

LES ENQUETES DE MIGRATION  
DANS LES PAYS DEVELOPPES

*Daniel COURGEAU*

*Institut National d'Etudes Démographiques, Paris*

INTRODUCTION

Dans le passé, l'espace était presque totalement absent des préoccupations des démographes. Aussi la migration interne apparaissait-elle beaucoup plus comme un phénomène perturbateur, dont il fallait éliminer l'effet, que comme un phénomène démographique digne d'intérêt. Au cours des dernières années, ce point de vue a été battu en brèche par de nombreux chercheurs. Ils montraient l'importance du rôle de l'espace dans l'évolution d'une société et l'effet prépondérant de la migration comme phénomène d'ajustement aux nombreuses disparités spatiales secrétées par ces sociétés.

Pour saisir un tel phénomène, les chercheurs avaient accès à deux types de sources classiques : les données de recensement et les données de registre de population. Nous allons d'abord montrer les raisons pour lesquelles ces sources se sont révélées progressivement insuffisantes pour analyser de façon satisfaisante un phénomène aussi complexe que la mobilité spatiale.

Les recensements fournissent essentiellement des informations sur les migrants, en posant une question sur le lieu de résidence à une date antérieure, qui peut être la date de naissance, ou sur les dernières migrations, en posant une double question sur la date de la dernière migration et sur le lieu de résidence antérieur. Ces questions, utiles pour donner une vue d'ensemble des migrations internes d'un

pays, sont largement insuffisantes dès que l'on veut analyser plus finement le processus. Ainsi la question qui permet de détecter les migrants omet les migrations multiples que peut effectuer un individu et ne compte pas comme migrant un individu ayant fait une migration de retour. Cet inconvénient est encore plus grave lorsque l'on utilise une question sur le lieu de naissance. Les questions sur la dernière migration donnent une vue tronquée de la mobilité, difficile à utiliser correctement sans un modèle adéquat, que seules des enquêtes peuvent fournir.

La mise au point, dans certains pays, de panels issus de ces données censitaires et de l'état civil fournit une information plus complète. Il faut voir cependant que, si ces panels permettent de suivre un même individu au cours de plusieurs périodes intercensitaires, cela ne résoud pas pour autant le manque d'informations sur la mobilité multiple au cours de chacune de ces périodes. De plus, ces panels ne remontent pas très avant dans le temps, en particulier pour la migration. Il faut voir qu'il s'agit là d'une source beaucoup plus prospective que rétrospective.

Ainsi l'utilisation des données de recensements est très limitée et ne permet généralement pas une analyse très fine de la mobilité.

Les données de registres de population semblent plus aptes à fournir une information détaillée sur ces migrations. On peut en effet dire que, lorsqu'ils sont correctement tenus, ils constituent la source la plus précise pour saisir l'événement qu'est une migration. Les déclarations de changement de résidence devraient théoriquement permettre de suivre un individu d'un lieu à l'autre tout au long de sa vie. D'un autre côté, les statistiques que publient les administrations locales ou nationales donnent une vue d'ensemble de l'évolution des migrations d'une année sur l'autre.

Ce double suivi, d'un individu d'un côté, d'une population de l'autre, fort utile pour l'analyse de la mobilité, va cependant révéler certaines lacunes lorsque l'on veut pousser plus avant cette analyse. En premier lieu, la notion de résidence prise pour définir la migration peut laisser échapper des formes de mobilité spatiale dont le rôle peut être prépondérant dans l'évolution d'une société. Ainsi, en Italie du Sud, de nombreux migrants vers le triangle industriel du Nord du pays restaient inscrits sur les registres de leur commune d'origine pendant parfois plusieurs années, masquant de ce fait des migrations fondamentales pour comprendre l'évolution de ce pays. On peut également citer le cas d'une mobilité temporaire (1) de ruraux vers les

---

(1) Dans certains pays, l'importance de cette mobilité est

villes, indispensable à saisir lorsque l'on veut comprendre l'évolution du monde rural et urbain d'un pays.

Dans d'autres pays également, les migrations à l'intérieur d'une unité administrative de base ne sont pas enregistrées, ou le sont mal. Cela introduit un arbitraire dans la définition d'une migration, qui peut, dans certains cas, fausser l'observation. En particulier, si la taille des unités administratives est importante, on omettra des migrations d'assez longue distance du rural vers le chef-lieu de cette unité, alors qu'on enregistrera des migrations de courte distance entre zones rurales de différentes unités administratives.

Mais le principal inconvénient de ces statistiques vient du petit nombre de renseignements publiés sur les caractéristiques individuelles des migrants. Ainsi en Belgique, les seules variables prises en compte dans ces statistiques sont le sexe et l'appartenance ou non à la population belge (Poulain, 1978). On pourrait cependant se reporter aux déclarations de changement de résidence pour avoir une information plus riche. Bien qu'un peu plus détaillées, ces déclarations ne relèvent qu'un petit nombre de caractéristiques : sexe, date de naissance, nationalité, état civil ; plus rarement : profession, lieu de mariage... De plus, de nombreux pays ont introduit récemment des législations sur la consultation des registres, qui en rendent l'utilisation très coûteuse en temps, sinon impossible (Woll, 1982).

Ainsi l'utilisation approfondie des registres de population se trouve-t-elle de plus en plus limitée, surtout lorsque l'on désire suivre un individu tout au long de sa vie avec le maximum d'informations sur d'autres domaines de son existence.

On voit donc que, pour analyser en profondeur le phénomène migratoire en liaison avec le cycle de vie et la place de l'individu dans la société, les données de recensement ou de registre sont en grande partie inadéquates. C'est la raison pour laquelle des chercheurs de plus en plus nombreux, travaillant dans des pays avec ou sans registres, ont dû entreprendre des enquêtes pour mieux saisir les mécanismes migratoires.

Il importe maintenant de mettre au clair les principaux objectifs de ces enquêtes, avant d'en présenter les modalités d'exécution et d'analyse.

---

reconnue : ainsi en Hongrie, à tout instant, un individu peut avoir simultanément une résidence permanente et un logement temporaire (Courgeau, 1980).

## 1. OBJECTIFS DES ENQUETES DE MIGRATION

En dépit de la diversité apparente des objectifs de ces enquêtes, S. Findley (1982) a montré que l'on pouvait les classer en trois grands types. Nous les présentons rapidement, avant de voir plus en détail celui que nous développerons ici.

Le premier de ces objectifs ne présente un intérêt que pour les pays sans registres de population. Il s'agit en fait de relever sur un échantillon suffisamment important d'individus les caractéristiques que fournirait un registre. Aux Etats-Unis, par exemple, l'enquête annuelle "Current Population Survey" pose un petit nombre de questions sur la mobilité géographique, en particulier sur le lieu de résidence un an auparavant, à un échantillon d'environ 65 000 ménages. Elle fournit une vue d'ensemble de la mobilité en fonction d'un certain nombre de caractéristiques des personnes interrogées (Long & Boertlein, 1981) : sexe, âge, état matrimonial, niveau d'éducation, statut professionnel...

Le second de ces objectifs consiste à enregistrer comme dans les recensements les migrants récents (au cours de la dernière année, par exemple) ou la dernière migration, en vue de comparer le comportement de ces migrants avec celui des sédentaires. La seule différence avec les recensements vient de ce que l'on relève de nombreuses autres caractéristiques de ces individus. Ces enquêtes ne sont donc pas centrées sur la migration, mais sur d'autres phénomènes qu'elle peut influencer. Elles sont aussi essentiellement transversales en ce qu'elles cherchent les raisons d'un comportement dans des variables du moment.

Le troisième objectif, beaucoup plus audacieux que les précédents, est d'essayer de saisir les enchaînements qui conduisent certains individus à migrer. C'est cet objectif que nous développerons ici.

Notons d'abord qu'il nous paraît illusoire de rechercher des "causes" de migration, car une telle causalité ne peut jamais être prouvée avec certitude. Nous préférons parler ici d'interaction entre migration et autres phénomènes sociaux, dont nous allons chercher à mettre en évidence les rouages. Voyons plus en détail les bases d'une telle recherche.

Nous partons de l'observation que toute société humaine peut être considérée comme composée de divers systèmes de relations en interaction les uns avec les autres. Ces systèmes peuvent être définis par les fonctions qu'ils réalisent et par les rapports qui s'établissent entre leurs membres. Les exemples suivants éclaireront ce que nous entendons par ces systèmes.

Pour ce faire, prenons un certain nombre de caractéristiques d'un individu, en vue de le placer dans différents systèmes de relations. Ainsi il se situera dans le système familial en tant que marié et père de deux enfants ; dans le système économique en tant qu'ingénieur dans une entreprise de travaux publics ; dans le système politique en tant qu'officier de réserve et adjoint au maire de sa commune ; dans le système d'éducation en tant que diplômé d'une école d'ingénieurs et préparant un doctorat d'Etat ; dans le système associatif en tant que membre d'une équipe de football... On voit dès lors plus clairement comment un changement dans un de ces systèmes pourra induire des changements dans les autres.

Pour étudier ces interrelations, nous allons devoir considérer leur expression au cours du temps et dans l'espace au travers d'événements. Notons ici que ce temps et cet espace ne sont pas premiers, mais au contraire engendrés par ces systèmes de relations. Ainsi l'âge auquel un enfant passe dans le groupe des adultes varie suivant les sociétés. De même, l'espace d'une ville est déterminé par la culture qui l'a créé. Il importe donc de définir plus précisément l'espace et le temps que les enquêtes vont saisir.

L'utilisation de registres de population permettait de considérer le temps comme une variable certaine, dont le déroulement se produit de façon continue. De même, le lieu de résidence d'un individu était à chaque instant défini de façon unique. Cette vue parfaitement déterministe va être modifiée par les enquêtes.

En premier lieu, le temps saisi par de telles enquêtes est beaucoup plus flou que le temps saisi par les registres. En effet, ces derniers prennent la date de déclaration de changement de résidence comme date de migration. Une telle date sera forcément très précise. Mais en fait, une migration n'est jamais un acte décidé à un instant précis et immédiatement accompli, comme le laissent penser les registres. Entre le moment où une migration est envisagée, celui où elle se réalise, celui où l'individu considère avoir rompu définitivement avec son lieu de résidence antérieur, une durée non négligeable peut s'être écoulée. On conçoit dès lors pourquoi une personne aura des difficultés à situer cette migration à un instant précis, alors qu'elle est liée à une période de son existence. Ainsi, conjointement aux problèmes de mémoire que nous soulevons plus loin, ceux plus fondamentaux de définition d'un événement dans le temps expliquent les différences que l'on peut observer entre registres et enquêtes. L'introduction d'un "temps flou" (2) nous semble utile à considérer. C'est là une voie de recherche encore peu explorée.

---

(2) Cette notion s'est dégagée de discussions sur ce thème avec A. Sorensen et d'autres chercheurs, au cours du mois de janvier 1983.

Beaucoup plus qu'un temps arbitrairement précis, ces enquêtes doivent chercher à saisir au mieux les interrelations temporelles entre événements. C'est donc là un point de vue essentiellement longitudinal. Il importe de déceler comment les événements d'un cycle de vie sont organisés au cours du temps, et quel rôle y joue la mobilité spatiale.

En second lieu, de telles enquêtes vont considérer un espace beaucoup plus complexe que le lieu de résidence saisi par les registres. Nous avons déjà indiqué que cette notion laissait échapper une mobilité dont le rôle pouvait être prépondérant sur l'ajustement d'une population aux disparités spatiales d'une société. Les enquêtes peuvent considérer la notion de changement de logement, qui nous semble beaucoup plus riche. La différence entre les deux concepts vient de ce que la résidence est définie comme "un lieu où l'individu a coutume d'habiter", alors que le logement représente tout "local d'habitation" où l'individu peut loger de façon beaucoup plus temporaire. Bien entendu, un certain nombre de règles permettront d'éviter de noter des changements de logement de peu d'intérêt pour l'étude des liens entre mobilité géographique et cycle de vie. Ainsi les séjours de vacances peuvent-ils, dans un premier temps, être omis. Par contre, des séjours temporaires ou saisonniers importants pour l'économie d'un pays seront saisis par cette notion.

Elle est cependant encore insuffisante en ce qu'elle rattache l'individu à chaque moment à un lieu unique, alors que son activité l'amène à fréquenter un espace beaucoup plus complexe. Pour dépasser cet espace ponctuel, nous avons proposé d'introduire la notion de changement d'espace de vie (Courgeau, 1980), l'espace de vie étant défini comme la portion d'espace où l'individu peut effectuer toutes ses activités. Cependant, devant les difficultés d'appréhender et de mesurer un tel espace, nous proposons pour le moment de le remplacer par un plus petit nombre de lieux auxquels l'individu est rattaché. Ainsi, lors de l'enquête sur la triple biographie, avons-nous simultanément saisi le lieu de travail et le logement de l'individu.

Même simplifiée et réduite à un petit nombre de lieux, cette notion entraîne une vue de la mobilité très différente de la simple migration. En effet, si l'espace de vie peut encore changer par un déplacement d'ensemble dans l'espace, de nouvelles formes de mobilité apparaissent : déplacement partiel, extension, restriction... C'est là une voie nouvelle de recherche, encore peu explorée mais riche de possibilités.

Après ce tour d'horizon des nouveaux concepts introduits par un tel type d'enquêtes, il est possible de dégager plus clairement ses objectifs. Nous avons déjà indiqué que son objectif principal était de saisir, de façon approfondie,

les enchaînements qui conduisent certains individus à migrer ou à modifier leur espace de vie. Seule une telle connaissance peut orienter des politiques visant à modifier une répartition spatiale jugée insatisfaisante. Pour remplir cet objectif, il est apparu nécessaire d'avoir une vue longitudinale des migrations des personnes interrogées. De plus, une telle enquête ne peut se contenter uniquement d'éléments sur la mobilité spatiale, mais doit relever les principaux éléments du cycle de vie des individus (familiaux, économiques, politiques...) qui lui sont reliés. Beaucoup plus que de rechercher une causalité entre ces divers phénomènes et d'interroger les personnes sur les raisons de leurs migrations, elle devrait permettre de déceler les principales interactions entre eux. Enfin elle devrait replacer ces migrations dans l'espace, en essayant, lorsque c'est possible, de donner une vue simplifiée, mais suffisamment claire, de l'espace de vie dans lequel circule la personne interrogée.

Nous avons ainsi mis en évidence comment les enquêtes pouvaient éclairer un phénomène aussi complexe que la mobilité spatiale, dont on ne saisissait par les sources classiques qu'une partie infime, le changement de résidence.

## 2. DES ENQUETES ESSENTIELLEMENT RETROSPECTIVES

En fait, deux types principaux d'enquêtes peuvent être distingués : les enquêtes à passages répétés et les enquêtes rétrospectives. En réalité, les enquêtes à passages répétés, très utilisées dans les pays en voie de développement, le sont très peu dans les pays développés. A notre connaissance, une seule enquête récente, réalisée aux Etats-Unis, suit vraiment ce schéma, avec cependant l'existence d'une enquête rétrospective faite lors du premier passage : la "Wisconsin Basic Needs Study" (3) entreprise en 1981 par D. Featherman.

Diverses raisons expliquent le manque d'intérêt pour ce type d'enquête. Lorsque l'on veut étudier de façon longitudinale la mobilité, il n'est plus possible de se contenter d'enquêtes enregistrant uniquement les départs et les arrivées des aires de l'échantillon, mais il faut suivre les individus initialement sélectionnés dans tous leurs déplacements. Ce suivi, pour être correct, nécessite une participation assidue des personnes interrogées (D. Featherman a prévu une compensation monétaire) ou l'existence de moyens

---

(3) Voir en Annexe 1 une présentation des diverses enquêtes citées dans cette communication. Bien entendu, cette liste d'enquêtes de migration réalisées dans les pays développés est loin d'être exhaustive.

d'information sur le devenir des migrants. P. Collomb (1982), bien que n'ayant pas réalisé une enquête à passages répétés, a travaillé sur un échantillon tiré des recensements de la population de 1954 et 1962 pour retrouver ces individus près de vingt ans après. Il a utilisé, pour ce faire, des informations fournies par diverses personnes du village d'origine de ces migrants (membres de la famille, postiers, maires...) et ensuite des informations fournies par les émigrants qu'il avait réussi à retrouver lors d'un premier passage. La réussite d'une telle opération (1,8 % de son échantillon seulement n'a pu être retrouvé) semble liée au fait qu'il travaillait dans un milieu rural. On peut en effet penser que, en milieu urbain, les contacts gardés avec le lieu d'origine seront beaucoup moins durables.

Un autre inconvénient des enquêtes à passages répétés vient de ce qu'elles nécessitent une longue observation pour recueillir suffisamment d'informations sur les interférences entre différents phénomènes du cycle de vie. De plus, ces informations seront recueillies au cours d'intervalles souvent à cheval sur le début et la fin de l'observation. On verra plus loin la quasi-impossibilité de traiter des intervalles à cheval sur le début de l'observation ; d'où la nécessité de faire une enquête rétrospective sur la période antérieure à ce début d'observation, pour éviter un tel inconvénient. Nous retombons à nouveau sur l'autre type d'enquête qu'il faut maintenant considérer.

Les enquêtes rétrospectives sont beaucoup plus faciles à réaliser en ce qu'elles ne nécessitent pas le suivi de l'échantillon. Elles posent cependant un certain nombre de problèmes difficiles à résoudre.

Le premier problème vient de ce que l'on ne peut évidemment enquêter que des personnes survivantes et présentes dans le pays. Même si l'échantillon est parfaitement tiré, il n'est généralement pas représentatif des cohortes considérées. Pour qu'il le soit, il faudrait que la mortalité (ou l'émigration vers l'étranger) soit indépendante du comportement migratoire de l'individu. Seule la présence de registres de population dans le pays où on réalise l'enquête permet de vérifier cette hypothèse (Hoem, 1983). Dans ce cas, il est possible de tirer un échantillon parmi les individus des cohortes observées qui sont morts ou qui ont émigré vers l'étranger avant l'enquête. La comparaison de leur comportement migratoire jusqu'à leur sortie de la population avec celui de l'échantillon de survivants est alors possible. A notre connaissance, aucun test de ce type n'a été réalisé pour l'analyse des migrations.

Le second problème vient de ce qu'un certain nombre d'individus choisis dans l'échantillon refusent de répondre au questionnaire. Ce taux de non-réponse a été de l'ordre de

11 % pour l'enquête "triple biographie" de l'I.N.E.D., il s'élève à 24 % pour l'enquête "Hopkins Life History Study" de J. Coleman et P. Rossi, et peut être encore plus important. Pour que l'échantillon observé soit toujours représentatif de la population dans son ensemble, il faut que la probabilité de répondre au questionnaire soit indépendante de l'histoire des migrations de la personne interrogée, et même, plus généralement, de son cycle de vie. On voit que cette hypothèse a peu de chances d'être réalisée et que des taux de non-réponse variables selon la catégorie socio-professionnelle, l'âge, la mobilité spatiale..., vont introduire des biais dans les résultats de ces enquêtes. A nouveau, l'existence de registres de population ou d'autres sources statistiques permet d'évaluer les biais introduits. Nous pouvons citer ici les résultats d'une comparaison de ce type effectuée par I. Lyberg (1983). Bien qu'elle porte sur une enquête de fécondité, elle fournit des éléments utiles pour une enquête de type plus général. Elle a montré que, bien que les taux de non-réponse variaient dans des proportions non négligeables selon l'âge et l'étape du cycle de vie familiale, ceux-ci avaient un effet négligeable sur les taux et les quotients que les données permettaient de calculer. Il serait dès lors intéressant de prolonger ce type d'étude sur des données concernant d'autres aspects de la vie, en particulier les migrations.

Nous aborderons maintenant les erreurs dans les déclarations des individus interrogés : événements non cités, erreurs sur les dates de migration, erreurs sur les lieux de résidence... C'est là un des problèmes les plus délicats de telles enquêtes, et l'on doit chercher, lors de leur réalisation, à les réduire au minimum : encourager la consultation de documents permettant de préciser ces dates (contrats de location, titres de propriété, livret de famille...), laisser les autres membres du ménage assister à l'entretien et favoriser les confrontations entre eux et la personne interrogée, essayer de relier au mieux vie familiale, vie professionnelle et migrations de façon à situer correctement les événements les uns par rapport aux autres...

La comparaison des dates et des lieux donnés lors d'une telle enquête avec des renseignements comparables obtenus à l'aide d'autres sources est dans ce cas fort utile. Citons une telle comparaison effectuée pour l'enquête norvégienne "Norwegian Life History Study" avec des données de registres militaires et de recensement, qui montre une bonne fiabilité de la plupart des informations rétrospectives recueillies par cette enquête. Seuls des renseignements recueillis dans les domaines de l'assistance publique et de la santé seraient moins fiables. Nous renvoyons surtout le lecteur au test réalisé en Belgique par J. Duchêne et M. Poulain, qui est présenté à cette même Chaire Quetelet, sur les réponses des personnes interrogées (hommes et femmes interrogés

séparément) à l'enquête "triple biographie" de l'I.N.E.D., comparées aux données des registres de population. Notons cependant ici que le registre de population n'est pas forcément à considérer comme la source "certaine" à laquelle comparer les données d'enquêtes. Nous avons déjà indiqué précédemment que le concept de migration des registres pouvait être différent de celui d'une enquête. En fait, comme l'observe J. Hoem (1983) "problems of unreported events are not restricted to interview data or to survey samples. In fact, in some aspects they may be worse in 'complete' population registers, which only record events subject to official registration, while unrecorded events may be as significant for demographic behaviour".

Enfin un dernier problème vient des questions rétrospectives sur les motivations des migrations. On peut craindre que la personne interrogée, en rationalisant a posteriori ces motifs, donne une vue très faussée des raisons qui l'ont conduite à migrer. Il semble en général préférable de noter un certain nombre de faits ayant accompagné une migration (changement du statut d'occupation, changement professionnel ou familial simultané à une migration...) pour lesquels les risques d'erreur sont déjà loin d'être négligeables. Il est alors possible de relier entre eux ces faisceaux d'événements que l'on peut comparer aux raisons données par les personnes interrogées.

### 3. PROBLEMES D'ECHANTILLONNAGE

Il est utile maintenant de distinguer les enquêtes de migration en deux types : les enquêtes nationales et les enquêtes portant sur une sous-population particulière. Cette sous-population est souvent définie par une migration d'un type spécifique, effectuée lors d'une période bien précise.

Le premier type d'enquêtes ne pose pas de problèmes très délicats à résoudre. Dans de nombreux cas, l'échantillon national ne portera pas sur l'ensemble des cohortes, mais seulement sur un petit nombre : l'enquête "triple biographie" n'interroge que des individus nés entre 1911 et 1936, l'enquête "Norwegian Life History" des hommes nés en 1921, 1931 et 1941, l'enquête "Hopkins Life History" des hommes nés entre 1929 et 1939. Dans la mesure où l'on peut tirer un échantillon à l'aide d'un registre de population, il est intéressant de tirer des cohortes très précises dont les membres auront connu les mêmes événements historiques (crises économiques, grèves...) au même âge, comme c'est le cas pour l'enquête norvégienne. Cela permet de contrôler un certain nombre de variables, qui peuvent échapper lorsque l'analyse est plus diffuse dans le temps. Dans les pays sans

registre, la taille de l'échantillon à tirer pour ce faire rend cette opération très coûteuse. Citons le cas de l'enquête "triple biographie" dont le jumelage avec une autre enquête couvrant un champ presque entièrement disjoint a permis d'éviter un grand nombre d'interviews sans objet.

Le second type d'enquêtes sera beaucoup plus délicat à réaliser, surtout dans les pays ne disposant pas de registre de population. L'échantillon est généralement défini à partir d'un ensemble d'individus ayant émigré ou immigré vers une zone du territoire national, au cours d'une période donnée. Ainsi pour l'enquête danoise, il s'agit des individus ayant émigré ou immigré dans la municipalité de Copenhague en 1978 ; pour l'enquête réalisée dans le Lauragais, des individus présents dans ces communes en 1954 et ne s'y trouvant plus en 1962 ; pour l'enquête sur les Parisiens, des individus radiés ou nouvellement inscrits dans les registres électoraux de cette région en 1978... Il importe de voir la fiabilité de ces sources en vue de retrouver des migrants et les moyens d'y arriver.

Dans les pays disposant de registres, on peut penser que l'on a là une source très fiable permettant de retrouver sans trop de problèmes les migrants, même après une durée importante. Notons cependant qu'au Danemark 12 % de ces migrants n'ont pu être retrouvés pour diverses raisons : absence temporaire (1,2 %), adresse fictive (1,1 % des individus étaient bien enregistrés, mais n'étaient pas résidents), adresse inconnue (5 %) ou autres raisons (4,7 %). De plus, la très forte mobilité de la population a rendu nécessaire le suivi d'une part non négligeable de l'échantillon (12,4 %), alors que l'enquête a eu lieu en moyenne un an après la déclaration de la première migration sur le registre.

Dans les pays sans registre, divers substituts sont utilisables. On peut, en premier lieu, faire une enquête juste après un recensement. Les délais de sortie des résultats détaillés sur les migrants rendent cette solution rarement utilisable, car trop tardive. En effet, en l'absence de registres, le suivi dans l'espace de l'échantillon est souvent difficile. Rappelons cependant les résultats très précis obtenus par P. Collomb en utilisant divers réseaux de relations. Rappelons qu'une telle méthode, utile pour suivre les émigrants d'un milieu rural, est certainement moins précise pour suivre les émigrants d'un milieu urbain.

On peut encore se reporter à d'autres sources moins exhaustives. Ainsi l'utilisation du fichier électoral en France pose un certain nombre de problèmes. Ce fichier ne porte que sur la population en âge et en droit de voter. Il exclut donc les étrangers, les personnes déchues de leur droit de vote... Et, bien que l'inscription y soit obligatoire, une proportion non négligeable, et surtout variable d'une

année sur l'autre, de Français n'est pas inscrite sur ces registres. Enfin, ce qui est plus gênant pour une enquête sur les migrations, rien n'oblige un individu à s'inscrire sur le fichier d'une commune où il migre. Il peut rester inscrit dans une autre commune, où il possède, par exemple, une résidence secondaire. Ces nombreuses conditions rendent l'utilisation de cette source très discutable. D'autres sources, peut-être plus fiables mais souvent d'accès plus délicat, peuvent être utilisées. Citons les fichiers de contribuables, ceux de sécurité sociale, ceux des abonnés à divers services (eau, électricité, gaz, téléphone...).

#### 4. STRUCTURE DU QUESTIONNAIRE

La formulation des questions et la structure du questionnaire peuvent avoir un rôle important sur les taux de non-réponse et sur la qualité de l'information recueillie.

Un premier point est d'essayer de favoriser au mieux la mémorisation de divers événements du cycle de vie et surtout leur position relative au cours du temps. Généralement on préfère partir de l'enfance de la personne interrogée pour lui faire suivre le cours du temps dans son déroulement réel. Il est cependant possible de partir du présent et de remonter progressivement vers le passé. Un test de la valeur relative de ces deux méthodes serait utile à réaliser. Cependant le problème le plus important est d'essayer de relier au maximum, lors de l'interview, les divers événements vécus par la personne, en particulier ceux de la vie migratoire et de la vie professionnelle. En effet, interroger successivement une personne sur la première, puis sur la seconde, conduit à de nombreuses répétitions qui font rapidement tomber son attention. Il est également utile de recourir aux dates de la vie familiale pour mieux situer les événements professionnels et migratoires. Un questionnaire qui suit au mieux le récit de la vie du répondant serait plus apte à replacer les divers événements dans l'ordre vécu.

Nous avons indiqué précédemment la variété des définitions du déplacement d'un individu dans l'espace. En vue d'avoir une information cohérente et utilisable, il est nécessaire de préciser à la personne interrogée les déplacements ou les changements professionnels à prendre en compte et ceux à ne pas considérer. La formulation des questions doit éviter toute ambiguïté et ne pas conduire à des interprétations incorrectes.

Il est également utile de ne pas mettre en début de questionnaire des questions dont la réponse peut être gênante pour la personne interrogée ou faire appel à des souvenirs

flous. Ainsi il est préférable de ne pas commencer une telle enquête par des questions sur la mobilité des parents.

## 5. METHODES SPECIFIQUES D'ANALYSE

Les données d'enquête vont poser des problèmes d'analyse, à bien des égards différents de ceux soulevés par les données de recensement ou de registre : problèmes liés au fait que l'effectif des personnes interrogées est généralement faible, que l'observation porte sur une période de temps limitée et que l'enquête permet de relier une grande variété de variables aux migrations. En revanche, les méthodes démographiques classiques, qui visaient à estimer les taux et quotients à partir des statistiques par âge sur les nombres d'événements survenus au cours d'une année, ne présentent plus d'intérêt ici. En effet, les enquêtes fournissent la date, supposée exacte, de chaque événement individuel, évitant ainsi ces problèmes.

En n'observant qu'un petit nombre d'individus, ces enquêtes vont nécessiter la mise au point de méthodes statistiques nouvelles pour assurer leurs conclusions. En effet, en démographie classique, on travaillait le plus souvent sur des populations d'effectif important où seul le calcul de taux ou de quotients présentait un intérêt. Mais dès que la taille de l'échantillon observé devient faible, les variations aléatoires jouent un rôle croissant. Il est alors nécessaire d'introduire cet aléa comme un élément des mesures effectuées, pour calculer la variance des estimations et mettre au point des tests qui permettent de conclure lorsqu'on désirera comparer des sous-populations.

En second lieu, ces enquêtes vont souvent fournir des intervalles entre une migration et la suivante à cheval sur la date d'observation. Bien que l'on ne sache pas la date de fin de cet intervalle, le fait qu'au moment de l'enquête une nouvelle migration ne se soit pas encore produite fournit une information dont il faut tenir compte. Nous verrons que, dans le cas où l'on n'observe que la fraction gauche de cet intervalle, il est aisé d'utiliser cette information. Cependant certaines enquêtes, en limitant dans le passé la période d'observation (4), vont introduire des intervalles dont on n'observe pas le début ; dans certains cas, on n'en observe ni le début ni la fin s'ils se situent hors de la fenêtre

---

(4) Ainsi l'enquête sur les "mobiles de migration entre Paris et la province" n'enregistre pas les migrations faites avant le 1.1.1946 et part du lieu de résidence à cette date pour les individus nés antérieurement.

d'observation. Pour pouvoir traiter de tels intervalles, il est nécessaire de faire l'hypothèse que l'histoire antérieure à la date limite d'observation n'influe en rien sur son futur. On voit qu'une telle hypothèse est généralement non vérifiée, ce qui conduit à rendre l'analyse de telles données très difficile. Nous renvoyons le lecteur à R. Ginsberg (1979) pour une discussion détaillée de ces problèmes.

Du fait que lors d'une enquête on observe un grand nombre de variables caractérisant l'individu à chaque moment de sa vie, on peut être tenté de chercher des relations de causalité entre ces divers événements. Dans la mesure où une causalité ne peut jamais être montrée avec une certitude absolue, il nous paraît préférable de parler ici d'interférences entre ces divers événements. Dans la mesure où chaque individu est suivi tout au long de sa vie, il sera possible de mettre en évidence des dépendances locales (Aalen et al. 1980), qui sont beaucoup plus précises qu'une simple corrélation. Ainsi, en suivant le déroulement au cours du temps de deux séries d'événements, il se peut que l'arrivée du premier change la probabilité d'arrivée du second, alors que l'inverse ne produit aucune modification. On aura ainsi mis en évidence une dépendance locale du second phénomène par rapport au premier, alors que dans l'autre sens on a indépendance locale.

Un certain nombre de nouvelles méthodes statistiques sont utilisables pour traiter de tels cas. Les plus simples vont relier la durée de séjour prise comme variable dépendante à d'autres variables caractérisant l'étape du cycle de vie dans laquelle se trouve l'individu. Mais on voit facilement qu'un tel ordre est totalement arbitraire et que l'on aurait aussi bien pu considérer la durée de séjour dans une étape du cycle familial comme une variable dépendante des migrations faites par l'individu. Nous essayerons d'échapper à cet artifice en considérant des variables plus complexes mêlant divers états du cycle de vie familiale, migratoire, économique...

Le dernier problème que nous traiterons ici sera de généraliser à des données d'enquête les méthodes d'analyse spatiale des migrations mises au point pour des données de recensement ou de registre.

Nous laissons ici de côté d'autres approches dont l'objectif est très différent. Nous pensons en particulier aux méthodes d'analyses factorielles et surtout d'analyse harmonique, qui peuvent être utilisées pour étudier la migration. Ces méthodes consistent à rechercher un petit nombre de facteurs, déterminés de façon théorique, dont les contributions à la variance du processus sont maximales. Une fois ces axes théoriques trouvés, on cherche les variables les plus fortement corrélées avec eux afin de leur donner une

signification pratique. Nous renvoyons le lecteur, pour une présentation de ces méthodes, à J.C. Deville (1974).

### 5.1. Généralisation de méthodes démographiques classiques

Il s'agit ici de développer des méthodes d'analyse non paramétriques qui généralisent les méthodes d'analyse longitudinale en démographie. Il faut tenir compte des petits effectifs observés, de l'existence d'intervalles ouverts et d'interférences plus complexes entre événements.

Le premier type d'interférences entre phénomènes démographiques est celui où l'un des phénomènes empêche l'autre de se produire. C'est l'effet classique de la mortalité sur la migration, et l'effet, plus intéressant dans le cas d'enquêtes rétrospectives, des sorties d'observation sur les migrations observées. On élimine cet effet en faisant l'hypothèse que les individus sortis d'observation auraient migré, s'ils avaient été observés, comme le font ceux que l'on observe aux mêmes durées.

Voyons plus en détail la méthode mise au point par Kaplan et Meier (1958) pour traiter ce cas.

Soient  $t_1 < t_2 \dots < t_k$  les durées de présence des divers individus dans leur premier logement, par exemple. Supposons que  $C_{1-1}$  personnes sortent d'observation au cours de l'intervalle  $(t_{1-1}, t_1)$  et que  $M_1$  individus émigrent à la durée  $t_1$  (5). Peu importe d'avoir les dates précises de sortie d'observation dans l'intervalle  $(t_{1-1}, t_1)$ , car elles ne jouent pas sur la population soumise au risque de migrer juste avant  $t_1$  :

$$N_1 = (M_1 + C_1) + \dots + (M_k + C_k)$$

On pourra donc représenter de la façon suivante les divers événements considérés :

Dates	$t_0$	$t_1$	$t_2$	$t_3$	...
Pop. soumise au risque	$N_0$	$N_1$	$N_2$	$N_3$	...
Nombre de migrations		$M_1$	$M_2$	$M_3$	...
Nombre sorties d'observ.	$C_0$	$C_1$	$C_2$		...

(Note 5 : page suivante).

Sous ces conditions, la méthode du maximum de vraisemblance conduit à l'estimation suivante de la probabilité de rester sédentaire avant la date  $t_1$  :

$$\hat{S}(t_1) = \prod_{r=1}^1 \left(1 - \frac{M_r}{N_r}\right)$$

Les individus qui sortent d'observation sans avoir préalablement migré sont pris en compte dans cette estimation, tant qu'ils sont observés comme sédentaires.

La formule suivante donne la variance asymptotique de  $S(t_1)$  :

$$\hat{V}\{\hat{S}(t_1)\} = \hat{S}^2(t_1) \sum_{r=1}^1 \frac{M_r}{N_r (N_r - M_r)} = \hat{S}^2(t_1) \sum_{r=1}^1 \left(\frac{1}{N_r - M_r} - \frac{1}{N_r}\right)$$

Notons que cette formule peut donner des intervalles de confiance qui sortent des limites de définition de  $S(t)$ , lorsque cette valeur est proche de zéro ou de l'unité. On peut éviter cet inconvénient en estimant la variance de  $\text{Log} [-\text{Log} \hat{S}(t_1)]$  qui fournira toujours un intervalle de confiance dans les limites (0,1).

Supposons maintenant que l'on veuille tester l'égalité des probabilités de rester sédentaire dans diverses sous-populations. Par exemple, celles-ci peuvent être définies par le nombre de migrations antérieures effectuées. Soient toujours  $t_1 < t_2 \dots < t_k$  les durées de présence dans la population globale, réunion de toutes les sous-populations. Dans cette population  $M_1$  individus migrent à la durée  $t_1$  tandis que  $N_1$  sujets sont soumis au risque. Dans la  $m$ -ième des  $r$  sous-populations,  $M_{m1}$  individus migrent à cette même durée avec  $N_{m1}$  sujets soumis au risque. Sous l'hypothèse que la probabilité de migrer est la même dans toutes les sous-populations, on peut montrer (Kalbfleisch & Prentice, 1980) que l'on a les estimations suivantes pour la moyenne et la variance de  $M_{m1}$  :

$$\mu_{m1} = N_{m1} \frac{M_1}{N_1}$$

$$V_{mm1} = N_{m1} \frac{(N_1 - N_{m1})}{N_1} \frac{M_1}{N_1} \frac{(N_1 - M_1)}{N_1} \frac{N_1^{-2}}{(N_1 - 1)^{-1}}$$

---

(5) Il s'agit là d'une convention pour présenter les données. Dans le cas où des sorties d'observation se produisent à la même durée que des migrations, on fait généralement l'hypothèse que les sorties d'observation interviennent après les migrations.

La covariance de  $M_{ml}$  et  $M_{nl}$  est aussi égale à :

$$V_{mnl} = - \frac{N_{ml} N_{nl} M_l (N_l - M_l) N_l^{-2} (N_l - 1)^{-1}}{N_l}$$

Il s'ensuit que le vecteur :

$$v_l = \begin{pmatrix} M_{1l} - \mu_{1l} \\ \vdots \\ M_{rl} - \mu_{rl} \end{pmatrix}$$

devrait avoir une moyenne nulle et comme matrice des variances et covariances  $V_l$ . En sommant sur toutes les durées de présence, on obtient un test permettant de comparer les différences entre le nombre observé de migrations dans chaque sous-population et le nombre attendu, sous l'hypothèse d'un comportement migratoire identique. Ecrivant

$$v = \sum_{l=1}^k v_l \quad \text{et} \quad v = \sum_{l=1}^k v_l$$

ce test peut s'obtenir en calculant  $v'V^{-1}v$  (où  $v'$  est le vecteur transposé de  $v$ ) qui devrait, si l'hypothèse est vérifiée, être un  $\chi^2$  à  $r-1$  degrés de liberté.

Cherchons maintenant à généraliser cette approche au cas où les migrations se font vers divers types de destinations. Par exemple, prenons un individu logé chez ses parents : il peut suivre ses parents vers un autre logement ; il peut au contraire devenir locataire ou propriétaire de son propre logement ; il peut être logé par son employeur ou dans d'autres conditions.

Supposons qu'il y ait  $r$  types de destinations possibles, dont on considère d'abord la  $j$ -ième. On pourrait, comme précédemment, calculer un indice du type  $\hat{S}_j(t_1)$ , mais on ne pourra plus lui donner la signification d'une probabilité de rester sédentaire. Seule la fonction

$$\hat{S}(t_1) = \prod_{j=1}^r \hat{S}_j(t_1)$$

sera la probabilité de rester sédentaire dans la zone initiale. Il est dès lors nécessaire de modifier cette approche.

Observons d'abord uniquement la  $j$ -ième destination, en considérant toutes les autres migrations comme des sorties d'observation. L'estimation de la probabilité de ne pas migrer vers  $j$  est alors possible, sous l'hypothèse qu'il y a indépendance entre les attractions des diverses destinations. Cette hypothèse est souvent contredite par les observations.

Ainsi il est difficile de considérer, dans l'exemple donné précédemment, les probabilités de suivre ses parents, devenir locataire... comme des risques indépendants. Un changement de point de vue s'impose.

Partons, pour ce faire, du quotient instantané de migration vers  $j$ ,  $m_j(t)$ , qui peut être défini comme la probabilité conditionnelle d'une migration vers la zone  $j$ , ayant toute l'information sur le passé du processus. On peut alors définir une intensité cumulée  $M_j(t)$  par la formule :

$$M_j(t) = \int_0^t m_j(s) ds$$

Cette intensité cumulée peut s'estimer (Aalen, 1978) à l'aide de la population soumise au risque juste avant une migration vers l'état  $j$  ayant lieu à l'instant  $t_r$ , ( $N_{rj}$ ) :

$$\hat{M}_j(t_1) = \sum_{r=1}^1 \frac{1}{N_{rj}}$$

Une telle estimation est correctement centrée sur  $M_j(t_1)$  et on peut estimer sa variance :

$$\hat{v}\{\hat{M}_j(t_1)\} = \sum_{r=1}^1 \frac{1}{(N_{rj})^2}$$

Lorsque plusieurs événements peuvent survenir à la même date, Aalen et al. (1980) proposent de les compter avec la même population soumise au risque, ce qui conduit aux estimateurs suivants :

$$\hat{M}_j(t_1) = \sum_{r=1}^1 \frac{M_{rj}}{N_{rj}} \quad \text{et} \quad \hat{v}\{\hat{M}_j(t_1)\} = \sum_{r=1}^1 \frac{M_{rj}}{(N_{rj})^2}$$

Il paraît cependant utile de rechercher des estimations plus correctes en travaillant, non plus sur un processus stochastique dont les sauts sont unitaires, mais sur des processus qui permettent des sauts multiples, dans la voie suivie par Jacobsen (1982).

Notons que ces intensités sont toujours bien définies, même lorsque les risques dépendent les uns des autres. Il en résulte une information plus précise que celle fournie par les probabilités de survie, lorsque l'on supprime toutes les destinations hormis la  $j$ -ième. Les courbes représentant les valeurs de l'intensité cumulée en fonction du temps fournissent les renseignements suivants :

- leurs pentes donnent une estimation des valeurs des quotients instantanés ;
- on peut interpréter  $M_j(t)$  comme le nombre attendu de mi-

- grations vers  $j$  qui auraient pris place s'il y avait eu en permanence un seul individu soumis au risque de migrer ;
- l'examen des courbes permet de proposer des modèles paramétriques s'adaptant bien aux observations. Ainsi, si l'intensité cumulée est approximativement une droite, on peut proposer un quotient instantané constant ;
  - on peut considérer les  $M_j(t)$  comme des processus approximativement indépendants. Il en résulte que l'on peut traiter séparément les courbes correspondant à diverses estimations.

La comparaison entre diverses destinations se réalise de la façon suivante. Soient deux destinations possibles  $i$  et  $j$ ,  $\hat{M}_i(t_r)$  et  $\hat{M}_j(t_r)$  les intensités cumulées de migration à l'instant  $t_r$ . Désignons par  $\Delta\hat{M}_i(t_r)$  et  $\Delta\hat{M}_j(t_r)$  les sauts réalisés à la date  $t_r$  par  $\hat{M}_i$  et  $\hat{M}_j$ . Aalen (1978) a montré que le test général d'égalité de probabilité de migrer vers  $i$  et  $j$  pouvait s'écrire :

$$T = \sum_{r=1}^1 W_r \{ \Delta\hat{M}_i(t_r) - \Delta\hat{M}_j(t_r) \}$$

où  $W_r$  est une pondération choisie en fonction de ce qui a été observé avant l'instant  $t_r$ . Le test de Gehan prend pour pondération :

$$W_r = N_{ri} N_{rj}$$

ce qui conduit au test suivant :

$$\hat{T} = \sum_{r=1}^1 (N_{rj} I_r - N_{ri} J_r)$$

où  $I_r = 1$  si le saut à l'instant  $t_r$  est en  $i$ , sinon  $I_r = 0$  ;  
et  $J_r = 1$  si le saut à l'instant  $t_r$  est en  $j$ , sinon  $J_r = 0$ .

La variance de  $T$  est :

$$\hat{V}(\hat{T}) = \sum_{r=1}^1 N_{ri} N_{rj}$$

Il en résulte que la variable  $\hat{T}/\hat{V}(\hat{T})^{1/2}$  est asymptotiquement normale, de moyenne 0 et d'écart type unitaire si les intensités de migration vers  $i$  et  $j$  sont les mêmes. On peut ainsi tester cette égalité. Ces tests peuvent être réalisés entre deux durées quelconques d'observation, plutôt qu'entre l'instant initial et celui de fin d'observation.

Le deuxième type d'interférence se produit lorsque l'un des phénomènes permet l'apparition de l'autre. Ainsi, en distinguant les migrations selon leur rang, on voit qu'il est nécessaire d'avoir déjà connu  $n-1$  migrations pour entrer

dans la population soumise au risque d'en faire une n-ième. On a alors une dépendance complète entre phénomènes. Il est utile de décomposer la population selon l'âge auquel les individus ont connu le phénomène permettant l'apparition de l'autre. La comparaison se fait comme précédemment en faisant intervenir, pour les migrations, l'âge à la migration antérieure et la durée qui la sépare de celle que l'on étudie.

Le troisième type d'interférence intervient lorsqu'aucun des deux phénomènes ne permet ni n'empêche l'arrivée de l'autre. Il est dès lors utile de déceler si l'arrivée d'un des phénomènes modifie la probabilité d'arrivée de l'autre.

Prenons le cas concret d'interférence entre migration d'une zone rurale et nuptialité. Ce cas est intéressant à étudier, car en France rurale le mariage est beaucoup plus tardif qu'en milieu urbain. Il est dès lors utile de comparer le comportement matrimonial des sédentaires en milieu rural à celui des émigrants vers les villes. Dans l'autre sens, il est intéressant d'observer le passage vers l'urbain des ruraux selon qu'ils sont célibataires ou mariés.

Pour répondre à de telles questions, il faut généraliser les méthodes suivantes. Dans l'exemple donné, deux types d'événements peuvent survenir à un individu, indépendamment du cas où il sort du champ d'observation, soit parce qu'il n'est plus observé par l'enquête, soit parce qu'il entre dans une autre catégorie. Deux variables aléatoires peuvent représenter les dates d'arrivée de ces événements :  $T_1$ , date de sortie du rural,  $T_2$ , date de mariage. La distribution simultanée de ces deux variables donne lieu à quatre types de quotients instantanés, où l'état initial, célibataire en milieu rural, est représenté par l'état 0 :

$$m_{0,1}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T_1 < t + \Delta t \mid t \leq T_1, t \leq T_2)$$

$$m_{0,2}(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T_2 < t + \Delta t \mid t \leq T_1, t \leq T_2)$$

$$m_{1,2}(t|u) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T_2 < t + \Delta t \mid T_1 = u, t \leq T_2) \quad u < t$$

$$m_{2,1}(t|u) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T_1 < t + \Delta t \mid t \leq T_1, T_2 = u) \quad u < t$$

Dans ce cas, on peut montrer que la condition nécessaire et suffisante pour que les deux variables  $T_1$  et  $T_2$  soient indépendantes est que :

$$m_{1,2}(t|u) = m_{0,2}(t) \quad \text{et} \quad m_{2,1}(t|u) = m_{0,1}(t)$$

Cependant, pour estimer ces quotients ou les intensités cumulées correspondantes, des hypothèses simplificatrices sont souvent nécessaires, étant donné le faible nombre de personnes interrogées. Supposons, par exemple, que les probabilités conditionnelles dépendent peu de la date  $u$ , mais seulement de  $t$ . Il convient de noter que, dans ce cas, tous les individus sont dans l'état 0 à l'instant initial. Ce n'est que lorsque des émigrations du rural ou des mariages commenceront à se produire que l'on pourra calculer des quotients du type  $m_{1,2}(t|u)$  ou  $m_{2,1}(t|u)$ . De plus, les populations soumises au risque seront, contrairement à la population célibataire dans le milieu rural, initialement croissantes. Les intensités cumulées de l'état  $i$  vers l'état  $j$  ne peuvent donc être définies sur l'intervalle  $(0,t)$  mais sur un intervalle  $(t_1,t)$  plus réduit. La date  $t_1$  correspond à l'instant où suffisamment d'individus se trouvent dans l'état  $j$  pour que le calcul de cette intensité ait un sens. Avec ces conventions, on peut, comme précédemment, estimer les intensités cumulées et leur variance. La comparaison se réalise avec les mêmes tests.

On peut cependant penser que les probabilités conditionnelles dépendent également de l'instant auquel le premier événement est arrivé. Dans ce cas, on décompose la population observée en sous-populations ayant connu cet événement entre diverses dates, choisies de telle sorte que l'on ait suffisamment d'individus pour estimer ces quotients instantanés.

Voyons enfin comment faire intervenir l'espace dans cette analyse. Nous avons déjà montré (Courgeau, 1980) qu'un quotient de type classique était de peu d'intérêt lorsque l'on désirait comparer des flux entre diverses zones du territoire. La probabilité de migrer d'une zone  $i$  vers une zone  $j$  dépend non seulement de la population  $i$ , mais également de la population  $j$ . Les méthodes proposées par Aalen s'étendent facilement à ce cas, car elles sont basées sur un modèle d'intensité multiplicative. Soient deux zones  $i$  et  $j$  (rural et agglomération parisienne, par exemple) dans lesquelles on observe, à la date  $t_1$ ,  $N_{1i}$  et  $N_{1j}$  individus ; on pourra, de la même façon que plus haut, estimer une intensité cumulée de migration de  $i$  vers  $j$  :

$$\hat{M}_{ij}(t_1) = \sum_{r=1}^1 \frac{M_{rij}}{N_{ri} N_{rj}}$$

et une variance :

$$\hat{V}\{\hat{M}_{ij}(t_1)\} = \sum_{r=1}^1 \frac{M_{rij}}{(N_{ri} N_{rj})^2}$$

Les tests de comparaison des divers flux seront les mêmes que précédemment.

## 5.2. Analyse paramétrique ou semi-paramétrique

On voit facilement que les méthodes non-paramétriques utilisées au paragraphe précédent vont rapidement être limitées par les effectifs enquêtés, et que des analyses très fines ne pourront être réalisées avec ces méthodes. On va maintenant chercher à utiliser des structures plus contraignantes, mais qui permettront de mettre en évidence les effets d'un grand nombre de variables.

Plaçons-nous directement dans le cas où l'on cherche à relier l'apparition d'un phénomène, ici la migration, à un certain nombre d'éléments de la vie des personnes interrogées. Nous laissons de côté des modèles plus purement paramétriques, qui cherchent à résumer une situation complexe par un petit nombre de paramètres dont la signification est souvent arbitraire.

Nous partons du quotient instantané de migration, qui est supposé lié à un certain nombre de caractéristiques de l'individu mesurées en début de séjour. On les représente par un vecteur de variables  $Z$ . Ce vecteur peut inclure des variables quantitatives (nombre d'enfants nés en début de période) et des variables qualitatives introduites sous la forme de variables binaires dont la valeur est 0 si l'individu n'est pas dans l'état considéré, 1 dans le cas contraire. Il est également utile d'introduire la durée de séjour.

Diverses formes peuvent être proposées pour donner le quotient instantané de migration en fonction de ces variables et de la durée de séjour. L'analyse non-paramétrique peut fournir des éléments sur cette forme. Nous proposons d'utiliser ici, à titre d'exemple, la forme suivante :

$$m_j(t; Z) = \exp(\alpha_j Z_1 + \beta_j Z_2 t)$$

où  $Z_1$  et  $Z_2$  sont des vecteurs de variables fournis par l'enquête, et  $\alpha_j$  et  $\beta_j$  les paramètres à estimer pour la destination  $j$ .

Cette estimation se fait par la méthode du maximum de vraisemblance. L'utilisation de cette méthode fournit toutes les informations utiles pour tester l'effet d'une ou plusieurs variables prises simultanément. En effet, la matrice des dérivées secondes de la fonction de vraisemblance par rapport aux paramètres du modèle donne une estimation par défaut de la matrice des variances et covariances des coefficients estimés. Dès lors, on pourra construire des tests sur l'effet d'une variable ou sur l'effet conjoint d'une série de variables.

Tableau 1

Effet de l'âge, de la durée de séjour, du cycle familial et du statut d'occupation du logement (a)

Variabiles	Modèle 1 (âge seul)	Modèle 2 (âge et durée de séjour)	Modèle 3 (âge, durée séj. et cycle famil.)	Modèle 4 (toutes variables)
Moins de 20 ans	4,350**	4,600**	3,002**	1,273
20-24 ans	5,865**	5,824**	4,317**	1,653
25-29 ans	2,856**	3,413**	3,294**	1,381
30-34 ans	2,921**	3,331**	3,267**	1,425
35-39 ans	2,953**	3,183**	3,184**	1,462
40-44 ans	1,801*	2,043**	2,018**	1,162
45-49 ans	1,607	1,695	1,632	1,079
Durée de séjour		-0,061**	-0,055**	-0,032**
Etat matrimonial			0,552**	0,486**
Nombre d'enfants			1,058	1,107*
Logé chez parents				0,732**
Logé par employeur				1,189*
Propriétaire				0,141**
Constante	0,057**	0,071**	0,108**	0,262**
$\chi^2$ (degrés de liberté)	181,41 (8)	270,69 (9)	303,04 (13)	494,57 (16)
Diff. de $\chi^2$ par rapport au précédent modèle (degrés de liberté)		89,28 (1)	32,35 (4)	191,53 (3)

(a) Les coefficients donnés ici pour les diverses variables, hormis la durée de séjour, sont les valeurs exp ( $\alpha$ ) qui donnent un effet multiplicatif pour un changement unitaire dans les variables. Un coefficient égal à 1 indique que la variable n'a aucun effet. Les valeurs marquées par \*\* ou par \* sont les coefficients dont la probabilité d'être différents de 1 est supérieure à 99 % ou à 95 %. Hormis la durée de séjour (en années) et le nombre d'enfants, toutes les autres variables sont des variables binaires.

Le tableau 1 fournit l'application de modèles de ce type (6) aux données de l'enquête "triple biographie". Un premier modèle met en évidence l'effet classique de l'âge sur la probabilité d'effectuer une nouvelle migration. Il s'agit d'un modèle du type précédent où  $Z_2 = 0$ . Nous avons porté, non pas les coefficients  $\alpha_j$  estimés, mais la valeur de  $\exp(\alpha_j)$ , qui donne l'effet multiplicatif d'une variable sur le quotient instantané, supposé ici indépendant du temps. Ainsi, par rapport aux personnes interrogées de plus de 50 ans, celles dont le séjour débute entre 20 et 24 ans ont un quotient instantané près de six fois plus élevé.

Un second modèle met en évidence un effet de durée de séjour en posant  $Z_2 = 1$ . La valeur du coefficient  $\beta$  est négative, comme on s'y attendait, et l'introduction de cette variable améliore fortement le modèle ( $\chi^2 = 89,29$ , avec 1 degré de liberté). On voit qu'après une durée de dix ans, la probabilité de migrer est divisée par deux. Notons que l'effet de l'âge reste, à peu de chose près, le même.

Le troisième modèle introduit des variables familiales, essentiellement l'état matrimonial en début de période, ainsi que le nombre d'enfants. Le fait d'être marié divise par deux le quotient, toutes choses étant égales par ailleurs, tandis que le nombre d'enfants n'a un effet significatif que dans le modèle suivant.

Le quatrième modèle fait intervenir le statut d'occupation du logement. Une fois cette variable prise en compte, on voit que le fait d'avoir sept enfants multiplie par deux le quotient de mobilité et a un effet significatif. Le statut d'occupation a un rôle très important : le fait de devenir propriétaire divise par sept le quotient. Dans ce modèle, l'effet de l'âge disparaît complètement du fait de l'intervention des nouvelles variables introduites.

Il est possible de tester l'identité des coefficients correspondant aux divers groupes d'âges en utilisant la matrice des variances et covariances  $A$  de ces paramètres  $\alpha$ . Pour le deuxième modèle, nous avons un  $\chi^2$  à sept degrés de liberté, calculé comme le produit  $\alpha' A^{-1} \alpha$  où  $\alpha'$  est le transposé du vecteur-colonne  $\alpha$ , égal à 140,61, qui est tout à fait significatif. Pour le quatrième modèle, ce  $\chi^2$  n'est plus égal qu'à 9,84, valeur non significative pour 7 degrés de liberté.

Ces modèles peuvent s'étendre dans deux directions. La première consiste à introduire une part non-paramétrique qui

---

(6) Nous avons utilisé pour ce faire le programme Rate, mis au point par N.B. Tuma, Stanford University.

permet de tenir compte de l'effet des variables non explicitement considérées. On peut les écrire sous la forme :

$$m_j(t;Z) = \lambda_0(t) \exp(\alpha_j Z_1 + \beta_j Z_2 t)$$

où  $\lambda_0(t)$  est une fonction identique pour toutes les personnes interrogées. L'estimation des paramètres  $\alpha$  et  $\beta$  est encore possible en utilisant une méthode de vraisemblance partielle mise au point par Cox (1972). La seconde extension consiste à décomposer la durée de séjour en courtes périodes successives, de telle sorte que les variables introduites puissent changer d'une période à la suivante. Cette méthode fournit une grande souplesse et permet d'introduire des dépendances plus fines entre états.

## CONCLUSION

Nous avons essayé de soulever, et parfois de résoudre, le plus grand nombre de problèmes posés par les enquêtes de migration dans les pays développés. Nous avons cependant dû aborder de façon sommaire certaines questions qui auraient souvent nécessité, à elles seules, un article détaillé. Nous pensons en particulier aux problèmes liés à la pondération de l'échantillon, pour lesquels nous renvoyons à l'article de J. Hoem (1983), aux non-réponses (voir à ce sujet l'article de I. Lyberg, 1983), aux méthodes d'analyse, dont nous avons volontairement limité le champ...

Parmi les questions soulevées auxquelles aucune réponse nette n'a pu être donnée, certaines apparaissent comme fondamentales à résoudre dans un proche avenir, car elles conditionnent le devenir de ces enquêtes.

En premier lieu, les problèmes de fiabilité nécessitent encore des tests très approfondis des réponses données par les personnes interrogées, des comparaisons avec diverses sources statistiques. La mise au point de procédures d'enquêtes garantissant au mieux cette fiabilité (confrontation des réponses données par plusieurs membres du ménage, utilisation de l'information écrite dont peut disposer le répondant sur ces événements...) devrait être un des objectifs prioritaires de recherche.

En second lieu, la mise au point de méthodes d'analyse permettant d'utiliser au mieux l'information fournie et éliminant au maximum le bruit de fond lié au temps flou des personnes interrogées, est un autre point important. Il se peut qu'un certain nombre de méthodes nouvelles, que nous avons présentées ici, soient invalidées par l'existence de cette incertitude.

En dépit de ces questions encore ouvertes et auxquelles, je pense, cette Chaire Quetelet apportera, sinon une réponse définitive, au moins un début de réponse, les enquêtes, réalisées avec le maximum de précautions, continueront certainement à fournir les éléments nécessaires à la compréhension de notre société.

### Annexe 1

Présentation rapide des principales enquêtes citées dans cette communication

#### Danemark

Enquête sur les motifs de migration, réalisée en 1979 par l'Office de Statistiques de Copenhague et l'Institut de Recherches Sociales (C. Woll), en vue d'analyser les émigrations et les immigrations dans la municipalité de Copenhague. Un échantillon de 3 200 ménages ayant effectué une telle migration a été tiré en 1978 de registres de population. L'enquête a été réalisée, soit par des entretiens personnels (911 questionnaires), soit par des questionnaires envoyés par la poste : 75 % des questionnaires ont été renvoyés, 62 % seulement étaient correctement remplis.

#### Etats-Unis

Hopkins Life History Study, réalisée en 1968 par l'Université Johns Hopkins (J. Coleman et P. Rossi) sur un échantillon représentatif de cohortes nées en 1929 et 1939 aux Etats-Unis. Un échantillon de 1 589 entretiens personnels a été obtenu avec des taux de non-réponse de 23,9 % pour la population noire (738 questionnaires) et 21,8 % pour la population blanche.

Wisconsin Basic Need Study, réalisée en 1981 par le laboratoire de recherches du Wisconsin (D. Featherman) sur un échantillon représentatif de la population présente dans l'état de Wisconsin. Un échantillon de 2 000 familles a été initialement enquêté, et des entretiens ont ensuite eu lieu par la poste ou le téléphone. Le suivi s'est fait à la fois sur les logements et sur les personnes interrogées lors du premier passage.

#### France

Enquête sur les émigrants de l'Ouest audois, réalisée en 1977 par l'I.N.E.D. (P. Collomb) sur un échantillon

d'individus présents dans la région en 1954, ayant changé de commune de résidence entre les recensements de 1954 et 1962. Une liste de 2 041 migrants, encore en vie en 1977, a été établie à partir des listes nominatives des deux recensements. Le pourcentage de migrants non retrouvés s'élève à 1,8 %. Tous les autres ont été interrogés par des enquêteurs.

Enquête 'triple biographie', réalisée par l'I.N.E.D. en 1981 (D. Courgeau et B. Riandey) sur un échantillon représentatif d'individus âgés de 45 à 69 ans et présents en France. Un échantillon de 4 571 questionnaires a été obtenu avec un taux approximatif de non-réponse de 11 %. Il a également été demandé aux enquêteurs de répondre aux questionnaires (358 réponses). Le même questionnaire a été utilisé en Belgique (J. Duchêne et M. Poulain) pour tester la fiabilité des réponses (50 questionnaires).

Enquête sur les mobiles de migration entre Paris et la province, réalisée en 1981 par l'I.A.U.R.I.F. (J. Foucher et M. Fleury) sur un échantillon de 3 000 personnes ayant été radiées des listes électorales de Paris en 1978 ou s'étant inscrites sur une liste de la région parisienne au cours de la même année. La méthode de sondage avec remplacement conduit à réaliser le nombre d'enquêtes initialement prévu, mais le taux de remplacement a été élevé.

### Norvège

Norwegian Life History Study, réalisée en 1971 par l'Université d'Oslo (N.R. Ramsoy) sur un échantillon d'hommes appartenant aux cohortes nées en 1921, 1931 et 1941. Cet échantillon a été tiré du registre de population, et les taux de non-réponse ont été respectivement de 17 %, 13,8 %, et 16,8 %.

### Bibliographie

AALEN, O. (1978), Nonparametric inference for a family of counting processes, The Annals of Statistics, 4, 701-726.

AALEN, O., O. BORGAN, N. KEIDING & J. THORMAN (1980), Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring, Scand. J. Stat., 7, 161-171.

- COLEMAN, J. (1981), Longitudinal Data Analysis, Basic Books, New York.
- COLLOMB, P. (1982), The outmigrants of West Aude, 19 years later, 13e réunion I.A.R.U.S., III, 3-29.
- COURGEAU, D. (1980), Analyse quantitative des migrations humaines, Masson, Paris.
- COX, D. (1972), Regression models and life tables, J. Roy. Stat. Soc., B 34, 187-220.
- DEVILLE, J.C. (1974), Méthodes statistiques et numériques de l'analyse harmonique, Annales de l'I.N.S.E.E., 15.
- DUCHENE, J. & M. POULAIN (1982), Interaction de la mobilité résidentielle et du cycle de vie des individus et des ménages, Communication au 7e Colloque national de démographie 'Les âges de la vie', Strasbourg, 5-7 mai 1982.
- FINDLEY, S. (1982), Migration survey methodologies : a review of design issues, I.U.S.S.P. Papers, 20.
- FLEURY, M. (1982), Présentation générale de l'enquête sur les mobiles de migration, 13ème réunion I.A.R.U.S., III, 187-206.
- GINSBERG, R. (1979), Timing and duration effects in residence histories and other longitudinal data, Reg. Sci. Urb. Econ., 311-331.
- HOEM, J. (1983), Weighting, misclassification, and other issues in the analysis of survey samples of life histories, Stockholm Res. Rep. in Demo., 11.
- JACOBSEN, M. (1982), Maximum-likelihood estimation in the multiplicative intensity model, Preprint, 4, University of Copenhagen.
- KALBFLEISCH, J. & R. PRENTICE (1980), The Statistical Analysis of Failure Time Data, Wiley & sons, New York.
- KAPLAN, E. & P. MEIER (1958), Nonparametric estimation from incomplete observation, J. Am. Stat. Assoc., 53, 457-481.
- LONG, J. & C. BOERTLEIN (1981), Using Migration Measures Having Different Intervals, en cours de publication.
- LYBERG, I. (1983), Nonresponse effects on survey estimates in the analysis of competing exponential risks, Stockholm Res. Rep. in Demo., 12.

- POULAIN, M. (1978), Du registre de population aux statistiques de migration interne en Belgique, Population et Famille, 45, 1-45.
- RAMSOY, R. (1977), Social mobilitet i Norge, Oslo.
- TUMA, N., M. HANNAN & L. GROENEVELD (1979), Dynamic analysis of event histories, Amer. J. Sociol., 84, 820-854.
- WOLL, C. (1982), The accomplishment of the migration survey 1979 with the greatest importance attached to the elaboration of the questionnaire, 13e réunion I.A.R.U.S., III, 31-66.