L'ANALYSE DES ENQUETES RETROSPECTIVES

Daniel COURGEAU

Institut National d'Etudes Démographiques
Paris, France

1. Introduction

Les enquêtes rétrospectives ont déjà une longue existence en démographie. Ainsi, en 1961, G. Pourcher (1964) avait interrogé un échantillon d'individus présents à Paris sur diverses étapes de leur existence passée. Il avait demandé des éléments de la vie familiale (origines familiales, dates de mariage, de naissance des enfants successifs, etc.), du cheminement migratoire vers Paris (date et destination des diverses migrations) et de la vie professionnelle (professions successives et dates des changements). Son exploitation a permis de fournir des informations très précieuses sur les causes, les circonstances et les conséquences des migrations vers Paris.

Cependant, cette exploitation réalisée à l'aide des méthodes classiques d'analyse, telles que l'utilisation de tableaux croisés, faisait perdre une partie de l'information recueillie : la date des divers événements et les interactions entre les diverses étapes de la vie des enquêtés. L'utilisation de cette information va, en fait, nécessiter une nouvelle réflexion sur les méthodes d'analyse utilisées. Plusieurs raisons expliquent ce fait.

En premier lieu, le démographe avait l'habitude d'utiliser des données exhaustives de recensement ou d'état civil, et les méthodes d'analyse ont été élaborées pour de telles données. Il s'était assigné la tâche d'isoler chaque phénomène démographique "à l'état pur", c'est-à-dire de séparer l'influence du phénomène étudié de celle des phénomènes perturbateurs (Henry, 1972). Cette attitude est tout à fait légitime lorsque le phénomène perturbateur est la mortalité. Il élimine, en effet, toute information sur ce qu'aurait pu être le comportement des individus s'ils avaient survécu. On peut, dans ce cas, faire l'hypothèse d'indépendance entre la mortalité et le phénomène étudié. Ainsi, pour estimer des quotients de nuptialité, on fait l'hypothèse que les individus qui meurent célibataires auraient eu le même comportement matrimonial que ceux qui survivent (Pressat, 1966). Cependant, lorsque le phénomène perturbateur n'est pas la mortalité, on a la possibilité, à l'aide d'enquêtes, par exemple, d'observer le comportement des individus qui ont déjà connu ce phénomène. Prenons le cas de l'émigration internationale, souvent considérée comme phénomène perturbateur, de la même façon que la mortalité, lorsque l'on analyse la nuptialité, par exemple. On a, en fait, la possibilité d'observer la nuptialité de ces émigrants. Le problème se pose alors "de substituer leur nuptialité à l'étranger à celle qu'ils auraient eue s'ils étaient restés, alors que cette nuptialité à l'étranger dépend de conditions qui peuvent être différentes" (Henry, 1972, p. 77). La solution de ce problème se trouve dans la possibilité de tester l'hypothèse d'indépendance et de lui substituer, lorsqu'elle n'est pas vérifiée, d'autres formes de dépendance, comme nous le verrons plus loin.

Or la mise en place de tels tests nécessite des méthodes statistiques dont la démographie classique ne s'était guère préoccupée. Il faut voir que, travaillant sur des populations exhaustives et sur des phénomènes "à l'état pur", la taille des effectifs soumis au risque enlevait tout intérêt au calcul de la variance des taux ou quotients estimés. L'utilisation d'enquêtes portant sur de faibles échantillons va rendre nécessaire le calcul de cette variance. Bien plus, si l'on désire analyser les dépendances entre phénomènes démographiques, certains taux ou quotients auront des dénominateurs qui pourront être très faibles, même en travaillant sur des populations exhaustives. Ainsi, en reprenant l'exemple précédent de la nuptialité des émigrants internationaux, cette émigration pourra porter sur de très faibles effectifs. Dans ce cas, le calcul de quotients de nuptialité des émigrants va nécessiter simultanément le calcul d'une variance de ces quotients.

On voit donc que les données recueillies par des enquêtes biographiques rétrospectives nécessitent une nouvelle réflexion sur les méthodes d'analyse démographique aptes à saisir correctement les interactions entre les divers phénomènes.

2. Bases théoriques de cette analyse

En premier lieu, l'objet propre de l'analyse va changer. En effet, les méthodes démographiques classiques isolaient chaque *phénomène* "à l'état pur", en vue de l'analyser indépendamment des autres. Cela paraît très clairement dans tous les manuels classiques (Pressat, 1966 ; Henry, 1972), où chaque phénomène fait l'objet d'un chapitre séparé : nuptialité, fécondité, mortalité, migration. L'objet de l'analyse proposée ici sera tout autre. L'observation rétrospective fournie par l'enquête permet de recueillir la *biographie* des enquêtés, succession d'événements survenus à un même individu tout au long de son existence. Il s'agira, dès lors, d'analyser comme un tout cohérent de telles biographies.

Dans la mesure où l'objet de l'analyse a changé, il est évident qu'il faudra introduire des modifications dans les méthodes d'analyse. Pour ce faire, nous allons considérer une biographie comme le résultat d'un processus stochastique complexe. Lors du déroulement du temps, l'individu connaît successivement divers événements dont le type peut être différent : par exemple, l'individu peut d'abord effectuer une migration avant de prendre un premier emploi, suivi par un mariage et une seconde migration simultanée, puis il aura successivement deux enfants avant d'effectuer une troisième migration, etc. On voit donc que l'on peut caractériser chacun de ces événements par la date à laquelle il se produit et par le type d'événement dont il s'agit (mariage, naissance, migration, changement professionnel, etc.). On peut, dès lors, représenter l'existence d'un individu par une série de couples de variables aléatoires (T_i, J_i) où T_i est l'instant d'arrivée du i^e événement de type J_i. On voit que les divers instants vérifient la relation $T_1 \leq T_2 \leq T_3$... On peut, dès lors, définir le quotient instantané pour le i^e événement de type j connu pour un individu, à l'instant t :

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(T_i < t + \Delta t, J_i = j \mid T_i \geq t \,;\, t_1, j_1; \ldots ; t_{i-1}, j_{i-1}) \qquad [1]$$

où t_k représente l'instant d'arrivée du k^e événement de type j_k. Selon les individus observés, l'ordre d'arrivée des divers types d'événements va changer, de sorte que l'on peut calculer le quotient instantané pour le i^e événement, quel que soit son type, sous la forme suivante :

$$h_i(t) = \sum_{j=1}^{m} h_{i,j}(t) \qquad [2]$$

où m est le nombre total de types d'événements considérés.

On peut, dès lors, pour un individu observé, calculer de proche en proche une fonction de vraisemblance correspondant aux divers événements de sa biographie. Pour la première période observée [0, t_1], nous devons introduire la probabilité qu'il n'ait connu aucun événement. On vérifie aisément que cette probabilité est égale à :

$$\exp\left[- \int_0^{t_1} h_1(t)\, dt \right].$$

Ensuite l'individu connaît l'événement de type j_1 à l'instant t_1, dont la probabilité est $h_{1,j_1}(t)\, dt$. A partir du couple (t_1, j_1), on peut calculer la probabilité de ne connaître aucun événement au cours de la période (t_1, t_2), qui est égale à :

$$\exp\left[- \int_{t_1}^{t_2} h_2(t)\, dt \right].$$

On peut continuer cette procédure jusqu'à l'instant de l'enquête.

La contribution à la vraisemblance d'un individu qui a connu m événements est donc égale à :

$$\prod_{i=1}^{m} \left[\exp\left(- \int_{t_{i-1}}^{t_i} h_i(t)\, dt\right)\right] \left[h_{1j_1}(t_i)\right]^{\delta_i} \qquad [3]$$

où t_m est la durée totale de l'observation de l'individu, δ_i une variable égale à l'unité sauf pour la fin du dernier intervalle ouvert t_m, où elle est égale à zéro. En faisant le produit de toutes les vraisemblances individuelles, on obtient la vraisemblance de l'échantillon observé. Pour rendre l'estimation des divers quotients possible, on va supposer qu'ils gardent la même valeur au cours d'un intervalle de temps le plus souvent pris égal à une année. L'application de la méthode du maximum de vraisemblance conduit alors à des estimations de $h_{ij}(\theta)$ et de la variance de ces quotients pour l'année θ.

En fait, étant donné le nombre limité d'individus enquêtés et surtout le nombre très élevé de quotients à estimer, l'analyste ne peut, pour le moment, estimer cette distribution conjointe dans son ensemble. Il se contentera d'approches partielles qui, sous certaines hypothèses, permettent l'estimation de ces quotients.

Nous allons maintenant présenter deux de ces approches, directement applicables à des données d'enquêtes rétrospectives.

3. Le modèle bivarié et ses généralisations

Plutôt que de faire intervenir l'ensemble des événements connus par les enquêtés, on va ici se concentrer sur l'interaction entre deux types d'événements. On verra dans la section suivante comment faire intervenir l'effet des autres événements en utilisant un modèle de régression.

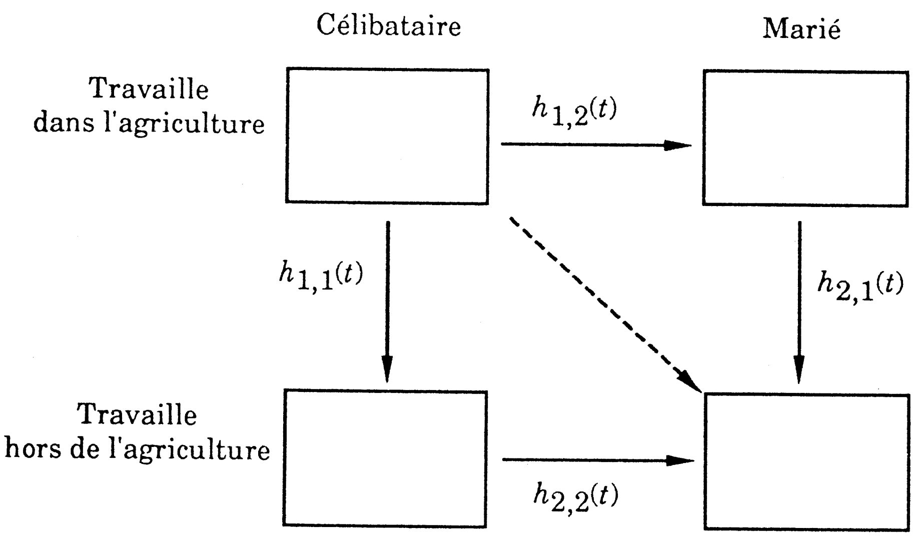

Figure 1. Modèle bivarié

La figure 1 montre un exemple de cette analyse appliquée à l'interaction entre le mariage et le départ du monde agricole (Courgeau et Lelièvre, 1986). Avec les notations de la section précédente, on voit que l'on n'a plus que quatre séries de quotients à estimer pour les deux événements : départ de l'agriculture ($j = 1$), mariage ($j = 2$). Ainsi, le quotient instantané de sortie de l'agriculture des célibataires s'écrit :

$$h_{1,1}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(T_1 < t + \Delta t, J_1 = 1 \mid T_1 \geq t). \qquad [4]$$

Le quotient de sortie de l'agriculture des individus déjà mariés s'écrit :

$$h_{2,1}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(T_2 < t + \Delta t, J_2 = 1 \mid T_2 \geq t\,;\, t_1, 1) \quad \text{où } t_1 \leq t. \qquad [5]$$

Des quotients de nuptialité similaires peuvent s'écrire pour la nuptialité des individus restés dans le monde agricole, $h_{1,2}(t)$, ou sortis de ce monde $h_{2,2}(t)$. Si l'on suppose ces divers quotients constants au cours d'une année d'observation et si l'on suppose $h_{2,1}(t)$ et $h_{1,2}(t)$ indépendants de la date d'arrivée de l'événement précédent, l'estimation de ces quotients et de leur variance est aisée (Courgeau et Lelièvre, 1986). On pourra, dès lors, tester les deux égalités :

$$h_{1,1}(t) = h_{2,1}(t) \qquad [6]$$

$$h_{1,2}(t) = h_{2,2}(t). \qquad [7]$$

Nous avons ainsi pu montrer, dans l'exemple considéré, que, pour les hommes, la première égalité n'est pas vérifiée, alors que la seconde l'est. On met ainsi en évidence une *dépendance locale* d'un des phénomènes sur l'autre. Pour les hommes, le processus fondamental est leur vie professionnelle. Leur nuptialité dépendra alors du choix qu'ils ont fait de rester dans le monde agricole ou d'en partir. Le fort taux de masculinité du milieu agricole et l'absence d'échanges matrimoniaux entre femmes des autres milieux et agriculteurs va, dès lors, expliquer la très faible nuptialité des agriculteurs et l'augmentation de cette nuptialité lorsqu'ils quittent le milieu agricole. Dans l'autre sens, le fait qu'ils soient mariés ou célibataires n'influe pas sur le départ du monde agricole. On voit donc que ce concept de dépendance locale formalise la notion intuitive qu'un processus stochastique peut influer localement sur le développement d'un autre processus à la date t, le second pouvant n'avoir aucune influence sur le premier. Il est intéressant de voir que, pour les femmes, la dépendance est encore locale, mais joue dans l'autre sens : la vie familiale l'emporte, chez les femmes, sur un projet professionnel.

Une telle analyse va donc permettre de montrer divers niveaux de dépendance entre deux phénomènes. Le niveau le plus bas correspond à la vérification des deux égalités [6] et [7] pour toutes les durées d'observation. On peut conclure, dans ce cas, à une *indépendance totale* entre les deux phénomènes. Pour le moment, aucune des analyses que nous avons réalisées ne nous a montré une telle indépendance. On voit donc le danger de traiter les phénomènes démographiques séparément, sous l'hypothèse d'une indépendance entre eux. Le second niveau correspond à la *dépendance locale*, que nous avons déjà montrée. Dans ce cas, une seule des deux égalités [6] ou [7] est vérifiée pour toutes les durées d'observation. Le troisième niveau correspond à une *dépendance réciproque* des deux phénomènes. Nous avons mis en évidence une telle dépendance réciproque en analysant la fécondité des femmes en liaison avec la migration vers des zones très urbanisées (Courgeau, 1987). La venue d'un enfant de rang deux ou plus est réduite lorsque la femme effectue une migration vers une zone très urbanisée. La migration vers des zones très urbanisées est réduite après la naissance, quel que soit son rang. Dans ce cas, aucune des égalités [6] et [7] n'est plus vérifiée pour toutes les durées d'observation.

Bien entendu, des cas intermédiaires peuvent se produire, où l'on peut avoir une dépendance locale jusqu'à un certain âge, suivie d'une dépendance réciproque. Egalement, lorsque l'on fait intervenir la durée entre deux événements, cette dépendance peut n'être vérifiée que lorsque les événements sont proches et disparaître après une durée plus longue. Enfin, une telle analyse permet de mettre en évidence d'autres formes de dépendance. Nous avons déjà indiqué que la fécondité des femmes migrant vers des zones très urbanisées était plus faible que celle des femmes vivant en milieu peu urbanisé. On peut, dès lors, se demander si la fécondité des femmes vivant en milieu peu urbanisé, mais qui migreront à l'avenir vers une zone très urbanisée, n'est pas déjà différente de celle des femmes qui vivront toute leur vie en milieu peu urbanisé. Un tel test est tout à fait réalisable avec nos données d'enquêtes rétrospectives. Nous avons ainsi pu montrer que les femmes (de moins de 35 ans pour les naissances de rang trois) qui émigreront vers des zones très urbanisées ont, avant leur migration, un calendrier de fécondité significativement différent de celui des femmes qui restent sédentaires en zone peu urbanisée. Bien plus, ce calendrier n'est pas différent de celui des femmes ayant déjà émigré vers ces zones (Courgeau, 1987). On met ainsi en évidence une *dépendance a priori* de la fécondité des femmes sur leur migration à venir. Dans ce cas, on peut avancer une hypothèse de sélectivité des migrantes qui ont déjà, parmi d'autres caractéristiques, une faible fécondité, comparée à celle des femmes qui resteront toute leur vie en milieu peu urbanisé. Il est intéressant de voir que, pour la migration dans l'autre sens, des zones très urbanisées vers les zones peu urbanisées, les femmes ont la même fécondité que celles qui resteront toute leur vie en zone très urbanisée. Ce n'est qu'une fois leur migration effectuée que l'on observe une forte remontée de leur fécondité.

Un autre problème est également intéressant à considérer. Il intervient lorsque les événements simultanés sont en nombre non négligeable. Il est représenté par la flèche diagonale dans la figure 1. On pourrait dire qu'il suffit de prendre un découpage temporel suffisamment fin pour supprimer ces cas de simultanéité : en considérant le mois, le jour, l'heure, la minute, etc., de chaque événement, il serait possible de distinguer celui qui précède l'autre. A notre avis, une telle distinction paraît vaine, car l'important n'est pas l'instant précis où un événement survient, mais, bien plus, l'instant où la décision a été prise d'effectuer cet acte. Il n'est guère possible, dans une enquête rétrospective, de demander cet instant de décision que seule une analyse psychologique très poussée pourrait mettre en évidence. C'est la raison pour laquelle nous avons dû introduire la notion de "temps flou", pour indiquer l'impossibilité de discerner si l'instant de décision d'un événement a précédé, a été simultané ou a suivi l'instant de décision d'un autre événement. Ainsi, en définissant comme simultanés deux événements qui se produisent la même année dans l'analyse précédente de la nuptialité et du départ de l'agriculture, les simultanéités touchent 6,7 % des mariages et 8,4 % des sorties de l'agriculture chez les hommes, et 5,1 % des mariages et 10,3 % des sorties de l'agriculture chez les femmes (Courgeau et Lelièvre, 1986).

En éliminant ces simultanéités de l'analyse des quatre séries de quotients, on peut penser être mieux en mesure de montrer les comportements des individus pour lesquels les décisions de quitter l'agriculture et de se marier n'ont pas été simultanées. Bien entendu, on peut faire varier cette période de simultanéité (six mois, un an, un an et demi, deux ans, etc.) pour voir la stabilité des comportements.

A partir du modèle bivarié, il est possible de le généraliser de diverses façons. Une première généralisation va considérer, non plus deux événements, mais deux séries d'événements. On peut, par exemple, suivre, à partir du mariage d'un individu, ses migrations successives et les naissances de ses enfants. On voit facilement que, même en ne tenant pas compte des dates des événements antérieurs, le nombre des séries de quotients à estimer va vite devenir très élevé. Si l'on considère m migrations et n enfants, il faudra estimer $m(n+1) + (m+1)n$ séries de quotients. Dans ce cas, les faibles effectifs enquêtés ne permettent l'estimation de tous ces quotients qu'avec une précision très grossière. Il sera donc nécessaire de faire des hypothèses pour n'avoir à estimer qu'un petit nombre de séries de quotients. Ainsi, dans l'exemple des naissances et migrations successives, on peut faire l'hypothèse que la migration d'un couple dépend du nombre d'enfants déjà nés, mais non des dates des naissances antérieures, ni du nombre de migrations antérieures ni de leur date (Courgeau, 1985). Dans ce cas, on met en évidence une pression de la taille de la famille sur la migration vers un logement plus grand. De même, on peut faire l'hypothèse que, dans l'autre sens, la naissance d'un nouvel enfant dépend du nombre de migrations déjà faites, mais non des dates de ces migrations, ni du nombre d'enfants nés antérieurement, ni des dates de naissance de ces enfants. Dans ce cas, on peut mettre en évidence des migrations vers un logement plus grand qui permette la venue de nouveaux enfants. Avec ces hypothèses, on voit que le nombre de séries de quotients à estimer est fortement réduit à $(m+n)$ au lieu de $(2mn + m + n)$. Une telle analyse a permis de montrer que, pour les femmes mariées avant 23 ans, les deux pressions existent, mais que seule l'existence des migrations prévoyant l'accroissement de la taille de la famille joue pour les femmes mariées plus âgées. Ces femmes, de milieu social plus élevé, ont la possibilité de prévoir un logement de taille suffisante pour leur descendance désirée.

Bien entendu, ces hypothèses semblent trop restrictives, et il serait utile de lever certaines d'entre elles pour avoir un modèle plus réaliste (Courgeau, 1985). Malheureusement, la taille de l'échantillon vient limiter ces investigations.

Une autre généralisation est de considérer des modèles trivariés qui permettent de mieux approcher le modèle multivarié présenté précédemment. A nouveau, le nombre élevé de quotients à estimer (12) vient limiter la poursuite des recherches dans cette direction. C'est la raison pour laquelle nous nous tournons maintenant vers un autre type de modèles.

4. Modèles de régression

Les modèles présentés dans la section précédente sont intéressants à considérer dans la mesure où ils n'imposent aucune forme a priori aux quotients estimés. Nous avons cependant vu que, lorsqu'on essaie de faire intervenir des interactions plus complexes que celles du modèle bivarié, la taille de l'échantillon vient vite rendre sans intérêt les quotients que l'on pourrait estimer.

On voit, dès lors, l'intérêt d'utiliser des modèles plus formalisés, qui permettent de faire intervenir un grand nombre de caractéristiques individuelles, qui pourront d'ailleurs varier au cours du temps. Il ne peut être question de présenter ici ces modèles dans leur ensemble, qui ont connu un fort développement au cours des années récentes (Kalbfleisch et Prentice, 1980 ; Cox et Oakes, 1984).

Nous allons poursuivre ici la généralisation du modèle bivarié présenté dans la section précédente, de façon à faire intervenir un grand nombre de caractéristiques individuelles.

Reprenant l'exemple du mariage et du départ du monde agricole, nous allons supposer que les quotients estimés vont dépendre de diverses caractéristiques individuelles, de façon multiplicative. Cela permet de tenir compte de la diversité des situations dans lesquelles se trouvent les divers enquêtés et de la variété des trajectoires suivies dans les autres domaines de leur biographie. Ces caractéristiques se présentent sous la forme d'un vecteur $z = (z_1, ..., z_n)$ qui pourra s'étendre sur un plus grand nombre de variables une fois que l'individu se sera marié ou aura quitté le monde agricole : $z' = (z_1, ..., z_n, z_{n+1}, ..., z_s)$. Ces variables pourront être des variables binaires indiquant, par exemple, que l'individu est dans un état donné ou non lors du mariage ou du départ du monde agricole, ou des variables quantitatives donnant, par exemple, la taille de sa fratrie.

Le modèle proposé ici est un modèle à *risques proportionnels* : le quotient instantané d'un individu donné est proportionnel à un certain quotient applicable à l'ensemble de la population. Le coefficient de proportionnalité est lié aux diverses caractéristiques de l'individu.

Pour formaliser un tel modèle, prenons le cas du départ de l'agriculture selon que l'individu est célibataire ou marié. Dans le premier cas, le quotient d'un individu ayant les caractéristiques z peut s'écrire :

$$h(t, z) = h_0(t) \exp (z\, \beta_1) \qquad [8]$$

où $h_0(t)$ est le quotient applicable à toute la population, et β_1 est un vecteur colonne de paramètres à estimer, mesurant l'effet des diverses caractéristiques sur le quotient $h(t, z)$.

Si l'individu est maintenant marié, nous pouvons écrire, pour la même durée t, son quotient de départ de l'agriculture :

$$h'(t, z) = h_0(t) \exp (z\, \beta_1 + \beta_0 + z'\, \beta_2) \qquad [9]$$

où $h_0(t)$ est le même quotient que le précédent, β_1, β_0 et β_2 des paramètres à estimer mesurant l'effet des diverses caractéristiques avant ou après le mariage. On voit que ce modèle peut s'écrire plus simplement sous la forme :

$$h(t, z, z') = h_0(t) \exp [z \beta_1 + H(t - t_1) (\beta_0 + z' \beta_2)] \qquad [10]$$

où t_1 est la date de mariage de l'individu et $H(t - t_1)$ est la fonction de Heaviside, nulle si $t \leq t_1$ et égale à l'unité si $t > t_1$. Les paramètres β_0, β_1 et β_2 peuvent être estimés par les méthodes de vraisemblance partielle proposées par Kalbfleisch et Prentice (1980).

Un autre type de modèle peut être intéressant à considérer ici. Dans ce modèle, on va se centrer sur l'arrivée d'un événement de la biographie et faire intervenir les diverses étapes du cycle de vie comme des variables dépendant du temps $z(t) = [z_1(t), ..., z_s(t)]$. Dans ce cas, le modèle peut s'écrire :

$$h[t, z(t)] = h_0(t) \; \exp [z(t) \beta]. \qquad [11]$$

A nouveau, les paramètres β peuvent être estimés par les méthodes de vraisemblance partielle proposées par Kalbfleisch et Prentice (1980). On voit que ces paramètres mesurent les modifications dans la probabilité de connaître un événement donné, lorsque l'on traverse les diverses étapes du cycle de vie. Les hypothèses faites ici sont cependant très fortes. Ainsi, ces modifications sont indépendantes de l'ordre d'arrivée des autres événements, elles ont un effet indépendant de l'âge auquel elles surviennent, et elles agissent de façon multiplicative sur le quotient. Seules les recherches à venir sur des échantillons suffisamment importants permettront de voir si ces hypothèses sont vraiment réalisées et, sinon, quelles autres hypothèses prendre.

5. Conclusions

Ce bref survol des méthodes d'analyse des enquêtes biographiques rétrospectives nous a permis de montrer le vaste champ de recherche qu'elles ont ouvert. L'exploration de ce champ en est à ses débuts, mais des résultats déjà très prometteurs ont été obtenus.

Elles nous permettent, en premier lieu, de prolonger l'analyse démographique classique des phénomènes "*à l'état pur*" en introduisant une analyse des phénomènes "*en interaction*". Il est possible, à l'aide d'enquêtes rétrospectives, de vérifier si, par exemple, la fécondité des sédentaires est la même que celle des migrantes. L'*hypothèse d'indépendance* entre phénomènes, que l'on avait dû faire en utilisant des données démographiques classiques, s'est révélée non vérifiée dans de nombreux cas. Nous avons alors pu mettre en évidence des *dépendances* plus complexes entre phénomènes : dépendance locale, dépendance réciproque, dépendance a priori. Cette dernière forme de dépendance, lorsqu'elle existe, pose le problème de l'antériorité causale. Dans ce cas, en effet, ce n'est plus un ordre temporel qui est pris en considération, mais beaucoup plus un ordre logique. Ainsi, nous avons montré que les femmes qui ont migré ou qui migreront vers des zones très urbanisées ont une fécondité plus faible que celles qui resteront toute leur vie dans des zones peu urbanisées. Notons également que toutes ces dépendances ne sont pas *déterministes*, mais *stochastiques*. Cela nous évite d'introduire une causalité en sciences sociales, qui ne peut en aucun cas être démontrée.

Nous avons également pu aborder le problème difficile du lien entre la date de l'événement observé et la date de la prise de décision, qui intervient en particulier dans le cas des événements dits *simultanés*. Nous avons proposé diverses solutions approchées de ce problème, qui ne peut être traité qu'en liaison avec des psychologues.

Lorsque nous avons essayé de faire intervenir des variables plus nombreuses pour voir leur influence sur un phénomène donné, nous avons été contraint d'utiliser un modèle plus restrictif. Le modèle multiplicatif proposé ici n'est que l'un des nombreux types de modèles que l'on peut proposer. Il présente certains avantages (quotients estimés toujours positifs) et a montré sa supériorité sur d'autres types de modèles dans des cas particuliers. Nous avons ainsi effectué des comparaisons avec un modèle de type additif, qui ont montré un moins bon ajustement des données (résultats non publiés). Il est évident que la recherche des modèles les mieux adaptés aux observations est un objectif essentiel, qui ne pourra être réalisé qu'avec une confrontation permanente entre modèles non-paramétriques et modèles paramétriques. L'utilisation de modèles semi-paramétriques, proposée ici, permet de résoudre, en partie seulement, ces problèmes. A nouveau, une approche pluridisciplinaire, avec des sociologues, en particulier, enrichirait cette analyse où de nombreuses hypothèses faites restent à vérifier.

Références

COURGEAU, D. (1985), "Interaction between spatial mobility, family and career life cycle", *European Sociological Review*, 1, n° 2, p. 139-162.

COURGEAU, D. (1987), "Construction de la famille et urbanisation", *Population*, 42, n° 1, p. 57-82.

COURGEAU, D. et E. LELIEVRE (1986), "Nuptialité et agriculture", *Population*, 41, n° 2, p. 303-326.

COX, D. et D. OAKES (1984), *Analysis of survival data*, Londres, Chapman and Hall.

HENRY, L. (1972), *Démographie. Analyse et modèles*, Paris, Larousse.

KALBFLEISCH, J. et R. PRENTICE (1980), *The statistical analysis of failure time data*, New York, John Wiley and Sons.

POURCHER, G. (1964), *Le peuplement de Paris*, INED, Travaux et Documents, Cahier n° 43, Paris, PUF.

PRESSAT, R. (1966), *Principes d'analyse*, Paris, INED.