ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE

CLAUDE THÉLOT

L'analyse statistique des tables de mobilité à l'aide du modèle quasisymétrique et de ses dérivés

Annales de la faculté des sciences de Toulouse 6e série, tome 11, nº 4 (2002), p. 575-585

<http://www.numdam.org/item?id=AFST_2002_6_11_4_575_0>

Annales de la Faculté des Sciences de Toulouse Vol. XI, n° 4, 2002
pp. 575–585

L'analyse statistique des tables de mobilité à l'aide du modèle quasisymétrique et de ses dérivés (*)

CLAUDE THÉLOT (†)

RÉSUMÉ. — Cet article retrace l'utilisation du modèle quasi-symétrique pour l'analyse statistique des tableaux de mobilité géographique et sociale, telle que l'auteur l'a conduite en France, de 1972 à 1982. On y présente quelques généralisations du modèle, quelques applications et quelques conclusions importantes, sociologiques ou économiques, tirées de cet usage, et l'on indique pourquoi, après dix ans d'utilisation, l'auteur a abandonné ce modèle, pourtant très fécond.

ABSTRACT. — This article describes the use of the quasi-symmetry model for analyzing geographical and social mobility tables, as carried out by the author in France between 1972 and 1982. Some extensions of the baseline model, some applications and some substantive economic and sociological conclusions drawn from its consideration are presented. The reasons why the author has discontinued the use of this model, in spite of its potential, are given.

1. Introduction

C'est en 1972, je crois, que j'ai lu le travail séminal d'Henri Caussinus [2], publié en 1966 dans les Annales de la faculté des sciences de l'université de Toulouse. J'étais, au sein de la division « Emploi » de l'INSEE, responsable de l'enquête « Formation-Qualification professionnelle » qui venait d'être faite en 1970, et, tout en organisant l'apurement du fichier, je réfléchissais aux modalités d'analyse des résultats de cette enquête, c'est-à-dire, pour une grande part, des tableaux carrés décrivant la mobilité professionnelle

(*) Reçu le 18 septembre 2001, accepté le 18 septembre 2002
(†) Conseiller maître à la Cour des comptes, Cour des comptes, Paris.

et la mobilité sociale des Français : la « mobilité professionnelle », c'est-à-dire le croisement entre les deux situations professionnelles occupées par les actifs à quelques années d'intervalle, et la « mobilité sociale », c'est-à-dire le croisement entre la position sociale et l'origine sociale – cette dernière étant mesurée par la position sociale du père – des personnes.

La publication de l'université de Toulouse circulait dans les bureaux, et je me souviens avec émotion, pratiquement trente ans après, de l'impression que m'a faite la lecture de ce travail, et en particulier d'une des relations fondamentales dégagées par Caussinus : $(QS) \cap (IM) \leftrightarrow (S)$. Il y avait de nombreux modèles possibles d'analyse des tableaux carrés, ou rectangulaires (les tableaux étaient rectangulaires si l'on choisissait des nomenclatures différentes pour l'origine et la position par exemple, ou si l'on voulait analyser la relation entre formation et emploi), en particulier dans la littérature américaine, et j'en connaissais et avais utilisé plusieurs sur des résultats d'enquêtes antérieures. Mais j'ai eu immédiatement l'impression que le travail et les résultats de Caussinus étaient exactement adaptés, adéquats à l'analyse de tableaux carrés où l'évolution macroéconomique ou macrososiale faisait que les deux marges ne pouvaient être identiques. Prenons la mobilité professionnelle par exemple. À cinq ans d'intervalle, la structure de l'emploi, ou la structure sociale n'est pas la même : le développement lui-même fait évoluer les emplois vers les secteurs tertiaires, vers les situations de cadres, etc. Donc les deux marges du tableau diffèrent nécessairement, et les sociologues appellent « mobilité structurelle » ce que cette évolution macroscopique, globale, impose comme mouvement aux personnes : dans ce contexte, par exemple, une bonne partie des paysans qui quittent leur ferme au cours de leur vie pour travailler à l'usine le font non par choix mais parce que le pays s'industrialise. La relation précédente entre le modèle quasi-symétrique de Caussinus, l'identité des distributions marginales (IM) et le modèle symétrique (S) revenait à dire que les flux – flux professionnels, flux sociaux, mais aussi flux migratoires entre régions par exemple – étaient « aussi symétriques que possible compte tenu de cette contrainte structurelle ».

À partir de là, j'ai utilisé, essayé d'approfondir, puis, en fin de compte, abandonné le modèle quasi-symétrique sur une période, au total, de 10 ans, de 1972 à 1982, et je voudrais, dans ce bref article écrit en hommage à H. Caussinus, relater cette histoire, en distinguant trois phases : l'utilisation et une extension initiales du modèle quasi-symétrique (1972-1974 : section 2), puis une recherche visant à généraliser certains principes de ce modèle (1978-1980 : section 3), enfin l'application du modèle, suivie de son abandon, aux tables de mobilité sociale (1979-1982 : section 4).

2. Application et extension initiales : l'analyse des migrations géographiques

La première analyse que j'ai publiée et qui utilisait les travaux de Caussinus portait sur les migrations géographiques des ménages entre 1962 et 1968 (entre les deux recensements de la population [5] [1]) : la « répulsion » de la région de départ, l' « attraction » de la région d'arrivée, la « proximité » entre régions étaient concrétisées et mesurées à travers différents modèles estimés sur les flux migratoires, qu'on se fixe ces mesures *a priori* (par exemple en retenant les proximités géographiques dans un modèle gravitationnel) ou qu'on les estime *a posteriori* dans le cadre du modèle quasi-symétrique.

Des différents modèles estimés, c'est le quasi-symétrique qui s'ajustait le mieux, et cela quel que soit l'âge des ménages. Il était même souvent admissible au regard des critères usuels, bien que les effectifs fussent élevés (plusieurs milliers dans chaque catégorie d'âge). Mais les différentes tentatives pour spécifier les paramètres du modèle (répulsion, attraction, proximité) ou pour relier après coup leurs estimations à des grandeurs caractéristiques de la région (salaire moyen, développement, structure de l'emploi, tension sur le marché du travail régional ...) s'avéraient plutôt décevantes. On aboutissait alors à trois conclusions intéressantes :

- la « flexibilité » du modèle quasi-symétrique, en ce sens qu'il s'adaptait bien aux données ;
- le fait que l'estimateur d'attractivité des régions, obtenu dans ce modèle, était très différent selon l'âge des migrants, ce qui, à la fois, n'était pas une nouveauté et renforçait la « crédibilité » du modèle : la région parisienne s'avérait en effet très attractive aux âges jeunes, puis n'attirait plus les migrants au-delà de 35 ans ; l'Ouest, le Sud-ouest et la Méditerranée attiraient – durant les années soixante, déjà – les personnes âgées ;
- enfin la difficulté à interpréter ces estimations à l'aide de variables structurelles, ce qui là encore pouvait avoir deux conséquences : susciter des travaux plus approfondis, mais aussi générer un certain scepticisme sur le modèle lui-même : il s'adaptait bien, mais si on ne pouvait aller au-delà d'une description « apurée » des migrations, quel sens réel devait-on lui conférer ?

---

(1) La date de publication, 1976, ne reflète pas le moment où ce travail a été fait. Comme pratiquement tous les autres articles de cette livraison d'*Annales de l'INSEE*, il avait été présenté à un colloque que j'avais organisé à Nantes, avec l'unité recherche de l'INSEE, sur l'analyse des tableaux carrés en 1974.

Quoi qu'il en soit, à l'issue du colloque de Nantes de 1974 où j'avais présenté ces résultats, nous avons décidé, H. Caussinus et moi-même, d'approfondir la description que le modèle quasi-symétrique permettait en posant et estimant un modèle sur les tableaux migratoires « entassés » par âge, qui postulait une constance de certains paramètres : soit l'attractivité et la répulsion des régions d'une part, soit la proximité entre régions d'autre part pouvaient-elles être supposées indépendantes de l'âge des migrants ? On voit ici apparaître une généralisation extrêmement importante des modèles applicables aux tables de mobilité : ils doivent pouvoir permettre de conclure si « quelque chose se conserve » selon une troisième dimension. Ici, c'est l'âge du migrant, plus tard, et ce cas sera encore plus important, ce sera le temps : on cherchera alors à répondre à la question : dans la succession des tables de mobilité au cours du temps, quelque chose – par exemple les proximités entre catégories sociales – se conserve-t-il (cf. ci-dessous) ?

En tout cas, dans le cas des migrations géographiques que nous avons étudié à l'aide du modèle « QSC » posé pour l'occasion – « quasi-symétrique à effets de distance constants » – nous posions bien la question : « si la quasi-symétrie à distances constantes est un modèle acceptable, cela signifie que l'on peut exhiber des indices de proximité entre régions indépendants de l'âge des migrants, les pouvoirs d'attraction-répulsion de chaque région variant par contre, en général, en fonction de l'âge des individus » (voir [3] [2]).

Et l'on a conclu, en effet, qu'en première approximation le modèle (QSC) était acceptable, ce qui enrichissait l'analyse en conférant en quelque sorte une signification intrinsèque, indépendante des caractéristiques du migrant (ici, de son âge), aux indices de proximité entre régions.

3. Généralisation théorique

Lors du colloque de Nantes de 1974, un participant, Jean-Claude Deville, avait fait le commentaire suivant à partir des indices de proximité du modèle quasi-symétrique : pourquoi les supposer symétriques ? Et de citer la marche en montagne : la « proximité » entre deux points de montagne, A et B, d'altitudes distinctes est très différente si l'on monte de A à B ou si l'on descend de B à A, du moins si on mesure cette proximité par le temps de la course par exemple.

Cette remarque m'a conduit, quelques années après (en 1978-1980), à essayer de généraliser le modèle quasi-symétrique. Je saisis l'occasion offerte par le présent article pour indiquer de façon synthétique les quelques résultats que j'ai obtenus et qui n'ont jamais été publiés.

(2) Cet article a, en réalité, été conçu et écrit juste après le colloque de Nantes, c'est-à-dire durant l'été 1974.

Le cadre général est le suivant : on se donne une fonction continue involutive ϕ de $\mathbb{R}^+$ dans $\mathbb{R}^+$, ce qui veut dire que $\phi[\phi(x)] = x$. Et le modèle qu'on cherche à étudier puis tester sur des données est le suivant, noté $(PQ\Phi)$: il existe un vecteur positif (p_i) et un vecteur positif (q_j), et il existe une matrice de nombres positifs d_{ij} vérifiant

$$d_{ij} = \phi(d_{ji}),$$

tels que [3]

$$p_{ij} = P[X = i, Y = j] = p_i q_j d_{ij}.$$

Ceci généralise bien le modèle quasi-symétrique. D'ailleurs, ce modèle général a deux formes « polaires », obtenues respectivement pour $\phi(x) = x$ – c'est justement le modèle (QS) – et pour $\phi(x) = 1/x$ – ce qui peut s'interpréter comme ces proximités de montagne d'autant plus grandes dans un sens qu'elles sont petites dans l'autre. Mais ϕ pourrait être homographique (avec des conditions pour qu'elle soit positive), ou construite à partir de n'importe quelle fonction f strictement monotone de $[0, a[$ sur $]a, \infty[$ en posant $\phi(x) = f(x)$ pour $x \leqslant a$ et $\phi(x) = f^{-1}(x)$ pour $x \geqslant a$.

La diagonale – c'est-à-dire les « immobiles » – peut ou non intervenir dans l'analyse, soit que le modèle ne stipule rien de particulier (si ϕ a au moins un point fixe), soit qu'il ne puisse être postulé pour elle (si ϕ n'a pas de point fixe), auquel cas on n'étudie que les « mobiles ». On se limitera à ce dernier cas.

Quatre propriétés peuvent être énoncées :

1. Le modèle $(PQ\Phi)$ est log-linéaire si et seulement si ϕ est la fonction identité ou la fonction inverse.

2. Pour chercher à généraliser la relation fondamentale rappelée plus haut, $(QS) \cap (IM) \Leftrightarrow (S)$, on peut définir, une fois la fonction ϕ fixée, les modèles :

   - $(\Phi) : p_{ij} = \phi(p_{ji})$
   - $(M\Phi L) : \forall\, i : \sum_{j \neq i} p_{ij} = \sum_{j \neq i} \phi(p_{ji})$
   - $(M\Phi C) : \forall\, j : \sum_{i \neq j} p_{ij} = \sum_{i \neq j} \phi(p_{ji})$

---

(3) Je retiens ici une formulation multiplicative, ce qui est dans la ligne de ce qu'écrivait H. Caussinus. Dans une formulation log-linéaire, le modèle s'écrit:

$$\log p_{i_1 i_2} = u + u_1(i_1) + u_2(i_2) + u_{12}(i_1 i_2) \text{ où } u_{12}(i_1 i_2) = \log \phi[\exp\{u_{12}(i_2 i_1)\}].$$

Pour ce type d'écriture, cf. l'article de Erosheva, Fienberg et Junker ([4]) dans ce volume.

On somme donc en ligne (L) ou en colonne (C). Dans le cas où ϕ est la fonction identité, les deux modèles ($M\Phi L$) et ($M\Phi C$) sont identiques et identiques à l'égalité des distributions marginales (IM).

Dans ces conditions $(M\Phi L) \cap (PQ\Phi) \Leftrightarrow (\Phi)$ d'une part et $(M\Phi C)\cap(PQ\Phi) \Leftrightarrow (\Phi)$ d'autre part, qui généralisent donc la formule fondamentale de Henri Caussinus, semblent vraies si et seulement si ϕ est l'identité ou la fonction inverse $\phi(x) = \frac{k}{x}$, où k est une constante positive quelconque. On retrouve donc les deux cas polaires [4].

3. Quelle que soit ϕ, les estimateurs du maximum de vraisemblance, f_{ij}, des effectifs n_{ij} du tableau sous le modèle ($PQ\Phi$) sont astreints à satisfaire les mêmes égalités :

$$\forall\, i : \sum_{j\neq i} f_{ij} = \sum_{j\neq i} n_{ij} \quad \text{et} \quad \forall\, j : \sum_{i\neq j} f_{ij} = \sum_{i\neq j} n_{ij}.$$

   Ainsi, quelle que soit ϕ, le tableau estimé sous le modèle ($PQ\Phi$) a les mêmes marges (limitées aux mobiles) que le tableau observé, ce qui revient à dire qu'on cherche avec le tableau estimé f_{ij} à remplir un tableau ayant les mêmes marges – donc les mêmes contraintes structurelles – que les observations et satisfaisant, case par case, ($PQ\Phi$).

4. En se concentrant sur le « dual » du modèle quasi-symétrique, celui où ϕ est la fonction inverse et en notant le modèle correspondant (PQI), I comme inverse, on a : $(PQI) \cap (QS) \Leftrightarrow (H)$ où (H) est l'hypothèse de quasi indépendance, c'est-à-dire où $\exists\ (p_i)$et (q_j) tels que $p_{ij} = p_i q_j$ (où $i \neq j$). Si l'on inclut la diagonale – les immobiles – et si l'égalité vaut, (H) est l'hypothèse d'indépendance. Ainsi, en un sens, le moèle (PQI) est l'« inverse » du modèle (QS).

   L'estimation de ce modèle (PQI) sur les tableaux de mobiles – hors diagonale –, que ce soit les migrations géographiques entre 1962 et 1968, les mariages par catégorie sociale (des deux beaux-pères ou du mari et du beau-père), enfin les tableaux de mobilité sociale croisant l'origine et la position sociales a été décevante : le modèle était très mal ajusté aux données. Ainsi, en un sens, le modèle (PQI) est l'« inverse » du modèle (QS).

Au total, la généralisation ne donnait pas de résultats théoriques très intéressants, sauf pour la fonction inverse, mais l'usage de cette dernière conduisait à des estimations très éloignées des observations. J'ai donc abandonné ces tentatives de généralisation.

---

(4) La condition est suffisante. Ce qui laisse penser que la condition est aussi nécessaire, c'est qu'avec une fonction ϕ simple comme $\phi(x) = k - x$, ces égalités ne sont pas vérifiées.

Mais lors de l'estimation de tous ces modèles sur les tableaux de mariage croisant les catégories sociales des deux beaux-pères, je me suis aperçu que le modèle quasi-symétrique était très bien ajusté aux données, sans que le modèle symétrique le soit, ce qui signifiait que l'identité des distributions marginales ne pouvait être retenue. Cela peut être illustré par la valeur des χ^2 reproduites dans le tableau suivant :

	(QS)	(S)	(IM)
X2	332,1	1198,2	866,1
ddl	231	253	22

Analyse des 416 000 mariages de l'année 1972 en France (pour 197 000 d'entre eux les deux catégories sociales des deux beaux-pères étaient différentes, la nomenclature distinguant 23 catégories sociales). D'où les degrés de liberté (ddl).

On obtient alors un résultat très intéressant du point de vue non de l'analyse des mariages *stricto sensu* mais, ce qui est également très pertinent pour un sociologue, des déclarations, donc des images : alors que les deux distributions marginales devraient être égales puisqu'il n'y a pas de raison que l'on ait, dans telle ou telle classe sociale, plus de filles ou plus de garçons, elles ne le sont pas : on ne peut pas du tout accepter (IM), et c'est même ce refus qui empêche le tableau des mariages d'être symétrique puisque (QS) est pratiquement acceptable (et cela en dépit de l'énormité des effectifs : il y a eu 197 000 mariages de catégories sociales différentes en 1972). Donc, seule explication possible, les garçons et les filles ne déclarent pas de la même manière, au moment de leur mariage, la position sociale de leur père. Résultat capital. Il peut s'interpréter de deux façons. Ou bien en termes de carrière des pères, puisque, en France en 1972, les garçons qui se sont mariés avaient, en moyenne, deux ans de plus que les filles qui se sont mariées, leurs pères étaient plus avancés dans leur propre carrière que les pères de ces dernières. Ou bien en termes purs de différence de déclaration, donc d'image que se font et ont ces jeunes mariés et mariées de leur origine sociale. Sans trancher entre les deux interprétations – qui doivent être toutes deux partiellement vraies – on pouvait constater que les différences n'étaient pas aléatoires, les garçons tendant à déclarer leurs pères « plus haut » et plus « indépendants » dans la hiérarchie sociale que les filles. En effet, il y avait plus d'agriculteurs, d'artisans, de commerçants, de cadres supérieurs, d'instituteurs parmi les pères des époux, et plus d'ouvriers (qualifiés ou non, agricoles ou non) et de personnels de service parmi les pères des épouses, et à peu près autant de cadres moyens et d'employés dans les deux cas.

Ainsi l'examen du tableau des mariages à l'aide des modèles de Henri Caussinus ($QS/IM/S$) débouchait sur des conclusions inattendues mais très intéressantes concernant la déclaration, et donc la perception, de son origine sociale.

4. L'analyse des tables de mobilité sociale et de leur succession dans le temps

En même temps que je cherchais à généraliser le modèle (QS), j'engageais une étude d'ensemble sur la mobilité sociale en France, qui devait aboutir en septembre 1980 à un gros document, de type universitaire, *Le poids d'Anchise* [6], puis en 1982 à mon premier livre *Tel père, tel fils ?* [7].

Cette fois-ci, ce n'était plus un article ou une étude isolée, mais un ouvrage que je préparais, avec différents chapitres articulés. L'un de ces chapitres intitulé « Le jeu stratégique » consistait à utiliser différents modèles pour, en plus d'une analyse descriptive, interpréter et comprendre non seulement le lien entre origine et position sociales à un moment donné, mais l'évolution de ce lien au cours du temps. On disposait en effet de quatre observations des tables de mobilité sociale : en 1953, en 1964, en 1970 et, pour *Tel père, tel fils ?*, d'une quatrième faite en 1977.

L'analyse faite dans *Le poids d'Anchise*, se déroule en deux étapes.

En premier lieu, on se demande si, compte tenu des contraintes structurelles, la table de mobilité a le moins évolué possible au cours du temps. On retrouve ici l'idée d'une mobilité structurelle, « minimale » en quelque sorte, imposée par le développement lui-même qui fait que, de 1953 à 1970 (ou 1977), les marges des tables de mobilité – aussi bien celle du père que celle du fils – ne peuvent être les mêmes, et que la différence entre ces marges paternelle et filiale s'est approfondie. La question que l'on se pose est finalement de savoir si l'évolution observée de la mobilité se ramène à cette seule évolution structurelle, imposée, ou si en plus il y a une évolution de la mobilité nette, « indépendante » de l'évolution des structures. Formellement, on sait que cela revient à se demander si le modèle

$$p_{ijt} \quad = \quad a_{it} b_{jt} c_{ij}$$

est acceptable, lequel contient des interactions ou des proximités (mais qui ne sont pas symétriques) c_{ij} constantes dans le temps, ou encore si les coefficients concurrentiels, c'est-à-dire les quantités $\frac{p_{ijt}}{p_{ikt}} \frac{p_{\ell kt}}{p_{\ell jt}}$, sont constants. Il faut donc ajuster ce modèle log-linéaire d'ordre 2 sur l'entassement des tables de mobilité [5].

(5) Sur les modèles log-linéaires, le livre majeur à l'époque était celui de Yvonne Bishop, Stephen Fienberg, Paul Holland [1].

La conclusion était sans ambiguïté : on pouvait accepter, comme une première approximation assez grossière, ce modèle, donc considérer que dans l'évolution observée de la mobilité une grande part tenait aux contraintes structurelles. Mais, et c'était plus qu'une nuance, cette analyse ne rendait pas compte de toute la réalité, en particulier parce que la baisse de l'immobilité des années cinquante aux années soixante-dix avait été plus forte que ce qu'aurait imposé la seule évolution des contraintes structurelles : dans toutes les catégories sociales, la proportion d'immobiles (p_{iit}) avait beaucoup baissé, beaucoup plus que les estimations (f_{iit}/n), suggérant une augmentation de la fluidité sociale qui aurait été au-delà du strict nécessaire.

D'où la deuxième étape de l'analyse : on se limite aux seuls mobiles, la table privée de sa diagonale et l'on analyse ainsi les seuls flux de mobilité où les fils sont dans une autre position sociale que leur père. L'on pouvait alors poser deux questions :

- premièrement, les flux de mobiles vérifient-ils la constance des proximités ? Il s'agissait d'estimer et de tester le même modèle

$$p_{ijt} \quad = \quad a_{it} b_{jt} c_{ij}$$

  sur les mobiles ; si la réponse était oui, l'analyse de la mobilité était assez simple : l'évolution d'ensemble aurait excédé la contrainte structurelle du seul fait d'un excès de baisse de l'immobilité, mais dès lors qu'on quittait son milieu d'origine, la fluidité ne se serait pas accrue ;

- deuxièmement, les flux de mobiles vérifient-ils des proximités qui peuvent évoluer dans le temps mais qui soient symétriques ? On retrouve le modèle quasi-symétrique appliqué aux différentes tables de mobilité,

$$p_{ijt} \quad = \quad p_{it} q_{jt} d_{ijt} \quad \text{avec} \quad d_{ijt} = d_{jit} \ ;$$

  dans ce cas, l'interprétation devient : à chaque période, les flux de mobilité ne peuvent être symétriques puisque des pères aux fils (en une génération) la structure sociale a profondément changé, mais sont-ils aussi symétriques que possible au sein de cette contrainte ?

Les deux questions, si on les marie, conduisent à l'interaction des deux modèles précédents, c'est-à-dire au modèle (QSC) présenté plus haut : si les flux de mobilité p_{ijt} ($i \neq j$) vérifiaient

$$p_{ijt} \quad = \quad p_{it} q_{jt} d_{ij} \quad \text{avec} \quad d_{ij} = d_{ji},$$

on pourrait considérer que les proximités étaient restées stables dans le temps et qu'elles étaient symétriques.

Je concluais que les deux modèles s'ajustaient de façon assez satisfaisante aux données, mais que le modèle de constance des proximités dans le temps était « meilleur », c'est-à-dire plus proche des données, que celui de proximités symétriques, notamment en regardant les cases correspondant à la mobilité sociale des fils de cadres et de gros indépendants. D'autre part, supposer les proximités à la fois constantes dans le temps et symétriques était une hypothèse trop forte [6].

Pour passer d'un gros travail de recherche à un ouvrage destiné à une plus grande diffusion, il faut réécrire, synthétiser, simplifier, raccourcir. C'est la raison pour laquelle, dans *Tel père, tel fils ?* l'utilisation du modèle quasi-symétrique ne figure pas. Seule est conduite l'analyse à partir des proximités constantes et c'est sur elle qu'est fondée la conclusion générale : « dans l'ensemble, les flux de passage entre catégories sociales (une fois l'immobilité mise à part) ont changé sous le seul fait de l'évolution de la structure sociale, le réseau de proximités restant, lui, à peu près stable au cours de ces vingt-cinq ans ». Et j'ajoutais en note : « D'autres hypothèses peuvent être éprouvées à propos de ces proximités, en particulier celle de savoir si elles sont symétriques. On n'a présenté ici que celle qui s'adaptait le mieux aux observations ».

C'est en effet la première raison pour laquelle le modèle quasi-symétrique n'apparaît pas dans mon livre, et pour laquelle je ne l'utiliserai plus après : il m'était apparu moins bien ajusté aux données de mobilité sociale. Mais il y en a une seconde : à la réflexion, pourquoi supposer symétriques les proximités entre catégories sociales, telles qu'elles seraient manifestées de façon implicite (si le modèle (QS) était vrai) par les flux de mobilité ?

N'y a-t-il pas là un reste de la distance géographique qui, elle, est bien symétrique (du moins la distance kilométrique) ? Pourquoi la distance père ouvrier → fils cadre supérieur serait-elle la même que la distance père cadre supérieur → fils ouvrier ? J'avais justement, dans un autre chapitre de mon livre, appelé « effet cliquet » le fait qu'il y avait très peu de ces descentes vers la classe ouvrière, qu'une fois une famille parvenue à la position de cadre, elle cherchait à s'y maintenir et si un des descendants devait quitter cette position, c'était pour devenir employé plutôt qu'ouvrier. Autrement dit, l'analyse sociologique rendait l'hypothèse de symétrie des proximités discutable et peut-être même peu pertinente. De ce point de vue, le modèle log-linéaire avec interaction d'ordre 2 ($a_{it}b_{jt}c_{ij}$) était nettement supérieur : l'hypothèse d'une constance dans le temps de proximités non symétriques était plus intéressante que celle d'une symétrie de proximités variables dans le temps.

---

(6) Par exemple, l'indice d'ajustement global $(\chi^2 - ddl)/N$ s'établissait à $0,8 \times 10^{-2}$ (proximités constantes), $2,0 \times 10^{-2}$ (proximités symétriques) et $2,4 \times 10^{-2}$ (proximités symétriques et constantes).

Ainsi, après avoir abandonné l'inégalité triangulaire que respectent les distances géographiques, mais que ne respectaient déjà pas les proximités de Henri Caussinus dans son modèle (QS), il fallait, dans l'analyse de la mobilité sociale (ce n'était peut-être pas aussi évident dans l'analyse des migrations géographiques), abandonner aussi l'hypothèse de symétrie.

5. Conclusion

De 1972 à 1982, j'ai ainsi découvert avec émotion et admiration, puis utilisé et étendu, enfin abandonné le modèle quasi-symétrique. Ce modèle (et ses dérivés), qu'il faut savoir gré à Henri Caussinus d'avoir « inventé » ou « découvert » – comme toujours en mathématiques, quel verbe faut-il employer ? –, a ainsi été une grille de lecture ou d'analyse très féconde, très précieuse, des tables de mobilité ou de migrations, même si je lui ai, ensuite, préféré d'autres instruments. On découvrira dans ce volume bien d'autres extensions du modèle de quasi-symétrie.

Bibliographie

[1] BISHOP (Y. M. M.), FIENBERG (S. E.), HOLLAND (P. W.). — *Discrete multivariate analysis: theory and practice*, MIT Press, Cambridge, Mass, 1975.

[2] CAUSSINUS (H.). — Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de Toulouse*, 29 (année 1965), (1966), 77–183.

[3] CAUSSINUS (H.), THÉLOT (C.). — Note complémentaire sur l'analyse statistique des migrations, *Annales de l'INSEE*, 22 (1976), 135–146.

[4] EROSHEVA (E. A.), FIENBERG (S. E.), JUNKER (B. J.). — Alternative statistical models and representations for large sparse multi-dimensional contingency tables, *Annales de la Faculté des Sciences de Toulouse*, 2002, dans ce volume.

[5] THÉLOT (C.). — Analyse statistique des migrations, *Annales de l'INSEE*, 22–23 (1976), 101–134.

[6] THÉLOT (C.). — Le Poids d'Anchise, Document de travail interne à l'INSEE, Nantes, 1980.

[7] THÉLOT (C.). — *Tel père, tel fils ?*, Dunod, Paris, 1982.