiques, lors de la recombinaison longitudinale on risque d'affecter à une même cohorte plusieurs observations se référant à un même âge et de surestimer ainsi son intensité.

Ainsi, si une année donnée n le dernier groupe ouvert est 49 ans ou plus et que l'année suivante $n+1$ il passe à 50 ans ou plus, on attribuera à la génération née en $n-49$ les deux groupes ouverts, le groupe 49+ de l'année n et le groupe 50+ de l'année $n+1$. De ce fait les âges 50 et suivants seront comptés deux fois.

C'est pourquoi il convient que dans le programme de calcul des tables un intervalle d'âge invariant soit fixé pour chacun des phénomènes. Il pourrait être de 15-49 ans pour toutes les manifestations de la fécondité ainsi que pour la nuptialité, de 0-30 ans pour la divortialité et de 0-99 ans pour la mortalité.

3- Autres indicateurs dérivés

Les données statistiques observées la même année ou sur la même génération peuvent être combinées de multiples manières pour conduire à des indicateurs dérivés.

a) Les taux bruts

Un *taux brut* est un rapport entre un *nombre absolu annuel d'événements* et l'effectif moyen de la population résidente⁴³ : taux brut de *natalité* (nombre absolu annuel de *naissances vivantes*), de *mortalité* (nombre absolu annuel de *décès*), de *nuptialité* (nombre absolu annuel de *mariages*), de *divortialité* (nombre absolu annuel de *divorces*) ; le *solde migratoire relatif* est aussi un taux brut.

⁴³ De façon plus précise, mais équivalente si les soldes naturel et migratoire n'ont pas de mouvement saisonnier, le dénominateur est le nombre de personnes-années d'exposition au risque.

L'effectif moyen de la population résidente qu'il convient de retenir est la *demi-somme* des effectifs de la population en début et en fin d'année. Cependant certains pays européens, qui disposent d'un registre de population (dont ils tirent des données *trimestrielles*, voire *mensuelles*), prennent pour population moyenne de l'année la moyenne arithmétique des effectifs trimestriels ou mensuels. Nous pensons qu'il serait éminemment souhaitable de les convaincre de renoncer à ce raffinement, vu le caractère très grossier d'un taux brut comme indicateur démographique et les très faibles différences séparant généralement les taux bruts obtenus.

b) L'effectif moyen des générations soumises au risque

Dans le cas d'un phénomène *renouvelable*, le nombre absolu d'événements observés une année donnée est le produit de l'*indicateur conjoncturel* par l'*effectif moyen des générations soumises au risque*. Ainsi, en matière de fécondité, cet effectif moyen est la moyenne pondérée des effectifs des diverses générations féminines qui, cette année-là, sont d'âge fécond (15-49 ans, pour fixer les idées), le coefficient de pondération de l'effectif d'un âge donné étant le taux de fécondité à cet âge⁴⁴.

Cet indicateur est une mesure de la *taille* du pays eu égard au phénomène considéré. Il possède cette propriété d'être à variations *lentes*, à la façon d'une moyenne mobile, ce qui facilite son interpolation et surtout son extrapolation. Grâce à son extrapolation, on pourra estimer l'indicateur conjoncturel dès qu'on disposera d'une estimation du nombre absolu d'événements, sans attendre de disposer des effectifs par âge et donc de la table correspondante.

Il est souhaitable de disposer pour chaque pays et chaque événement renouvelable (fécondité) ou traité comme tel (primo-nuptialité, divorce) de cet indicateur dérivé, extrapolé à horizon de quelques années. Il appartiendra ensuite à Eurostat de déterminer si cet indicateur doit être publié.

c) Recombinaison longitudinale des taux et quotients : les indicateurs longitudinaux

Les collections annuelles de taux et de quotients conduisent, après reclassement selon l'année de naissance (ou le couple d'années de naissance dans les cas des taux et quotients établis à partir de données en carrés dans le diagramme de Lexis) aux collections longitudinales correspondantes. Ces collections longitudinales se résument à l'aide d'indicateurs dérivés analogues aux indicateurs dérivés des collections transversales.

Une question qui se pose dans le cas longitudinal et qui n'a pas d'équivalent transversal est la suivante : faute, pour une génération presque parvenue au *terme* de sa période d'exposition au risque, de disposer encore des *derniers* taux ou quotients, on ne peut résumer le comportement de cette génération au moyen des indicateurs dérivés habituels. Or les derniers taux ou quotients sont souvent *petits* et leur estimation permettrait de compléter la collection et d'estimer sans grand risque d'erreur l'indicateur dérivé correspondant.

Plusieurs méthodes peuvent être envisagées pour procéder à cette estimation.

La première est celle du *gel* : on estime le taux ou quotient manquant, quelle que soit la génération considérée, par le taux ou quotient le plus *récent* observé au *même âge* (ou à la même durée). Il conviendra naturellement de *limiter* l'emploi de cette méthode : par exemple, si on estime la descendance finale, on se limitera aux générations dont la somme des taux manquants estimés ne dépasse pas 15% de la descendance finale estimée. Il sera d'autre part souhaitable d'assortir l'estimation de la descendance finale d'une *lettre*

⁴⁴ Pour plus de détail, voir en annexe, *L'analyse démographique conjoncturelle*.

indicatrice (par exemple * si moins de 1% a été estimé, # si entre 1% et 5% a été estimé, etc.). L'usage d'une lettre indicatrice est également souhaitable lorsqu'on utilise l'une ou l'autre des méthodes ci-dessous.

La deuxième méthode est celle de l'*extrapolation* des taux et quotients manquants. Cette extrapolation peut être opérée sur une base longitudinale ou sur une base transversale.

- Par exemple sur une base longitudinale, on connaît les taux de fécondité d'une génération donnée jusqu'à l'âge $i-1$: la partie manquante de la descendance finale est la somme des taux de l'âge i à l'âge w . On considère les générations *antérieures* à la génération considérée dont on connaît *toute* l'histoire et on extrapole la série des descendance partielles à i ans ou plus par un procédé donné (par exemple : ajustement *linéaire* sur les *dix* générations les plus récentes dont on connaît toute l'histoire).
- Par exemple sur une base transversale, on estime le taux de fécondité non encore observé d'une génération donnée à un âge donné par extrapolation des taux observés à ce même âge au fil des années. On additionne les estimations ainsi obtenues pour évaluer la somme des taux jusqu'à l'âge w des générations non encore parvenues au terme de leur vie féconde.

Une troisième méthode, applicable aux seuls indicateurs longitudinaux d'*intensité* fondés sur des *taux* (événements renouvelables ou traités comme tels), est la suivante. La somme longitudinale $\Sigma(n,i)$ des taux manquants de la génération g depuis l'âge i jusqu'à l'âge w est approximativement égale à la somme *transversale* $S(n,i)$ des taux entre ces mêmes âges pour l'année d'observation n lorsque $n = g + AMT(n, i)$ où $AMT(n, i)$ est l'âge moyen transversal à la survenance pour les événements se produisant l'année n à i ans ou plus⁴⁵. On peut montrer que, dans l'ensemble, cette troisième méthode est notablement plus précise que les deux premières.

d) La construction d'indicateurs dérivés mensuels

La connaissance des nombres absolus d'événements à l'échelle *mensuelle*, tous âges réunis, et pas seulement *annuelle*, permet d'établir des indicateurs conjoncturels mensuels en matière de fécondité, de primo-nuptialité par sexe et de mortalité par sexe. Ces indicateurs mensuels décrivent plus *finement* l'évolution chronologique des phénomènes étudiés que les indicateurs annuels, et permettent de prouver le cas échéant la simultanéité (donc la liaison causale vraisemblable) entre une anomalie dans l'évolution de l'indicateur et la survenance de tel ou tel événement (d'ordre politique, social ou autre). La méthodologie de la construction des indicateurs mensuels par l'ODE a été présentée dans l'article de Gérard Calot, reproduit en l'annexe, *L'analyse démographique conjoncturelle*, contribution au volume *The joy of demography* publié en l'honneur de Dirk J. van de Kaa, par Anton Kuijsten, Henk de Gans et Henk de Feijter, NethurD Publications, La Haye, 1999.

e) L'extrapolation à court terme des indicateurs transversaux annuels

Lorsque, pour un pays donné, les données se rapportant à une année (récente) donnée ne sont pas encore disponibles, on ne peut produire d'indicateur pour ce pays cette année-là, ni estimer quoi que ce soit au niveau de n'importe quelle agrégation géographique contenant ce pays-là. C'est pourquoi, surtout si la population du pays en question est d'effectif relativement faible par rapport à celui de la zone agrégée, il est souhaitable d'estimer, même grossièrement, les données manquantes.

⁴⁵ Appliquée à $i = \bullet$, cette relation ne fournit que des résultats très approximatifs : elle consiste, par exemple en matière de fécondité, à assimiler l'indicateur conjoncturel à la descendance finale, moyennant le décalage de l'âge moyen transversal à la maternité. Cette relation devient de plus en plus précise à mesure que i se rapproche de w .

Les effectifs de population seront estimés en supposant l'invariance, à âge égal, des rapports apparents de survie ; les nombres absolus d'événements par âge seront estimés en supposant l'invariance, à âge égal, des taux, ces deux invariances étant appréciées par rapport à l'année la plus récente pour laquelle l'information est disponible.

Si on dispose en outre d'une estimation du nombre annuel d'événements, on corrigera proportionnellement les nombres annuels d'événements par âge pour amener leur total à coïncider avec le nombre annuel estimé.

Avec ces données ainsi grossièrement estimées, on procédera comme s'il s'agissait de données observées.

<i>Formulation des indicateurs dérivés</i>			
	Type d'âge	Evénement renouvelable	Evénement non renouvelable
$\hat{L}\left(a+\frac{1}{2}\right)$	Atteint	///	$\hat{L}\left(a+\frac{1}{2}\right) = 1 - \prod_{i=a}^{i=a} (1 - p_i^{G+i})$
$\hat{L}(a)$	Révolu	///	$\hat{L}(a) = 1 - \prod_{i=a}^{i=a-1} (1 - p_i^{G+i})$
$\hat{D}(G)$	Atteint ou révolu	$\sum_{i=a}^{i=W} p_i^{G+i}$	$1 - \prod_{i=a}^{i=W} (1 - p_i^{G+i})$
$\bar{x}(G)$	Atteint	$\frac{\sum_{i=a}^{i=W} i p_i^{G+i}}{\sum_{i=a}^{i=W} p_i^{G+i}}$	$\frac{\sum_{a=a}^{a=W} a \left[\hat{L}\left(a-\frac{1}{2}\right) - \hat{L}\left(a+\frac{1}{2}\right) \right]}{1 - \hat{L}(W)}$
	Révolu	$\frac{\sum_{i=a}^{i=W} \left(i + \frac{1}{2}\right) p_i^{G+i}}{\sum_{i=a}^{i=W} p_i^{G+i}} = \frac{1}{2} + \frac{\sum_{i=a}^{i=W} i p_i^{G+i}}{\sum_{i=a}^{i=W} p_i^{G+i}}$	$\frac{\frac{1}{2} + \sum_{a=a}^{a=W-1} a \left[\hat{L}(a-1) - \hat{L}(a+1) \right]}{1 - \hat{L}(W)}$
$\hat{D}(a, b)$	Atteint ou révolu	$\sum_{i=a}^{i=b-1} p_i^{G+i}$	$1 - \prod_{i=a}^{i=b-1} (1 - p_i^{G+i})$

Avec $\hat{L}(a)$ = probabilité que l'événement ne se produise pas avant l'âge exact a ,

$\hat{D}(G)$ = intensité finale de survénance de l'événement,

$\bar{x}(G)$ = âge moyen à la survénance de l'événement,

$\hat{D}(a, b)$ = intensité de survénance de l'événement entre les âges exacts a et b .

Conclusion

La définition de la méthodologie à suivre en matière de calcul des indicateurs démographiques est naturellement un point crucial dans l'organisation d'un système international de calcul d'indicateurs démographique, car sans une méthodologie appropriée il n'est pas possible d'aboutir à une bonne comparabilité des données.

Certaines propositions pourraient apparaître superflues ou trop complexes pour un contexte peu fréquent, mais il faut garder à l'esprit qu'un système de cette nature ne se refait pas régulièrement et qu'il doit pouvoir faire face à toutes les situations. Naturellement la qualité a un coût, mais si l'on veut assurer la pérennité d'un tel système ce n'est qu'en étant exigeant sur la qualité.

Mais si une bonne méthodologie est une condition nécessaire à un système performant, ce n'est pas une condition suffisante. Il existe de nombreux autres points qui peuvent s'avérer aussi important dans la réussite d'une telle entreprise.

Ainsi convient-il :

- de mettre en œuvre des outils de contrôle de la qualité et de la cohérence interne des données incorporées dans la base, qui doivent aboutir à des purges régulières de la base ;
- de créer les conditions qui permettront de traiter de manière unique l'ensemble des informations disponibles, quelle que soit leur structure (ce que nous avons traité dans le premier chapitre) ;
- de concevoir l'articulation entre tous les modules ainsi définis afin d'arriver à un système cohérent et complet qui parcourt, sans interruption, toutes les étapes de l'incorporation des données de bases à la production des indicateurs dérivés ;
- de définir une structure suffisamment souple pour permettre l'incorporation facile de tous les modules supplémentaires dont la nécessité ne manquera pas de se faire sentir pour satisfaire la demande de besoins nouveaux.

ANNEXE 1

The construction of life tables

Reproduit de :Gérard Calot et Ana Franco

“The construction of life tables”

in Guillaume Wunsch, Michel Mouchart et Josianne Duchêne (éditeurs),

Life Tables : Data, Methods, Models, Kluwer, 2001,

pp. 31-75

THE CONSTRUCTION OF LIFE TABLES

by Gérard CALOT⁴⁶ and Ana FRANCO⁴⁷

The paper first recalls the probabilistic background of the estimation of mortality intensity by age and sex during a year of observation. It leads to the fundamental pair of relationships – hereafter numbered (24) and (32) – yielding the number of deaths inside each triangle of the Lexis diagram.

Three methods (the last one with two additional sub-variants) of practical computation of the probability of dying between two consecutive ages are considered : the first one (*A*) is based on the number of deaths inside each *triangle*, taking into account the distribution of *birthdays* within each of the two birth-cohorts concerned, the second one (*B*) is based on the number of deaths inside each *square* and the distribution of birthdays, the third one (*C*) simply consists in the computation of *rates* : ratio of the number of deaths in the square to the *mid-year* population. Method *C* takes the *observed rate* as an estimate of the *risk of mortality* q , while method C_2 takes it as an estimate of the *probability of dying* Q . Method C_1 is intermediate between *C* and C_2 ⁴⁸.

It is shown that, except for ages 0 and 1 and for higher ages (above 80), and except for cohorts born in a period of *abrupt* changes in the birth-rate, method *C*, and even methods C_1 and C_2 , yield satisfactory results. But for cohorts born during world wars, it is really worthwhile to use methods *A* or *B* and – at higher ages – to use method *A*. Moreover if methods *C* and C_1 yield quite comparable results, method C_2 becomes inaccurate above age 60.

I – The estimation of a constant risk

Let us consider a set of N individuals exposed to the occurrence of a *non repeatable* event, denoted E . Individual j , $j = 1, 2, \dots, N$, is exposed to E between times b_j and e_j , beginning and end of the *exposure* period, respectively.

Let us assume that the *intensity* of occurrence of E , denoted q , also called the *risk*⁴⁹ or the *instantaneous quotient* of occurrence of E , is *constant* over time and *identical* for all individuals.

Under these assumptions, the probability that event E occurs to individual j between times t and $t + dt$, given that it did not occur to that individual before time t , is, for any j and t :

$$q dt. \tag{1}$$

⁴⁶ Observatoire Démographique Européen, Saint-Germain-en-Laye, France

⁴⁷ Eurostat, Statistical Office of the European Communities, Luxembourg.

⁴⁸ Methods *C*, C_1 and C_2 consist in deriving the estimate \hat{Q} of the probability of dying, Q , from the observed *rate* r :

$$r = \frac{2D}{P_1 + P_2}$$

by :

$$\text{Method } C : \hat{Q} = 1 - e^{-r}$$

$$\text{Method } C_1 : \hat{Q} = \frac{r}{1 + r/2}$$

$$\text{Method } C_2 : \hat{Q} = r$$

⁴⁹ If event E is *death*, q is called the *force of mortality*.

Let $P(b, t)$ denote the probability that event E does not occur to a given individual exposed to E between times b and t . We may write :

$$P(b, t + dt) = P(b, t) (1 - q dt), \quad (2)$$

expressing that the probability $P(b, t + dt)$ of no occurrence between times b and $t + dt$ is :

- the probability $P(b, t)$ of no occurrence between times b and t multiplied by
- the probability $1 - q dt$ of no occurrence between times t and $t + dt$, given its non-occurrence between times b and t .

From (2), it follows that P satisfies :

$$\frac{\mathfrak{N}\{\text{Log}[P(b, t)]\}}{\mathfrak{N}t} = -q, \quad (3)$$

which implies, taking into account that $P(b, b) = 1$, that :

$$P(b, t) = e^{-q(t-b)}. \quad (4)$$

Therefore, the probability that E occurs to individual j between times t and $t + dt$ is :

$$e^{-q(t-b_j)} q dt. \quad (5)$$

The probability of occurrence of E during a *unitary* period – i.e. of length equal to 1 – is denoted Q . Using (4), the probability of no occurrence during a unitary period, $1 - Q$, is :

$$1 - Q = e^{-q},$$

so that Q is related to q by :

$$Q = 1 - e^{-q} \approx \frac{q}{1 + \frac{q}{2}} \text{ if } q \text{ is small.} \quad (6)$$

Let us now consider a set of N individuals, *independently* exposed to the occurrence of event E . The probability that E occurs to individuals j_1, j_2, \dots, j_n at times $t_{j_1}, t_{j_2}, \dots, t_{j_n}$ and *not* to the other $N - n$ individuals, is :

$$L = \prod_{j=j_1}^{j_n} \left[e^{-q(t_j - b_j)} q \right] \prod_{j \notin (j_1, \dots, j_n)} \left[e^{-q(e_j - b_j)} \right]. \quad (7)$$

Let T_j be, for individual j , the time elapsed from the beginning b_j of the exposure period until :

- the time of occurrence of E , if E did occur to individual j ;
- the end e_j of the exposure period, if E did not occur to individual j .

T_j is called the *duration of the period at risk*⁵⁰ for individual j . Likelihood L thus satisfies :

$$\frac{\mathfrak{N}[\text{Log}(L)]}{\mathfrak{N}q} = n \text{Log}(q) - q \sum_{j=1}^N T_j. \quad (8)$$

The maximum likelihood estimate of q , derived from the sample, is therefore :

⁵⁰ The *exposure period* of individual j is (b_j, e_j) , but his *period at risk* is (b_j, e_j) if event E does not occur to him, (b_j, t_j) if event E occurs to him at time t_j .

$$\hat{q} = \frac{n}{\sum_{j=1}^N T_j} \quad (9)$$

Thus, the risk, when *constant over time* and *identical among individuals*, is estimated, on the basis of a sample of *independent* exposures, by the ratio of the *number of events recorded* to the *sum of the durations of the periods at risk*.

If the available information results from *several* independent samples satisfying the same assumptions with identical risk, the overall estimate of this common risk is the ratio of the *total number of events recorded* to the *total sum of the durations of the periods at risk*, *total* meaning computed on the *sum* of the various samples.

It then follows that the overall estimate of q is the *weighted harmonic mean* of the estimates yielded by the different samples, the weights being the *numbers of events recorded* :

$$\hat{q} = \frac{\sum_{s=1}^m n_s}{\sum_{s=1}^m \left(\sum_{j=1}^{N_s} T_{j,s} \right)} = \frac{\sum_{s=1}^m n_s}{\sum_{s=1}^m \left(\frac{n_s}{\hat{q}_s} \right)} \quad (10)$$

If the N individuals are exposed to E during periods of *unitary* duration i.e. $\forall j : e_j - b_j = 1$, then the maximum likelihood estimate \hat{q} of q differs from the *proportion* f of individuals recording event E . In fact, we have :

$$f = \frac{n}{N} \quad \text{with} \quad E(f) = 1 - e^{-q} = Q,$$

which shows that f is an unbiased estimate of Q , while :

$$\hat{q} = \frac{n}{\sum_{j=1}^N T_j} = \frac{n}{N \bar{T}} = \frac{f}{\bar{T}}$$

is a slightly biased⁵¹ estimate of q :

$$E(\hat{q}) = E\left(\frac{f}{\bar{T}}\right) \neq \frac{E(f)}{E(\bar{T})} = q$$

since :

$$E(\bar{T}) = E(T) = \int_0^1 e^{-qx} qx dx + e^{-q} = \frac{1 - e^{-q}}{q} = \frac{E(f)}{q}$$

⁵¹ It can be shown that, for large N , the mathematical expectation of \hat{q} is :

$$E(\hat{q}) \approx q \left[1 + \frac{1}{N} \frac{1 - e^{-q}(1+q)}{(1 - e^{-q})^2} \right] \approx q \left(1 + \frac{1}{2N} \right) \quad \text{if } q \text{ is small.}$$

Considering a sample of N individuals, if *only* the number n of occurrences of E is known, i.e. if n is known, but not \bar{T} , we can base an estimate of q on f . We have :

$$E(f) = 1 - e^{-q},$$

that is :

$$q = -\text{Log}[1 - E(f)].$$

Since f converges in probability (towards $1 - e^{-q}$), we can base the following estimates of q and Q on f :

$$\mathfrak{q} = -\text{Log}(1-f) \approx \frac{f}{1 - \frac{f}{2}} \quad \text{if } f \text{ is small} \quad (11)$$

$$\mathfrak{Q} = 1 - e^{-\mathfrak{q}} = f$$

An equivalent means of computing $\mathfrak{q} = -\text{Log}(1-f)$ is derived as the limit of the convergent sequence defined by the iterative relationship :

$$q_{k+1} = f \frac{q_k}{1 - e^{-q_k}} \quad (12)$$

starting from the limit of the right hand side of (12) when $q \rightarrow 0$:

$$q_0 = f. \quad (13)$$

As an application of these results, let us consider the estimation of the probability of dying between exact ages i and $i + 1$ on a *closed* cohort – i.e. *without* migration – (Figure 1).

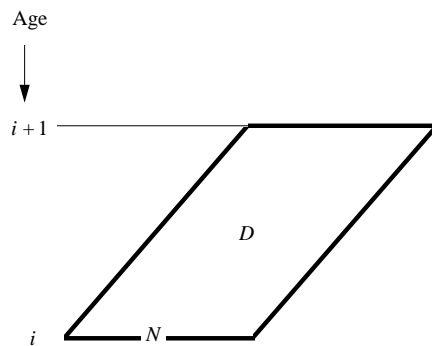


Figure 1. Estimation of the probability of dying in a parallelogram with horizontal sides.

The duration of the exposure period, between birthdays i and $i + 1$, is one year for *each* individual. The risk – also called the *force* – of mortality, q , is assumed to be *constant* with age – at least when age ranges between i and $i + 1$ – and to have no *seasonal* variation : q is the same for all individuals during the *whole* of their period at risk, whatever the location of their birthday within the year. If D is the number of deaths recorded and N the number of members of the cohort who were present at their i^{th} birthday, then q and Q are respectively estimated by :

$$\mathfrak{q} = -\text{Log}\left(1 - \frac{D}{N}\right) \approx \frac{\frac{D}{N}}{1 - \frac{D}{2N}} \quad \text{if } \frac{D}{N} \text{ is small}$$

$$\mathfrak{Q} = 1 - e^{-\mathfrak{q}} = \frac{D}{N}$$

II. – Variable risks

If we drop the assumption that q is *constant* over time, relation (4) yielding the probability that event E does not occur between times b and t , becomes :

$$P(b, t) = e^{-\int_b^t q(x) dx} \quad (14)$$

and quotient Q_i , between exact ages i and $i + 1$, i.e. in a unitary period, is given by :

$$Q = 1 - e^{-\int_0^1 q(x) dx} \quad (15)$$

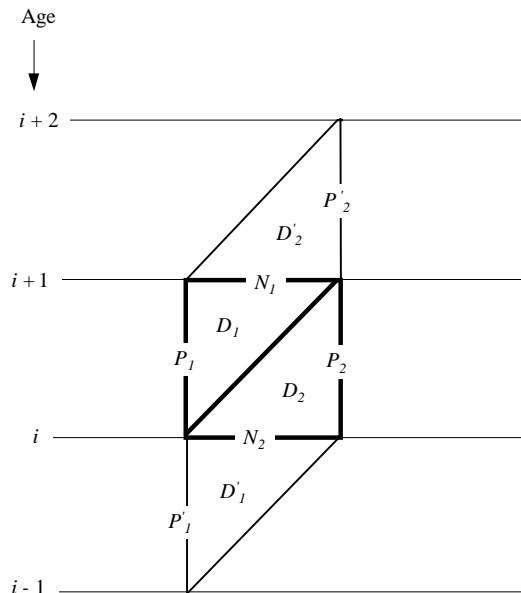
The estimation problem of q and Q , when exposure periods differ among individuals, remains parametrical only if we specify the form of risk function $q(t)$. In the following developments, we shall assume that risk q varies *linearly*⁵² with time, with a derivative q' remaining *small* compared to q , and we shall consider the case of *mortality* within triangles of the Lexis diagram. Ages are denoted i and the calendar year under consideration is T .

In this specific case, it is assumed that, except for the beginning of life (i.e. for ages $i = 0$ or $i = 1$), the *force of mortality* $q(y)$ at *exact age* y , $i < y < i + 1$, for persons belonging to the infinitesimal cohort born between g and $g + dg$, is a *linear* function of y but does not depend on g . Mortality is thus free from *seasonal* variations.

Furthermore, the following assumptions are made concerning migrations :

(i) migrations are *uniformly* distributed inside each triangle. More precisely, the *net migratory balance* between ages y and $y + dy$, among the infinitesimal cohort born between g and $g + dg$, is assumed to be $s dg dy$, where s is *constant* inside a parallelogram with vertical sides, made up of two adjacent triangles (Figure 2).

(ii) immigrants, from the time of their arrival, and emigrants, until the time of their departure, are exposed to the *same* risk of death as the non migrants of the same age.



⁵² An alternative assumption, in line with Gompertz's law, which is generally close to reality at adult ages above 30 or 40 years of age, consists in assuming that it is not q' but q'/q , i.e. the *logarithmic* derivative of q , that is constant.

Figure 2. Notation of population numbers at year ends and on birthdays and notation of numbers of deaths

These two assumptions regarding migrations are obviously *simplistic*. They enable to take an extremely *rough* account of mobility. But if, at all ages, the magnitude of migrations is small compared to the size of the resident population, the impact on mortality is necessarily small and the correction for migrations derived from these assumptions is to be viewed as a *crude* correction for the *volume* of net migrations. In fact, alternative assumptions concerning the seasonal pattern of net migrations or differential mortality between migrants and non-migrants would be very difficult to formulate.

Inside the upper parallelogram of Figure 2, the value of s can be derived from the populations present at the end of each year and recorded deaths in the relevant triangles :

$$s_1 = P'_2 - P_1 + D_1 + D'_2 \quad (16)$$

and the number of $(i + 1)^{\text{th}}$ birthdays celebrated during year T is :

$$\begin{aligned} N_1 &= P_1 - D_1 + \frac{s_1}{2} \\ &= \frac{P_1 + P'_2}{2} + \frac{D'_2 - D_1}{2} \end{aligned} \quad (17)$$

Similarly, inside the lower parallelogram, the value of s is :

$$s_2 = P_2 - P'_1 + D'_1 + D_2 \quad (18)$$

and the number of i^{th} birthdays celebrated during year T is :

$$\begin{aligned} N_2 &= P_2 + D_2 - \frac{s_2}{2} \\ &= \frac{P_2 + P'_1}{2} + \frac{D_2 - D'_1}{2} \end{aligned} \quad (19)$$

In each square of the Lexis diagram, two sorts of triangles will be considered : the *upper* triangle and the *lower* triangle. Let us start with the lower one (Figure 3).

Relationship in the lower triangle

Let $dN_2 = N_2 g_2(u) du$ denote the number of persons celebrating their i i^{th} birthday between times u and $u + du$, time 0 being January 1st of year T . The *density* of the distribution of life lines at time u and birthday i is thus $g_2(u)$, among the N_2 members of the annual cohort born during year $T - i$, present at their i^{th} birthday.

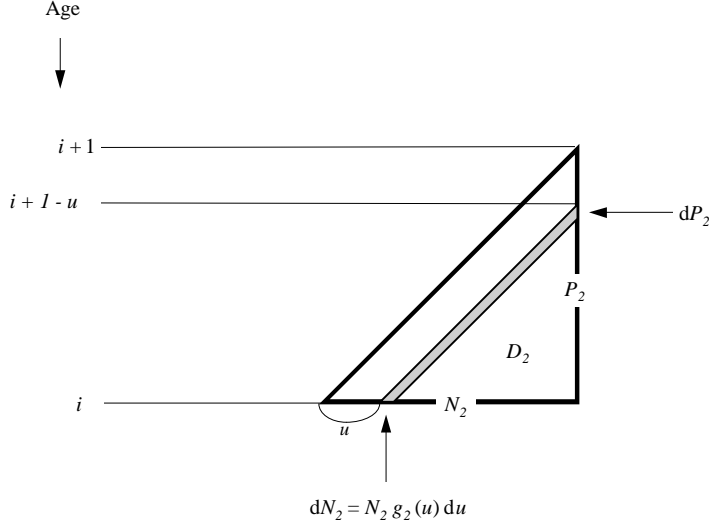


Figure 3. Lower triangle of a square in the Lexis diagram.

Among these dN_2 persons, the number of survivors on December 31st, year T , is :

$$\begin{aligned}
 N_2 g_2(u) du e^{-\int_{x=0}^{1-u} q(x) dx} &= N_2 g_2(u) du e^{-\int_{x=0}^{1-u} \left[q + \left(x - \frac{1}{2}\right) q' \right] dx} \\
 &= N_2 g_2(u) du e^{-q(1-u) - \frac{q'}{2} \left[\left(\frac{1}{2} - u\right)^2 - \frac{1}{4} \right]} \quad (20) \\
 &= N_2 g_2(u) du e^{-q(1-u) + \frac{q'}{2} u(1-u)}
 \end{aligned}$$

where q and q' refer to exact age $i + \frac{1}{2}$.

On the other hand, $s_2 du dy$ net immigrants, belonging to the same infinitesimal cohort (i.e. the one born in year $T - i$, between times u and $u + du$), arrived between ages y and $y + dy$. Among them, the number of survivors on December 31st of year T , is :

$$\begin{aligned}
 s_2 du dy e^{-\int_{x=y}^{1-u} q(x) dx} &= s_2 du dy e^{-\int_{x=y}^{1-u} \left[q + \left(x - \frac{1}{2}\right) q' \right] dx} \\
 &= s_2 du dy e^{-q(1-u-y) - \frac{q'}{2} \left[\left(\frac{1}{2} - u\right)^2 - \left(y - \frac{1}{2}\right)^2 \right]} \quad (21) \\
 &= s_2 du dy e^{-q(1-u-y) + \frac{q'}{2} [u(1-u) - y(1-y)]}
 \end{aligned}$$

Finally, the number of persons aged i to $i + 1$ on December 31st, is the sum of the integrals corresponding to the two categories :

$$\begin{aligned}
P_2 = N_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} g_2(u) du \\
+ s_2 \int_{u=0}^1 \left[\int_{y=0}^{1-u} e^{-q(1-u-y) + \frac{q'}{2}[u(1-u)-y(1-y)]} dy \right] du
\end{aligned} \tag{22}$$

As shown in Annex 2⁵³, this equality is approximately :

$$P_2 \approx N_2 e^{-q(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{V_2}{2}(q^2 - q') \right] + \frac{S_2}{2} e^{-\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) \tag{23}$$

where m_2 and V_2 are the *mean* and the *variance*, respectively, of the distribution of the dates of the i^{th} birthdays that are celebrated during calendar year T .

In terms of deaths, (23) can be written, using equation (19), as :

$$\begin{aligned}
D_2 = N_2 - P_2 + \frac{S_2}{2} \\
\approx N_2 \left\{ 1 - e^{-q(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{V_2}{2}(q^2 - q') \right] \right\} + \frac{S_2}{2} \left[1 - e^{-\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) \right]
\end{aligned} \tag{24}$$

Relationship (24) can be expressed in the following form :

$$q \approx \frac{D_2}{\frac{N_2}{q} \left\{ 1 - e^{-q(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{V_2}{2}(q^2 - q') \right] \right\} + \frac{S_2}{2q} \left[1 - e^{-\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) \right]} \tag{25}$$

which is an equation in q for given q' , D_2 , N_2 , s_2 , m_2 and V_2 . This equation can be solved by applying the rapidly converging iteration relationship :

$$q_{k+1} = \frac{D_2}{\frac{N_2}{q_k} \left\{ 1 - e^{-q_k(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{V_2}{2}(q_k^2 - q') \right] \right\} + \frac{S_2}{2q_k} \left[1 - e^{-\frac{q_k}{3}} \left(1 + \frac{q_k^2}{36} \right) \right]} \tag{26}$$

starting from the limit of the right-hand side of (26) when $q \rightarrow 0$:

$$q_0 = \frac{D_2}{N_2 (1 - m_2) + \frac{S_2}{6}} \tag{27}$$

Thus, based on the assumed prior knowledge of q' , D_2 , N_2 , s_2 , m_2 and V_2 , we obtain the estimation of risk q corresponding to age $i + 1/2$. Note that, if q is assumed to be constant ($q' = 0$), the denominator in relationship (26) is equal to the mathematical expectation of the sum of the durations of the periods at risk. The value of q_0 , given by (27), is an approximate solution⁵⁴ of (24) viewed as an equation in q .

Relationship in the upper triangle

If we consider the *upper* triangle of the square in the Lexis diagram (Figure 4), we have a series of relationships that are similar to relationships (20) to (27).

⁵³ For a better understanding of Annex 2, Annex 1 should be read prior to Annex 2.

⁵⁴ Relationship (27) means that the population at risk approximately consists of two groups :

- N_2 persons, exposed during an *average* period of $1 - m_2$ year,
- $\frac{S_2}{2}$ net immigrants, exposed during an *average* period of $\frac{1}{3}$ year.

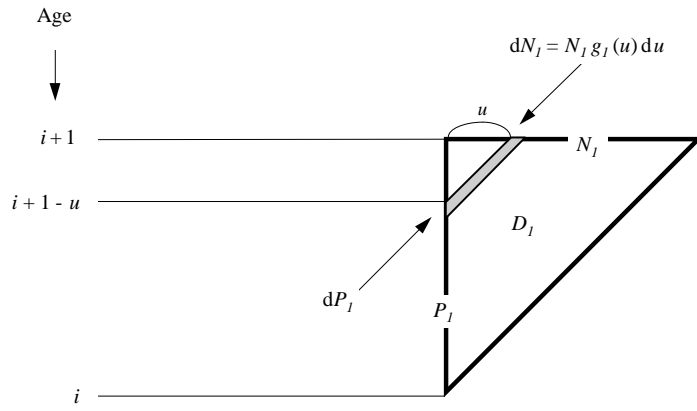


Figure 4. Upper triangle of a square in the Lexis diagram.

Let $dN_1 = N_1 g_1(u) du$ denote the number of persons celebrating their $(i + 1)^{\text{th}}$ birthday between times u and $u + du$ of year T . Thus $g_1(u)$ is the density of the distribution of $(i + 1)^{\text{th}}$ birthdays among the surviving and present members of the cohort born during year $T - i - 1$.

These persons belong to *two* categories :

- those already present on the 1st of January, still surviving,
- net immigrants, aged y to $y + dy$ on arrival, who have not died between their immigration and their $(i + 1)^{\text{th}}$ birthday.

The size of the first subgroup is :

$$\begin{aligned} dP_1 e^{-\int_{y=1-u}^1 q(y) dy} &= dP_1 e^{-\int_{x=1-u}^1 \left[q + \left(x - \frac{1}{2}\right) q' \right] dx} = dP_1 e^{-qu - \frac{q'}{2} \left[\frac{1}{4} - \left(\frac{1}{2} - u\right)^2 \right]} \\ &= dP_1 e^{-qu - \frac{q'}{2} u(1-u)} \end{aligned} \quad (28)$$

while that of the second subgroup is :

$$\begin{aligned} s_1 du dy e^{-\int_y^1 q(x) dx} &= s_1 du dy e^{-\int_y^1 \left[q + \left(x - \frac{1}{2}\right) q' \right] dx} = s_1 du dy e^{-q(1-y) - \frac{q'}{2} \left[\frac{1}{4} - \left(y - \frac{1}{2}\right)^2 \right]} \\ &= s_1 du dy e^{-q(1-y) - \frac{q'}{2} y(1-y)} \end{aligned} \quad (29)$$

Thus we have :

$$N_1 g_1(u) du = dP_1 e^{-qu - \frac{q'}{2} u(1-u)} + s_1 du dy e^{-q(1-y) - \frac{q'}{2} y(1-y)}$$

that is :

$$dP_1 = N_1 e^{qu + \frac{q'}{2} u(1-u)} g_1(u) du - s_1 du dy e^{qu + \frac{q'}{2} u(1-u)} e^{-q(1-y) - \frac{q'}{2} y(1-y)}$$

which leads to :

$$\begin{aligned} P_1 &= N_1 \int_{u=0}^1 e^{qu + \frac{q'}{2} u(1-u)} g_1(u) du \\ &\quad - s_1 \int_{u=0}^1 e^{qu + \frac{q'}{2} u(1-u)} \left[\int_{y=1-u}^1 e^{-q(1-y) - \frac{q'}{2} y(1-y)} dy \right] du \end{aligned} \quad (30)$$

As shown in Annex 3, this equality is approximately :

$$P_1 \approx N_1 e^{q m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (q^2 - q') \right] - \frac{S_1}{2} e^{\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) \quad (31)$$

or, in terms of deaths :

$$D_1 = P_1 - N_1 + \frac{s_1}{2} \quad (32)$$

$$\approx N_1 \left\{ e^{q m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (q^2 - q') \right] - 1 \right\} - \frac{s_1}{2} \left[e^{\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) - 1 \right]$$

This latter relationship can be written in the following form :

$$q \approx \frac{D_1}{\frac{N_1}{q} \left\{ e^{q m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (q^2 - q') \right] - 1 \right\} - \frac{s_1}{2 q} \left[e^{\frac{q}{3}} \left(1 + \frac{q^2}{36} \right) - 1 \right]} \quad (33)$$

which is an equation in q for given q' , D_1 , N_1 , s_1 , m_1 and V_1 . This equation can be solved by applying the rapidly converging iteration relationship :

$$q_{k+1} = \frac{D_1}{\frac{N_1}{q_k} \left\{ e^{q_k m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (q_k^2 - q') \right] - 1 \right\} - \frac{s_1}{2 q_k} \left[e^{\frac{q_k}{3}} \left(1 + \frac{q_k^2}{36} \right) - 1 \right]} \quad (34)$$

starting from the limit of the right hand of (34) when $q \rightarrow 0$, which constitutes an approximate solution of (32) :

$$q_0 = \frac{D_1}{N_1 m_1 - \frac{s_1}{6}} \quad (35)$$

Thus, based on the assumed prior knowledge of q' , D_1 , N_1 , s_1 , m_1 and V_1 , we obtain the estimation of risk q corresponding to age $i + 1/2$. Again here, if q is assumed to be *constant* ($q' = 0$), the denominator in relationship (34) is equal to the mathematical expectation of the sum of the durations of the periods at risk⁵⁵.

Taken together, (24) and (32) constitute the *fundamental relationships* yielding the number of deaths inside each triangle of the Lexis diagram under the assumptions that q' is small compared to q and q^2 is small compared to q' . Except for age 0 – and, to a lesser extent, for age 1 – for which q' is large, such an assumption is close to reality for any age i .

If the relevant birth-cohorts both correspond to a *uniform* density of life lines, we have $m_1 = m_2 = 1/2$ and $V_1 = V_2 = 1/12$. In this case, (24) and (32) become (36) and (37), respectively :

$$D_2 \approx N_2 \left[1 - e^{\frac{q'}{12}} \frac{1 - e^{-q}}{q} \right] - s_2 \frac{e^{-q} - 1 + q - \frac{q^2}{2}}{q^2} \quad (36)$$

or approximately :

$$q \approx \frac{2 D_2}{N_2 + \frac{s_2}{3}}$$

and :

⁵⁵ Similarly to (27), the expression of q_0 given by (35) means that the population at risk approximately consists of :

- N_1 persons, exposed during an *average* period of m_1 year, *minus*
- $\frac{s_1}{2}$ net immigrants, exposed during an *average* period of $\frac{1}{3}$ year.

$$D_1 \approx N_1 \left[e^{\frac{q'}{12}} \frac{e^q - 1}{q} - 1 \right] - s_1 \frac{e^q - 1 - q - \frac{q^2}{2}}{q^2}. \quad (37)$$

or approximately :

$$q \approx \frac{2 D_1}{N_1 - \frac{s_1}{3}}$$

It is only for birth-cohorts affected by a major event, such as a war (especially the outbreak of a war), that the differences between (24) and (32) on the one hand, and (36) and (37) on the other hand, can be substantial. The values of m_1 , m_2 , V_1 and V_2 can then be derived from the monthly distribution of births during the *year of birth* of the cohort, under the assumption that mortality and migration have not substantially altered the distribution of birthdays at later ages.

III. – Construction of a period life table (period of one single calendar year)

Let us assume that the following statistical information is available for the population *present* in a given territory, during a given year T , by sex and single year of age :

- (i) population numbers, on January 1st and December 31st;
- (ii) deaths by year of birth, i.e. by *triangle* in the Lexis diagram.

Each square of the Lexis diagram yields *two* estimates of the *force* of mortality q at age $i + \frac{1}{2}$: q_1 and q_2 following the iterative sequences (26) and (34). To summarise these two estimates of q , we shall retain the *weighted harmonic mean*, as would be the case if q was assumed to be constant, according to (10). Consequently, the estimate of q in the square is given by the limit of the convergent sequence defined by adding numerators and adding denominators in (26) and (34) :

$$q_{k+1} = \frac{\text{Numerator}}{\text{Denominator}} \quad (38)$$

where :

Numerator = $D = D_1 + D_2 =$ Number of deaths in the square

$$\begin{aligned} \text{Denominator} = & \frac{N_1}{q_k} \left\{ e^{q_k m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (q_k^2 - q') \right] - 1 \right\} \\ & + \frac{N_2}{q_k} \left\{ 1 - e^{-q_k (1 - m_2) + \frac{q'}{2} m_2 (1 - m_2)} \left[1 + \frac{V_2}{2} (q_k^2 - q') \right] \right\} \\ & - \frac{s_1}{2 q_k} \left[e^{\frac{q_k}{3}} \left(1 + \frac{q_k^2}{36} \right) - 1 \right] + \frac{s_2}{2 q_k} \left[1 - e^{-\frac{q_k}{3}} \left(1 + \frac{q_k^2}{36} \right) \right] \end{aligned} \quad (39)$$

An iterative process is necessary to assume that q' is *known* : at the first step, q' is taken as equal to zero ; at the second step, q' is taken as equal to half the difference⁵⁶ between the

⁵⁶ This estimation of q' corresponds to a parabolic fit over *three* consecutive values of q :

$$q'_i = \frac{q_{i+1} - q_{i-1}}{2}$$

A parabolic fit over *five* consecutive values of q , yielding smoother results, is provided by :

$$q'_i = \frac{2 (q_{i+2} - q_{i-2}) + q_{i+1} - q_{i-1}}{10}$$

estimates of q yielded by the first step at ages $i + 1$ and $i - 1$, ..., etc. The process is brought to an end when estimates for *all* ages vary little from one step to the next. If it is the *logarithmic* derivative that is assumed to be constant, its value at a given step is taken as half the difference between the logarithms of the estimates of q at ages $i + 1$ and $i - 1$ obtained at the previous step, the first step using zero as an estimate⁵⁷.

If the numbers of deaths D are *only* known by *age at last birthday* – i.e. inside each *square*, and not inside each *triangle*, of the Lexis diagram – another iterative process is needed (before the one just mentioned) : at the first step, D_1 and D_2 are taken as equal to $D/2$ and the migration intensities s_1 and s_2 are estimated accordingly. At the second step, D_1 and D_2 are taken as equal to the value derived from the first step by (32) and (24) respectively, and s_1 and s_2 are re-estimated, ..., etc. This process is stopped when D_1 and D_2 are stabilised for *all* ages.

Let us observe that an approximation of the solution of (38) is :

$$\mathcal{Q} \approx \frac{D_1 + D_2}{N_1 m_1 + N_2 (1 - m_2) + \frac{s_2 - s_1}{6}}$$

which shows that the approximate impact of migrations on the number of deaths inside a square of the Lexis diagram is related to the *difference* in the migratory balances inside the two triangles of that square.

Special case of ages 0 and 1

The above procedures apply to any age other than 0 and 1, because of large q' and the impossibility of deriving q' (or q'/q) from the values of q at *both* ages $i - 1$ and $i + 1$. At age 0 for example, we make the three following assumptions :

- there are no migrations,
- during year T , births and first birthdays are *identically* distributed over time within the year,
- D_1 and D_2 are known.

Then Q is estimated by :

$$\mathcal{Q} = 1 - \left(1 - \frac{D_1}{P_1}\right) \left(1 - \frac{D_2}{N_2}\right) \quad (40)$$

where $N_2 = B_T$ refers to the number of live births recorded during year T .

Relationship (40) can be proved as follows.

⁵⁷ The parabolic fit over *five* consecutive values of q yields smoother results :

$$\frac{q'_i}{q_i} = \frac{2 \operatorname{Log} \left(\frac{q_{i+2}}{q_{i-2}} \right) + \operatorname{Log} \left(\frac{q_{i+1}}{q_{i-1}} \right)}{10}$$

Under the three afore-mentioned assumptions, we have :

$$\left(1 - \frac{D_1}{P_1}\right) \left(1 - \frac{D_2}{N_2}\right) = \frac{P_2 / N_2}{P_1 / N_1} = \frac{\int_0^1 e^{-\int_0^{1-u} q(x) dx} g_2(u) du}{\int_0^1 e^{\int_{1-u}^1 q(x) dx} g_1(u) du}$$

that is :

$$\frac{\int_0^1 e^{-\int_0^{1-u} q(x) dx + \int_{1-u}^1 q(x) dx} g_2(u) du}{\int_0^1 e^{\int_{1-u}^1 q(x) dx} g_1(u) du} = (1-Q) \frac{\int_0^1 e^{\int_{1-u}^1 q(x) dx} g_2(u) du}{\int_0^1 e^{\int_{1-u}^1 q(x) dx} g_1(u) du} = 1 - Q \quad \text{if} \quad \forall u : g_1(u) = g_2(u)$$

This relationship holds for *any* risk function $q(x)$, provided that *migrations* are zero and that birthday distributions g_1 and g_2 are *identical* at ages i and $i + 1$. We shall rewrite it as follows:

$$\hat{Q} = 1 - \left(1 - \frac{D_1}{P_1}\right) \left(1 - \frac{D_2}{N_2}\right) \text{ with } a = \frac{D_1}{P_1}, \quad b = \frac{D_2}{N_2}.$$

If net migrations are not zero inside the *first* square but if birthday distributions at ages i and $i + 1$ are identical, let us denote respectively D_1^0 and D_2^0 (instead of D_1 and D_2) the numbers of deaths in the *upper* and the *lower* triangles of the *first* square (i.e. between exact ages 0 and 1), D_1^1 and D_2^1 the similar numbers of deaths in the upper and lower triangles of the *second* square (i.e. between exact ages 1 and 2), D_1^2 and D_2^2 those in the triangles of the *third* square (i.e. between exact ages 2 and 3). Let us also denote respectively P_1^0 and P_2^0 (instead of P_1 and P_2) the population numbers of completed age 0 on January 1st and December 31st, year T . Similarly, P_1^1 and P_2^1 are the population numbers of completed age 1 on January 1st and December 31st, year T , P_1^2 and P_2^2 those of completed age 2 on January 1st and December 31st, year T .

Two estimates of each term a and b can be calculated. Concerning a , i.e. the *upper* triangle, the first estimate assumes that migrations during year T within cohort born year $T - 1$ take place *exclusively* on January 1st, the second one that migrations during year T within cohort born year T take place *exclusively* on December 31st. Under each of these assumptions, net migrations are zero, but in the first one, the population number on January 1st is not P_1^0 but $P_2^1 + D_1^0 + D_2^1$:

$$a_1 = \frac{D_1^0}{P_2^1 + D_1^0 + D_2^1} \quad a_2 = \frac{D_1^0}{P_1^0}$$

Concerning b , i.e. the *lower* triangle, the first estimate assumes that migrations during year T within cohort born year $T - 1$ take place *exclusively* immediately after birth, the second one that migrations during year T within cohort born year T take place *exclusively* on December 31st:

$$b_1 = \frac{D_2^0}{P_2^0 + D_2^0} \quad b_2 = \frac{D_2^0}{B_T}$$

It follows that the lower and upper bounds of \hat{Q} are:

$$1 - [1 - \min(a_1, a_2)] [1 - \min(b_1, b_2)] \text{ and } 1 - [1 - \max(a_1, a_2)] [1 - \max(b_1, b_2)]$$

If both net migrations during year T respectively inside the vertical-sided parallelogram corresponding to cohort born year $T - 1$, $s_U = P_2^1 - P_1^0 + D_1^0 + D_2^1$, and inside the lower triangle, $s_L = P_2^0 - B_T + D_2^0$, are small, the two estimates are naturally close to each other. For example, in France in 1990, the lower and upper bounds of the estimated infant mortality rate were respectively 8.41 and 8.48 per 1,000 *male* live births and 6.24 and 6.31 per 1,000 *female* live births.

It must be observed that a satisfactory estimate of the infant mortality rate (\hat{Q} at age 0) requires the availability of the numbers of deaths by *triangle* in the Lexis diagram. If the information available is limited to the *total* number of deaths, SQ_0 , inside the *first* square, a rough estimate of the infant mortality rate is still made possible via the preliminary estimation of the breakdown of SQ_0 between D_1^0 and D_2^0 . In contemporary Europe, the proportion of D_2^0 to SQ_0 is approximately 90%.

The same procedure can be applied to the estimation of the probability of dying between exact ages 1 and 2. The lower and upper bounds of that probability are again:

$$1 - [1 - \min(a_1, a_2)] [1 - \min(b_1, b_2)] \text{ and } 1 - [1 - \max(a_1, a_2)] [1 - \max(b_1, b_2)]$$

with :

$$a_1 = \frac{D_1^1}{P_2^2 + D_1^1 + D_2^2}, \quad a_2 = \frac{D_1^1}{P_1^1}, \quad b_1 = \frac{D_2^1}{P_2^1 + D_2^1}, \quad b_2 = \frac{D_2^1}{P_1^0 - D_1^0}$$

The lower and upper bounds of the probability of dying between exact ages 1 and 2 are closer (in absolute terms and, frequently, in relative terms) to each other than the lower and upper bounds of the infant mortality rate. For example, in France in 1990, the lower and upper bounds of the probability of dying between exact ages 1 and 2 were respectively 6.97 and 7.02 per 10,000 *male* live births and 5.00 and 5.02 per 10,000 *female* live births.

In contemporary Europe, the share of the lower triangle in the *second* square (between exact ages 1 and 2) is only slightly over 0.5 (around 52 to 55%).

Crude estimate of the probability of dying

If we assume that :

- birthdays are *uniformly* distributed during the year, whatever the birthday considered
- the force of mortality $q(x)$ is *small* and *constant* between exact ages i and $i + 1$
- migration intensities, s_1 and s_2 , inside the two triangles of the same square, are *equal*
- iterative relationship (38) is applied only *once*, starting with $q \approx 0$
- numbers of deaths D_1 and D_2 are *close* to each other

then q is estimated by :

$$\hat{q} \approx \frac{2D}{N_1 + N_2} = \frac{2D}{P_1 - D_1 + P_2 + D_2} \approx \frac{D}{(P_1 + P_2)/2} \quad (41)$$

and Q by :

$$\mathcal{Q} = 1 - e^{-\hat{q}} = 1 - e^{-\frac{D}{(P_1 + P_2)/2}} \quad (42)$$

or even, still *more* crudely :

$$\mathcal{Q} \approx \frac{D}{(P_1 + P_2)/2} \quad (43)$$

Note that estimating the force of mortality q by :

$$\hat{q} = \frac{D}{(P_1 + P_2)/2}$$

means estimating an *instantaneous quotient* by the corresponding *observed rate*, i.e. the ratio of the number of events recorded to the *average* exposed population. This *average* exposed population is approximately the *mid-year* population of the age under consideration.

Number of survivors and life expectancy

The estimation \hat{q} of q between ages i and $i + 1$, based on the data available, leads to the following estimate of the probability of dying :

$$\begin{aligned} \mathcal{Q} &= 1 - e^{-\int_0^1 q(x) dx} \\ &= 1 - e^{-\int_0^1 \left[\hat{q} + \left(x - \frac{1}{2}\right) q' \right] dx} \\ &= 1 - e^{-\hat{q}} \end{aligned} \quad (44)$$

On the other hand, the average time lived between birthdays i and $i + 1$ by survivors at age i is :

$$\begin{aligned} \int_0^1 e^{-\int_0^x q(x) dx} dx &= \int_0^1 e^{-\int_0^x \left[\hat{q} + \left(x - \frac{1}{2}\right) q' \right] dx} dx = \int_0^1 e^{-\hat{q}x + \frac{q'}{2}x(1-x)} dx \\ &\approx e^{-\frac{\hat{q}}{2} + \frac{q'}{8}} \left[1 + \frac{\hat{q}^2 - q'}{24} \right] \approx e^{-\frac{\hat{q}}{2} + \frac{q'}{12}} \left[1 + \frac{\hat{q}^2}{24} \right] \\ &\approx e^{\frac{q'}{12}} \frac{1 - e^{-\hat{q}}}{\hat{q}} = e^{\frac{q'}{12}} \frac{\mathcal{Q}}{\hat{q}} \end{aligned} \quad (45)$$

If \mathcal{Q}_i denotes the estimate of the probability of dying at age i , then the number of survivors $S(j|i)$ at age j , for 1 survivor at age i , $i < j$, and the life expectancy $e(i)$ at age i are as follows :

$$\begin{aligned} S(j|i) &= \prod_{k=i}^{j-1} (1 - \mathcal{Q}_k) \quad \text{with} \quad S(i|i) = 1 \\ e(i) &= \sum_{j \geq i} \left[S(j|i) e^{\frac{q'_j}{12}} \frac{\mathcal{Q}_j}{\hat{q}_j} \right] \end{aligned} \quad (46)$$

IV. – Construction of a cohort life table (cohort born during one single year)

If, instead of working on the data for a single calendar year, we use the information concerning the two adjacent triangles of a parallelogram with horizontal sides, for the *same*

annual birth-cohort but for *two* consecutive calendar years (Figure 5), then the distribution of birthdays inside the cohort does not play any role.

In this case, the relationship between N_1 and N_2 , as can be demonstrated using the results of Annex 1, is :

$$N_1 \approx N_2 e^{-q} + \frac{s_1}{2} e^{-\frac{q}{3} - \frac{q'}{12}} \left(1 + \frac{q^2}{36}\right) + \frac{s_2}{2} e^{-\frac{2q}{3} - \frac{q'}{12}} \left(1 + \frac{q^2}{36}\right) \quad (47)$$

or in terms of deaths :

$$D = N_2 - N_1 + \frac{s_1 + s_2}{2} \quad (48)$$

$$\approx N_2 \left(1 - e^{-q}\right) + \frac{s_1}{2} \left[1 - e^{-\frac{q}{3} - \frac{q'}{12}} \left(1 + \frac{q^2}{36}\right)\right] + \frac{s_2}{2} \left[1 - e^{-\frac{2q}{3} - \frac{q'}{12}} \left(1 + \frac{q^2}{36}\right)\right]$$

Relationship (48) leads to the converging iterative sequence :

$$q_{k+1} = \frac{D}{\frac{N_2}{2q_k} (1 - e^{-q_k}) + \frac{s_1}{2q_k} \left[1 - e^{-\frac{q_k}{3} - \frac{q'}{12}} \left(1 + \frac{q_k^2}{36}\right)\right] + \frac{s_2}{2q_k} \left[1 - e^{-\frac{2q_k}{3} - \frac{q'}{12}} \left(1 + \frac{q_k^2}{36}\right)\right]} \quad (49)$$

where $D = D_1 + D_2$ is the number of deaths in the parallelogram.

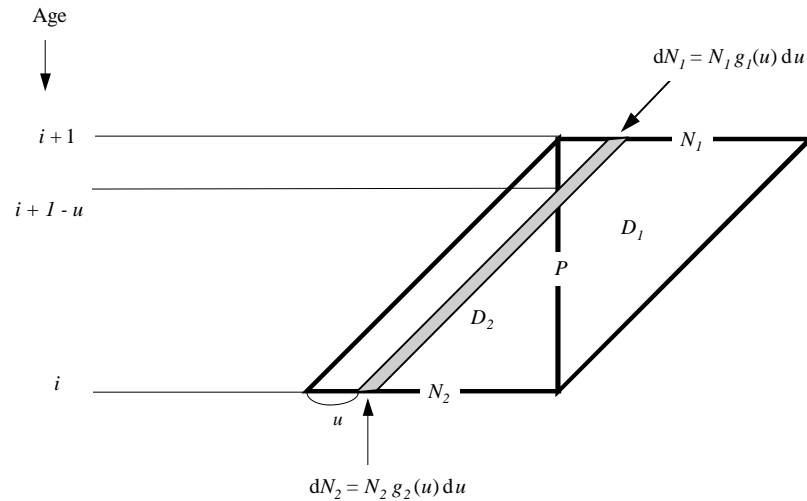


Figure 5. Parallelogram with horizontal sides.

V. – Comparison of methods used to compute life tables

In the case of France, in 1950 and 1990, we have applied the results presented in section III, on the basis of population figures by single year of age at year ends :

- *Method A* : application of iterative relationship (38), knowing the number of deaths by *triangle* in the Lexis diagram and the distribution of birthdays within each annual birth-cohort ;
- *Method B* : application of *double* iterative relationship (38), knowing the number of deaths by *square* in the Lexis diagram and the distributions of birthdays ;
- *Method C* : application of (42), knowing *only* the number of deaths by *square* in the Lexis diagram : the *observed rate* :

$$r = \frac{D}{(P_1 + P_2)/2}$$

is used as an estimate of the *risk of mortality* and the probability of dying is derived :

$$\hat{q} = r \quad \mathcal{Q} = 1 - e^{-r}$$

Two variants of method *C* are considered : methods C_1 and C_2 use the second and the first order approximations, respectively :

$$\mathcal{Q} = \frac{r}{1 + r/2} \text{ for method } C_1 \quad \mathcal{Q} = r \text{ for method } C_2$$

For age 0 and age 1 only, all three methods apply the same procedure : that expressed by relationship (40), assuming that deaths by *triangle* are available for these two special ages.

For the computation of life expectancy at all ages, all four methods assume that, above age 90, probabilities of death increase at a constant rate with age (linear adjustment of their logarithms).

Method A : the recalculation of the number of deaths in each triangle

To assess the validity of Method A, which is the most accurate among the three under consideration (if the underlying assumptions hold), we may compare the *observed* numbers of deaths D_1 and D_2 with their *recalculated* values \hat{D}_1 and \hat{D}_2 , derived from \hat{q} according to (32) and (24) :

$$\hat{D}_1 = N_1 \left\{ e^{\hat{q} m_1 + \frac{q'}{2} m_1 (1 - m_1)} \left[1 + \frac{V_1}{2} (\hat{q}^2 - q') \right] - 1 \right\} - \frac{S_1}{2} \left[e^{\frac{\hat{q}}{3}} \left(1 + \frac{\hat{q}^2}{36} \right) - 1 \right]$$

and :

$$\hat{D}_2 = N_2 \left\{ 1 - e^{-\hat{q} (1 - m_2) + \frac{q'}{2} m_2 (1 - m_2)} \left[1 + \frac{V_2}{2} (\hat{q}^2 - q') \right] \right\} + \frac{S_2}{2} \left[1 - e^{-\frac{\hat{q}}{3}} \left(1 + \frac{\hat{q}^2}{36} \right) \right]$$

Since, by construction, we have :

$$D = D_1 + D_2 = \hat{D}_1 + \hat{D}_2$$

it is sufficient to compare D_1 and \hat{D}_1 .

Figure 6 shows how the ratio D_1 / \hat{D}_1 varies according to age, for males as well as for females, both in 1950 and 1990. It appears that, in 1990, the ratio fluctuates randomly around 1 below age 50. But at higher ages, there is a clear *upward trend* : D_1 is *systematically larger* than it ought to be if the assumptions were fully valid. The same upward trend is also visible in 1950 above age 50, but a *downward trend* seems to exist below 50. In fact, in 1950, the ratio D_1 / \hat{D}_1 fluctuates around (slightly *above*) 1 only in the neighbourhood of age 50. The ratio exceeds 1 for all ages more often than not.

The reason for these biases is attributable⁵⁸ to the *seasonality* of mortality : the upper triangle contains deaths that mostly occur during the first part of the year while the lower

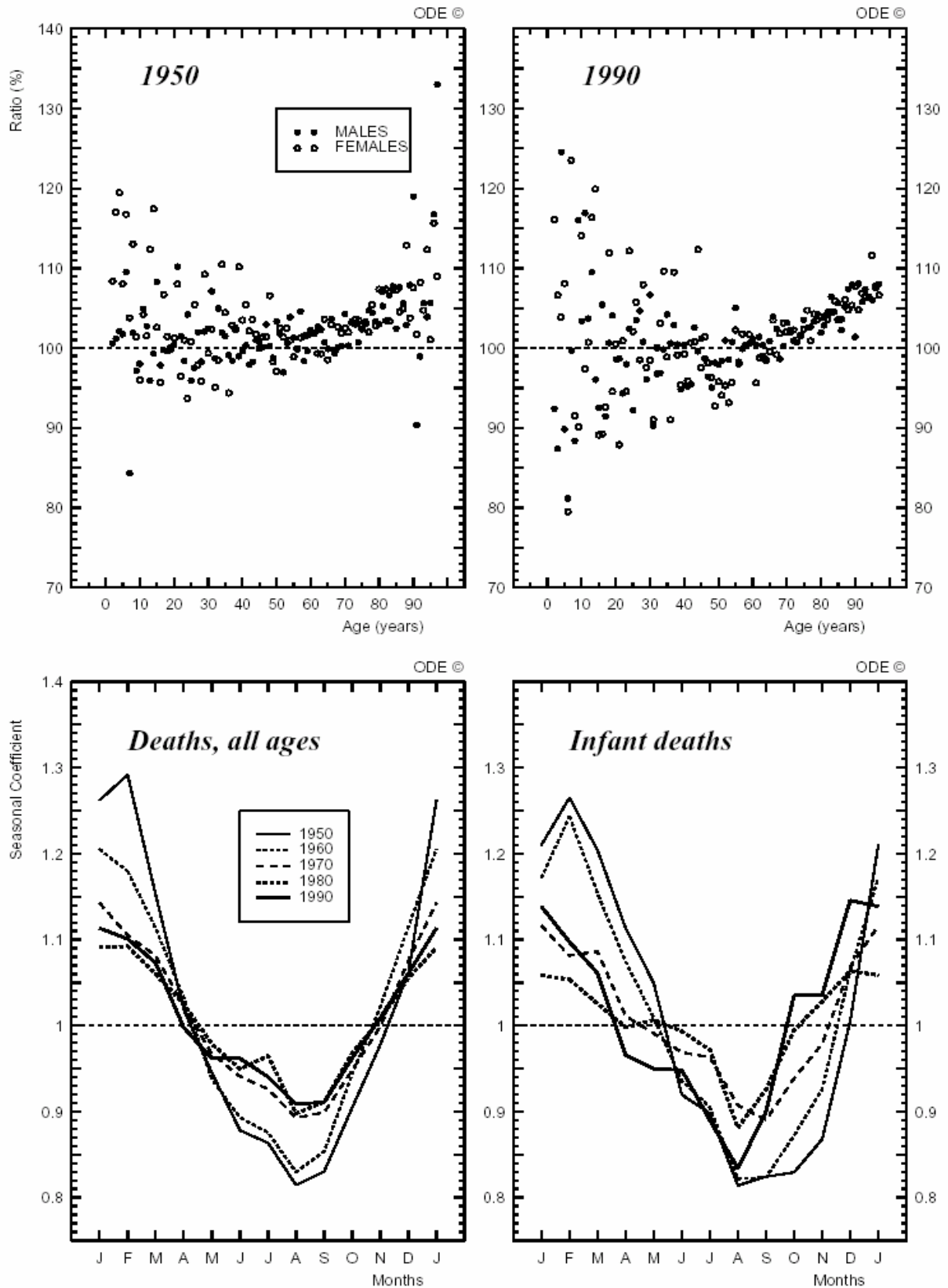
⁵⁸ It might also be due to a wrong allocation, by triangle, of deaths located in a given square of the Lexis diagram.

Figure 6 FRANCE

Upper Part : RATIO, by SEX and AGE, of the OBSERVED NUMBER of DEATHS to the RECALCULATED NUMBER of DEATHS

Method A, UPPER TRIANGLE of each square, 1950 and 1990

Lower Part : SEASONAL PROFILE of the MONTHLY NUMBER of DEATHS DEATHS for ALL AGES and INFANT DEATHS, 1950, 1960, 1970, 1980, 1990



triangle contains deaths that mostly occur during the latter part of the year. If, at a given exact age, the seasonality of mortality is such that the force of mortality, q , is *higher* in the first months than in the last months of the year, then D_1 tends to exceed the value \hat{D}_1 estimated under the assumption of absence of seasonality.

If this explanation is true, it means that mortality showed some seasonality – especially above age 50 – at almost all ages in 1950, but this seasonality has greatly declined, if not disappeared, under age 50 in 1990.

For France, deaths by month are available from 1946 to 1991 (for all ages, and for infant deaths), by broad age-groups from 1946 to 1962 and by detailed age-groups from 1981 to 1991. The seasonality of mortality can thus be estimated, for all ages and for age 0 from 1950 to 1991, for age-group 80+ from 1950 to 1958⁵⁹ and for detailed age-groups from 1981 to 1991.

Indeed Figure 7 shows that the seasonality of deaths (all ages and age 0) has greatly declined in the past 40 years and that the months with the highest seasonal coefficients are January and February, with November having a seasonal coefficient almost equal to 1. At higher ages, the seasonality is more pronounced than for all ages (January and February had seasonal coefficients for ages 80+ close to 1.5 in the early 1950's). Figure 8 shows the seasonal profile of mortality in selected years. Figure 9 compares the proportion of deaths in the upper triangle of the Lexis diagram to 50 %, using the following index :

$$I = \frac{1}{6} \sum_{m=1}^{12} \left[\left(1 - \frac{2m-1}{24} \right) c_m \right]$$

Index I would be equal to 1 if seasonal coefficients c_m – and death numbers adjusted for seasonal variations – were constant over months. The magnitude of this index and its decline to almost 1 in the 1980's is consistent with the discrepancies shown by Figure 6.

Method B compared to Method A : does the breakdown, by triangle, of the number of deaths in the squares of the Lexis diagram improve the estimation of the probabilities of dying ?

Figure 10 (upper part) shows the percentage difference between the probability of dying estimated according to Method B and its equivalent estimated according to Method A. In 1950 as well as in 1990, the percentage difference hardly ever reaches 1 %. Below age 70, it is systematically smaller than 0.1 %, and below 85 smaller than 0.5 %. The similar figure for females leads to the same conclusion.

These discrepancies result from the differences between the observed and the recalculated numbers of deaths inside each triangle. It is therefore due to the seasonality of mortality again, but the impact is only noticeable at higher ages.

The conclusion is that the breakdown of the numbers of deaths D by triangle adds nothing below 70 years of age – and hardly anything between 70 and 80 years of age – to the quality of the estimation of the probabilities of dying. It is only at ages 80 and over that an appreciable improvement is noted.

However, to establish life tables by birth-cohorts, the breakdown is clearly necessary.

⁵⁹ The method used to estimate the seasonal coefficients "loses" four years at each extremity of the period for which monthly data are available. It takes in due consideration the number of days in each month.

Figure 7 FRANCE, 1950-1987
Monthly seasonal coefficients of the number of DEATHS
DEATHS ALL AGES (full lines) and INFANT DEATHS (dotted lines)

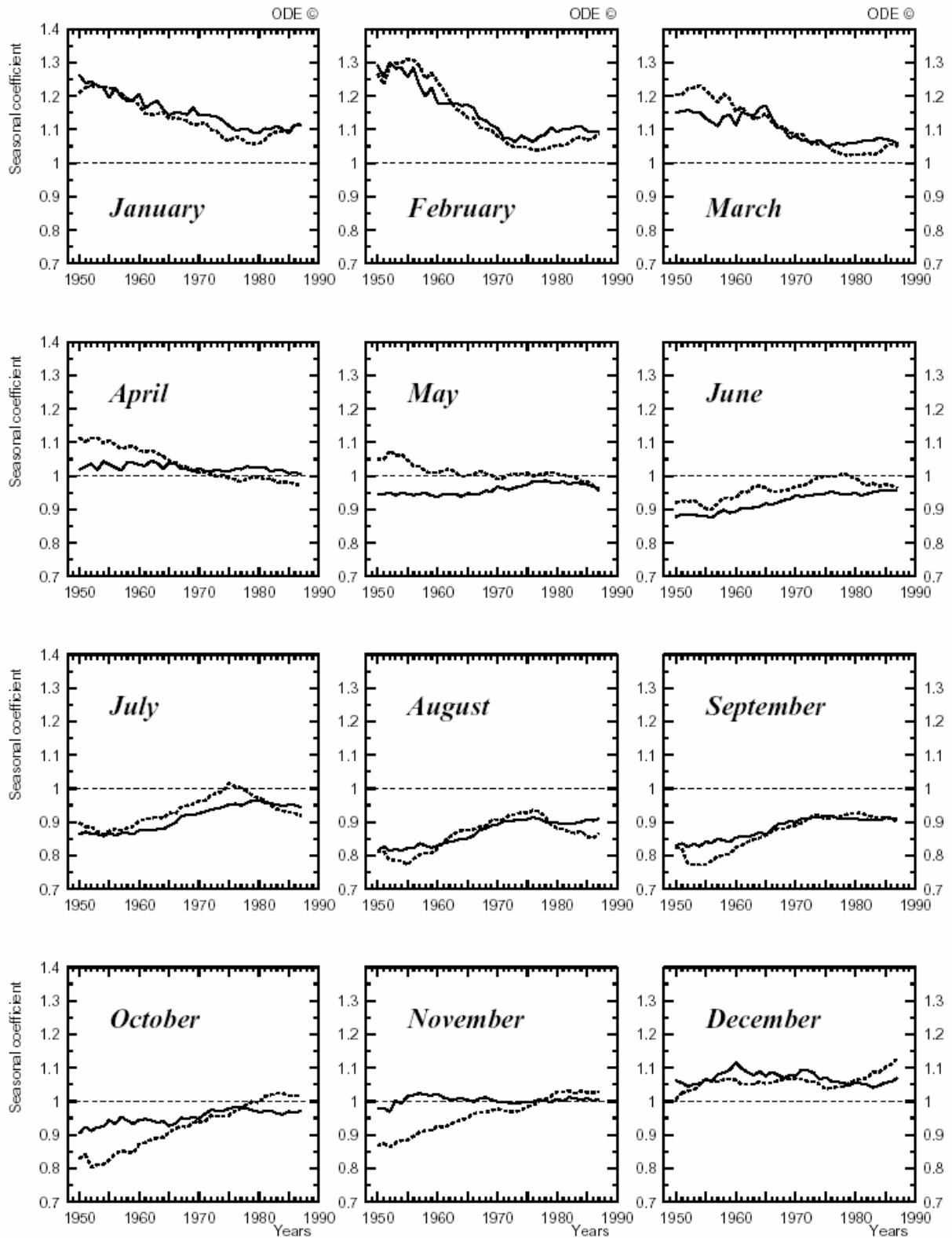


Figure 8 FRANCE, 1950, 1960, 1970, 1980, 1990
SEASONAL PROFILE of the MONTHLY number of DEATHS
DEATHS for ALL AGES, INFANT DEATHS

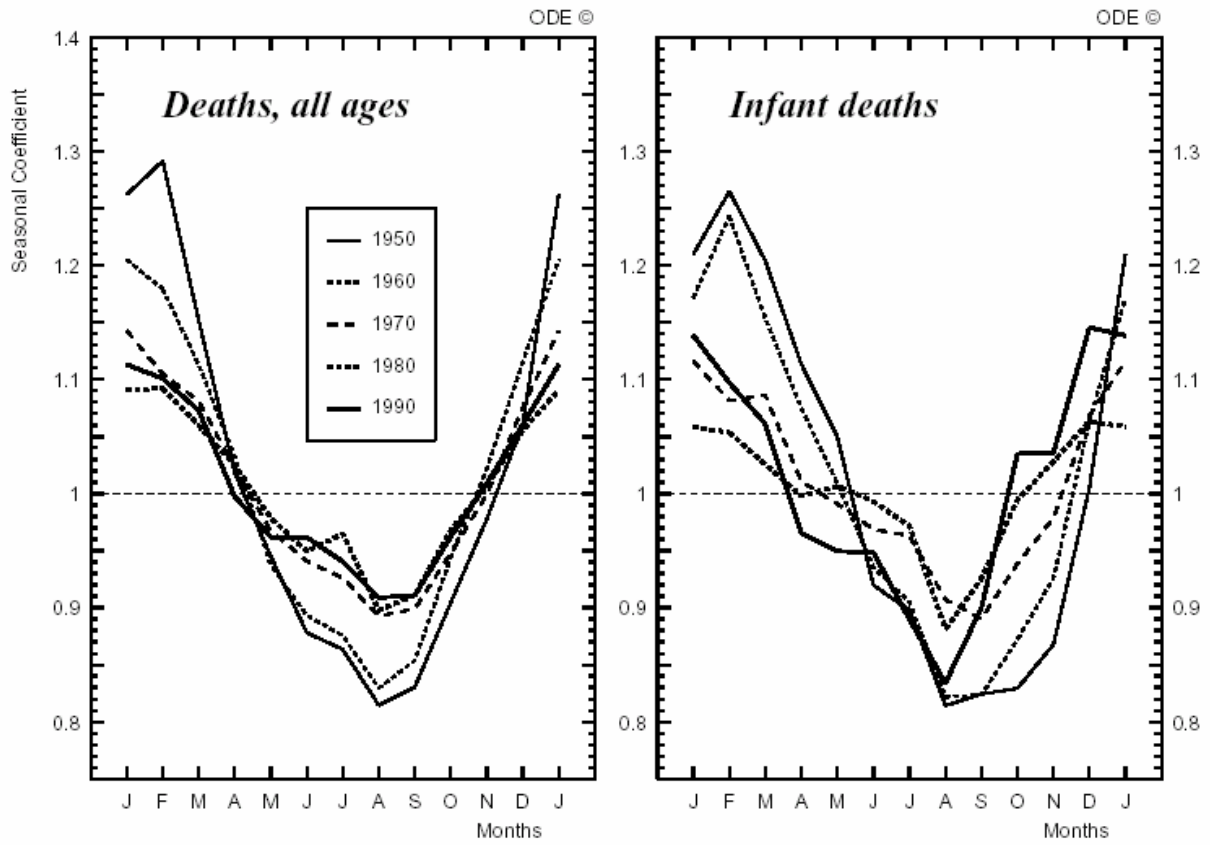


Figure 9 FRANCE, 1950-1987
Ratio of the proportion of DEATHS in the UPPER TRIANGLE to 50 %
(based on seasonal coefficients)

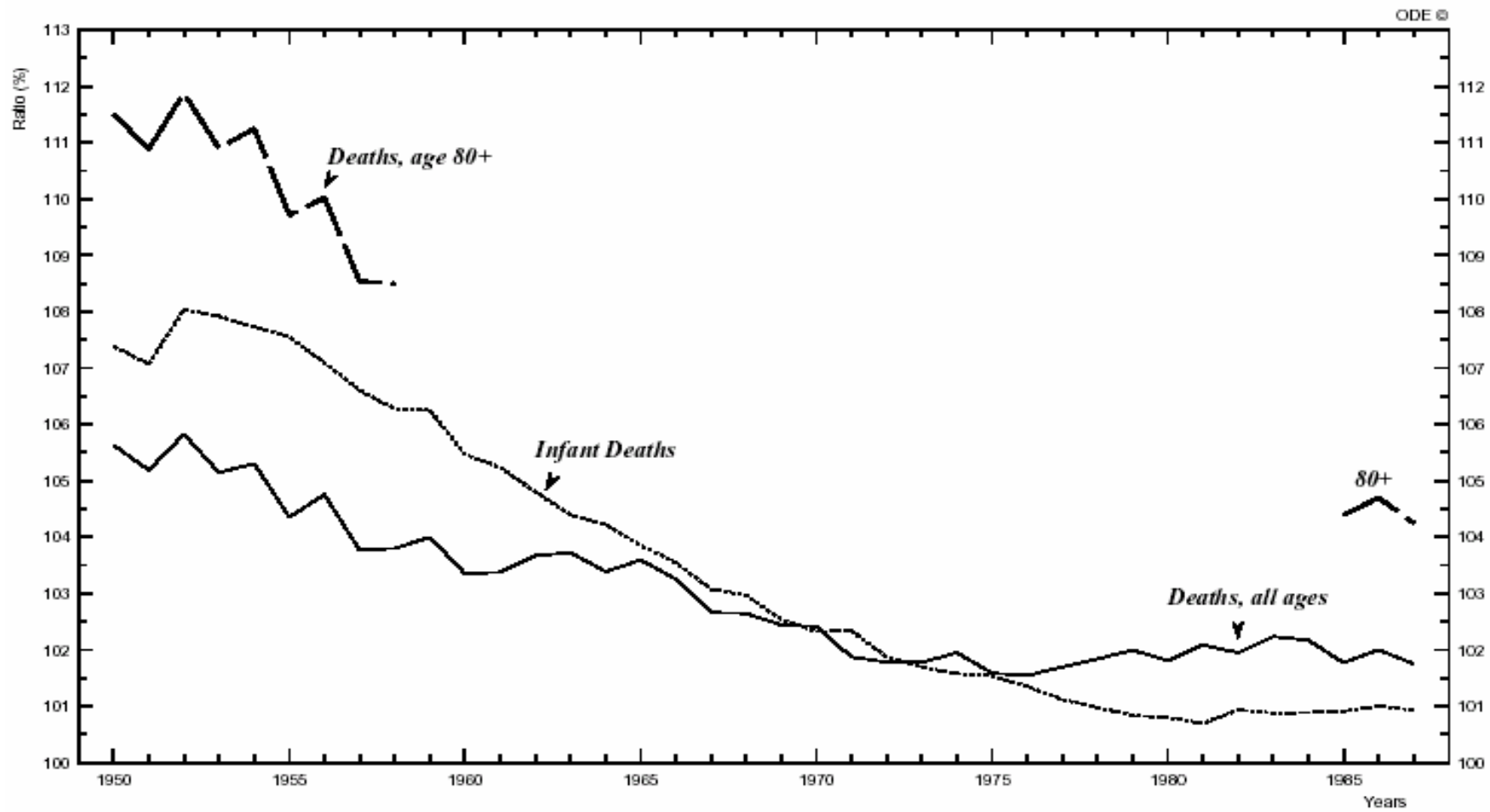
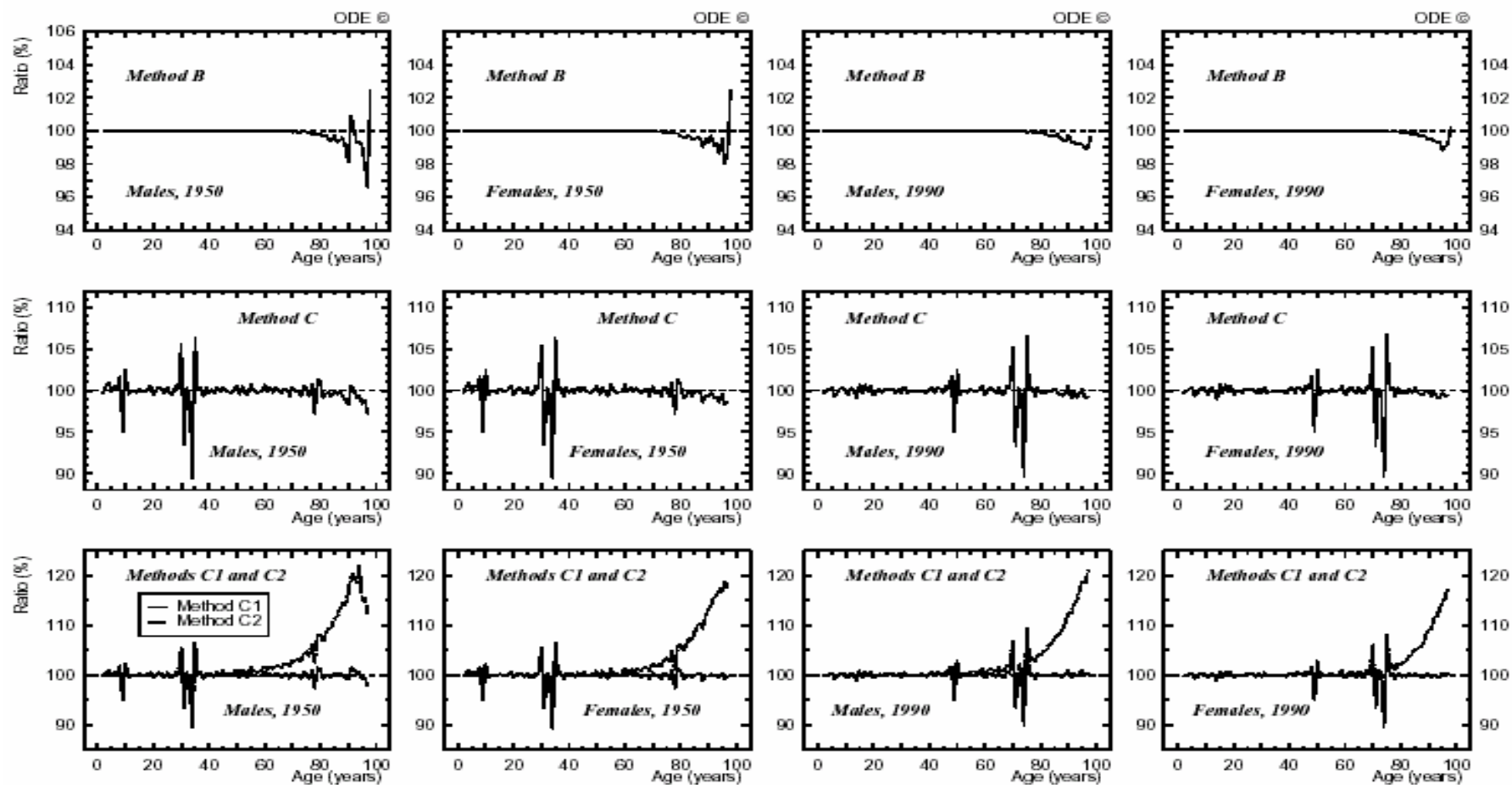


Figure 10 FRANCE, 1950 and 1990
RATIO, by SEX and AGE, of the PROBABILITY of DYING
estimated according to Methods B, C, C1 and C2
to its equivalent estimated according to Method A



Method C and variants C₁ and C₂ compared to Method A : is it worthwhile to take into account the distribution of birthdays within cohorts ?

First, the correction made to remove the effect of the non-uniformity of the distribution of birthdays within the year *can be large*, as is shown in Figures 10 (central and lower parts). For instance, the percentage difference between probabilities of dying according to Method A and Method C exceeds 10 % at age 34 in 1950 and at age 74 in 1990 (i.e. the 1915-1916 birth-cohorts). With Method C₂, the magnitude of risks (because of the difference between $1 - e^{-\mathcal{Q}}$ and \mathcal{Q}) adds its own effect, bringing the percentage difference with method A to 20 % at age 90. But methods C and C₁ yield very similar results. The conclusion would be again identical for females.

Second, this correction, which – for a given pair of two consecutive cohorts – is very close to a *constant percent* correction, clearly *attenuates* most of the anomalies which appear in the curves of the probabilities of dying according to age : the correction *smooths* the curves, as is clear in Figure 11 (upper part).

Third, the magnitude of the correction to apply to the probabilities of dying is large in itself, but also compared to the variations of the latter *over time*. Figure 11 (lower part) shows that several of the erratic fluctuations of Q through time are eliminated by the correction.

Finally, it is worthwhile to take the distribution of birthdays into account if the country considered has recorded abrupt changes in its birth-rate, as is the case of most European countries because of the two world wars.

The following question may now be posed : given that the distribution of birthdays is generally only known for the *year of birth*, i.e. for age *zero*, can we assume that mortality and migrations do not substantially modify these distributions for later ages ? This question has been examined in Calot and Caselli (1990)⁶⁰, on the basis of census population figures by sex and *month* of birth. The conclusion was that there was good agreement between the corrections based on the distribution of births by month and the corrections based on the distribution of birthdays derived from the census.

VI. – Mortality data for European countries in Eurostat data banks

The analysis of demographic change and, in particular, the assessment of the most *recent* evolution, is a permanent concern for a wide range of observers, at national and international levels. Despite national differences in definitions, *comparable* data are needed – over long periods in the past, but also for most recent periods, even insights into the future through population projections.

This led the then director of the Institut National d'Etudes Démographiques to set up, in the 1980's, a computerised system for the analysis of demographic data in the field of fertility, which was later extended to nuptiality, and more recently to mortality. This system, called "Projet international d'analyse démographique conjoncturelle", was underpinned with information kindly provided by 40 national (or sub-national) statistical offices in Europe and Northern America.

⁶⁰ G. Calot and G. Caselli, Détermination d'une table de mortalité : la conversion des taux en quotients, *Population*, 6, November-December 1991, I.N.E.D., Paris, pp 1441-1490.

Figure 11 FRANCE, 1990

Upper Part : PROBABILITIES of DYING between 65 and 80 years of age, by SEX and AGE, estimated by Methods A and C

Lower Part : PROBABILITY of DYING at AGE 50 and AGE 60, by SEX

Comparison between Method A and Method C

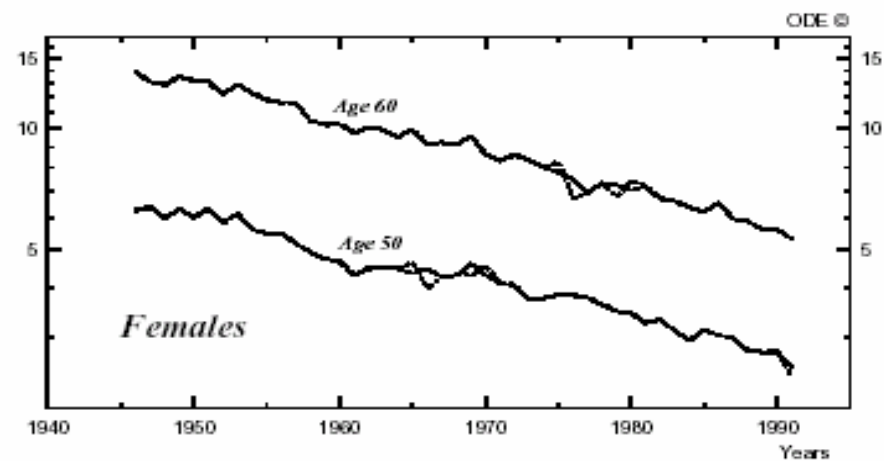
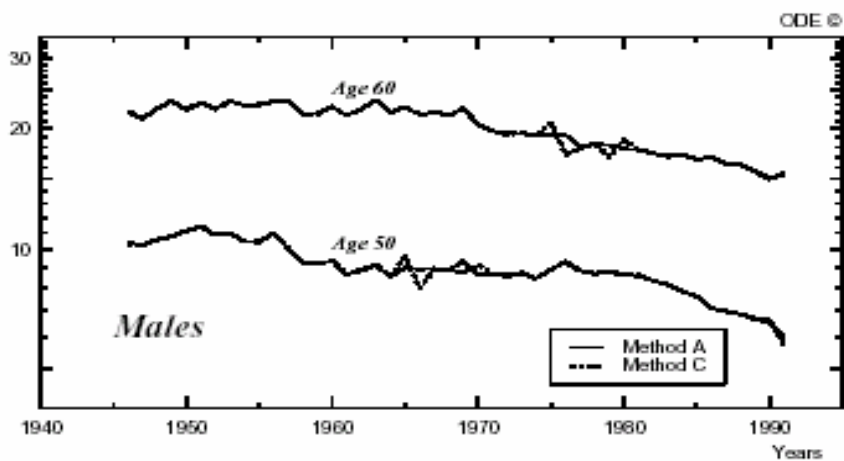
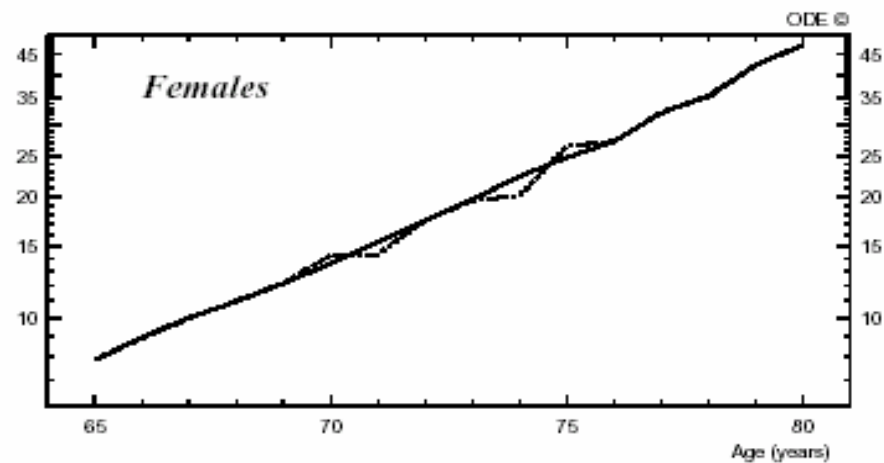
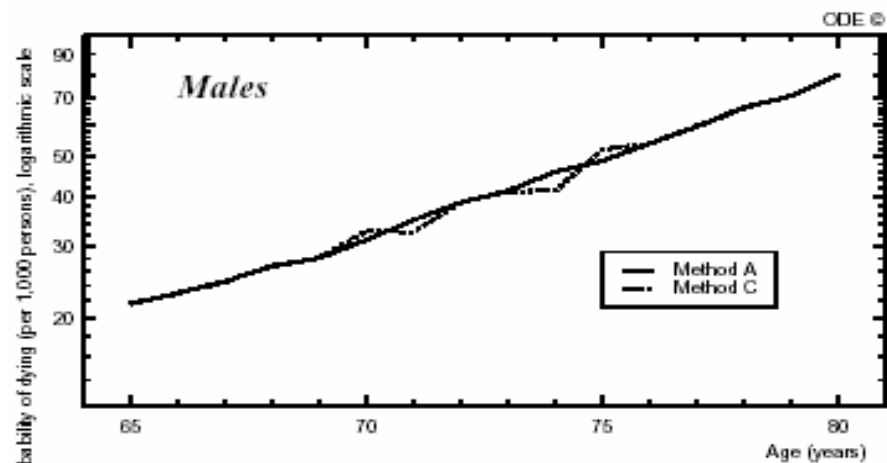


Figure 12 FRANCE, 1950, 1970 et 1990
PROBABILITY of DYING by AGE, according to SEX (Method A)

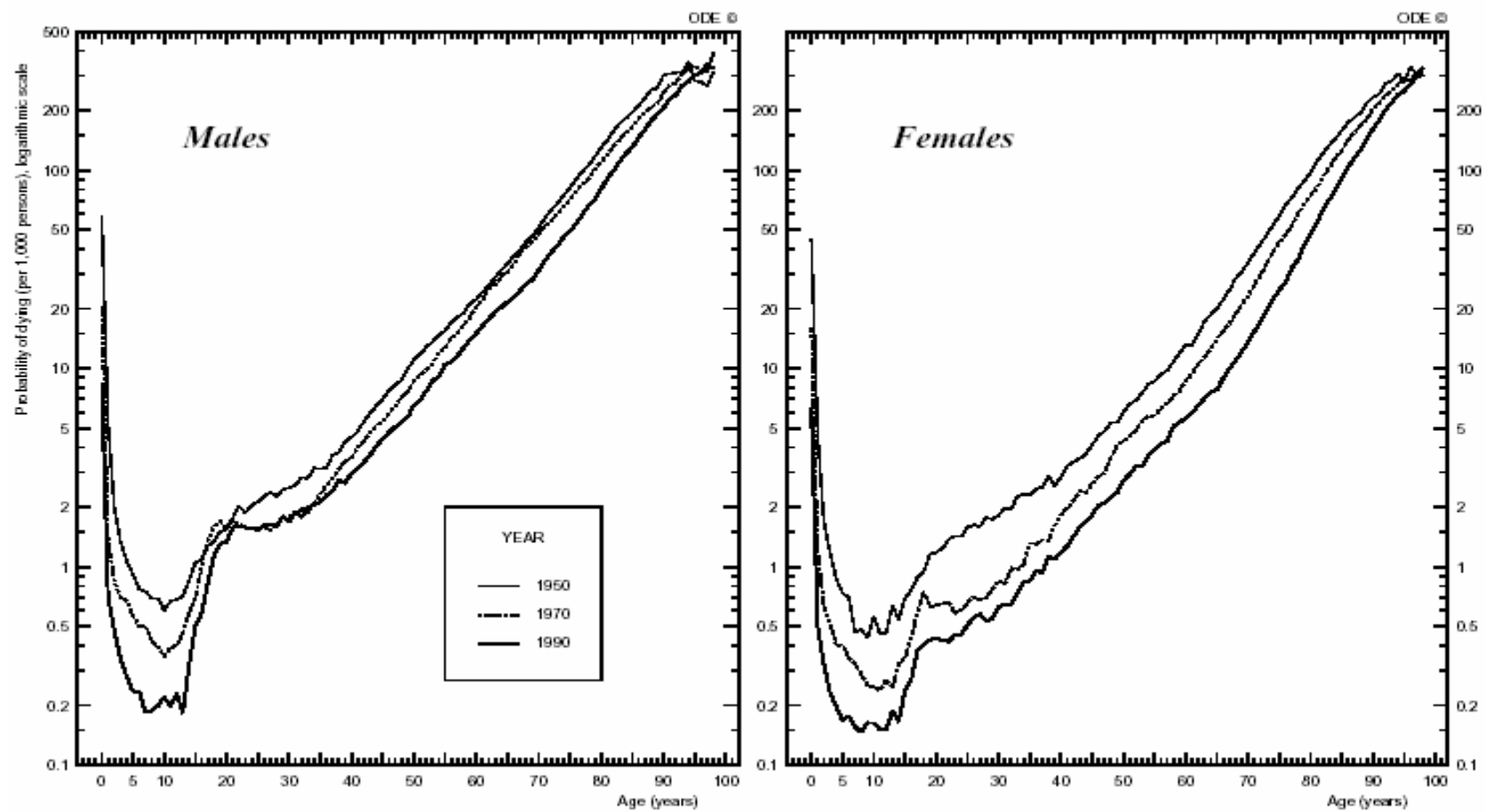
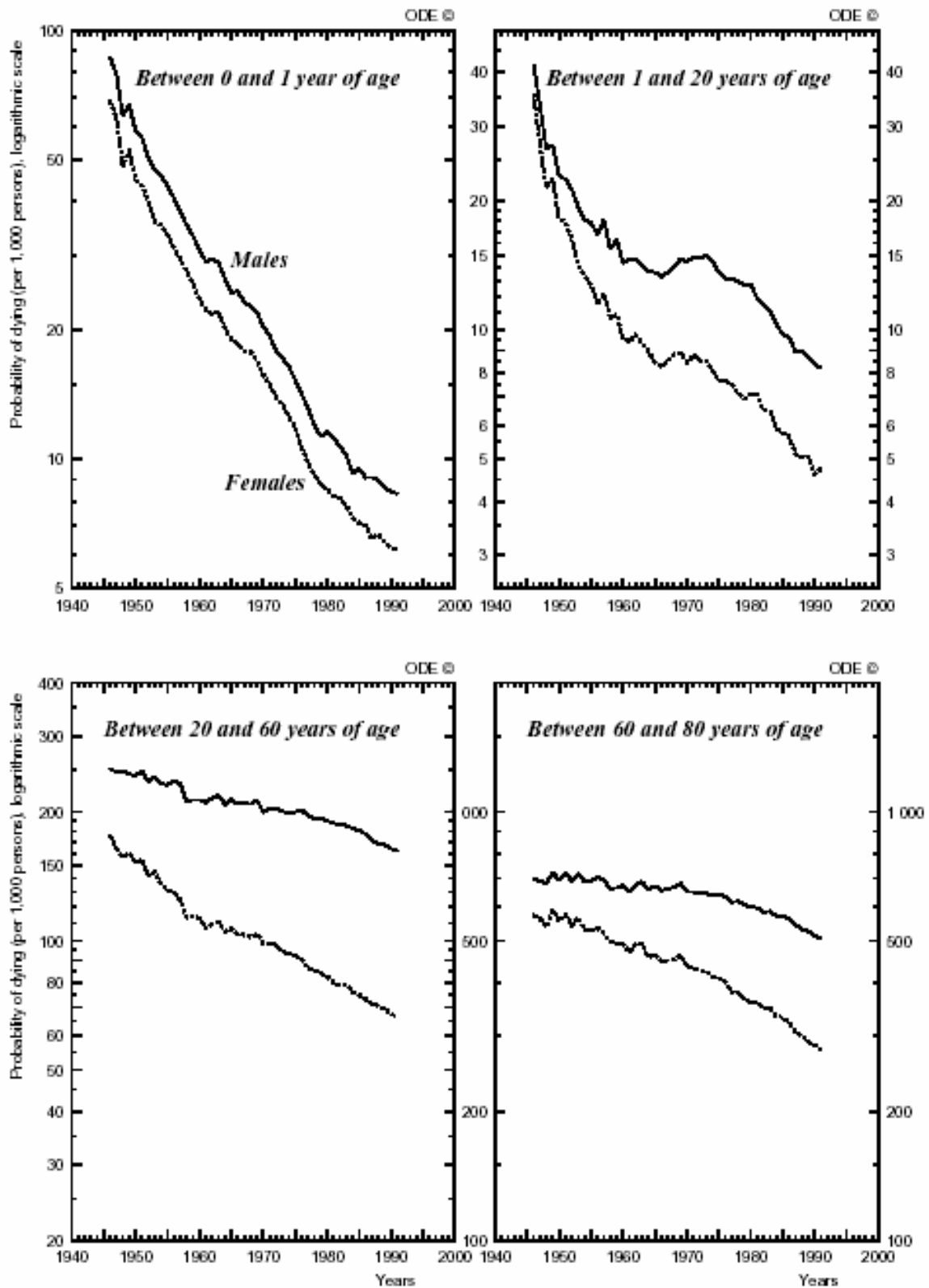


Figure 13 FRANCE, 1946-1991
PROBABILITY of DYING between various ages



The system, together with the data supplied by countries, was transferred to Eurostat, whose responsibility is preparing yearbooks with comparable and up-to-date information on the member States of the European Union, and also keeping the different services of the Commission, the governments and the public at large permanently informed on current changes.

Within Eurostat, the system was renamed *Syscodem* ("Système communautaire d'observation démographique") and the yearly supply of basic information by statistical offices was put on a systematic footing. A software upgrade has since been written. It contains *utility* programs for the acquisition and correction of basic data, *processing* programs which produce comparable outputs, and programs for the *presentation* of the results, in terms of tables and graphs. Examples of the latter⁶¹ are appended to this paper. At the same time, the scope of the system is being extended (in terms of geographical coverage, time periods and fields of study – for example divorces by marriage cohorts).

As far as *mortality by sex and age* is concerned, the basic data supplied by countries for each calendar year are the following :

- *Population by sex and age at year ends*

The data supplied refer to the resident population on January 1st, by sex and single year of age, up to an open-ended age-group (generally 100+). A few countries produce their data for a different date in the year (for example Ireland on April 15). In this case, estimates are made of the sex-age distributions on January 1st.

- *Deaths by sex, age and year of birth*

In most countries deaths are classified by age at last birthday *and* year of birth, i.e. by triangle in the Lexis diagram. For a few of them, available data only refer to age at last birthday, i.e. to *squares* in the Lexis diagram.

- *Births by year and month for existing birth-cohorts*

Live births, not only by year, but also by month, over long periods of the past are available in many countries. These data provide estimates of the distributions of birthdays within cohorts. Census data of total enumerated population by year *and* month of birth can also be used when the monthly distribution of live births is not available.

VII. – Conclusion

The article presents a new approach to the construction of life tables.

It shows that the conventional computation of *rates* based on population numbers at year ends and death numbers by *squares* in the Lexis diagram leads to sizeable *biases* in the probabilities of dying when the birth-rate recorded *abrupt* changes at the time the corresponding cohorts were born⁶² (such is the general case in Europe due to the two world wars). On the basis of the *monthly* distributions of live births, commonly available in European countries over long periods of the past, it is possible to correct these biases satisfactorily.

Moreover, the methods proposed *smooth* the probabilities of dying, offering the possibility to establish life tables *annually* (and not only for periods of several years), in spite of the relatively large random fluctuations of the rates due to the small magnitude of death numbers

⁶¹ Figures 6 to 13 were drawn using the Syscodem software for graphs.

⁶² These biases are similar to those which occur in the computation of rates for *non-repeatable* events, such as age-specific *fertility* rates.

at most ages. With modern computers, these methods do not require significantly more resources than conventional ones but greatly improve the results.

It is advisable to record annual deaths by sex and single year of age not simply in *squares* of the Lexis triangle, but also in *triangles*. For ages 0 and 1 year of age, it is still more strongly advisable than for other ages.

Annex 1 : Approximations for $\int_a^b e^{j(x)} g(x) dx$ and $\int_a^b x^k e^{j(x)} g(x) dx$

Let us consider, on the one hand, a continuous random variable X , ranging from a to b , with density $g(x)$, mean m , variance V , moments around the mean m_k and, on the other hand, a continuous function $j(x)$.

The first derivatives of $y = e^{j(x)}$ are :

$$\begin{aligned} y' &= e^{j(x)} j'(x) \\ y'' &= e^{j(x)} [j''(x) + j'(x)^2] \\ y''' &= e^{j(x)} [j'''(x) + 3j'(x)j''(x) + j'(x)^3] \end{aligned}$$

The development of $e^{j(x)}$ around any m is uniformly convergent. Therefore, denoting j' , j'' , j''' the successive derivatives of $j(x)$ at $x = m$, the mathematical expectation of $e^{j(x)}$ is :

$$\begin{aligned} E[e^{j(x)}] &= \int_a^b e^{j(x)} g(x) dx \\ &= \int_a^b e^{j(m)} \left[1 + (x-m)j' + \frac{(x-m)^2}{2!} (j'' + j'^2) + \frac{(x-m)^3}{3!} (j''' + 3j'j'' + j'^3) + \dots \right] g(x) dx \\ &\approx e^{j(m)} \left[1 + \frac{V}{2} (j'' + j'^2) + \frac{m_3}{6} (j''' + 3j'j'' + j'^3) \right] \end{aligned}$$

that is :

$$\int_a^b e^{j(x)} g(x) dx \approx e^{j(m)} \left[1 + \frac{V}{2} (j'' + j'^2) + \frac{m_3}{6} (j''' + 3j'j'' + j'^3) \right] \quad (1)$$

Similarly, k being a non negative integer, the mathematical expectation of $X^k e^{j(x)}$ is :

$$\begin{aligned} E[X^k e^{j(x)}] &= \int_a^b x^k e^{j(x)} g(x) dx \\ &= \int_a^b e^{j(m)} \left[m^k + (k m^{k-1} + m^k j') (x-m) + \left(\frac{k(k-1)}{2} m^{k-2} + k j' m^{k-1} + \frac{j'' + j'^2}{2} m^k \right) (x-m)^2 + \dots \right] g(x) dx \\ &\approx e^{j(m)} \left[m^k + V m^{k-2} \left(\frac{k(k-1)}{2} + m k j' + m^2 \frac{j'' + j'^2}{2} \right) \right] \end{aligned}$$

that is :

$$\int_a^b x^k e^{j(x)} g(x) dx \approx e^{j(m)} \left[m^k + V m^{k-2} \left(\frac{k(k-1)}{2} + m k j' + m^2 \frac{j'' + j'^2}{2} \right) \right] \quad (2)$$

For example, if X is a *uniform* random variable on $(0,1)$, $g(x)$ is equal to 1, m to $\frac{1}{2}$, V to $\frac{1}{12}$ and m_3 to 0.

We then have :

$$\int_0^1 e^{j(x)} dx \approx e^{j\left(\frac{1}{2}\right)} \left(1 + \frac{j'' + j'^2}{24} \right) \quad (3)$$

and :

$$\int_0^1 x e^{j(x)} dx \approx \frac{e^{j\left(\frac{1}{2}\right)}}{2} \left(1 + \frac{j'}{6} + \frac{j'' + j'^2}{24} \right) \quad (4)$$

$$\int_0^1 x^2 e^{j(x)} dx \approx \frac{e^{j(\frac{1}{2})}}{3} \left(1 + \frac{j'}{4} + \frac{j'' + j'''}{32} \right) \quad (5)$$

$$\int_0^1 x^3 e^{j(x)} dx \approx \frac{e^{j(\frac{1}{2})}}{4} \left(1 + \frac{j'}{4} + \frac{j'' + j'''}{48} \right) \quad (6)$$

An illustration of (3) is the following. Let us consider the cumulative function of the *normal* law :

$$\begin{aligned} \Pi(u) &= \int_{-\infty}^u \frac{1}{\sqrt{2p}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{2} + \frac{u}{\sqrt{2p}} \int_0^1 e^{-\frac{u^2 x^2}{2}} dx \end{aligned}$$

We here have, for any x and therefore for $x = 1/2$:

$$\begin{aligned} j(x) &= -\frac{u^2}{2} x^2 & j &= -\frac{u^2}{8} \\ j'(x) &= -u^2 x & j' &= -\frac{u^2}{2} \\ j''(x) &= -u^2 & j'' &= -u^2 \end{aligned}$$

After (3), we may write :

$$\Pi(u) \approx \frac{1}{2} + \frac{u}{\sqrt{2p}} e^{-\frac{u^2}{8}} \left[1 + \frac{u^2}{24} \left(\frac{u^2}{4} - 1 \right) \right]$$

This approximation of order 2 for $u = 1$ yields $\Pi(1) = 0.8411$ while the exact value is 0.8413. For $u = 2$, it yields $\Pi(2) = 0.984$ while the exact value is 0.978.

The accuracy of this approximation diminishes as u increases, but improves for fixed u if $e^{j(x)}$ is developed further. The reader may check that the fourth order approximation yields $\Pi(3) = 0.9987$ to four exact decimal places :

$$\Pi(2u) \approx \frac{1}{2} + u \sqrt{\frac{2}{p}} e^{-\frac{u^2}{2}} * z(u)$$

with $z(u)$ equal to :

$$1 + \frac{u^2}{3!} (u^2 - 1) + \frac{u^4}{5!} (u^4 - 6u^2 + 3) + \frac{u^6}{7!} (u^6 - 15u^4 + 45u^2 - 15) + \frac{u^8}{9!} (u^8 - 280u^6 + 210u^4 - 420u^2 + 105)$$

Annex 2 : Approximation for relationship (22)

Let us apply the results of Annex 1 to the two integrals appearing in relationship (22) :

$$P_2 = N_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} g_2(u) du \\ + S_2 \int_{u=0}^1 \left[\int_{y=0}^{1-u} e^{-q(1-u-y) + \frac{q'}{2}[u(1-u)-y(1-y)]} dy \right] du$$

under the assumption that q' and q^2 are small compared to q , q^3 being small compared to q' .

The first integral :

$$N_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} g_2(u) du$$

corresponds to :

$$j(u) = -q(1-u) + \frac{q'}{2}u(1-u) \\ j'(u) = q + q' \left(\frac{1}{2} - u \right) \approx q \\ j''(u) = -q'$$

On the basis of relationship (1) in Annex 1, we thus prove the first part of relationship (23) :

$$N_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} g_2(u) du \approx N_2 e^{-q(1-m_2) + \frac{q'}{2}m_2(1-m_2)} \left[1 + \frac{V_2}{2}(q^2 - q') \right]$$

The second integral :

$$S_2 \int_{u=0}^1 \left[\int_{y=0}^{1-u} e^{-q(1-u-y) + \frac{q'}{2}[u(1-u)-y(1-y)]} dy \right] du$$

can be written :

$$S_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} \left[\int_{y=0}^{1-u} e^{qy - \frac{q'}{2}y(1-y)} dy \right] du$$

To compute the inner integral in y , we apply relationship (3) of Annex 1, putting $y = (1-u)z$:

$$\int_{y=0}^{1-u} e^{qy - \frac{q'}{2}y(1-y)} dy = (1-u) \int_0^1 e^{q(1-u)z - \frac{q'(1-u)}{2}z + \frac{q'(1-u)^2}{2}z^2} dz \\ \approx (1-u) e^{\frac{q(1-u)}{2} - \frac{q'(1-u)}{4} + \frac{q'(1-u)^2}{8}} \left[1 + \frac{(q^2 + q')(1-u)^2}{24} \right] \\ = e^{\frac{q(1-u)}{2} - \frac{q'(1-u)}{4} + \frac{q'(1-u)^2}{6}} \left[1 - u + \frac{q^2}{24}(1-u)^3 \right]$$

with :

$$j(z) = q(1-u)z - \frac{q'(1-u)}{2}z + \frac{q'(1-u)^2}{2}z^2 \rightarrow j\left(\frac{1}{2}\right) = \frac{q(1-u)}{2} - \frac{q'(1-u)}{4} + \frac{q'(1-u)^2}{8} \\ j'(z) = q(1-u) - \frac{q'(1-u)}{2} + q'(1-u)z \rightarrow j'\left(\frac{1}{2}\right) = q(1-u) - \frac{q'(1-u)}{2} + \frac{q'(1-u)^2}{2} \\ j''(z) = q'(1-u)^2$$

It follows that the second term in (22) is equal to :

$$\begin{aligned}
s_2 \int_{u=0}^1 \left[\int_{y=0}^{1-u} e^{-q(1-u-y) + \frac{q'}{2}[u(1-u)-y(1-y)]} dy \right] du &= s_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} \left[\int_{y=0}^{1-u} e^{qy - \frac{q'}{2}y(1-y)} dy \right] du \\
&\approx s_2 \int_{u=0}^1 e^{-q(1-u) + \frac{q'}{2}u(1-u)} e^{\frac{q(1-u)}{2} - \frac{q'(1-u)}{4} + \frac{q'(1-u)^2}{6}} \left[1 - u + \frac{q^2}{24}(1-u)^2 \right] du \\
&= s_2 \int_{u=0}^1 e^{-\frac{q(1-u)}{2} + \frac{q'(1-u)}{4} - \frac{q'(1-u)^2}{3}} \left[1 - u + \frac{q^2}{24}(1-u)^2 \right] du
\end{aligned}$$

that is :

$$s_2 \int_{u=0}^1 \left[\int_{y=0}^{1-u} e^{-q(1-u-y) + \frac{q'}{2}[u(1-u)-y(1-y)]} dy \right] du \approx s_2 \int_{u=0}^1 \left(u + \frac{q^2}{24}u^3 \right) e^{-\frac{qu}{2} + q' \left(\frac{u}{4} - \frac{u^2}{3} \right)} du$$

To compute this latter expression, we apply relationships (4) and (6) of Annex 1 :

$$\begin{aligned}
s_2 \int_{u=0}^1 \left(u + \frac{q^2}{24}u^3 \right) e^{-\frac{qu}{2} + q' \left(\frac{u}{4} - \frac{u^2}{3} \right)} du &\approx \frac{s_2}{2} e^{-\frac{q}{4} + \frac{q'}{24}} \left(1 - \frac{q}{12} - \frac{q'}{72} + \frac{q^2}{96} - \frac{q'}{36} + \frac{q^2}{48} \right) \\
&\approx \frac{s_2}{2} e^{-\frac{q}{4}} \left(1 - \frac{q}{12} + \frac{q^2}{32} \right) \\
&\approx \frac{s_2}{2} e^{-\frac{q}{3}} \left(1 + \frac{q^2}{36} \right)
\end{aligned}$$

with :

$$\begin{aligned}
j(u) &= -\frac{qu}{2} + q' \left(\frac{u}{4} - \frac{u^2}{3} \right) \quad \rightarrow \quad j\left(\frac{1}{2}\right) = -\frac{q}{4} + \frac{q'}{24} \\
j'(u) &= -\frac{q}{2} + q' \left(\frac{1}{4} - \frac{2u}{3} \right) \quad \rightarrow \quad j'\left(\frac{1}{2}\right) = -\frac{q}{2} - \frac{q'}{12} \\
j''(u) &= -\frac{2q'}{3}
\end{aligned}$$

Annex 3 : Approximation for relationship (30)

Let us apply the results of Annex 1 to the computation of the integrals in (30) :

$$P_1 = N_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} g_1(u) du - s_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} \left[\int_{y=1-u}^1 e^{-q(1-y) - \frac{q'}{2}y(1-y)} dy \right] du$$

under the assumption that q' and q^2 are small compared to q , q^3 being small compared to q' .

The first integral :

$$N_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} g_1(u) du \quad \text{corresponds to :}$$

$$j(u) = qu + \frac{q'}{2}u(1-u)$$

$$j'(u) = q + q' \left(\frac{1}{2} - u \right) \approx q$$

$$j''(u) = -q'$$

Applying relationship (1) of Annex 1, we thus prove the first part of relationship (31) :

$$N_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} g_1(u) du \approx N_1 e^{qm_1 + \frac{q'}{2}m_1(1-m_1)} \left[1 + \frac{V_1}{2}(q^2 - q') \right]$$

The second integral is :

$$s_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} \left[\int_{y=1-u}^1 e^{-q(1-y) - \frac{q'}{2}y(1-y)} dy \right] du.$$

To compute the integral in y , we apply relationship (3) of Annex 1, putting $y = 1 - uz$:

$$\begin{aligned} \int_{y=1-u}^1 e^{-q(1-y) - \frac{q'}{2}y(1-y)} dy &= u \int_0^1 e^{-quz - \frac{q'}{2}uz + \frac{q'}{2}z^2} dz \\ &\approx u e^{-\frac{qu}{2} - \frac{q'u}{4} + \frac{q'u^2}{8}} \left[1 + \frac{(q^2 + q')u^2}{24} \right] \\ &= e^{-\frac{qu}{2} - \frac{q'u}{4} + \frac{q'u^2}{6}} \left(u + \frac{q^2}{24}u^3 \right) \end{aligned}$$

where :

$$\begin{aligned} j(z) &= -quz - \frac{q'}{2}uz + \frac{q'u^2}{2}z^2 \quad \rightarrow \quad j\left(\frac{1}{2}\right) = -\frac{qu}{2} - \frac{q'u}{4} + \frac{q'u^2}{8} \\ j'(z) &= -qu - \frac{q'u}{2} + q'u^2z \quad \rightarrow \quad j'\left(\frac{1}{2}\right) = -qu - \frac{q'u}{2} + \frac{q'u^2}{2} \\ j''(z) &= q'u^2 \end{aligned}$$

It follows that the second term in (30) is equal to :

$$\begin{aligned}
s_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} \left[\int_{y=1-u}^1 e^{-q(1-y) - \frac{q'}{2}y(1-y)} dy \right] du &\approx s_1 \int_{u=0}^1 e^{qu + \frac{q'}{2}u(1-u)} e^{-\frac{qu}{2} - \frac{q'u}{4} + \frac{q'u^2}{6}} \left(u + \frac{q^2}{24}u^3\right) du \\
&= s_1 \int_{u=0}^1 e^{\frac{qu}{2} + \frac{q'u}{4} - \frac{q'u^2}{3}} \left(u + \frac{q^2}{24}u^3\right) du \\
&= s_1 \int_{u=0}^1 \left(u + \frac{q^2}{24}u^3\right) e^{\frac{qu}{2} + q'\left(\frac{u}{4} - \frac{u^2}{3}\right)} du
\end{aligned}$$

To compute this latter expression, we apply relationships (4) and (6) of Annex 1 :

$$\begin{aligned}
s_1 \int_{u=0}^1 \left(u + \frac{q^2}{24}u^3\right) e^{\frac{qu}{2} + q'\left(\frac{u}{4} - \frac{u^2}{3}\right)} du &\approx \frac{s_1}{2} e^{\frac{q}{4} + \frac{q'}{24}} \left(1 - \frac{q}{12} - \frac{q'}{72} + \frac{q^2}{96} - \frac{q'}{36} + \frac{q^2}{48}\right) \\
&\approx \frac{s_1}{2} e^{\frac{q}{4}} \left(1 - \frac{q}{12} + \frac{q^2}{32}\right) \\
&\approx \frac{s_1}{2} e^{\frac{q}{3}} \left(1 + \frac{q^2}{36}\right)
\end{aligned}$$

with :

$$\begin{aligned}
j(u) &= \frac{qu}{2} + q'\left(\frac{u}{4} - \frac{u^2}{3}\right) \quad \rightarrow \quad j\left(\frac{1}{2}\right) = \frac{q}{4} + \frac{q'}{24} \\
j'(u) &= \frac{q}{2} + q'\left(\frac{1}{4} - \frac{2u}{3}\right) \quad \rightarrow \quad j'\left(\frac{1}{2}\right) = \frac{q}{2} - \frac{q'}{12} \\
j''(u) &= -\frac{2q'}{3}
\end{aligned}$$

ANNEXE 2

L'analyse démographique conjoncturelle

Reproduit de :Gérard Calot

L'analyse démographique conjoncturelle

in The joy of demography,

édité en l'honneur de Dirk J. van de Kaa,

par Anton Kuijsten, Henk de Gans et Henk de Feijter,

NethurD Publications, La Haye, 1999

pp. 295-323

L'analyse démographique conjoncturelle

1. Introduction

La production et la publication des statistiques démographiques courantes, aussi bien en matière de naissances que de mariages ou de décès, est généralement réalisée en plusieurs étapes.

Dans un premier temps, l'Office national de Statistique procède à un comptage des bulletins qu'il reçoit des bureaux locaux d'état civil. A un moment donné, une première estimation *provisoire*, généralement mensuelle, est publiée, qui se fonde sur un nombre de bulletins *déjà* reçus le plus souvent *incomplet* pour diverses sortes de raisons : par exemple parce que les bulletins reçus portent seulement sur une *partie du territoire* (tel est le cas lorsque la périodicité des transmissions de bulletins dépend de la taille de la municipalité) ou parce que le délai maximum légal entre le moment où se produit un événement d'état civil et celui de son enregistrement est de plusieurs mois. Des retards accidentels peuvent également affecter la transmission des bulletins de certains mois.

Dans tous les cas, il est souhaitable que l'Office de Statistique publie non seulement le nombre de bulletins qu'il a effectivement reçus à une date donnée pour un mois *m* donné, mais encore – et surtout – l'estimation (sans biais) qu'il en déduit du nombre *total* d'événements qui se sont produits ce même mois *m* dans l'*ensemble* du pays.

Le cas échéant, une deuxième estimation du nombre d'événements survenus le mois *m*, encore provisoire mais établie sur une base plus exhaustive, donc plus précise, est ultérieurement publiée.

Enfin, avec un décalage chronologique beaucoup plus grand, l'exploitation statistique complète des bulletins d'état civil de l'année conduira à des données *définitives*, en termes non seulement de période d'occurrence des événements mais surtout de *caractéristiques socio-démographiques* des personnes concernées par ces événements (sexe, âge, état matrimonial, ...).

C'est seulement lorsque ces données définitives sont disponibles qu'on peut établir différentes sortes d'indices élaborés permettant d'analyser le phénomène considéré. Un bon nombre de ces indices est obtenu par combinaison des données définitives de *flux* avec les évaluations de population résidente par sexe et âge : tel est le cas des *indicateurs conjoncturels* de *fécondité* ou de *primo-nuptialité* ou de l'*espérance de vie à la naissance*.

Le problème auquel est confronté l'analyste de la conjoncture démographique, dès qu'il dispose d'une estimation, fût-elle provisoire, d'un nombre *absolu*, mensuel ou annuel, d'événements est celui de la *conversion* de cette estimation en celle de l'indice élaboré correspondant. C'est à la solution de ce problème qu'est consacré le présent article.

2. L'effectif moyen des générations soumises au risque

Pour traiter cette question, nous supposons tout d'abord que l'événement étudié est *renouvelable* et nous prendrons l'exemple des *naissances*. Nous nous appuyerons sur le concept d'*effectif moyen des générations soumises au risque* que nous allons en premier lieu présenter.

Considérons, une année donnée, la répartition du nombre des naissances selon l'âge de la mère, en admettant que cet âge est défini comme étant celui atteint durant l'année civile de la naissance⁶³. Désignons l'année d'observation par n (indice supérieur), l'âge de la mère par i (indice inférieur), le nombre annuel de naissances de l'année n par N^n et le nombre d'enfants nés l'année n de mères d'âge i (ou, plus précisément, nées elles-mêmes l'année $n - i$) par N_i^n (événements situés dans un parallélogramme à côtés verticaux du diagramme de Lexis).

L'effectif des femmes nées l'année $n - i$ varie légèrement au cours de l'année n , sous l'effet de la mortalité et des migrations. Nous conviendrons d'en résumer la valeur pour l'ensemble de l'année n par la demi-somme de sa valeur au 1^{er} janvier, P_{i-1}^n (effectif d'âge révolu $i - 1$ au 1^{er} janvier n), et de celle, P_i^{n+1} , au 31 décembre :

$$F_i^n = \frac{P_{i-1}^n + P_i^{n+1}}{2}$$

Le taux de fécondité f_i^n à l'âge i atteint pendant l'année n est le rapport entre le nombre de naissances N_i^n et l'effectif correspondant soumis au risque F_i^n :

$$f_i^n = \frac{N_i^n}{F_i^n} = \frac{N_i^n}{(P_{i-1}^n + P_i^{n+1})/2}$$

L'indicateur conjoncturel de fécondité I^n de l'année n est la somme, étendue aux différents âges de la période féconde (15-49 ans, pour fixer les idées), des taux de fécondité par âge :

$$I^n = \sum_{i=15}^{49} f_i^n$$

Il s'ensuit que le rapport G^n du nombre annuel de naissances N^n à l'indicateur conjoncturel I^n est la moyenne pondérée des effectifs féminins F_i^n aux divers âges de fécondité, le coefficient de pondération de l'effectif d'âge i étant le taux de fécondité f_i^n à ce même âge i observé cette même année n :

$$G^n = \frac{N^n}{I^n} = \frac{\sum_{i=15}^{49} N_i^n}{\sum_{i=15}^{49} f_i^n} = \frac{\sum_{i=15}^{49} f_i^n F_i^n}{\sum_{i=15}^{49} f_i^n}$$

soit encore :

$$N^n = I^n \cdot G^n$$

⁶³ C'est seulement pour faciliter l'exposé que nous supposons que l'âge de la mère est celui atteint durant l'année civile de la naissance. Si l'âge de la mère est celui en années révolues au moment de la naissance, on aboutit à des résultats équivalents : le taux de fécondité f_i^n à l'âge i pour l'année n est alors le rapport entre le nombre N_i^n d'événements observés dans le carré du diagramme de Lexis et le nombre de femmes-années d'exposition au risque F_i^n , estimé par $(P_i^n + P_i^{n+1})/2$ en admettant que la densité des lignes de vie au sein des deux générations annuelles concernées est uniforme :

$$f_i^n = \frac{N_i^n}{(P_i^n + P_i^{n+1})/2}$$

Le nombre annuel N^n de naissances de l'année n apparaît ainsi comme le *produit* de l'indicateur conjoncturel I^n , résumé des comportements de fécondité *propres* à l'année n , par l'effectif moyen pondéré G^n des générations féminines qui, au cours de l'année n , appartiennent aux âges féconds. Cet effectif moyen G^n est l'*héritage* de la natalité des années de 15 à 49 ans antérieures à l'année n considérée, corrigé par la mortalité et les migrations intervenues depuis l'époque de la naissance.

En d'autres termes, le nombre de naissances de l'année n est la résultante *multiplicative* d'une *intensité*, caractéristique des comportements de fécondité de l'année n elle-même, intensité que mesure l'indicateur conjoncturel, et d'un *effectif*, hérité du passé, égal à la moyenne *pondérée* des effectifs féminins F_i^n de l'année considérée selon l'âge, le poids de l'effectif d'âge i étant le taux de fécondité f_i^n correspondant à cet âge et à cette année.

En toute rigueur, l'effectif moyen G^n , que nous qualifions d'*héritage* du passé, dépend un peu des comportements de fécondité de l'année n puisque les coefficients de pondération des effectifs F_i^n sont les taux de fécondité f_i^n de l'année n elle-même. Mais on sait que la moyenne pondérée d'éléments pas trop variables (les effectifs F_i^n) dépend assez peu des coefficients de pondération (les taux f_i^n) : si on modifie légèrement ceux-ci, on ne modifie guère la valeur de la moyenne pondérée. En particulier, lorsque les effectifs F_i^n de l'année n ne dépendent pas de l'âge i , c'est-à-dire lorsque les effectifs féminins aux divers âges de fécondité sont de *même taille*, l'effectif moyen pondéré coïncide avec la valeur commune des F_i^n , *quels que soient* les coefficients de pondération. Aussi ne commet-on pas grande erreur en raisonnant comme si l'effectif moyen G^n dépendait *exclusivement* des effectifs F_i^n et *aucunement* des taux f_i^n .

On aura une idée de l'effet des changements de coefficients de pondération en considérant la figure 1 qui décrit, pour la France, l'évolution depuis 1946 de l'effectif moyen des générations féminines en âge de fécondité, établi respectivement en utilisant pour coefficients de pondération les taux de l'année n elle-même (valeur exacte) et les taux d'une année *fixe* (successivement : 1950, 1960, 1970, 1980 et 1990). C'est dans la période 1965-1985, époque où sont parvenues aux âges de fécondité les générations du *baby-boom* et où, par conséquent, les effectifs des divers âges féconds étaient le plus *inégaux*, que l'effet est maximum. L'effet est sensiblement nul vers 1990 parce que les générations féminines d'âge fécond sont, à cette date, de taille très voisine.

Cas particulier :

mortalité et migrations nulles avant 50 ans et calendrier transversal de fécondité invariable

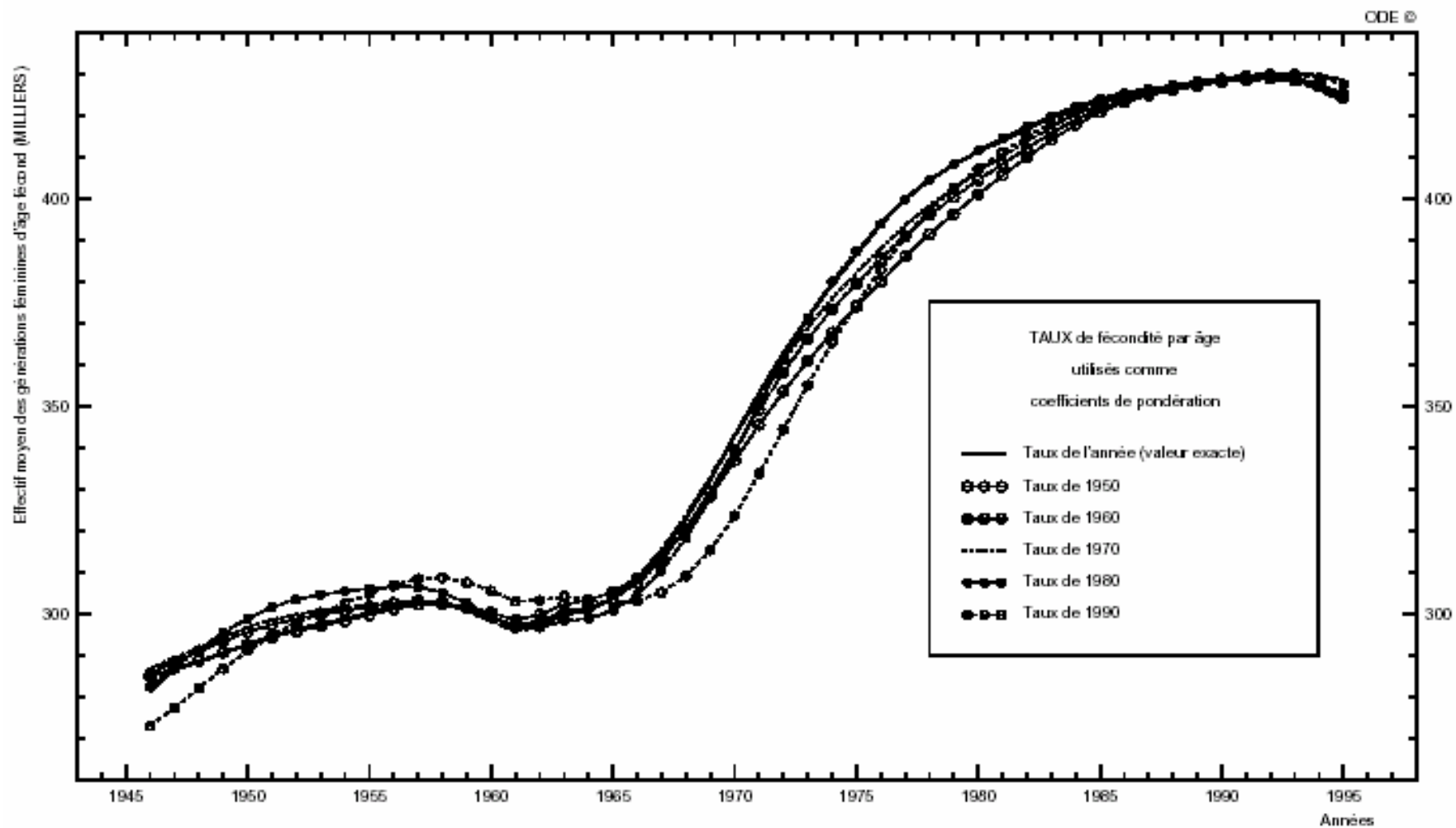
Supposons que la mortalité et les migrations sont *nulles* entre la naissance et la fin de la période féconde. Alors, le nombre F_i^n de femmes d'âge i l'année n est égal au nombre de naissances *féminines* de l'année $n - i$, soit encore φN^{n-i} , φ étant la proportion de filles parmi les nouveau-nés.

Supposons, de plus, que le *calendrier* de la fécondité transversale est *invariable* d'une année à l'autre :

$$\frac{f_i^n}{I^n} \text{ dépend de } i \text{ mais non de } n, \text{ soit } \frac{f_i^n}{I^n} = a_i, \text{ avec } \sum_{i=15}^{49} a_i = 1$$

et désignons par \bar{x} l'*âge moyen* constant à la maternité auquel correspond ce calendrier invariable :

Figure 1 FRANCE, 1946-1995
 Effectif moyen des générations féminines d'âge fécond
 Valeur exacte et valeurs estimées sur la base des taux d'une année donnée



$$\bar{x} = \sum_{i=15}^{49} a_i i$$

Sous ces deux hypothèses, la série G^n de l'effectif moyen des générations féminines d'âge fécond est, au coefficient ϕ près, la *moyenne mobile pondérée* de la série des naissances annuelles, *décalée* de l'âge moyen constant à la maternité \bar{x} :

$$G^n = j \sum_{i=15}^{49} a_i N^{n-i} = j \sum_{k=n-49}^{n-15} a_{n-k} N^k$$

$$n - \bar{x} = \sum_{i=15}^{49} a_i (n-i) = \sum_{k=n-49}^{n-15} a_{n-k} k$$

Ces deux hypothèses simplificatrices ne sont jamais très éloignées de la réalité, ce qui explique que l'évolution de la courbe G^n est généralement très *lisse*, comme l'est celle d'une moyenne mobile sur un assez grand nombre de termes (on a ici 35 termes : de 15 à 49 ans). Ce caractère lisse de l'évolution de G^n va faciliter les interpolations et les extrapolations⁶⁴.

Comme on le constate sur la figure 2, qui décrit l'évolution observée depuis la Seconde guerre mondiale dans six pays européens, les variations de l'effectif moyen G^n des générations féminines d'âge fécond sont effectivement régulières.

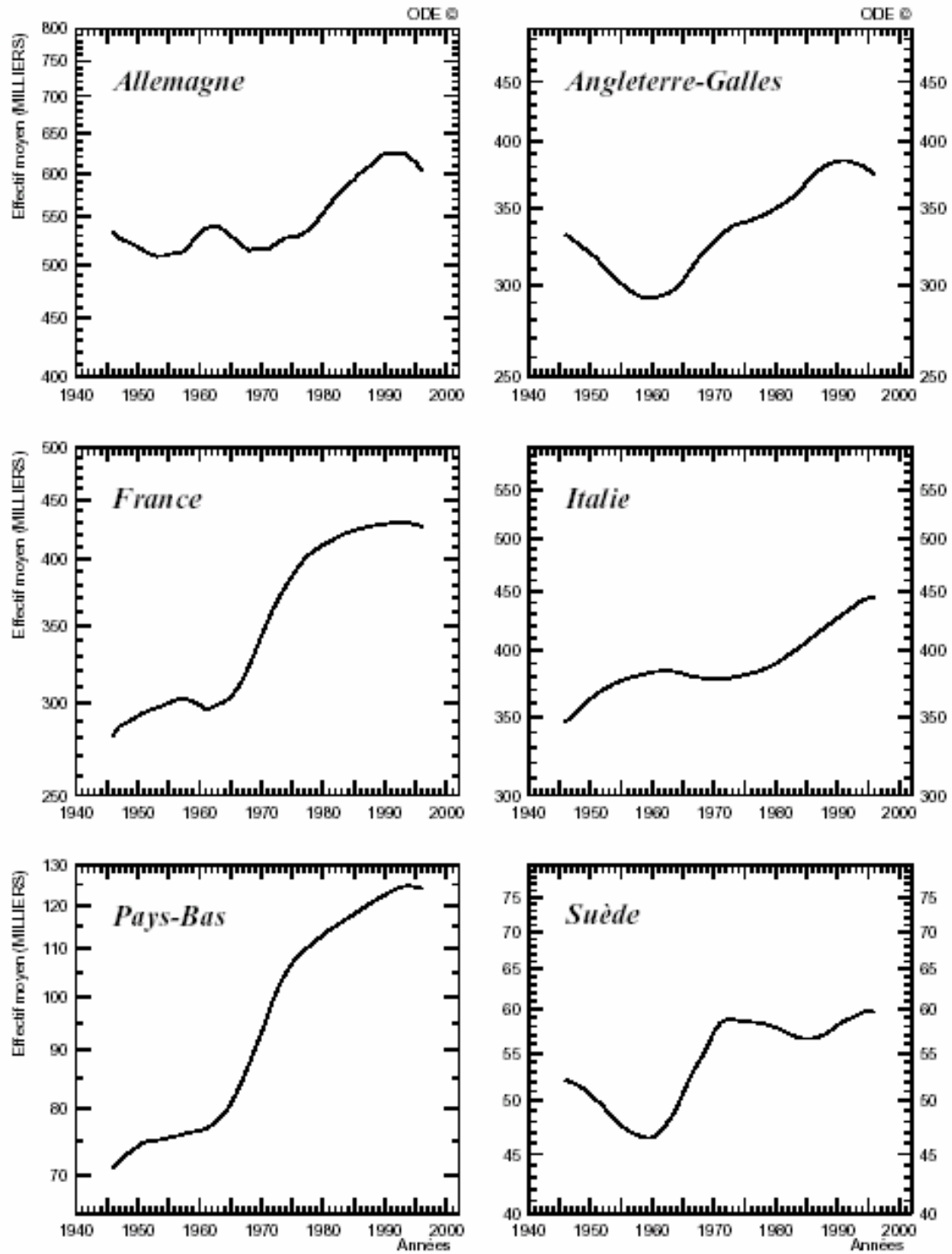
On notera l'effet, variable selon les pays – en ampleur et, dans une moindre mesure, en calendrier –, qu'a eu le *baby-boom* sur l'augmentation de cet effectif moyen. Par ailleurs, dans la plupart des pays considérés, l'effectif moyen a commencé à décroître vers 1990, en écho à la chute du nombre absolu des naissances à partir des années 1965-1970 : si, depuis vingt-cinq ans, le nombre absolu des naissances en Europe a décliné moins rapidement, en valeur relative, que l'indicateur conjoncturel de fécondité, dans les années futures, la diminution de l'effectif moyen des générations féminines d'âge fécond pèsera, à la baisse, sur l'évolution du nombre absolu des naissances.

Définition générale de l'effectif moyen des générations soumises au risque

D'une façon générale, considérons un flux annuel N^n d'événements renouvelables, classés selon l'âge i de la personne qui le subit, et définissons le taux f_i^n à l'âge i comme le rapport du nombre N_i^n d'événements au nombre de personnes-années F_i^n d'exposition au risque à l'âge i durant l'année n au sein de la population *totale*. Alors, le flux annuel N^n est le produit de la somme des taux f_i^n par la moyenne pondérée des nombres F_i^n , c'est-à-dire par l'*effectif moyen* G^n des générations soumises au risque. Observons que si l'événement considéré n'est pas renouvelable (ainsi, le *premier* mariage) mais traité comme renouvelable, les taux f_i^n sont les taux dits de *seconde catégorie*, qui utilisent comme dénominateur l'effectif *total* de la population résidente d'âge i et non l'effectif des *seules* personnes (dans l'exemple du premier mariage : les *célibataires*) qui n'ont pas enregistré l'événement et qui sont pourtant les seules à être effectivement soumises au risque.

⁶⁴ En matière de *projection* de population, on peut s'appuyer sur le caractère lisse de l'évolution du rapport entre le nombre absolu projeté de naissances et l'indicateur conjoncturel de fécondité également projeté, pour détecter d'éventuelles erreurs de calcul.

Figure 2
EFFECTIF MOYEN des GENERATIONS FEMININES
en AGE de FECONDITE dans six pays européens
Echelles des ordonnées logarithmiques de même module



3. L'effectif moyen des générations en âge de primo-nuptialité

En matière de primo-nuptialité, c'est-à-dire de mariages de célibataires, on peut déterminer, de façon analogue au cas de la fécondité, l'effectif moyen des générations masculines et l'effectif moyen des générations féminines en âge de primo-nuptialité, en traitant le premier mariage comme un événement renouvelable.

Ces deux effectifs moyens n'évoluent pas de manière rigoureusement simultanée (figure 3), bien que les variations des effectifs de population masculine et féminine soient généralement concomitantes et de même ampleur : la raison tient au fait que l'âge moyen au premier mariage des femmes est plus *précoce* (de l'ordre de deux à trois années) que celui des hommes. Il en résulte qu'une augmentation de natalité provoque, une vingtaine d'années après, une augmentation de l'effectif féminin moyen en âge de primo-nuptialité deux à trois ans *plus tôt* que celle de l'effectif masculin. La présence temporaire d'un plus grand nombre de candidates au premier mariage que de candidats a pour effet d'abaisser l'indicateur conjoncturel de primo-nuptialité féminine et de relever son homologue masculin (figure 4), la situation inverse produisant un effet inverse.

Sur la figure 4, on notera le cas particulier de l'Allemagne, où l'augmentation de la natalité de la période 1934-1944 a provoqué, vers 1960, une augmentation de l'effectif féminin en âge de premier mariage quelques années plus tôt que celle de l'effectif masculin. En outre, dans ce pays, les pertes militaires de la Seconde guerre mondiale sont à l'origine de l'excédent considérable de candidates au premier mariage de 1945 à 1955.

Les variations temporelles de l'effectif moyen des générations en âge de primo-nuptialité font ainsi apparaître les « tensions » qui se manifestent sur le « marché matrimonial » du fait des évolutions non rigoureusement parallèles des effectifs de l'un et l'autre sexe.

4. L'interpolation mensuelle de l'effectif moyen annuel

L'interpolation, à l'échelle mensuelle, de l'effectif moyen annuel des générations soumises au risque est facilitée par le caractère *lisse* de ce dernier. En convenant que le douzième de la valeur annuelle est la valeur mensuelle *typique* de l'année, qui se situe à mi-chemin entre juin et juillet, on peut définir une courbe *régulière* passant par ces valeurs typiques, puis lire les valeurs de chaque mois sur la courbe régulière ainsi déterminée.

C'est ce qui a été réalisé sur la figure 5 qui se rapporte à l'effectif moyen des générations féminines de la France en âge de fécondité. On a ajusté une courbe polynômiale de degré 5 sur six points typiques consécutifs (juin-juillet des années $n+1$ à $n+6$) et, pour les douze mois de la période *centrale* (qui va de juillet $n+3$ à juin $n+4$), on a retenu les valeurs mensuelles lues sur cette courbe polynômiale ajustée⁶⁵.

Grâce au degré élevé des polynômes utilisés, la courbe mensuelle ajustée passe exactement par les valeurs annuelles typiques observées et les raccords d'une période centrale à la suivante se font *sans* discontinuité, en termes aussi bien de valeurs que de dérivées d'ordre 1 ou 2. C'est seulement à chacune des *extrémités* de la période d'étude qu'on retient *aussi*, au moins *provisoirement*, les valeurs mensuelles lues sur la courbe

⁶⁵ De façon précise, considérons six années consécutives, soit une période de 72 mois. Prenons pour date-origine le 1er janvier de la troisième année et adoptons le mois comme unité de durée. Les valeurs typiques des six années se rapportent aux dates -30, -18, -6, 6, 18, 30 et les milieux des mois de la période centrale se situent aux dates -5,5 (juillet de la troisième année), -4,5, ..., 4,5, 5,5 (juin de la quatrième année). L'ajustement polynômial consiste à déterminer la courbe de degré 5 qui passe par les six points d'abscisses ± 6 , ± 18 , ± 30 et à retenir les douze valeurs correspondant aux abscisses $\pm 0,5$, $\pm 1,5$, ..., $\pm 5,5$.

polynômiale ajustée, pour les mois respectivement *antérieurs* à la *première* période centrale et *postérieurs* à la *dernière* période centrale. Lorsqu'on disposera ultérieurement d'informations supplémentaires, on *révisera* les valeurs correspondant aux nouvelles périodes centrales. Par ailleurs, de façon à améliorer la qualité de l'ajustement pour les mois du passé récent ou du futur proche, on peut procéder, préalablement à l'interpolation mensuelle, à une *extrapolation* des valeurs annuelles.

5. L'extrapolation de l'effectif moyen annuel

A un moment donné, désignons par a l'année *la plus récente* pour laquelle les taux de fécondité par âge f_i^n sont disponibles et par b (avec généralement $b \geq a+1$) l'année *la plus récente* pour laquelle on connaît les effectifs féminins au 1er janvier P_i^n par âge révolu.

Pour les années a et avant, on connaît le nombre absolu d'événements N^n , la somme I^n des taux f_i^n et donc le rapport $G^n = N^n / I^n$. Comment estimer l'effectif moyen G^n pour les années n postérieures⁶⁶ à a , en supposant que le décalage de $n-a$ années n'est pas trop grand (disons $n-a$ au plus égal à 5 ou 10 ans) ?

Une première méthode, purement graphique, consiste en une extrapolation *manuelle* à l'année n de la courbe *lisse* G^k connue jusqu'à l'année $k = a$.

On peut aussi procéder par le calcul et convenir de remplacer, dans l'expression de G^n :

$$G^n = \frac{\sum_{i=15}^{49} f_i^n F_i^n}{\sum_{i=15}^{49} f_i^n} = \frac{\sum_{i=15}^{49} f_i^n \frac{P_{i-1}^n + P_i^{n+1}}{2}}{\sum_{i=15}^{49} f_i^n}$$

le taux non encore observé f_i^n par le *dernier* observé *au même âge* (en notant que ce taux est *obsolète* de $n-a$ années), soit f_i^a , et l'effectif non encore observé F_i^n par le *dernier* observé *pour la même génération*, soit $P_{i-1-(n-b)}^b$ (cet effectif est *obsolète* de $n-b+1/2$ années) :

$$\mathcal{G}^n = \frac{\sum_{i=15}^{49} f_i^a P_{i-1-(n-b)}^b}{\sum_{i=15}^{49} f_i^a}$$

Le rapport de la valeur *estimée*, résultant de l'application de ce type de formule, à la valeur *exacte* peut être mesuré pour une année k quelconque, *antérieure* ou égale à $n-a$, sur la base de taux obsolètes de $n-a$ années et d'effectifs obsolètes de $n-b+1/2$ années :

$$\frac{\mathcal{G}^k}{G^k} = \frac{\sum_{i=15}^{49} f_i^{k-(n-a)} P_{i-1-(n-b)}^{k-(n-b)}}{\sum_{i=15}^{49} f_i^{k-(n-a)}} / \frac{\sum_{i=15}^{49} f_i^k \frac{P_{i-1}^k + P_i^{k+1}}{2}}{\sum_{i=15}^{49} f_i^k}$$

⁶⁶ Nous traitons ici de l'extrapolation (vers le *futur*) de l'effectif moyen G^n . Le problème serait très voisin si on voulait rétropolier (vers le *passé*) cet effectif moyen.

Figure 3
EFFECTIF MOYEN des GENERATIONS MASCULINES et FEMININES
en AGE de PREMIER MARIAGE dans six pays européens
Echelles des ordonnées logarithmiques de même module

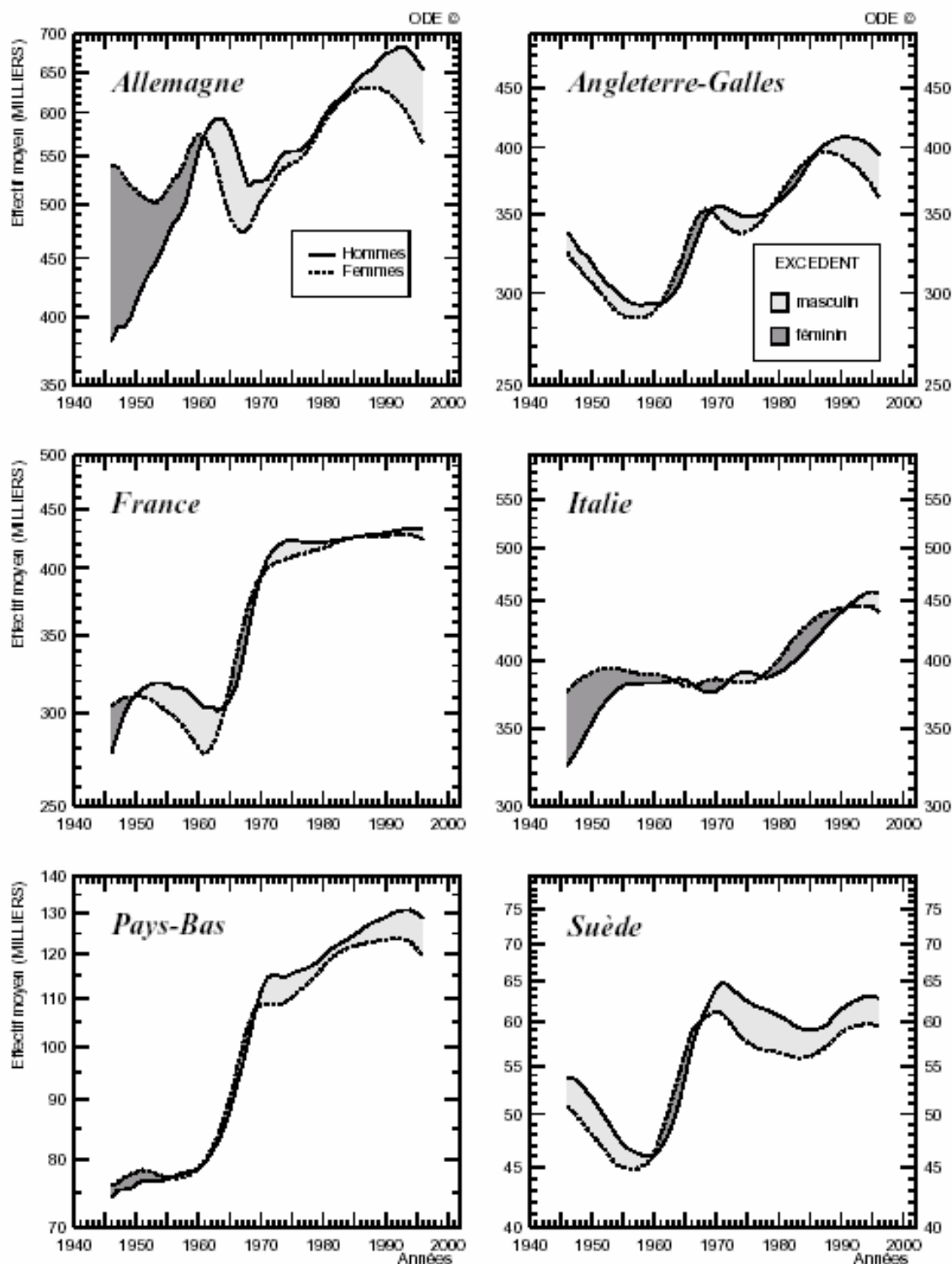


Figure 4 RAPPORT :

(1) en terme d'INDICATEUR CONJONCTUREL de PRIMO-NUPTIALITE de l'indicateur FEMININ à l'indicateur MASCULIN

(2) en terme d'EFFECTIF MOYEN des GENERATIONS en AGE de PREMIER MARIAGE de l'effectif MASCULIN à l'effectif FEMININ

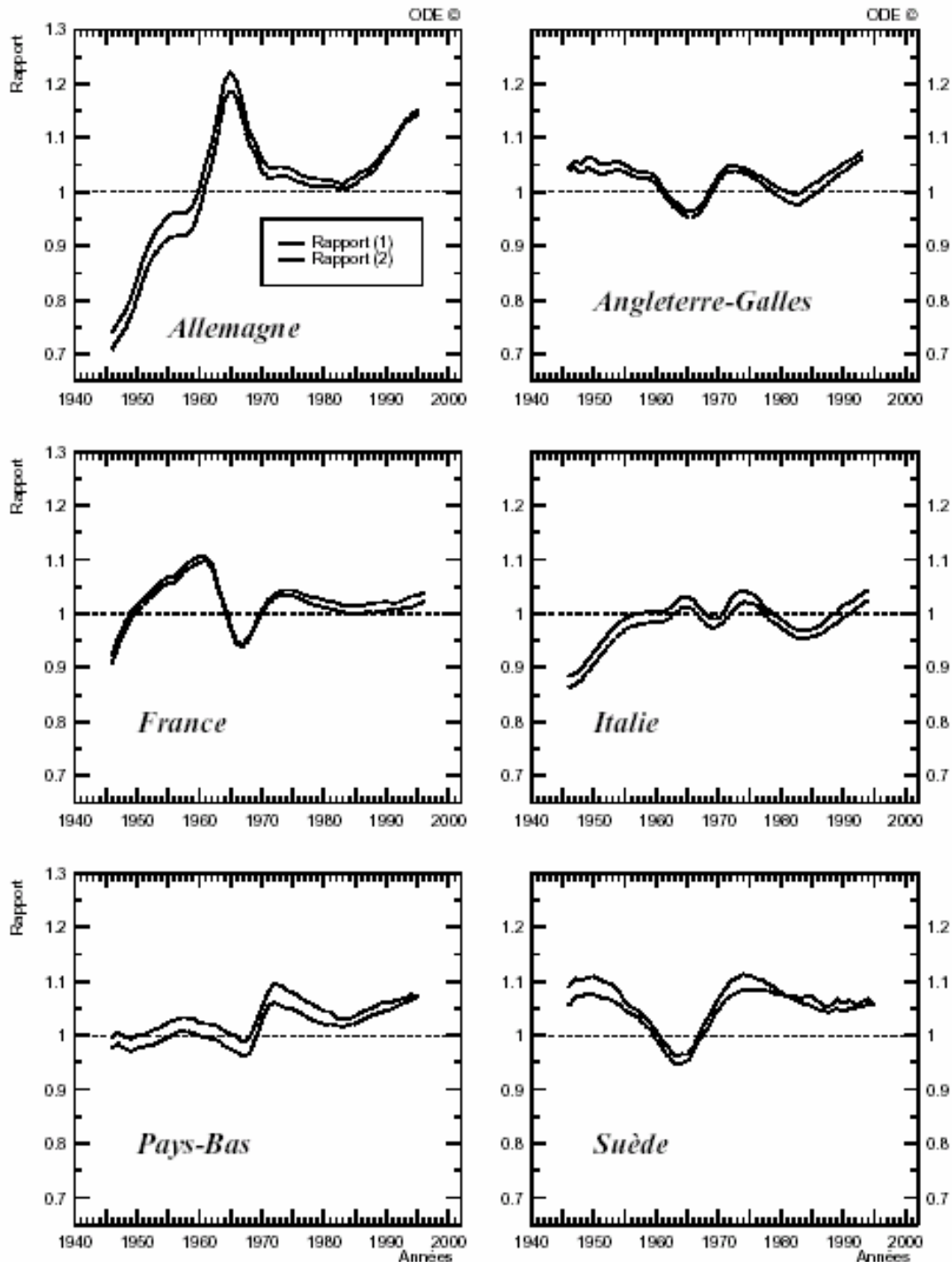
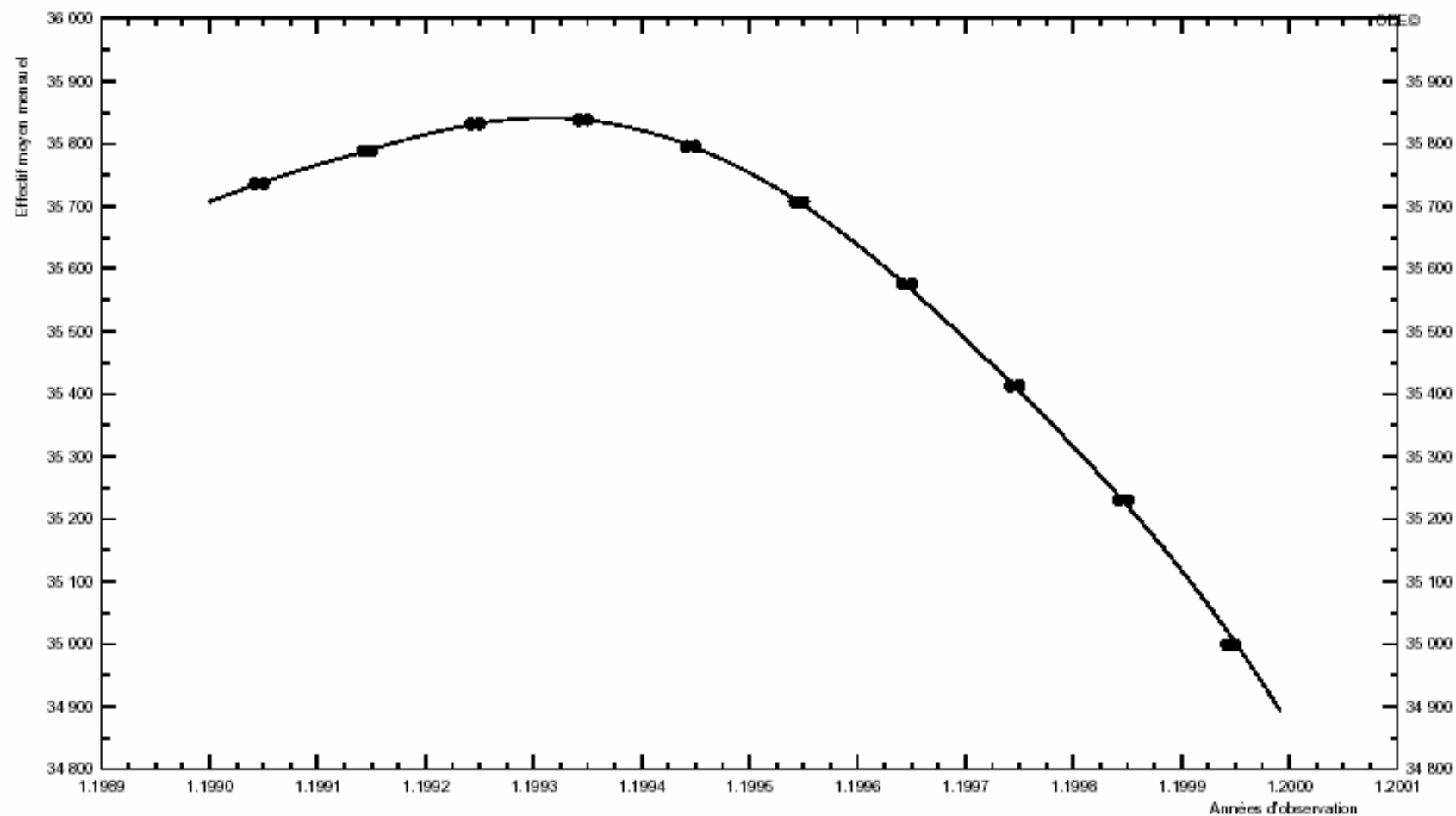


Figure 5 FRANCE, 1990-1999. EFFECTIF MOYEN des GENERATIONS FEMININES en AGE de FECONDITE
 Valeurs annuelles divisées par 12 et rapportées à JUIN et JUILLET,
 Valeurs MENSUELLES INTERPOLEES par polynômes de degré 5 sur 6 points consécutifs
 dont on retient l'année centrale



En extrapolant (par un procédé quelconque) à l'année n ce rapport, qui est connu jusqu'à l'année $k = n - a$, puis en divisant l'estimation \mathcal{G}^n par la valeur extrapolée à l'année n de ce rapport, on obtient une estimation améliorée de G^n .

L'extrapolation du rapport repose implicitement sur l'hypothèse de la stabilité des migrations et de la mortalité et, dans une moindre mesure, sur celle de la régularité de l'évolution des taux à âge égal.

Si on dispose d'une *projection* de population, au moins à l'horizon du 1er janvier $n+1$, on peut aussi utiliser les effectifs projetés P_{i-1}^n et P_i^{n+1} , c'est-à-dire estimer G^n par :

$$\mathcal{G}^n = \frac{\sum_{i=15}^{49} f_i^a \frac{P_{i-1}^n + P_i^{n+1}}{2}}{\sum_{i=15}^{49} f_i^a}$$

estimation qui peut à son tour être améliorée moyennant extrapolation, à l'année n , du rapport connu jusqu'à l'année $k = n - a$:

$$\frac{\mathcal{G}^k}{G^k} = \frac{\sum_{i=15}^{49} f_i^{k-(n-a)} \frac{P_{i-1}^k + P_i^{k+1}}{2}}{\sum_{i=15}^{49} f_i^{k-(n-a)}} / \frac{\sum_{i=15}^{49} f_i^m \frac{P_{i-1}^k + P_i^{k+1}}{2}}{\sum_{i=15}^{49} f_i^k}$$

6. L'estimation des indicateurs conjoncturels annuels et mensuels : comment convertir un nombre absolu d'événements en indicateur ?

Dès qu'on dispose d'une évaluation, même provisoire, d'un *nombre absolu* d'événements, on peut estimer l'indicateur conjoncturel qui lui correspond en divisant ce nombre absolu par la valeur de l'effectif moyen des générations soumises au risque. Ceci vaut aussi bien à l'échelle annuelle qu'à l'échelle mensuelle.

Toutefois, à l'échelle mensuelle, une opération préalable est nécessaire. Il convient en effet de corriger le nombre absolu observé de deux phénomènes perturbateurs : la *composition en jours* du mois (nombre de jours et, le cas échéant, nombres de lundis, de mardis, ..., de dimanches, si le phénomène étudié est soumis à une fluctuation hebdomadaire importante, comme il en va notamment en matière de mariages) et les *variations saisonnières* mensuelles.

On trouvera en annexe, deux tableaux donnant, pour la France, les résultats les plus récents dont on dispose à la date où nous écrivons (février 1998) sur la fécondité et la primo-nuptialité.

Dans les figures 6 et 7, on a représenté l'évolution mensuelle des indicateurs conjoncturels de fécondité et de primo-nuptialité en France, corrigés de la composition journalière du mois et des variations saisonnières. On a indiqué sur ces mêmes figures l'évolution des indicateurs conjoncturels *lissés* obtenus par application d'une formule de lissage, due à Jan Hoem (Université de Stockholm), qui fournit une valeur lissée jusqu'au dernier mois d'observation.

Pour estimer sur *longue période* l'évolution de l'effectif moyen des générations féminines d'âge fécond et par conséquent celle de l'indicateur conjoncturel de fécondité, il est *nécessaire* de disposer, *chaque année*, des effectifs de la population féminine par âge et il est *souhaitable* de disposer de taux de fécondité par âge qui ne soient pas trop obsolètes.

Cependant, même si les taux disponibles sont assez largement obsolètes, on peut quand même les utiliser, à moins qu'ils ne se rapportent à une année exceptionnelle ou qu'on veuille estimer l'indicateur conjoncturel d'une année elle-même exceptionnelle. La qualité de l'estimation obtenue pourra être appréciée en comparant, pour les années dont les taux par âge sont disponibles, l'indicateur conjoncturel estimé et l'indicateur conjoncturel observé.

On trouvera représentée dans la figure 8 l'évolution de l'indicateur conjoncturel de fécondité en Suisse, estimé à partir des nombres absolus de naissances totales et des effectifs féminins par âge au 1er janvier de chaque année depuis 1861, le jeu de taux par âge retenu étant invariablement celui observé en 1932. A partir de 1932, on dispose de la série annuelle des naissances par année d'âge de la mère, ce qui permet de calculer la valeur exacte de l'indicateur conjoncturel. On constate ainsi que, durant la période de 65 ans considérée (1932-1996), l'erreur maximale commise en utilisant le calendrier transversal de la fécondité de 1932 atteint 0,08 enfant pour une femme en 1968, l'erreur n'excédant 0,03 enfant pour une femme que de 1962 à 1975 et 0,05 enfant pour une femme que de 1964 à 1972. Rappelons que les décennies 1960 et 1970 correspondent, en Suisse comme dans le reste de l'Europe, à une époque où le calendrier de la fécondité était spécialement *précoce*, donc assez différent de celui de 1932 mais surtout où les générations en âge de fécondité étaient spécialement *inégaies* du fait de l'arrivée progressive à l'âge de la maternité des générations du *baby-boom*.

7. L'indicateur conjoncturel mensuel de primo-nuptialité

Les nombres mensuels d'événements dont on dispose en matière de *mariages* se rapportent généralement à l'*ensemble* des mariages (quels que soient les âges des époux et quels que soient leurs états matrimoniaux antérieurement au mariage), tandis que la *primo-nuptialité*, par exemple masculine, ne concerne par convention que les *premiers* mariages et, au surplus, d'hommes qui avaient *moins de 50 ans révolus* au moment de leur mariage.

Aussi, lorsqu'on dispose d'un nombre total de mariages, est-il nécessaire d'estimer, selon le sexe, le nombre de *premiers mariages avant 50 ans* qui lui correspond. Ceci peut se faire moyennant extrapolation, et interpolation si on travaille à l'échelle mensuelle, de la série annuelle observée du rapport entre le nombre de premiers mariages avant 50 ans et le nombre de mariages totaux. Ces extrapolations et interpolations peuvent être réalisées de la même façon que les opérations analogues effectuées sur l'effectif moyen des générations soumises au risque.

8. L'indicateur conjoncturel mensuel de mortalité

L'indicateur conjoncturel mensuel de *fécondité*, pour le mois m de l'année n , s'obtient en divisant le nombre absolu mensuel NM^m de naissances observé le mois m , préalablement corrigé en NM^{m*} pour tenir compte de la composition journalière du mois et des variations saisonnières, par l'effectif moyen des générations féminines d'âge fécond établi pour le même mois m . C'est aussi l'indicateur conjoncturel qu'on *aurait* obtenu pour l'ensemble de l'année n si les effectifs de naissances selon l'âge de la mère avaient été

égaux à $\frac{12NM^{m*}}{N^n} N_i^n$ au lieu de N_i^n .

Figure 6 FRANCE, 1977-1997
Evolution de l'INDICATEUR CONJONCTUREL mensuel de FECONDITE
Série corrigée des variations saisonnières
et série lissée par moyenne mobile Hoem sur 25 termes

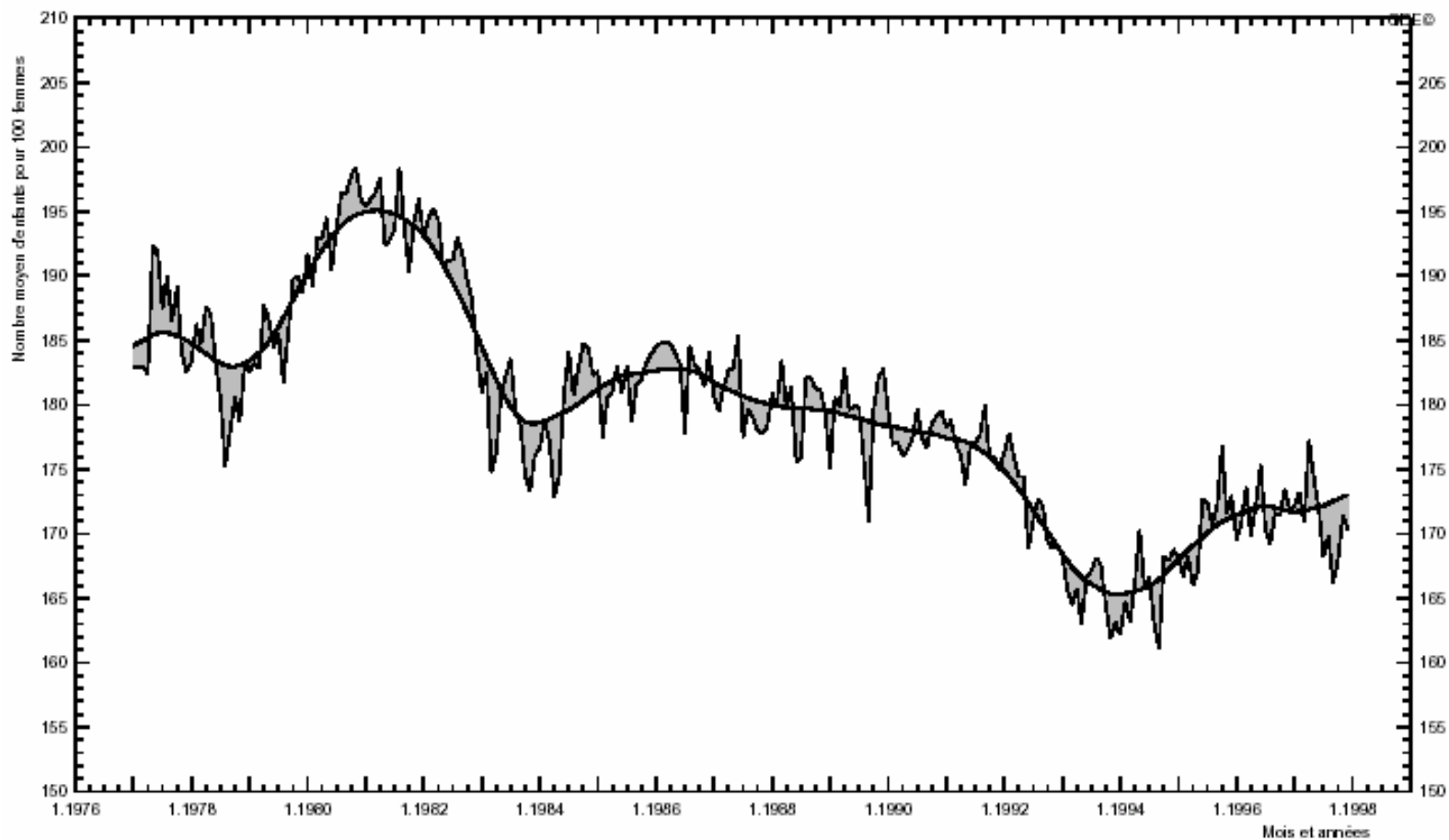
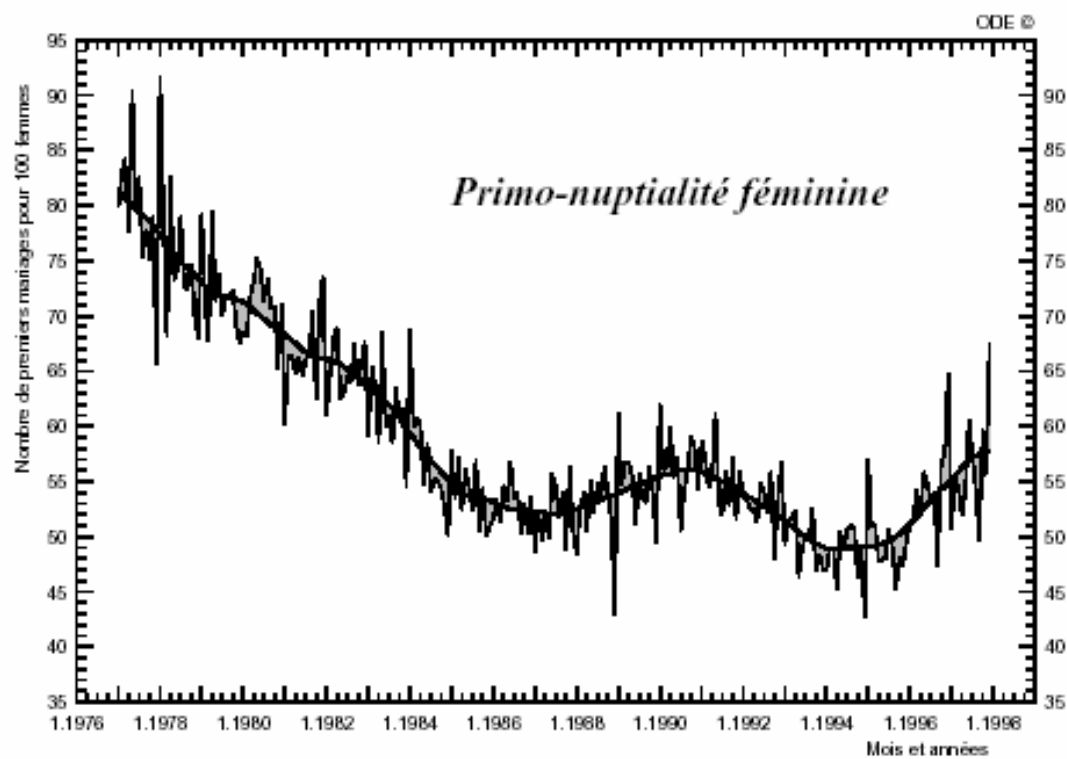
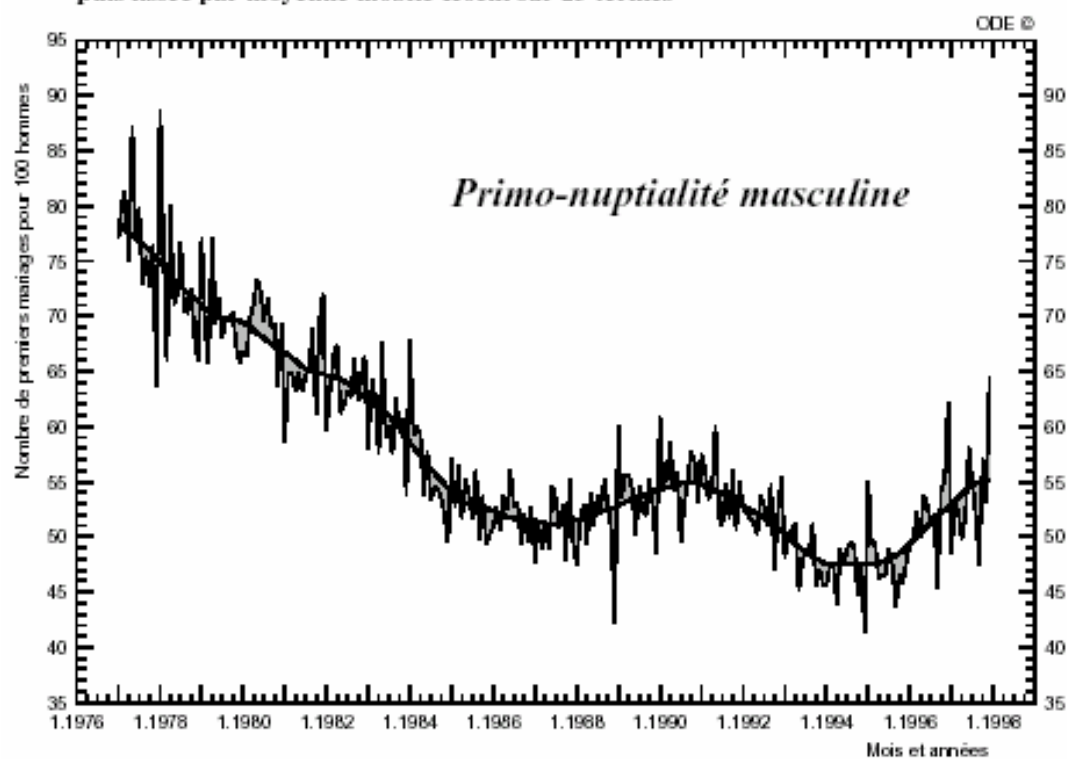


Figure 7 FRANCE, 1977-1997
Evolution de l'INDICATEUR CONJONCTUREL mensuel
de PRIMO-NUPTIALITE MASCULINE et FEMININE
 Série corrigée des variations saisonnières,
 puis lissée par moyenne mobile Hoem sur 25 termes



De façon analogue, on peut convenir de construire un indicateur conjoncturel mensuel de *mortalité*, exprimé en termes d'espérance de vie à la naissance masculine ou féminine, en déterminant l'espérance de vie qu'on *aurait* obtenue pour l'ensemble de l'année n si, au lieu des nombres de décès par sexe et âge réellement observés D_i^n , on avait enregistré ces mêmes nombres *multipliés* par $\frac{12DM^{m*}}{D^n}$, expression où DM^{m*} désigne le nombre mensuel de décès du mois m corrigé de la composition journalière du mois et des variations saisonnières. On trouvera représentées dans la figure 9 les évolutions des indicateurs conjoncturels mensuels de mortalité masculine et féminine en France depuis vingt ans.

Du fait que le mouvement saisonnier des décès n'est pas indépendant du sexe et surtout de l'âge, l'indicateur conjoncturel mensuel de mortalité ainsi défini diffère de celui qu'on aurait établi si on avait disposé des nombres *mensuels* de décès par sexe et âge et construit une table de mortalité *mensuelle*. Il fournit cependant une description de l'évolution mensuelle qui est *cohérente* avec l'évolution de l'indicateur annuel (la moyenne des douze indicateurs mensuels est sensiblement l'indicateur annuel) et qui reproduit les variations conjoncturelles du nombre absolu mensuel. En particulier, les mois marqués par une épidémie de grippe, qui correspondent à un indicateur mensuel relativement *faible*, apparaissent avec netteté.

9. Conversion d'un nombre annuel de décès en espérance de vie à la naissance

On a vu plus haut la manière de convertir un nombre absolu annuel de *naissances* ou de *mariages* en l'indicateur conjoncturel correspondant (indicateurs conjoncturels de fécondité et de primo-nuptialité masculine et féminine) : on divise le nombre absolu d'événements, qu'on a préalablement exprimé en nombre de premiers mariages avant 50 ans dans le cas de la primo-nuptialité, par l'estimation de l'effectif moyen des générations soumises au risque.

La même question se pose de convertir un nombre annuel de *décès*, portant généralement sur l'*ensemble* des *deux* sexes, en les espérances de vie, masculine et féminine, à la naissance. Voici un procédé permettant d'opérer cette conversion.

Soit a l'année la plus récente pour laquelle on dispose de la table de mortalité par sexe et âge, table dont les espérances de vie à la naissance, masculine, féminine et deux sexes, sont désignées respectivement par $Ev0H(a)$, $Ev0F(a)$ et $Ev0(a)$.

Soit, de même, b (avec $b \geq a+1$) l'année la plus récente pour laquelle on dispose des effectifs de population par sexe et âge au 1er janvier, c (souvent $c = a$) l'année la plus récente pour laquelle on dispose de la table de fécondité par âge et m (souvent $m = a$) l'année la plus récente pour laquelle on dispose des soldes migratoires par sexe et âge. Désignons par D^n l'évaluation du nombre absolu annuel de décès dont on dispose pour l'année n ($n > a$).

Figure 8 SUISSE, 1861-1996
Evolution de l'INDICATEUR CONJONCTUREL de FECONDITE
Données observées (1932-1996)
Données estimées (1861-1996) sur la base du calendrier de la fécondité observé en 1932

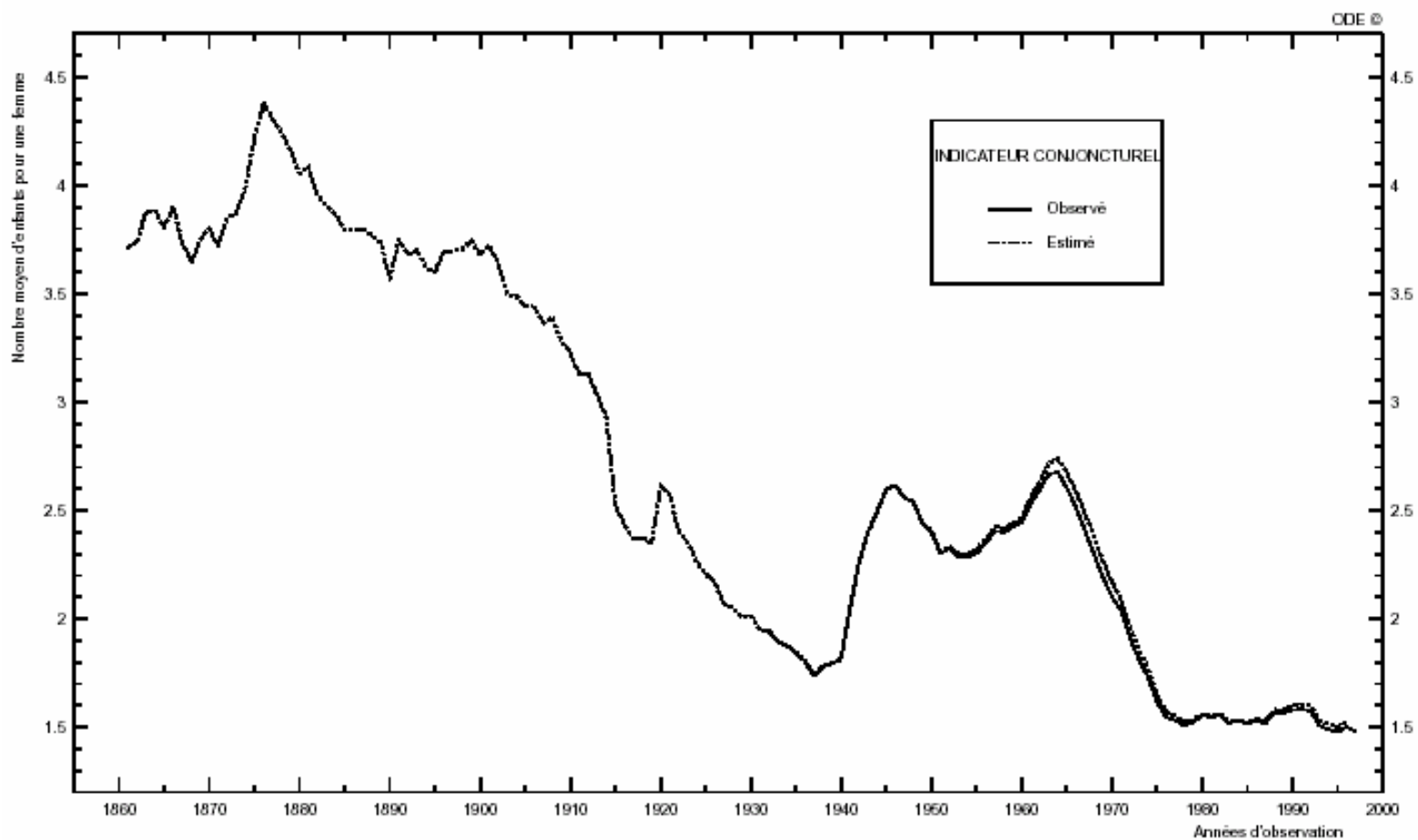
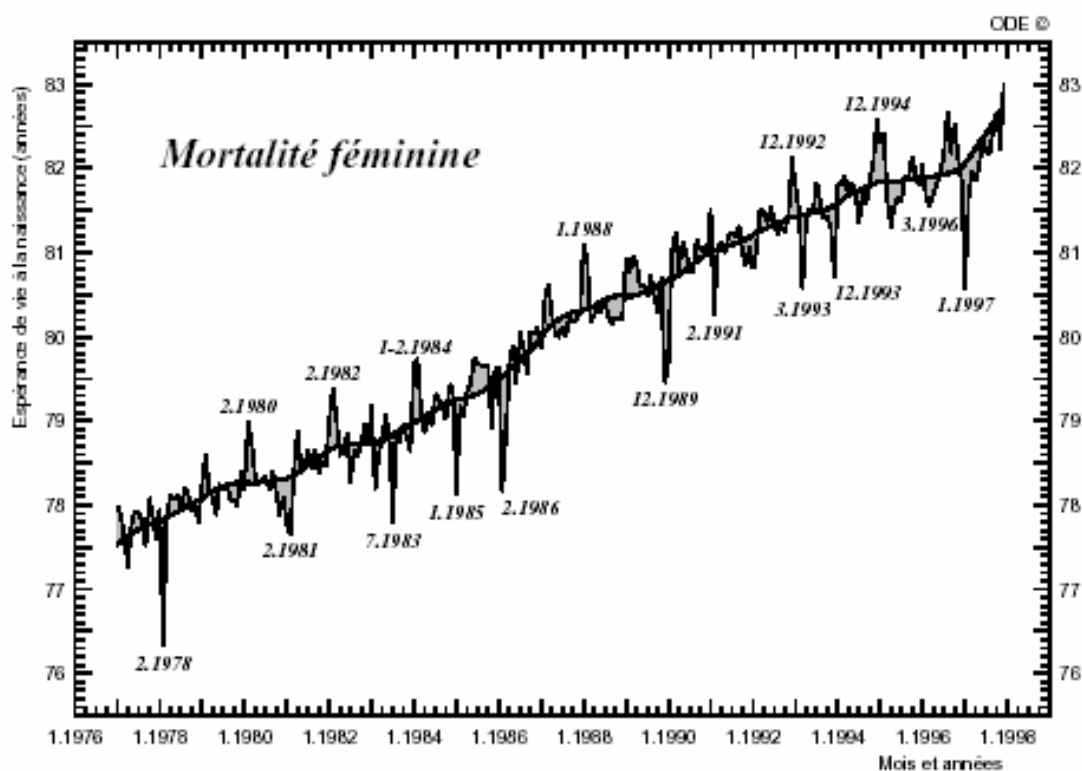
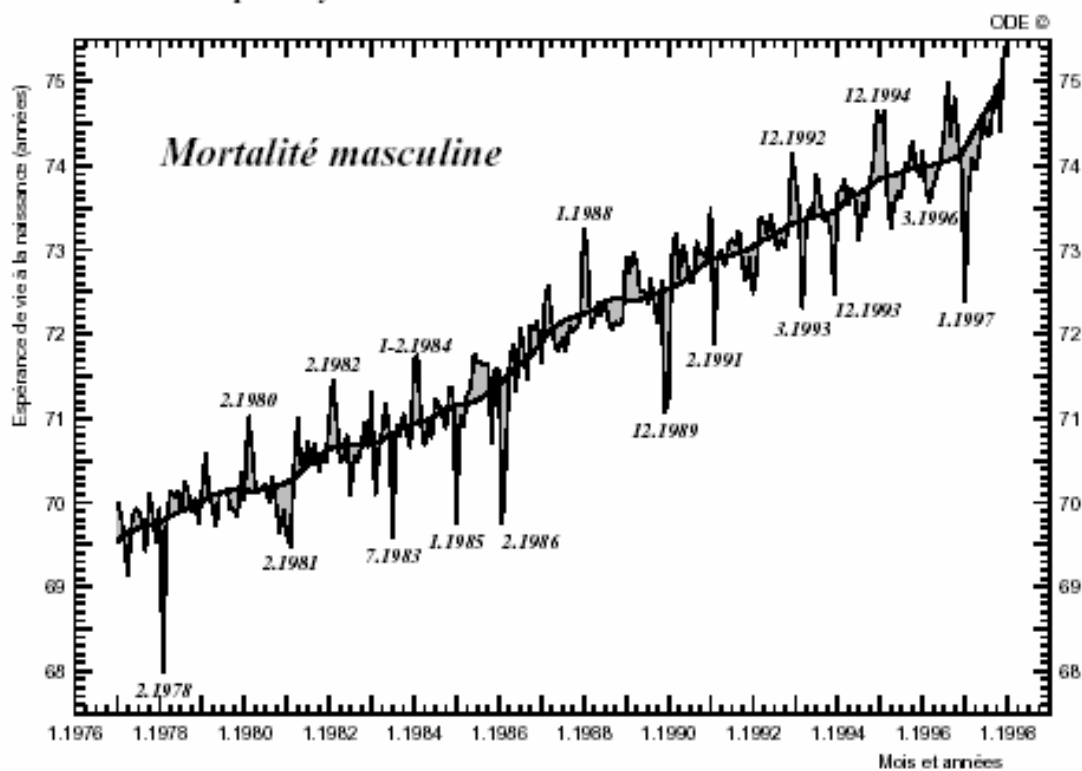


Figure 9 FRANCE, 1977-1997
Evolution de l'INDICATEUR CONJONCTUREL mensuel
de MORTALITE MASCULINE et FEMININE
 Série corrigée des variations saisonnières
 et série lissée par moyenne mobile Hoem sur 25 termes



On effectue une *série* de $p+1$ *projections* de population (p de l'ordre de 5 à 10) à mortalité *constante*, fécondité *constante* et soldes migratoires *constants* sous les hypothèses ci-après :

date de départ de la projection : 1er janvier $k-(n-b)$, c'est-à-dire 1er janvier $b-(n-k)$
effectifs initiaux : ceux *réellement* observés par sexe et âge au 1er janvier $b-(n-k)$
date-horizon : 1er janvier $k+1$, soit $n-b+1$ bonds d'un an dans le temps
table de mortalité par sexe et âge de l'année $a-(n-k)$,
table de fécondité par âge de l'année $c-(n-k)$
soldes migratoires par sexe et âge de l'année $m-(n-k)$

Ces projections de population sont réalisées successivement pour $k = b-p+1, b-p+2, \dots, b$, puis pour $k = n$. De ces $p+1$ projections, on retient le nombre de décès *projeté* D^{k*} que l'on rapporte au nombre de décès *observé* D^k . Le gain d'espérance de vie à la naissance au cours de la période qui va de l'année $a-n+k$ à l'année k , soit sur un intervalle de $n-a$ années, est égal à $Ev0H(k) - Ev0H(a-n+k)$ pour le sexe masculin, $Ev0F(k) - Ev0F(a-n+k)$ pour le sexe féminin et $Ev0(k) - Ev0(a-n+k)$ pour l'ensemble des deux sexes. La corrélation entre l'un ou l'autre de ces trois gains et le rapport du nombre de décès deux sexes projeté pour l'année k à mortalité constante (celle de l'année $a-n+k$) à celui observé, D^{k*}/D^k , est généralement étroite.

Sur le nuage, par exemple masculin, de p points d'abscisses D^{k*}/D^k et d'ordonnées $Ev0H(k) - Ev0H(a-n+k)$, on détermine, par la méthode des moindres carrés, le paramètre a de la relation statistique (droite ajustée passant par le point de coordonnées 0 et 1) :

$$D^{k*}/D^k = 1 + a [Ev0H(k) - Ev0H(a-n+k)]$$

unissant abscisses et ordonnées (Figure 10). Cette relation statistique est ensuite appliquée, pour $k = n$, au rapport D^{n*}/D^n , ce qui fournit $Ev0H(n) - Ev0H(a)$ et donc l'estimation cherchée de $Ev0H(n)$. On procède de la même façon pour le sexe féminin et pour l'ensemble des deux sexes.

On peut encore procéder de la même façon pour estimer les espérances de vie, masculine et féminine, non pas à la naissance, mais à un âge quelconque. La figure 11 est l'analogue de la figure 10 pour l'estimation de l'espérance de vie à 60 ans, sur la base du nombre absolu annuel de décès.

10. La signification d'un indicateur conjoncturel de fécondité ou de nuptialité

Le concept d'effectif moyen des générations soumises au risque, qu'il a été nécessaire d'adapter dans le cas de la mortalité, c'est-à-dire d'un événement *non renouvelable* dont les intensités sont mesurées par une série de *quotients* par âge, permet de préciser la signification d'un indicateur conjoncturel. Que l'événement soit renouvelable ou non, la démarche suivie pour apprécier la portée d'un nombre absolu d'événements est la même : elle consiste en une comparaison entre ce nombre absolu et un nombre de référence.

En matière de fécondité, on compare le nombre absolu des naissances, c'est-à-dire l'effectif de la génération née durant l'année, à l'effectif des générations adultes dont cette génération est issue. Cette comparaison est effectuée sur la base du sexe *féminin* : on rapporte l'effectif de la génération féminine née durant l'année à l'effectif moyen (pondéré) des générations féminines qui, cette année-là, ont l'âge d'avoir des enfants.

L'égalité entre ces deux effectifs, c'est-à-dire la valeur 1 de ce rapport, sert ainsi de *repère*, qu'on dénomme *remplacement* ou encore *strict remplacement*.

Etant donné que la proportion de filles à la naissance est invariablement de 100 *filles* pour 205 *naissances*, il est équivalent de considérer le nombre *total* de naissances de l'année et de prendre pour repère la valeur 2,05 du rapport – qui n'est alors autre que l'indicateur conjoncturel – c'est -à-dire de retenir 2,05 comme repère de l'indicateur conjoncturel de fécondité. Autrement dit, la valeur 2,05 enfants pour une femme, prise par l'indicateur conjoncturel de fécondité, signifie très exactement l'égalité entre le nombre de filles nées durant l'année et l'effectif moyen pondéré des diverses générations féminines qui, la même année, ont l'âge d'être mères.

On peut affiner très légèrement le repère en observant que la comparaison précédente porte, d'une part, sur des filles qui viennent de naître et, d'autre part, sur des femmes dont l'âge moyen est de l'ordre de 28 ans. En divisant la valeur-repère 2,05 par la proportion des filles qui atteindront à leur tour l'âge d'être mères, proportion de l'ordre de 0,985 si on se réfère aux tables de mortalité transversales actuelles, on aboutit à une nouvelle valeur-repère, égale à 2,08 et arrondie habituellement à 2,1 enfants pour une femme.

Dans ces conditions, la valeur 2,08 enfants pour une femme, prise par l'indicateur conjoncturel de fécondité, dont on dit qu'elle correspond au strict remplacement, signifie très exactement l'égalité entre l'effectif qui *sera*, en l'absence de migrations internationales, celui de la génération féminine née durant l'année, lorsqu'elle atteindra à son tour l'âge d'avoir des enfants, et l'effectif *moyen* des diverses générations féminines qui appartiennent *actuellement* au groupe d'âge fécond.

Plus généralement, le *rapport* de l'indicateur conjoncturel de fécondité à 2,08 est aussi le *rapport* entre l'effectif qui *sera*, en l'absence de migrations internationales, celui de la génération féminine née durant l'année, lorsqu'elle atteindra l'âge d'avoir des enfants, et l'effectif *moyen* des diverses générations féminines qui ont *actuellement* l'âge de la maternité.

Nous préférons cette définition de l'indicateur conjoncturel de fécondité à celle souvent donnée et que nous estimons critiquable, fondée sur l'artifice de la *cohorte fictive* : nombre moyen d'enfants auquel *parviendrait*, en fin de vie féconde, un ensemble de femmes qui, aux différents âges, *auraient* le même taux de fécondité que celui observé au même âge durant l'année – mais sur des générations réelles *différentes* –. En effet, cette définition repose implicitement sur l'hypothèse de la plausibilité de l'existence d'une telle génération. Or ce calcul peut fort bien être *irréaliste* dans la mesure où il est impossible d'imaginer qu'une génération *réelle* puisse avoir un tel comportement *tout au long de sa vie féconde*. Qu'on songe par exemple au cas de l'année 1916 en France : quel sens aurait le comportement d'une génération qui vivrait *toute* sa vie féconde dans les mêmes conditions, à âge égal, que celles qui prévalaient durant l'année 1916 où la quasi-totalité des jeunes hommes étaient au front ? En revanche, la valeur prise en 1916 par l'indicateur conjoncturel de fécondité, soit 1,21 enfant pour une femme, signifie que le nombre de filles nées en 1916 n'a atteint que $1,21/2,05 = 59\%$ d'une classe d'âge féminine moyenne alors en âge d'avoir des enfants.

En matière de primo-nuptialité, par exemple masculine, l'indicateur conjoncturel est le rapport entre le nombre absolu de mariages d'hommes célibataires de moins de 50 ans célébrés durant l'année et l'effectif moyen (pondéré) des générations masculines, qui cette année-là ont l'âge du premier mariage. Quand l'indicateur conjoncturel de primo-nuptialité masculine vaut par exemple 0,6 premier mariage pour un homme, cela signifie que le nombre de premiers mariages célébrés avant 50 ans représente 60% d'une classe d'âge masculine moyenne en âge de premier mariage.

Figure 10 FRANCE, 1988-1996

Estimation de l'ESPÉRANCE de VIE à la naissance de l'année n selon le sexe
 connaissant les effectifs par sexe et âge jusqu'au 1er janvier n
 et les tables de fécondité, de mortalité et les soldes migratoires jusqu'à l'année n-1
 En abscisses : rapport entre le nombre de décès projeté et le nombre observé
 En ordonnées : gain d'espérance de vie à la naissance (en années) entre n-1 et n

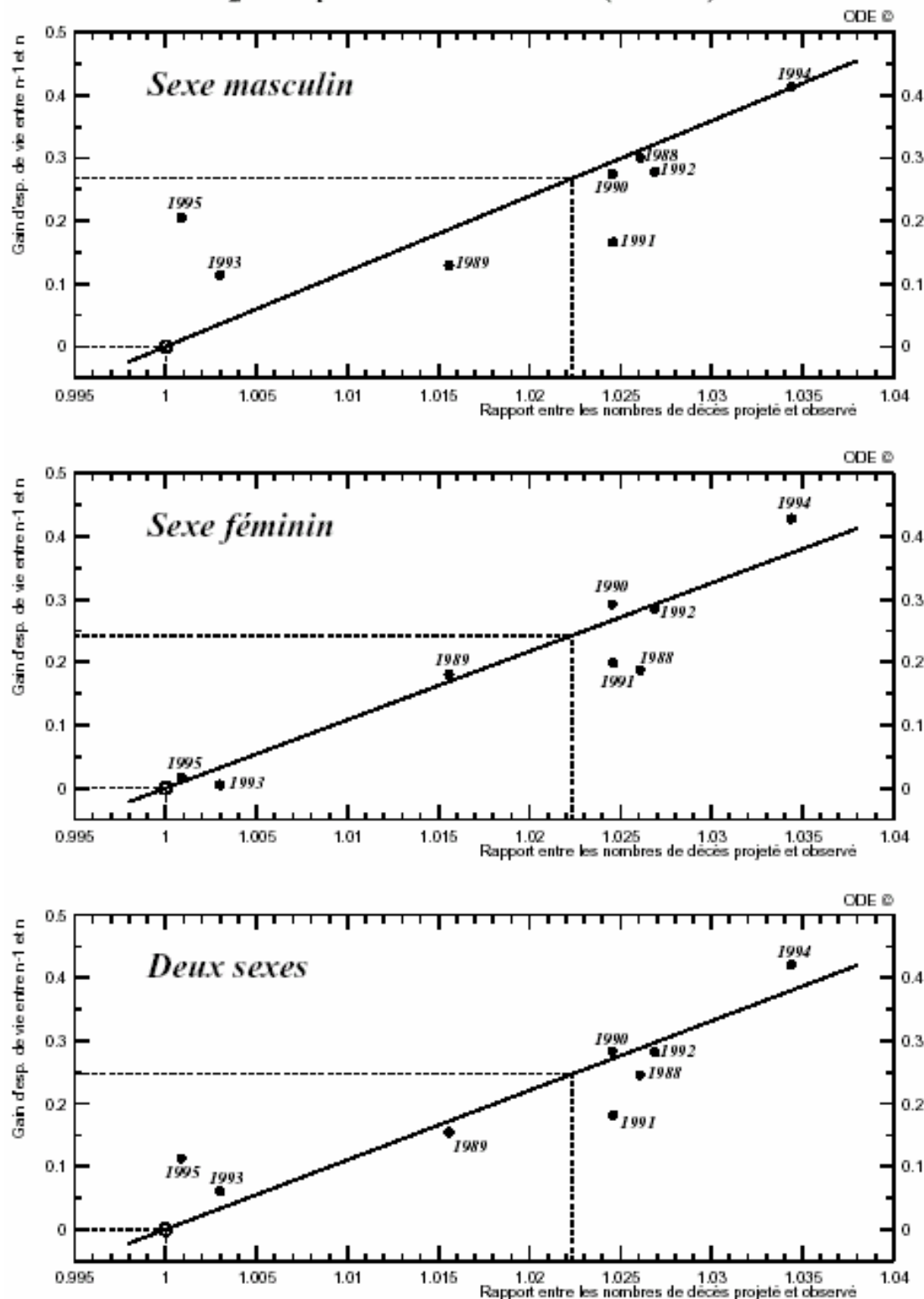
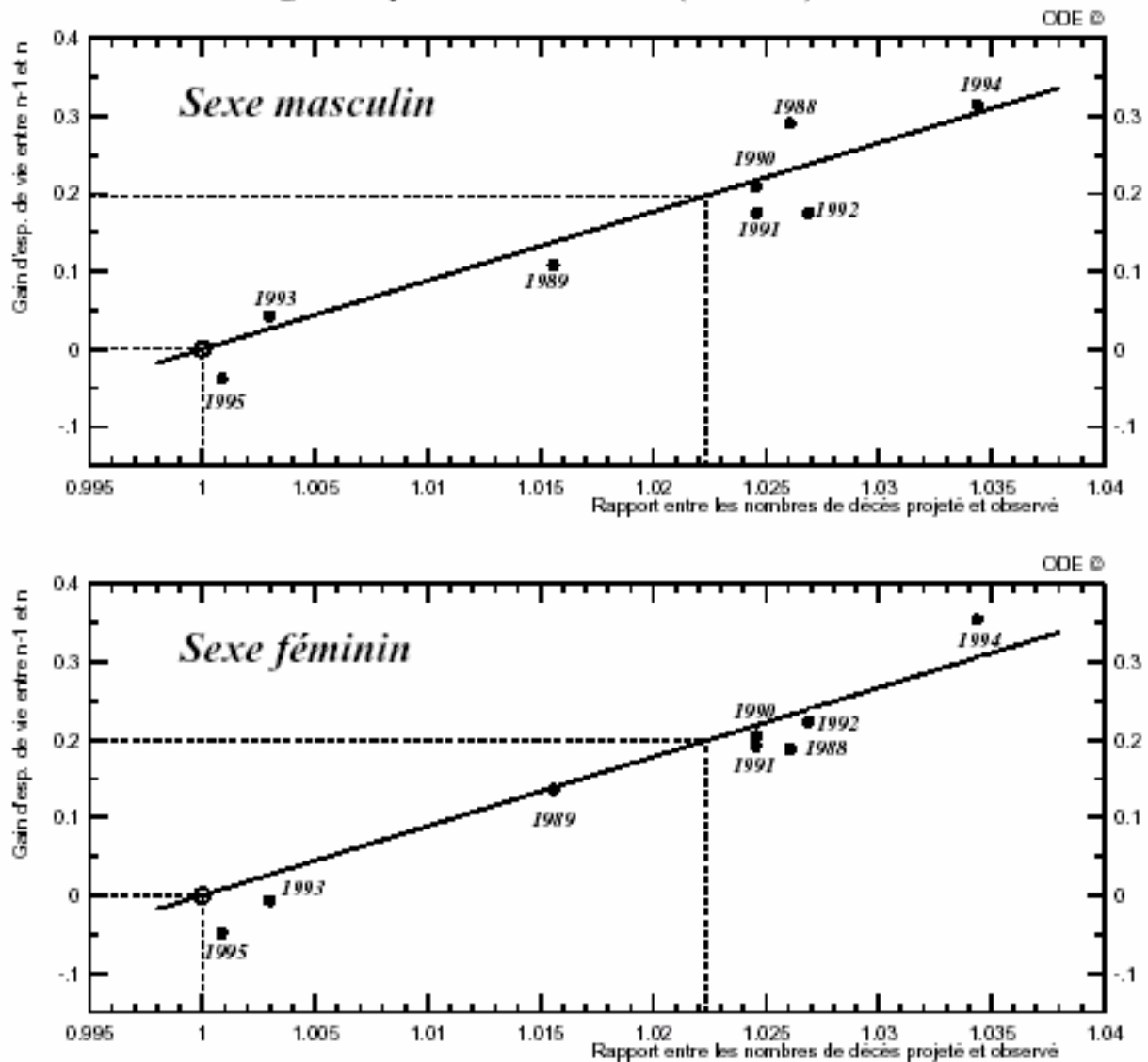


Figure 11 FRANCE, 1988-1996

Estimation de l'ESPERANCE de VIE à 60 ans de l'année n selon le sexe
 connaissant les effectifs par sexe et âge jusqu'au 1er janvier n
 et les tables de fécondité, de mortalité et les soldes migratoires jusqu'à l'année n-1
 En abscisses : rapport entre le nombre de décès projeté et le nombre observé
 En ordonnées : gain d'espérance de vie à 60 ans (en années) entre n-1 et n



11. Flux annuel d'événements et intensité/calendrier du phénomène

Revenons au nombre annuel de naissances. Nous avons vu plus haut que ce qui en conditionne la valeur l'année n , c'est *quasi*-exclusivement l'indicateur conjoncturel I^n de cette même année n , puisque l'effectif moyen des générations féminines d'âge fécond de l'année n est *quasi*-entièrement déterminé par l'évolution démographique au cours des années *antérieures* à n . Le *moteur* du nombre absolu des naissances n'est donc pas directement le niveau de fécondité des générations qui, l'année n , ont l'âge d'être mères, c'est-à-dire la *descendance finale moyenne* des générations qui, l'année n , appartiennent aux divers âges féconds. Observons que cette descendance finale moyenne peut être prise comme égale à la descendance finale $DF(n - \bar{x}^n)$ de la génération née en $n - \bar{x}^n$, où \bar{x}^n est l'âge moyen (transversal) à la maternité observé l'année n , compte tenu du fait que la descendance finale varie en général *lentement* d'une génération à l'autre.

Si on veut relier le nombre absolu des naissances observé l'année n au niveau de fécondité de la génération née en $n - \bar{x}^n$, il faut faire intervenir, comme facteur en quelque sorte *perturbateur*, le rapport entre I^n et $DF(n - \bar{x}^n)$. Or ce rapport varie lui-même de façon complexe, sous l'effet des variations du calendrier transversal de la fécondité au fil des années successives.

On peut montrer⁶⁷ que sous les hypothèses très particulières suivantes :

- l'indicateur conjoncturel I^n des années successives est *invariant*
- la distribution de l'âge transversal à la maternité varie de telle sorte que tous ses moments *centrés* demeurent cependant *invariants*
- l'âge moyen transversal à la maternité \bar{x}^n varie *linéairement* avec n

la descendance finale est *invariante* et, en désignant par \bar{x}' la dérivée *constante* de \bar{x}^n par rapport à n , le rapport de l'indicateur à la descendance finale est égal à :

$$\frac{I}{DF} = 1 - \bar{x}'$$

En d'autres termes, l'indicateur conjoncturel est la descendance finale d'un régime de fécondité *rigoureusement invariable* (taux de fécondité invariants à âge égal) qui conduirait, avec les *mêmes* effectifs féminins par âge F_i^n que ceux présents l'année n et avec les *mêmes* taux f_i^n que ceux enregistrés l'année n , au *même* nombre de naissances N^n que celui observé. Aux époques où le calendrier de la fécondité évolue rapidement, l'indicateur conjoncturel peut s'écarter notablement de la descendance finale moyenne des générations alors en âge de fécondité et une approximation, d'ailleurs assez grossière, de leur rapport est suggérée par la relation ci-dessus (indicateur *excédant* la descendance finale moyenne de l'ordre de 10% lorsque l'âge moyen à la maternité *s'abaisse* au rythme de 0,1 an par an, inversement descendance finale moyenne excédant l'indicateur de l'ordre de 10% lorsque l'âge moyen à la maternité *s'élève* au rythme de 0,1 an par an).

Ne perdons cependant pas de vue que ce qui importe en matière de fonctionnement de la machinerie démographique, ce n'est pas le *niveau de fécondité* (mesuré par la descendance finale) des générations qui se trouvent à l'époque considérée appartenir aux âges féconds, mais le *nombre absolu des naissances*. Or le nombre absolu des naissances dépend *quasi*-exclusivement, les effectifs féminins en âge de fécondité étant *donnés*, de l'indicateur conjoncturel. Même s'il est vrai que celui-ci constitue une image imparfaite de la

⁶⁷ Voir par exemple G. CALOT, Relationships between cohort and period demographic indicators, *Population*, An English selection, 5, 1993, 183-222.

descendance finale des générations qui sont alors d'âge fécond, en raison de l'effet perturbateur des variations du calendrier, c'est lui et non la descendance finale, qui détermine le nombre absolu des naissances.

Observons par ailleurs que ce que nous avons appelé indicateur conjonctuel est une mesure d'*intensité* dans le cas de la fécondité et de la primo-nuptialité, dont les événements sont (ou sont traités comme) *renouvelables*, tandis qu'en matière de mortalité, dont les événements sont *non renouvelables*, l'indicateur conjonctuel s'exprime en termes d'espérance de vie à la naissance, caractéristique non pas d'intensité mais de *calendrier*. C'est qu'en matière de mortalité, l'intensité ne soulève aucune question ; elle est *invariablement* égale à l'unité : à toute époque et dans toutes les générations, les humains sont *tous* mortels. Dans le cas des événements renouvelables, le flux annuel d'événements est en premier lieu sensible aux variations de l'intensité transversale ; dans le cas des événements non renouvelables dont l'intensité est égale à l'unité (événements qualifiés de *fatals* dans la littérature), il est en premier lieu sensible aux variations du calendrier transversal. Il n'en irait pas de même si on traitait la primo-nuptialité comme produisant des événements non renouvelables : le flux annuel serait alors sensible *à la fois* aux variations de l'intensité transversale et à celles du calendrier transversal : la corrélation sur laquelle nous nous sommes appuyés pour estimer l'espérance de vie à partir du nombre annuel de décès n'aurait pas de sens en matière de primo-nuptialité.

Index des indicateurs dont la méthodologie est fournie

Populations

Population moyenne :	p. 36
Estimation de la population au 1 ^{er} janvier à partir d'un effectif à une date quelconque :	p. 23
Effectif moyen des générations soumises au risque :	p. 66 et 70

Fécondité

Taux de fécondité par âge :	
• âge atteint :	p. 50
• âge révolu :	p. 42
• âge quinquennal :	p. 18
Indicateur conjoncturel de fécondité :	p. 66
Descendance finale :	p. 66
Age moyen transversal à la maternité :	
• pour l'ensemble des rangs :	p. 34 et 67
• de rang 1 :	p. 61
• de rang 2 :	p. 61
• de rang 3 :	p. 61
• de rang 4 et plus :	p. 61
Age moyen longitudinal à la maternité :	
• pour l'ensemble des rangs :	p. 34 et 67
• de rang 1 :	p. 61
• de rang 2 :	p. 61
• de rang 3 :	p. 61
• de rang 4 et plus :	p. 61
Infécondité définitive dans les générations :	p. 68
Proportion de femmes, dans les générations, avec :	
• 1 enfant :	p. 68
• 2 enfants :	p. 68
• 3 enfants :	p. 68
• 4 enfants ou plus :	p. 68
Probabilité d'agrandissement :	p. 68

Nuptialité

Taux de primo-nuptialité par âge et sexe :	
• âge atteint :	p. 50
• âge révolu :	p. 42
• âge quinquennal :	p. 18
Indicateur conjoncturel de primo-nuptialité :	p. 66
Intensité de la primo-nuptialité dans les générations :	p. 66
Age moyen transversal au 1 ^{er} mariage :	p. 34 et 67
Age moyen longitudinal au 1 ^{er} mariage :	p. 34 et 67

Divortialité

Indicateur conjoncturel de divortialité :	p. 62
Proportion de mariages dissous par divorce dans les promotions :	p. 62
Ancienneté moyenne du mariage au moment du divorce :	p. 62
Ancienneté moyenne du mariage au moment du divorce dans les promotions :	p. 62

Mortalité

Quotients de mortalité par âge et sexe :	p. 53
Taux de mortalité :	p. 53
Espérances de vie par âge et sexe :	p. 67