REVUE DE STATISTIQUE APPLIQUÉE

F. ROSENFELD
M. SALOMON

Utilisation de modèles markoviens et pseudo markoviens dans les études de marché

Revue de statistique appliquée, tome 10, nº 2 (1962), p. 67-74
<http://www.numdam.org/item?id=RSA_1962__10_2_67_0>

UTILISATION DE MODÈLES MARKOVIENS ET PSEUDO MARKOVIENS DANS LES ÉTUDES DE MARCHÉ (1)

F. ROSENFELD et M. SALOMON

Société d'Économie et de Mathématique Appliquée

1 - PRINCIPE DE LA METHODE -

La méthode analytique de prévision de la demande exposée dans la présente note, se fonde sur l'étude du comportement des unités de consommation, la demande totale étant la somme des demandes de toutes les unités de consommation. Ces unités, peuvent en général être réparties, suivant leurs caractéristiques, en un certain nombre de classes, ou strates.

Le modèle représentatif de l'évolution de la demande, envisagé, trouve des applications soit à l'étude des marchés de biens durables, soit à celle des matières consommées par ceux-ci, par exemple : équipements de chauffage domestique et combustibles ou autres sources d'énergie utilisés, automobiles et essence ou huile consommées.

Une unité de consommation (individu, ménage, logement, etc.) peut posséder un seul ou plusieurs des biens durables qui font l'objet de l'étude ; le mode d'équipement des unités de consommation est précisément un des critères de leur stratification.

Si l'on désigne par :

N_k le nombre d'unités de consommation de la strate $k(k = 1, 2, \ldots, z)$, et par :

c_k la consommation moyenne de matière au cours d'une période t pour les unités de cette strate, la consommation totale de matière de toute la population étudiée pendant la même période t s'écrit :

$$C = \sum_k N_k c_k$$

----

(1) Communication présentée à la 33e Session de l'Institut International de Statistique (Paris - 1961).

La variation de la consommation entre l'instant t et t + t est égale à :

$$\Delta C = \sum_k (N_k \, \Delta c_k + c_k \Delta N_k + \Delta \, c_k \Delta N_k).$$

Elle s'exprime en fonction des variations des consommations moyennes des strates et des variations des effectifs contenus dans chacune de celles-ci. Ces deux facteurs doivent être étudiés séparément.

L'étude des variations de la consommation, ou de la demande moyenne de chaque strate, peut être conduite suivant les méthodes classiques d'étude de la demande de catégories homogènes d'unités de consommation. On se limitera à dire ici que cette demande dépend notamment de l'évolution de facteurs techniques affectant les taux de rendement des équipements et de facteurs économiques, tels que les prix des équipements, le coût de leur utilisation, les prix des matières en cause et le revenu disponible par unité de consommation.

L'analyse des variations des effectifs des diverses strates appelle des considérations d'un caractère très différent que l'on se propose de développer ici.

2 - MODELE MARKOVIEN D'EVOLUTION DES EFFECTIFS DES CLASSES DE CONSOMMATEURS -

La variation de l'effectif ΔN_k de la classe k entre l'époque t et l'époque t + 1 est :

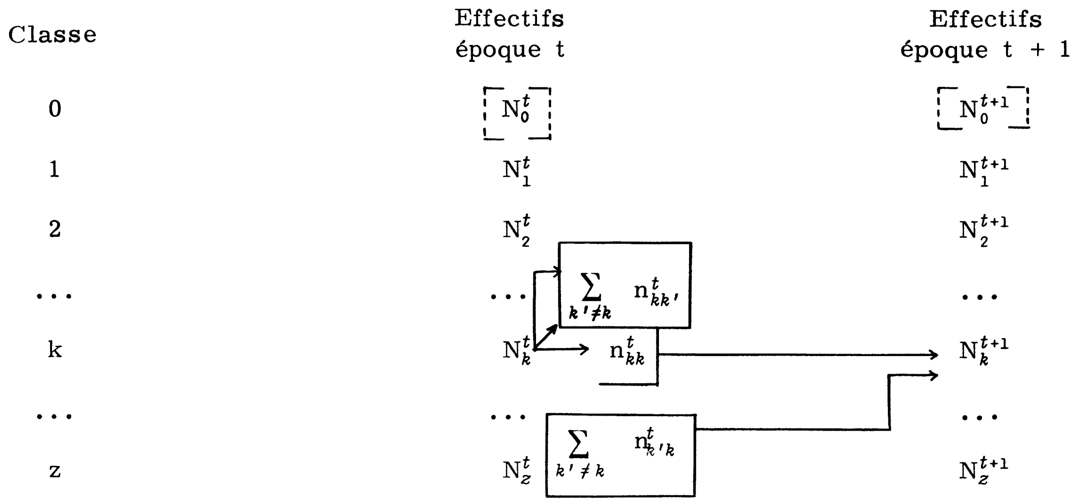
$$\Delta N_k = N_k^{t+1} - N_k^t \qquad (1)$$

Cette variation est égale à la différence entre le nombre des unités qui entrent dans la classe k entre l'époque t et l'époque t + 1 et celui des unités qui en sortent. Les unités qui entrent peuvent soit provenir des autres classes $k'(k' \neq k)$, soit avoir été créées pendant la période considérée ou être, en quelque sorte, nées avec les caractéristiques correspondant à la classe k. De même, les unités sortantes sont d'une part celles qui quittent la classe k pour passer dans une des classes k' et, d'autre part, celles qui disparaissent ou, en quelque sorte, meurent pendant l'intervalle de temps considéré.

On peut imaginer qu'il existe un réservoir fictif d'unités, appelé classe zéro, d'où proviendraient les unités naissantes et où iraient se placer les unités qui disparaissent, de sorte que toutes les unités entrant dans une classe k proviennent des classes k' et toutes les unités qui en sortent vont dans les classes k' avec

$$k' = 0, 1, 2, \ldots, z, \; k' \neq k.$$

Les effectifs des différentes classes aux époques t et t + 1 apparaissent alors comme suit :



Si l'on désigne par n_{gh}^t le nombre d'unités de la classe g passant dans la classe h entre l'instant t et l'instant t + 1, on a :

$$N_k^{t+1} = N_k^t + \sum_{k' \neq k} n_{k'k}^t - \sum_{k' \neq k} n_{kk'}^t$$

ou

$$N_k^{t+1} = n_{kk}^t + \sum_{k' \neq k} n_{k'k}^t = \sum_{k'=0}^{z} n_{k'k}^t$$

En d'autres termes, tout se passe comme si la classe k était à l'instant qui précède l'époque t + 1 une classe vide et qu'elle se remplissait avec les éléments provenant de toutes les classes 0, 1, ..., z, y compris la classe k elle-même, pour constituer le nouvel effectif N_k^{t+1}.

En posant
$$p_{k'k}^t = \frac{n_{k'k}^t}{N_{k'}^t} ,$$

(2) il vient
$$N_k^{t+1} = \sum_{k'=0}^{z} p_{k'k}^t \, N_k^t$$

Le rapport $p_{k'k}^t$ défini comme une fréquence, peut être considéré comme la probabilité pour qu'une unité de la classe k' passe dans la classe k entre l'époque t et l'époque t + 1.

L'ensemble de ces probabilités se dispose sous forme de matrice :

$$\begin{bmatrix} p_{00} & p_{01} & p_{02} & \dots & p_{0k} & \dots & p_{0z} \\ p_{10} & p_{11} & p_{12} & \dots & p_{1k} & \dots & p_{1z} \\ p_{20} & p_{21} & p_{22} & \dots & p_{2k} & \dots & p_{2z} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ p_{k0} & p_{k1} & p_{k2} & \dots & p_{kk} & \dots & p_{kz} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ p_{z0} & p_{z1} & p_{z2} & \dots & p_{zk} & \dots & p_{zz} \end{bmatrix}$$

où $$\sum_{k'=0}^{z} p^{kk'} = 1$$

On peut admettre qu'à toute époque t, la situation St de l'ensemble étudié (répartition des unités de consommation et leur comportement en ce qui concerne cette consommation) est définie par :

- la distribution des effectifs des classes N_k^t ;

- la distribution des autres caractéristiques des unités à l'intérieur des classes ;

- l'ensemble des caractéristiques du milieu économique dans lequel la population étudiée est plongée, notamment : système des prix des biens durables et des matières consommées étudiées et de leurs substituts, coûts de l'entretien et de la manutention.

Lorsque les probabilité de passage $p_{k'k}$ à l'époque t sont indépendantes des évènements antérieurs, le schéma représentant l'ensemble des effectifs N^t est un *schéma de Markov*. Si au contraire les probabilités de passage sont fonction des évènements antérieurs, le schéma est *pseudo-markovien*, c'est le cas par exemple d'unités de consommation dont les probabilités de passage dépendent des âges des équipements et dont les compositions par âges des équipements des diverses classes varient avec le temps. En pratique, il faut rechercher une stratification qui rende le schéma très voisin d'un schéma de Markov de manière à faire usage des propriétés de ce dernier.

3 - ANALYSE DES PROBABILITES DE PASSAGE D'UNE CLASSE A UNE AUTRE -

Désignons par σ_k la probabilité qu'une unité de la classe k existant à l'époque t survive à l'époque t + 1. Pendant l'intervalle de temps situé entre ces deux époques, trois *options* s'offrent à une telle unité :

- conserver l'équipement existant ;
- renouveler son équipement avec du matériel du même type ;
- changer de type d'équipement.

Désignons par α_k, β_k et γ_k, les probabilités respectives de ces trois options. On a :

$$\alpha + \beta + \gamma = 1$$

On désignera par $\gamma_{kk'}$, la probabilité pour qu'une unité de la classe k, qui survit entre t et t + 1, choisisse un équipement k' ; on a $\sum_k \gamma_{kk'} = \gamma_k$, avec $k' \neq k$.

Les probabilités de passage s'écrivent alors :

$$p_{kk} = \sigma_k (\alpha_k + \beta_k) = \sigma_k (1 - \gamma_k)$$

$$p_{kk'} = \sigma_k \gamma_{kk'}$$

$$p_{k0} = 1 - \sigma_k$$

On pourra désigner par V_k la probabilité p_{0k} pour qu'une unité nouvelle créée entre t et t + 1 appartienne à la classe k

$$p_{0k} = V = \frac{n_{0k}^t}{\sum_k n_{0k}^t} = \frac{n_{0k}^t}{N_0^t}$$

N_0^t étant le nombre d'unités nouvelles créées entre l'instant t et l'instant t + 1 et n_{0k} étant celles, parmi ces unités nouvelles, entrant dans la classe k.

En fonction de ces éléments et des effectifs connus pour l'époque t, les effectifs de la classe k à l'époque r + 1 s'écrivent :

$$N_k^{t+1} = V_k N_0^t N_k^t (1 - p_{k0}) (1 - \gamma_k) = \sum_{\substack{k' \neq 0 \\ k' \neq k}} N_{k'}^t (1 - p_{k'0}) \gamma_{k'k} \qquad (3)$$

Les éléments dont dépendent les probabilités de passage, ou ces probabilités elles-mêmes, de même que les probabilités de survie des unités, peuvent le plus souvent être déterminées de manière empirique, au moyen d'enquêtes par sondage ; ils peuvent aussi donner lieu à une analyse plus poussée quant aux motifs des options et aux seuils économiques et psychologiques à partir desquels les choix se portent vers chacune des diverses solutions possibles.

Par exemple, les probabilités d'option d'un ménage consommateur de charbon, entre le maintien de l'équipement de chauffage existant, le renouvellement immédiat de cet équipement et le passage à un équipement d'un autre type, varient suivant que le ménage demeure dans le logement qu'il occupe, ou qu'il s'installe dans un nouvel appartement. On conçoit également que ces probabilités soient fonction des revenus des ménages, des caractéristiques des logements qu'ils occupent, de l'âge et des types des équipements en concurrence ainsi que des coûts d'utilisation de ceux-ci, coûts qui comprennent notamment le prix des appareils, les frais d'installation, les dépenses d'entretien, les prix des combustibles et s'il y a lieu, les frais de leur manutention (cas des chauffages centraux collectifs).

L'analyse peut conduire au calcul des coûts totaux actualisés, portant sur une longue période ; coûts d'achat et d'installation des matériels et coûts d'utilisation. Une telle analyse doit se faire en se plaçant dans l'optique des utilisateurs qui généralement font *inconsciemment* un calcul analogue, le plus souvent en adoptant des hypothèses simplificatrices concernant les durées des équipements, la stabilité ou l'évolution des conditions économiques générales ultérieures, la valeur de cession du matériel à céder et même le taux d'intérêt de l'argent.

4 - EXEMPLE D'APPLICATION : DEMANDE DE CHARBON DANS LES FOYERS DOMESTIQUES(1).

Un modèle du type décrit a été utilisé pour essayer d'expliquer la demande de charbon dans les foyers domestiques et d'en prévoir l'évolution à moyen terme.

Le schéma du modèle retenu était le suivant :

----

(1) Les Charbonnages de France, pour qui l'étude de la demande de charbon dans les foyers domestiques a été réalisée, ont bien voulu autoriser la publication des informations contenues dans la présente note.

On a cherché à répartir les ménages en classes homogènes vis-à-vis de la consommation de charbon, à déterminer ensuite un ensemble de probabilités de passage des ménages entre ces différentes classes, enfin, compte tenu de l'évolution de la consommation moyenne de chacune d'elles, d'établir des perspectives de consommation pour les classes en question.

4.1 - Détermination de classes homogènes de ménages consommateurs de charbon

Les données de base disponibles ont été élaborées à partir d'une importante enquête réalisée par les Charbonnages de France au cours de la campagne de vente 1957-1958. Il était possible de distinguer les ménages suivant les caractéristiques de l'habitat (dimension de la commune), du logement (type et ancienneté), de l'équipement de chauffage, de la situation sociale du ménage, de la région et enfin de la consommation annuelle de combustible .

On a recherché l'influence sur la consommation annuelle de charbon, de ces différents facteurs ; on a relevé notamment que le type de logement (selon les 3 variantes : rural, urbain-collectif, urbain-individuel) avait une influence très importante, et qu'il était possible de distinguer des classes d'équipement de chauffage, correspondant à des consommations moyennes présentant des différences significatives entre elles. Il a même été possible de faire apparaître des différences significatives en fonction des autres caractéristiques, notamment en ce qui concerne les régions.

La faiblesse des effectifs de l'échantillon n'a pas permis de subdiviser les classes en combinant différents critères et il a été nécessaire de limiter l'étude à 6 classes de ménages caractérisées par leurs équipements de chauffage ($k = 1, 2, \ldots, 6$). Les quantités moyennes consommées par ces classes présentaient entre elles des écarts significatifs pour un seuil de 95 % ; de plus, la forme des distributions des consommations autour des moyennes de chaque classe, ainsi que la valeur des coefficients de variation permettait de considérer ces classes comme relativement homogènes.

4.2 - Etablissement d'une matrice de probabilités de passage

Une méthode directe pour l'estimation des probabilités de passage consisterait à mesurer dans le temps, les fréquences des différents comportements des ménages à l'égard des divers types d'equipement ; malheureusement, les données disponibles ne permettaient pas d'adopter une telle méthode.

Des enquêtes effectuées en 1954-1955 et 1957-1958, permettaient de connaître seulement la répartition des ménages entre les différentes classes à ces deux époques. Les probabilités de passage ont donc dû être estimées en exprimant le fait que la distribution de 1957-1968 se déduit de celle de 1954-1955 par l'application de ce système de probabilités de passage.

Le tableau ci-après donne, dans la première colonne, une description sommaire des 6 classes d'équipement retenues et indique dans les deux dernières colonnes la répartition des ménages au cours des deux enquêtes. Les fréquences des diverses classes, à l'époque 1954-1955, $f_k^{(1)}$, et à l'époque 1957-1958, $f_k^{(2)}$ peuvent s'écrire :

$$f_k^t = \frac{N_k^t}{\sum_{k=1}^{6} N_k^t}$$

L'insuffisance des informations sur la création de nouvelles unités de consommation, ainsi que sur la disparition de ménages au cours des périodes étudiées, a conduit à faire des hypothèses simplificatrices pour éviter de faire intervenir les probabilités p_{0k} et p_{k0}. Les passages des unités entre les diverses classes, d'une époque d'enquête à l'autre, s'expriment alors par les 6 équations suivantes :

$$f_k^{(2)} = \sum_{k'=1}^{6} p_{k'k}\, f_{k'}^{(1)} \qquad (k = 1, 2, \ldots, 6) \tag{4}$$

les probabilités $p_{k'k}$ étant les probabilités de passage entre deux époques, 1954-1955 et 1957-1958, distantes de trois ans.

Types d'équipements	N° d'ordre k	Fréquences ménages %	
		Enquête 1954-55 100 $f_k^{(1)}$	Enquête 1957-58 100 $f_k^{(2)}$
Chauffage central collectif avec ou sans cuisinière et poêle	1	1,1	1,8
Chauffage central individuel au charbon	2	6,8	7,8
Pas de chauffage central, pas de poêle, mais cuisinière	3	32,3	22,5
Pas de chauffage central, pas de cuisinière, mais poêle	4	7,1	7,2
Pas de chauffage central, mais cuisinière et poêle	5	25,0	30,2
Autres	6	27,7	30,5
Total		100	100

Les $f_k^{(1)}$ et $f_{k'}^{(2)}$ ne sont pas indépendantes ; elles sont liées par la relation :

$$\sum_k f_k^{(2)} = \sum_{k'} f_{k'}^{(1)} = 1.$$

D'autre part, les probabilités de passage sont liées par les 6 équations :

$$\sum_{k=1}^{6} p_{k'k} = 1 \qquad (k' = 1, 2, \ldots, 6) \tag{5}$$

On dispose ainsi de 11 relations indépendantes pour déterminer les 36 probabilités $p_{k'k}$. Il est donc nécessaire de fixer 25 d'entre eux au moyen d'hypothèses de base et de relations supplémentaires.

L'examen des divers types d'équipement montre que certains passages ne peuvent pratiquement jamais avoir lieu, ce qui conduit à un certain nombre de probabilités nulles. De cette manière, le nombre d'inconnues se trouve

Ancien équipement	Nouvel équipement					
	1	2	3	4	5	6
1	1	0	0	0	0	0
2	p_{21}	p_{22}	0	0	0	p_{26}
3	0	p_{32}	p_{33}	0	p_{35}	0_8
4	0	p_{42}	0	p_{44}	0	p_{46}
5	0	p_{52}	0	0	p_{55}	p_{56}
6	p_{61}	0	0	0	0	p_{66}

limité à 15, liées par 11 relations indépendantes. Le tableau ci-dessus donne la matrice de base avec les inconnues restant à déterminer.

Pour se donner 4 relations supplémentaires, permettant de résoudre le système, on considère les différentes inéquations auxquelles conduisent les inégalités exprimant que les $p_{k'k}$ sont des probabilités, c'est-à-dire $0 \leqslant p_{k'k}$; $\leqslant 1$, quand on les combine avec le système d'équations (4) et (5). On obtient ainsi des limites pour certaines probabilités et l'on considèrera diverses solutions obtenues en se plaçant à ces limites.

On obtient en définitive des matrices de probabilités "limites" encadrant les matrices cherchées et qui permettent de donner des fourchettes pour la répartition des ménages entre les différentes classes d'équipement.

Il a été possible de s'assurer de la validité des résultats obtenus en les recoupant par des informations sur les ventes annuelles d'équipements de chauffage. En définitive, 3 matrices de probabilités de passage ont été étudiées, dont 2 paraissent les plus vraisemblables et conduisent d'ailleurs à des résultats assez voisins.

4.3 - Etablissement de perspectives de consommation

Par la suite, à partir d'hypothèses sur la stabilité des probabilités de passage, ou sur leur évolution, il est possible d'obtenir des perspectives de répartition des ménages entre les 6 classes retenues. A l'aide d'éléments, ou d'hypothèses complémentaires sur l'évolution des quantités moyennes consommées à l'intérieur de chaque classe, en tenant compte notamment du progrès technique, du prix du combustible et des exigences de confort, on obtient en définitive des perspectives de consommation pour chaque classe et, par addition, pour l'ensemble des consommateurs. Cette méthode analytique a fourni une prévision globale pour 1965, qui a été confirmée par deux autres études basées sur des méthodes synthétiques.

Comme on s'en rend compte, le modèle analytique établi n'a pu être utilisé en l'occurence que de manière imparfaite, parce que le matériel statistique avait été recueilli antérieurement, à l'occasion d'enquêtes destinées à d'autres fins. L'un des avantages du modèle mathématique exposé réside notamment dans le fait qu'il fixe un cadre précis aux enquêtes tendant à recueillir l'information nécessaire aux études sur les perspectives de la demande. Sur le plan explicatif, il permet de rendre compte du mécanisme des mouvements des consommateurs entre les diverses classes ; en d'autres termes, il donne une image analytique de la fluidité d'un marché.