



**HAL**  
open science

# Estimation des probabilités de transition des chaînes de Markov

Francois Bonnieux

► **To cite this version:**

Francois Bonnieux. Estimation des probabilités de transition des chaînes de Markov. 1974, 15 p. <hal-01600049>

**HAL Id: hal-01600049**

**<https://hal.science/hal-01600049v1>**

Submitted on 2 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

INRA - RENNES  
ECONOMIE RURALE  
TRAVAILLEURS

INRA - ESK

RENNES

FB n° 3

f. bonnieux



estimation des probabilités de transition  
des chaînes de markov

1974

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE  
STATION D'ECONOMIE RURALE DE RENNES

Le mot modèle recouvre une large classe d'objets entendus comme des substituts des systèmes que nous étudions. Un modèle consiste en la représentation formelle d'idées ou de connaissances relatives à un système, c'est un objet artificiel, intégralement monté. Le raisonnement sur le modèle nous permet d'explorer les conséquences logiques des hypothèses retenues, de les confronter avec les résultats de l'expérience, d'arriver ainsi à mieux connaître la réalité, et à agir plus efficacement sur elle. Si nous nous référons au domaine plus particulier de l'économie, nous constatons que l'économétrie considère que les données économiques sont généralement engendrées par un système de relations simultanées, dynamiques et stochastiques. Les chaînes de Markov ont été utilisées pour analyser des données chronologiques où une observation à chaque instant consiste en la catégorie à laquelle appartient l'observation.

Une bibliographie partielle de travaux économétriques utilisant ce modèle comprendrait ceux de CHAMPERNOWNE (1953) et SOLOW (1951) sur la distribution des revenus et des salaires ; PRAIS (1955) et GOODMAN (1965) sur la mobilité sociale, HART et PRAIS (1956), ADELMAN (1958), PRESTON et BELL (1961) et STEINDL (1965) sur la distribution des firmes par taille et la concentration ; GALE (1960) et SMITH (1961) sur des modèles analysant les relations internationales et interrégionales ; TELSER (1962, 1963) sur la consommation ; COOTNER (1964) sur les mouvements de stocks et de prix ; BONNIEUX (1969, 1972) sur des modèles de gestion. Par ailleurs de nombreux travaux ont été entrepris

à partir de ce modèle en géographie, démographie. Il s'agit donc d'un modèle d'intérêt très général. Une des principales difficultés rencontrées dans son utilisation est d'ordre statistique et concerne l'estimation de ses paramètres.

### Définition du modèle

On considère une suite d'évènements indicés par  $t$  ( $t = 0, 1; \dots T$ ) ayant chacun  $r$  issues possibles  $s_i$  ( $i = 1, 2, \dots r$ ). Ce processus peut être représenté par une variable aléatoire  $x_t$  ( $t = 0, 1, \dots T$ ) prenant ses valeurs dans  $\{1, \dots r\}$  et telle que :

$$x_t = i$$

si  $s_i$  est l'issue du  $t^{\text{e}}$  évènement.

Les évènements sont liés de telle sorte que la probabilité pour que  $x_t$  soit égale à  $i$  est entièrement définie par sa valeur à l'instant précédent, ainsi :

$$P(x_t / x_{t-1}, x_{t-2}, \dots) = P(x_t / x_{t-1})$$

D'où la probabilité d'avoir la suite  $(x_0, x_1, \dots, x_t)$  :

$$P(x_0, x_1, \dots, x_t) = P(x_0) \prod_{t=1}^T P(x_t / x_{t-1})$$

Le processus est donc défini par la distribution initiale de probabilité  $P(x_0)$  et les probabilités conditionnelles  $P(x_t/x_{t-1})$ . Posons :

$$p_{ij} = p_{ij}(t) = P(x_t = j/x_{t-1} = i)$$

Les  $p_{ij}$  sont appelées probabilités de transition associées aux passages de  $s_i$  à  $s_j$ . Nous les supposons indépendantes de  $t$ , le processus est donc supposé homogène.

On définit la matrice carrée d'ordre  $r$ ,

$$P = [p_{ij}]$$

elle est appelée matrice de transition et :

$$0 \leq p_{ij} \leq 1$$

$$\sum_j p_{ij} = 1$$

Le processus que nous venons de définir est une chaîne de Markov (homogène) à  $r$  états.

Connaissant l'état initial  $x_0$  et la matrice  $P$  on peut déterminer la distribution de chaque variable aléatoire  $x_t$ , ou encore rechercher s'il existe une loi de  $x_t$  lorsque  $t$  tend vers l'infini. Dans ce contexte on dit que la chaîne est ergodique s'il existe un vecteur

$$\Pi = (\Pi_1 \dots \Pi_r)$$

tel que

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \Pi_j$$

où  $p_{ij}^{(t)}$  est l'élément d'ordre  $(i, j)$  de  $P^t$ , ayant les deux propriétés suivantes.

$$0 \leq \Pi_j \leq 1$$

$$\sum_j \Pi_j = 1$$

$$\text{et } \Pi = \Pi P$$

La difficulté fondamentale pour utiliser une chaîne de Markov comme modèle concerne l'estimation de  $P$ .

#### Problème de l'estimation de la matrice de transition

Considérons  $n$  individus dont l'évolution est régie par une chaîne de Markov ergodique de matrice  $P$  inconnue. On désigne par  $n_i(t)$  le nombre d'individus dans l'état  $i$  à l'instant  $t$ , et par  $n_{ij}(t)$  le nombre d'individus qui passent de l'état  $i$  à l'état  $j$  dans la transition qui s'effectue de l'instant  $t-1$  à l'instant  $t$ . L'état de l'échantillon de  $n$  individus peut être repéré à chaque instant par le vecteur d'état :

$$[n_1(t), n_2(t), \dots, n_r(t)]$$

et les transitions par les matrices des comptes de transition d'élément général  $[n_{ij}(t)]$ . On appellera microdonnées les  $n_{ij}(t)$ , par opposition aux macrodonnées qui désigneront les  $n_i(t)$ . Si l'on connaît les microdonnées il est possible d'estimer P en utilisant de la méthode du maximum de vraisemblance (ANDERSON et GOODMAN 1957).

Posons :

$$n_{ij} = \sum_t n_{ij}(t)$$

les  $n_{ij}$  forment une statistique exhaustive minimale. La fonction de vraisemblance associée à un échantillon de taille  $n$  observé pendant  $T$  périodes s'écrit :

$$\left[ \prod_{t,i} \frac{n_i(t-1)!}{\prod_j n_{ij}(t)!} \right] \times \left[ \prod_{i,j} p_{ij}^{n_{ij}} \right]$$

Pour dériver les estimateurs des  $p_{ij}$  on suppose fixés les  $n_i(0)$ , on maximise le logarithme de la fonction de vraisemblance sous les contraintes :

$$\sum_j p_{ij}^{-1} = 0$$

et on trouve la solution unique :

$$\widehat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \geq 0$$

Ces estimateurs sont convergents, sans biais de plus ils suivent asymptotiquement la loi multinormale \*.

ANDERSON et GOODMAN exposent par ailleurs un certain nombre de tests basés soit sur le principe du rapport de vraisemblance, soit sur le  $\chi^2$ . On trouvera dans BONNIEUX (1972) une généralisation de ces résultats au cas non-homogène. MARTIN (1967) examine le problème de l'estimation dans le cadre bayésien.

Le plus souvent en pratique les microdonnées (c'est-à-dire les trajectoires individuelles) sont inconnues, l'information porte uniquement sur les macrodonnées (c'est-à-dire les vecteurs et état), nous examinerons le problème de l'estimation en utilisant la méthode des moindres carrés.

#### Application de la méthode des moindres carrés simples

Définissons les probabilités absolues :

$$q_j(t) = P(x_t=j)$$

on a :

$$q_j(t) = \sum_i q_i(t-1) p_{ij}$$

---

\* Selon que  $n \rightarrow \infty$  ou que  $T \rightarrow \infty$ .

Nous observons le vecteur d'état d'éléments  $n_j(t)$ , divisons les par l'effectif total  $n$  afin de définir le vecteur des fréquences observées d'éléments  $y_j(t)$ . On a la relation

$$y_j(t) = \sum_i y_i(t-1) p_{ij} + u_j(t)$$

où les  $y_i$  sont observés,  $p_{ij}$  désignent les paramètres à estimer et  $u_j(t)$  une erreur aléatoire.

Sous forme matricielle nous avons :

$$y_j = X p_j + u_j$$

$$\begin{array}{l}
 y_j : T \times 1 \\
 p_j : r \times 1 \\
 u_j : T \times 1
 \end{array}
 \quad
 \begin{array}{c}
 X \\
 \text{TXr}
 \end{array}
 =
 \begin{bmatrix}
 y_1(0) & y_2(0) & \dots & y_r(0) \\
 \vdots & \vdots & & \vdots \\
 y_1(t-1) & y_2(t-1) & \dots & y_r(t-1) \\
 \vdots & \vdots & & \vdots \\
 y_1(T-1) & y_2(T-1) & \dots & y_r(T-1)
 \end{bmatrix}$$

regroupons les  $r$  relations :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} X & 0 & \dots & 0 \\ 0 & X & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_r \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix}$$

ou encore :  $y = \tilde{X} p + u$

supposons que Rang  $X = r$

$$E(u) = 0$$

$$E(uu') = \Sigma$$

$\Sigma$  est carrée d'ordre  $Tr$  et singulière (LEE et al., 1970, p.33)  
MILLER (1952) suggère d'estimer ce modèle par l'estimateur  
des moindres carrés simples :

$$\hat{p} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' y$$

On peut vérifier (LEE et al., p.34) que les estimateurs vérifient  
la relation :

$$\sum_j \hat{p}_{ij} = 1$$

par contre la contrainte de non-négativité n'est pas nécessairement satisfaite.

MADANSKY (1959) a montré que les  $u_j$  ne sont pas corrélés à  $X$ , il en résulte que  $\hat{p}$  converge vers  $p$ . Cependant les erreurs sont hétéroscédastiques donc l'estimateur n'est pas efficace.

Une solution proposée par LEE et al., (1970),  
THEIL et REY (1966) consiste à minimiser :

$$u'u = (y - \tilde{X} p)' (y - \tilde{X} p)$$

sous les contraintes

$$p \geq 0$$

$$Gp = \eta_r$$

G est la matrice  $r \times r^2$  définie par

$$G = (i_r \dots i_r)$$

en appelant  $i_r$  la matrice identité d'ordre  $r$ , et  $\eta_r$  le vecteur à  $r$  éléments tous égaux à un.

Sous cette forme le problème se ramène à la résolution d'un programme quadratique.

### Application de la méthode des moindres carrés généralisés

Nous avons donc le modèle :

$$y = \tilde{X} p + u$$

avec  $E(u) = 0$

$$E(uu') = \Sigma$$

$\Sigma$  est une matrice non-diagonale, singulière. Appliquons au modèle une matrice de pondération  $H$  ( $r^T \times r^T$ ), nous obtenons :

$$Hy = H \tilde{X} p + H u$$

d'où :

$$\hat{p} = (X' H' H X)^{-1} X' H' H y$$

Le problème posé est donc celui du choix de H. Comme  $\Sigma$  est singulière, il n'existe pas de matrice H qui transforme les erreurs hétéroscédastiques en erreurs homocédastiques, cependant on peut définir des matrices de pondération qui conduisent à des estimateurs plus efficaces que celui des moindres carrés simples. On trouvera dans THEIL et REY (1966) MADANSKY (1959), LEE et al., (1970 p.64-67) différentes pondérations.

Nous allons montrer que la singularité de  $\Sigma$  résulte d'une redondance parmi les paramètres. Après l'avoir éliminé on pourra expliquer la méthode des moindres carrés généralisés. Auparavant précisons les hypothèses et supposons que u est indépendant de  $\tilde{X}$  ce qui entraîne que y et u ont la même matrice des covariances. On peut par ailleurs préciser la distribution de y. Supposons que les fréquences observées y sont engendrées par une distribution multinomiale d'espérances  $q_j(t)$ , de variances  $q_j(t) [1 - q_j(t)] / N(t)$  et de covariances  $-q_i(t) q_j(t) / N(t)$ .

$$\Sigma = (\Sigma_{ij}) \quad i, j = 1 \dots r$$

et

$$\Sigma_{ij} = \begin{bmatrix} \frac{-q_i(1)q_j(1)}{N(1)} & & & \\ & \frac{-q_i(2)q_j(2)}{N(2)} & \dots & \\ & & \dots & \\ & & & \frac{q_i(T)q_j(T)}{N(T)} \end{bmatrix}$$

$i \neq j$



On constate que les  $r$  équations sont linéairement dépendantes, nous écrivons donc le modèle sous la forme :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{r-1} \end{bmatrix} = \begin{bmatrix} X & & & \\ & X & & 0 \\ & & \ddots & \\ & 0 & & X \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{r-1} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{r-1} \end{bmatrix}$$

ce qui donne la forme condensée

$$y_* = X_* p_* + u_*$$

avec  $E(u_*) = 0$

$$E(u_* u_*') = \Sigma_*$$

où  $\Sigma_*$  est une sous matrice régulière de  $\Sigma$ .

Nous sommes alors amenés à résoudre le problème suivant :

$$\text{minimiser } (y_* - X_* p_*)' \Sigma_*^{-1} (y_* - X_* p_*)$$

sous les contraintes  $R p_* \leq \eta_r$  et  $p_* \geq 0$

où  $R$  désigne une sous matrice de  $G$ .

Pour le faire on peut utiliser un algorithme de programmation quadratique.

Rappelons qu'en l'absence de contraintes nous aurions l'estimation suivante :

$$\hat{p}_* = (X_*' \Sigma_*^{-1} X_*)^{-1} X_*' \Sigma_*^{-1} y_*$$

$$\text{et } v(\hat{p}_*) = (X_*' \Sigma_*^{-1} X_*)^{-1}$$

$$\text{d'où } \hat{p}_r = \eta_r - R \hat{p}_*$$

Toutefois rien ne nous assure que cet estimateur satisfasse aux contraintes. Pour estimer  $\Sigma$  on remplacera les vrais fréquences  $q_j(t)$  par les fréquences observées  $y_j(t)$  ce qui conduit à une estimation convergente.

#### Autres méthodes, choix d'un estimateur

LEE et al., (1970) étudient en détail les différentes méthodes d'estimation, outre celles basées sur les moindres carrés, ils envisagent la méthode du  $X^2$  minimum, celle du maximum de vraisemblance et celle basée sur la minimisation de la somme des valeurs absolues des écarts. Par ailleurs ils envisagent le problème sous l'angle bayésien.

Les études de Monte-Carlo peuvent servir de base pour choisir une méthode d'estimation. Lorsque les microdonnées existent, la méthode du maximum de vraisemblance fournit les meilleurs résultats. Dans le cas général où seules les macrodonnées existent, les méthodes sous contraintes sont meilleures que les autres, on peut retenir celle des moindres carrés généralisés, toutefois lorsqu'on dispose d'informations a priori il faut choisir l'approche bayésienne. LEE et JUDGE (1972) ont tenté de généraliser cette approche au cas d'une chaîne non homogène.

#### BIBLIOGRAPHIE -

- ADALMAN I.-G., 1958 ; "A stochastic analysis of the size distributions of farms"  
JASA, 53, 893-904.
- ANDERSON T.-W., GOODMAN L.-A., 1957 ; "Statistical inference about Markov chains"  
AMS, 28, 89-110.
- BONNIEUX F., 1969 ; "Conduite d'un élevage en avenir aléatoire"  
Rech. Econ. Soc. Rur., 2, 29-40.
- BONNIEUX F., 1972 ; "Modèle de conduite d'un élevage de veaux de boucherie"  
Rev. Stat. Appli., 20, 71-84.
- CHAMPERNOWNE D.-B., 1953 ; "A model of income distributions"  
Economic Journal, 63, 318-351.
- COOTNER P.-H., 1964 ; The random character of stock market prices.  
MIT Press, Cambridge Mass.
- GALE D., 1960 ; The theory of linear economic models  
Mc Graw-Hill, New-York, pp.260-290.
- GOODMAN L.-A., 1965 ; "On statistical analysis of mobility tables"  
Am. J. of Sociology, 70, 564-585.
- HART P.-E., PRAIS S.-J., 1956 ; "The analysis of business concentration, a statistical approach"  
JRSSA, 119, 150-175.

- LEE T.-C., JUDGE G.-G., 1972 ; "Estimation of transition probabilities in a non stationary finite Markov chains"  
Metroeconomica, 34, 180-201.
- LEE T.-C., JUDGE G.-G., ZELLNER A., 1970 ; Estimating the parameters of the markov probability model from aggregate time series data.  
North Holland Publ. Co, Amsterdam.
- MADANSKY A., 1959., "Least squares estimation in finite Markov processes". Psychometrika, 24, 137-144.
- MARTIN J.-J., 1967 ; Bayesian decision problems and Markov chains. Wiley, New-York.
- MILLER G.-A., 1952 ; "Finite Markov processes in psychology"  
Psychometrika, 17, 149-167.
- PRAIS S.-J., 1955;"Measuring social mobility" JRSSA, 118,56-66.
- PRESTON L.-E., BELL J., 1961 ; "The statistical analysis of industry structure : an application to the food industry" JASA, 56, 925-932.
- SMITH P.-E., 1961 ; "Markov chains, exchange matrices and regional development". J. of Regional Science, 3, 27-36.
- SOLOW R., 1951 ; "Some long-run aspects of the distribution of wage incomes". Econometrica, 19,333-334.
- STEINDL J., 1965 ; Random processes and the growth of firms. Hafner Publ. Co, New-York.
- TELSER L. - G., 1962 ; "The demand for branded goods as estimated from consumer panel data"  
Rev. of Econ. d. Stat, 44, 300-324.
- TELSER L. - G., 1962 ; "Advertising and cigarettes". JPE, 70, 471-499.
- TELSER L. - G., 1963 ; "Least squares estimates of transition probabilities" in Measurement in economics, Stanford UP.
- THEIL H., REY G., 1966 ; "A quadratic programming approach to the estimation of transition probabilities" Management Science, 12, 714-721.