REVUE DE STATISTIQUE APPLIQUÉE

A. BERCHTOLD

Modélisation autorégressive des chaînes de Markov : utilisation d'une matrice différente pour chaque retard

Revue de statistique appliquée, tome 44, n° 3 (1996), p. 5-25

<http://www.numdam.org/item?id=RSA_1996__44_3_5_0>

Rev. Statistique Appliquée, 1996, **XLIV** (3), 5-25

MODÉLISATION AUTORÉGRESSIVE DES CHAÎNES DE MARKOV : UTILISATION D'UNE MATRICE DIFFÉRENTE POUR CHAQUE RETARD*

A. Berchtold

André Berchtold, Université de Genève
THEMES, Département d'Econométrie,
102, bd Carl-Vogt, CH-1211 Genève 4, Suisse.
e-mail : Andre.Berchtold@themes.unige.ch
URL :http ://www.unige.ch :80/ses/ecoth/staff/berchtol/welcome.html

RÉSUMÉ

La réduction du nombre de paramètres des chaînes de Markov d'ordre supérieur à 1 a déjà suscité plusieurs articles. En particulier, Raftery (1985) a proposé une modélisation autorégressive utilisant la même matrice de transition, affectée d'un coefficient, pour chaque retard. Dans cet article, nous montrons qu'un modèle du même type, mais utilisant une matrice différente pour chacun des retards, permet d'obtenir de meilleurs résultats, tout en restant simple à estimer, même lorsque l'on n'a que peu de données à disposition.

Mots-clés : *Chaîne de Markov, Matrice d'ordre (1,g), Théorème limite, Données manquantes, Qualité de la modélisation.*

ABSTRACT

The reduction of the number of parameters in high-order Markov chain already inspired several articles. In particular, Raftery (1985) proposed an autoregressive modelling using a same transition matrix, with a coefficient, for every lag. In this paper, we show that a model of the same type, but utilizing a different matrix for each lag, give best results and is not harder to estimate, even when the number of data is small.

Keywords : *Markov chain, Matrix of order (1,g), Limit theorem, Missing data, Quality of modelling.*

* Nous tenons à remercier les professeurs Claude Tricot et Gilbert Ritschard, de l'Université de Genève, pour leur soutien lors de la réalisation de ce travail.

1. Introduction

Le principal problème des chaînes de Markov réside dans leur très grand nombre de paramètres à estimer, ce qui a généralement empêché l'utilisation de chaînes d'ordre supérieur à 1, même lorsque les données étaient visiblement d'ordre supérieur à 1. En réponse à ce problème, plusieurs modèles alternatifs ont été proposés pour réduire le nombre de paramètres tout en conservant une bonne modélisation des données. Les principaux sont ceux de Jacobs & Lewis (1978), Pegram (1980), Logan (1981) et Raftery (1985). Dans cet article, après un bref rappel du modèle autorégressif de Raftery (aussi appelé «Mixture Transition Distribution Model»), nous proposerons deux modèles dérivés, permettant une meilleure représentation de la plupart des séquences de données de par l'utilisation de matrices de transition différentes pour chacun des retards. Nous appliquerons également le traitement des données manquantes proposé par Mehran (1989) à nos modèles, et nous terminerons par une illustration numérique de nos propos.

1.1 Le modèle de Raftery

Soit une variable discrète $\{X_t; t \in \mathbb{N}\}$ prenant ses valeurs dans l'ensemble $V = \{1; \ldots; m\}$ et soit une chaîne de Markov d'ordre l. Soit les états e_{t-1} et e_t, définissant respectivement les périodes $t-1$ et t d'une chaîne de Markov d'ordre l :

$$e_{t-1} = [X_{t-1} = i_1 \ ; \ X_{t-2} = i_2 \ ; \ \ldots \ ; \ X_{t-l+1} = i_{l-1} \ ; \ X_{t-l} = i_l]$$
$$e_t = [X_t = i_0 \ ; \ X_{t-1} = i_1 \ ; \ \ldots \ ; \ X_{t-l+2} = i_{l-2} \ ; \ X_{t-l+1} = i_{l-1}]$$

avec $i_0, i_1, \ldots, i_l \in V$.

L'approximation autorégressive proposée par Raftery consiste à écrire les probabilités de transition entre les états e_{t-1} et e_t comme :

$$\begin{aligned} P(X_t = i_0, \ldots, X_{t-l+1} = i_{l-1} | X_{t-1} = i_1, \ldots, X_{t-l} = i_l) &= P(X_t = i_0 | X_{t-1} = i_1, \ldots, X_{t-l} = i_l) \\ &\cong \varphi_1 P(X_t = i_0 | X_{t-1} = i_1) + \ldots + \varphi_l P(X_t = i_0 | X_{t-l} = i_l) \\ &\cong \varphi_1 q_{i_1 i_0} + \ldots + \varphi_l q_{i_l i_0} \\ &= \sum_{g=1}^{l} \varphi_g q_{i_g i_0} \end{aligned} \tag{1}$$

où $q_{i_1 i_0}, \ldots, q_{i_l i_0}$ sont des probabilités de transition appartenant à une matrice de probabilités de type «ligne» (ie : dont chaque ligne est une distribution de probabilité), de dimensions $(m \times m)$, notée Q :

$$Q = \begin{pmatrix} q_{1,1} & \cdots & q_{1,m} \\ \vdots & & \vdots \\ q_{m,1} & \cdots & q_{m,m} \end{pmatrix}$$

avec $\iota = Q\iota$, où ι est un vecteur de dimensions $(m \times 1)$ composé de 1.

Si l'on considère l'ensemble des valeurs possibles de la variable X à l'époque t, le modèle peut être récrit comme suit. Soit $\chi'_{t-g} = (x_{t-g}(1), \ldots, x_{t-g}(m))$, un vecteur où $x_{t-g}(j) = 1$ si $X_{t-g} = j$, et 0 sinon, $g = 1, \ldots, l$, $j = 1, \ldots, m$. Soit $\hat{\chi}'_t = (P(X_t = 1), \ldots, P(X_t = m))$, le vecteur des probabilités d'être dans chacune des m valeurs à l'époque t. Alors :

$$\hat{\chi}'_t = \sum_{g=1}^{l} \varphi_g \chi'_{t-g} Q \tag{2}$$

Nous pouvons remarquer que les vecteurs de sélection χ'_{t-g} sont écrits sans chapeau, car la valeur de la variable X aux périodes $t-1$ à $t-l$ est supposée connue, ce qui implique que ces vecteurs sont définis de façon unique et ne sont donc pas aléatoires. En revanche, il est entendu que si une valeur passée de X est inconnue, le vecteur χ'_{t-g} correspondant doit être remplacé par une distribution de probabilité notée $\hat{\chi}'_{t-g}$.

Les résultats de ce modèle autorégressif étant des probabilités, il est nécessaire d'une part que chacun de ces résultats soit compris entre 0 et 1 inclus et d'autre part que la somme des probabilités d'atteindre l'époque t à partir d'une combinaison quelconque des valeurs de $t-1$ à $t-l$ soit strictement égale à 1. Le second problème est facilement résolu en posant la contrainte suivante :

$$\sum_{g=1}^{l} \varphi_g = 1 \tag{3}$$

Pour contraindre les probabilités à être comprises entre 0 et 1 deux solutions ont été proposées : La première consiste à contraindre les paramètres $\varphi_1, \ldots, \varphi_l$ à ne prendre que des valeurs comprises entre 0 et 1 inclus. La seconde (Raftery & Tavaré (1994)) consiste à estimer le modèle sans contraintes de non-négativité des paramètres φ_g, puis à vérifier que la plus petite probabilité pouvant être calculée par le modèle n'est pas négative. Formellement, il suffit que les inégalités suivantes soient vérifiées :

$$T\, q_{-(j)} + (1 - T)\, q_{+(j)} \geqslant 0 \quad , \quad j = 1 \ldots, m \tag{4}$$

avec :

$$T = \sum_{\substack{g=1 \\ \varphi_g \geqslant 0}}^{l} \varphi_g$$

$$q_{-(j)} = \min_{1 \leqslant i \leqslant m} \{q_{ij}\}$$

$$q_{+(j)} = \max_{1 \leqslant i \leqslant m} \{q_{ij}\}$$

L'estimation proprement dite des paramètres $\varphi_1, \ldots, \varphi_l$ s'effectue par maximisation de la log-vraisemblance suivante :

$$L = \sum_{i_0=1}^{m} \cdots \sum_{i_l=1}^{m} n_{i_0,\ldots,i_l} \log \left(\sum_{g=1}^{l} \varphi_g q_{i_g i_0} \right)$$

où $n_{i_0,\ldots,i_l}$ est le nombre d'apparitions de la séquence :

$$X_t = i_0 \; ; \; X_{t-1} = i_1 \; ; \; \ldots \; ; \; X_{t-l} = i_l \quad , \quad \forall i_0, \ldots, i_l \in V \tag{5}$$

dans les données utilisées.

Cette modélisation autorégressive des chaînes de Markov d'ordre l ne nécessite l'identification que de $m[m-1] + [l-1]$ paramètres indépendants ($m[m-1]$ pour la matrice de transition Q et $[l-1]$ pour les paramètres φ_g, le l-ième étant dépendant par la contrainte $\sum_{g=1}^{l} \varphi_g = 1$), ce qui est généralement très inférieur aux $m^l[m-1]$ paramètres indépendants de la vraie matrice de transition d'ordre l.

1.2 Chaînes de Markov d'ordre (1,g)

De façon similaire aux probabilités de transition homogènes d'ordre 1, nous pouvons définir les probabilités de transition homogènes d'ordre $(1, g)$, $g \geqslant 1$, entre les périodes $t-g$ et t :

$$P(X_t = i_0 | X_{t-g} = i_g) = q_{(g)i_g i_0} \quad , \quad i_0, i_g \in V \quad , \quad g \in \mathbb{N}^* \quad , \quad \forall t$$

Pour chaque retard g, ces probabilités génèrent une matrice Q_g, de dimensions $(m \times m)$:

$$Q_g = \begin{pmatrix} q_{(g)1,1} & \cdots & q_{(g)1,m} \\ \vdots & & \vdots \\ q_{(g)m,1} & \cdots & q_{(g)m,m} \end{pmatrix}$$

avec $\iota = Q_g \iota$, où ι est un vecteur de dimensions $(m \times 1)$ composé de 1.

Les probabilités de transition d'ordre $(1, g)$ se relient aux probabilités de transition d'ordre 1 de la manière suivante :

- La matrice de transition homogène d'ordre (1,1), Q_1, est identique à la matrice des transition homogènes d'ordre 1, Q.
- Si l'hypothèse d'homogénéité des probabilités de transition d'ordre 1 est vérifiée, la relation suivante est vraie :

$$Q^g = Q_g \quad , \quad g \in \mathbb{N}^*$$

2. Les deux modèles dérivés

2.1 Les modèles

Nous présentons maintenant deux modèles dérivés de celui de Raftery, mais faisant intervenir une matrice différente pour chacun des retards :

- Modèle utilisant la matrice Q^g pour le retard d'ordre g :

$$\hat{\chi_t}' = \sum_{g=1}^{l} \varphi_g \chi'_{t-g} Q^g \tag{6}$$

- Modèle utilisant la matrice de transition d'ordre $(1, g)$ pour le retard d'ordre g :

$$\hat{\chi_t}' = \sum_{g=1}^{l} \varphi_g \chi'_{t-g} Q_g \tag{7}$$

Un commentaire s'impose à propos du nombre de paramètres de ces trois différents modèles. Nous savons déjà que le modèle (2), en plus des $l-1$ paramètres φ_g indépendants, ne nécessite l'estimation que des $m[m-1]$ paramètres indépendants de la matrice Q. Le modèle (6) diffère du (2) par l'utilisation des matrices Q^g, mais, comme ces matrices sont entièrement définies à partir de Q, le nombre total de paramètres indépendants de ce modèle est identique à celui du modèle précédent, soit $[l-1]+m[m-1]$. En revanche, le modèle (7) nécessite l'estimation des paramètres de l matrices de dimensions $m \times m$, soit en tout $[l-1]+lm[m-1]$ paramètres indépendants, ce qui semble aller à l'encontre de nos efforts visant à pouvoir estimer facilement l'ensemble des paramètres du modèle. Ce n'est cependant pas le cas, car les $lm[m-1]$ paramètres concernant les l matrices de transition ne s'estiment pas en une seule fois (comme dans le cas d'une matrice de transition d'ordre l), mais en l étapes indépendantes, l'ensemble des données pouvant être utilisé pour chacune d'elles.

Pour illustrer notre propos, prenons l'exemple d'une séquence de n=100 données sur m=4 valeurs différentes, avec un modèle d'ordre l=3. La matrice de transition d'ordre 3 nécessite l'identification de $4^3[4-1] = 192$ paramètres indépendants à l'aide de $100-3 = 97$ données utilisables, ce qui est manifestement irréalisable. En revanche, le modèle (7) nécessite, en plus de l'estimation des paramètres φ_g, l'identification de $4[4-1] = 12$ paramètres indépendants à l'aide de $100-1 = 99$ données utilisables pour la matrice Q_1, l'identification de $4[4-1] = 12$ paramètres à l'aide de $100-2 = 98$ données utilisables pour Q_2 et l'identification de $4[4-1] = 12$ paramètres à l'aide de $100-3 = 97$ données utilisables pour Q_3. Nous voyons ainsi que, lorsque les nombres de retards et de valeurs sont petits par rapport au nombre de données disponibles, l'utilisation des matrices Q_g est tout-à-fait possible.

2.2 Contraintes sur les paramètres

Comme dans le cas du modèle (2), les paramètres $\varphi_1, \ldots, \varphi_l$ doivent satisfaire à certaines contraintes pour garantir que les résultats obtenus soient bien des probabilités. La première contrainte donnée par l'équation (3) reste valable pour les deux nouveaux modèles. En revanche, pour le cas où l'on admet que certains φ_g peuvent être négatifs, les inégalités (4) doivent être récrites de la façon suivante :

PROPOSITION 1. *Les résultats du modèle (7) seront tous non-négatifs si les inéquations suivantes sont respectées :*

$$\sum_{\substack{g=1 \\ \varphi_g \geqslant 0}}^{l} \varphi_g q_{(g)-(j)} + \sum_{\substack{g=1 \\ \varphi_g < 0}}^{l} \varphi_g q_{(g)+(j)} \geqslant 0 \quad , \quad j = 1 \ldots, m \tag{8}$$

avec :

$$q_{(g)-(j)} = \min_{1 \leqslant i \leqslant m} \{q_{(g)ij}\}$$
$$q_{(g)+(j)} = \max_{1 \leqslant i \leqslant m} \{q_{(g)ij}\}$$

où $q_{(g)ij}$ représente la probabilité contenue dans la matrice de transition Q_g, d'ordre $(1, g)$, à l'intersection de la i-ième ligne et de la j-ième colonne.

Ce résultat est obtenu par une démonstration similaire à celle donnée dans Raftery & Tavaré (1994).

Si l'on remplace chaque matrice Q_g par Q^g, la proposition 1 s'applique également au modèle (6).

2.3 Théorème limite

Les modélisations proposées ne sont pertinentes que si leur distribution limite, c'est-à-dire la distribution stable atteinte lorsque le modèle est appliqué répétitivement sur un grand nombre de périodes consécutives, est identique à celle du modèle standard d'ordre 1. Nous donnons ci-dessous deux théorèmes limites pour les modèles (6) et (7).

PROPOSITION 2. *Soit la variable discrète $\{X_t; t \in \mathbb{N}\}$ prenant ses valeurs dans l'ensemble $V = \{1, \ldots, m\}$ et soit une séquence de données de taille n. Soit la matrice de transition Q, d'ordre 1, supposée régulière, et soit une matrice de transition Q_g, d'ordre $(1, g)$, $g > 0$, supposée régulière elle aussi. Alors, lorsque $n \to \infty$, Q_g a la même distribution limite que Q.*

DÉMONSTRATION DE LA PROPOSITION 2.

Soit $\pi = (\pi_1, \ldots, \pi_m)'$ la distribution limite de la matrice Q, $\pi_{(g)} = (\pi_{(g)1}, \ldots, \pi_{(g)m})'$ la distribution limite de la matrice Q_g, n_i le nombre d'apparitions de la valeur i dans les données et p_i la probabilité d'apparition de la valeur i dans les données :

$$p_i = \frac{n_i}{n} \quad , \quad i = 1, \ldots, m$$
$$p = (p_1, \ldots, p_m)'$$

La loi des grands nombres (Kemeny & Snell (1976)) dit que lorsque la matrice Q est régulière, nous avons :

$$\lim_{n \to \infty} p = \pi \tag{9}$$

De façon similaire, nous obtenons pour la matrice Q_g :

$$\lim_{n \to \infty} p = \pi_{(g)} \tag{10}$$

Finalement, en égalisant (9) et (10), nous obtenons :

$$\pi = \pi_{(g)}$$

□

PROPOSITION 3. *Supposons que* $Q(= Q_1)$, Q_2, $\ldots, Q_l$ *et* $\sum_{g=1}^{l} \varphi_g Q_g$, *sont des matrices de transition régulières, construites à partir d'une séquence de données de taille* $n \to \infty$. *Alors,* $\sum_{g=1}^{l} \varphi_g Q_g$ *a la même distribution limite que* Q.

DÉMONSTRATION DE LA PROPOSITION 3

Soit $\pi = (\pi_1, \ldots, \pi_m)'$, la distribution limite de Q. Si $n \to \infty$ et si les matrices $Q_1, \ldots, Q_l$ sont régulières, nous pouvons écrire, en appliquant la proposition 2 :

$$\begin{aligned} \pi' &= \pi' Q_1 \\ &\vdots \\ \pi' &= \pi' Q_l \end{aligned}$$

En multipliant chaque équation par le paramètre φ_g correspondant et en sommant, nous obtenons :

$$\varphi_1\pi' + \ldots + \varphi_l\pi' = \varphi_1\pi'Q_1 + \ldots + \varphi_l\pi'Q_l$$

$$\left(\sum_{g=1}^{l}\varphi_g\right)\pi' = \pi'\left(\sum_{g=1}^{l}\varphi_g Q_g\right)$$

$$\pi' = \pi'\left(\sum_{g=1}^{l}\varphi_g Q_g\right)$$

Comme nous supposons la matrice $\sum_{g=1}^{l}\varphi_g Q_g$ régulière, le vecteur π est bien sa distribution limite.

□

THÉORÈME 1. *Théorème limite pour le modèle (7)*

Soit la variable $\{X_t; t \in \mathbb{N}\}$ *prenant ses valeurs dans l'ensemble* $V = \{1, \ldots, m\}$ *et soit le modèle (7). Supposons que les matrices* $Q(= Q_1), Q_2, \ldots, Q_l$ *sont régulières, que* $\sum_{g=1}^{l}\varphi_g Q_g$ *est strictement positive, que* $\pi = (\pi_1, \ldots, \pi_m)'$ *est la distribution limite de* Q, $\pi_k > 0$, $\sum_{k=1}^{m}\pi_k = 1$ *et que la séquence de données utilisée est de taille* $n \to \infty$. *Alors :*

$$\lim_{t\to\infty} P(X_t = i_0 | X_1 = i_1, \ldots, X_l = i_l) = \pi_{i_0} \quad , \quad i_0, i_1, \ldots, i_l \in V$$

Ce résultat est obtenu par une démonstration similaire à celle donnée dans Raftery (1985) et en utilisant la proposition 3.

La proposition 3 utilisée dans la démonstration du théorème limite nécessite que la séquence de données soit de taille infinie. Il est évident que ceci n'est pas possible dans la réalité. Nous nous contenterons alors d'utiliser une séquence de données aussi grande que possible et nous vérifierons empiriquement que la distribution limite des matrices Q_g est suffisamment proche de la distribution limite de Q pour rendre le modèle pertinent.

Pour démontrer le théorème limite dans le cas du modèle (6), nous utiliserons la proposition suivante :

PROPOSITION 4. *Supposons que* Q *et* $\sum_{g=1}^{l}\varphi_g Q^g$ *sont des matrices de transition régulières. Alors,* $\sum_{g=1}^{l}\varphi_g Q^g$ *a la même distribution limite que la matrice* Q.

La démonstration de cette proposition est similaire à la démonstration de la proposition 3, à condition de remplacer les matrices Q_g par Q^g.

THÉORÈME 2. *Théorème limite pour le modèle (6)*

Soit la variable $\{X_t; t \in \mathbb{N}\}$ *prenant ses valeurs dans l'ensemble* $V = \{1, \ldots, m\}$ *et soit le modèle (6). Supposons que la matrice de transition* Q *est régulière, que* $\pi = (\pi_1, \ldots, \pi_m)'$ *est sa distribution limite,* $\pi_k > 0$, $\sum_{k=1}^{m} \pi_k = 1$, *et que la matrice* $\sum_{g=1}^{l} \varphi_g Q^g$ *est strictement positive. Alors :*

$$\lim_{t \to \infty} P(X_t = i_0 | X_1 = i_1, \ldots, X_l = i_l) = \pi_{i_0} \quad , \quad i_0, i_1, \ldots, i_l \in V$$

Ce résultat est obtenu par une démonstration similaire à celle donnée dans Raftery (1985) et en utilisant la proposition 4.

3. Détermination du modèle le plus performant

Nous allons à présent démontrer que le modèle (7) permet de représenter correctement un plus grand nombre de séquences de données que les modèles (2) et (6). Nous donnons tout d'abord les deux définitions sur lesquelles est basé le raisonnement.

DÉFINITION 1. *Qualité d'une modélisation autorégressive*

Dans la classe des modélisations autorégressives d'ordre l *utilisant des matrices de transition homogènes d'ordre 1, un modèle sera d'autant meilleur que la probabilité utilisée pour le passage d'une valeur* i_k *de l'époque* t-k *à une valeur* i_0 *de l'époque* t *est proche de la probabilité de transition homogène d'ordre* $(1, k)$ *pouvant être calculée d'après les données,* $\forall\, i_0, i_k \in V,\ k = 1, \ldots, l$.

DÉFINITION 2. *Modélisation autorégressive correcte*

Dans la classe des modélisations autorégressives d'ordre l *utilisant des matrices de transition homogènes d'ordre 1, un modèle sera dit «correct» si les probabilités de transition homogènes utilisées entre* t-k *et* t *sont exactement les probabilités homogènes d'ordre* $(1, k)$ *calculées à partir des données,* $k = 1, \ldots, l$.

Dans un premier temps (propositions 5 à 7), nous supposerons que les données utilisées sont strictement homogènes.

PROPOSITION 5. *Le modèle (2) ne permet une représentation autorégressive correcte des données que si la matrice* Q *est idempotente.*

DÉMONSTRATION DE LA PROPOSITION 5

Dans le modèle (2), la matrice Q est utilisée pour chacun des retards de 1 à l. Elle représente donc simultanément les probabilités de transition entre $t-1$ et t, $t-2$ et t, ..., $t-l$ et t. Or, nous savons par ailleurs que, si Q est la matrice des transition homogènes d'ordre 1 entre $t-1$ et t, alors la matrice des transitions homogènes entre $t-2$ et t est la matrice Q^2, ..., et la matrice des transitions homogènes entre $t-l$ et t est la matrice Q^l. Cela implique le système de contraintes suivant :

$$\begin{aligned} Q &= Q^2 \\ &\vdots \\ Q &= Q^l \end{aligned}$$

Autrement dit :

$$Q = Q^2 = \ldots = Q^l$$

ce qui signifie que la matrice Q est idempotente. Le modèle (2) ne permet donc une modélisation autorégressive correcte des données que si celles-ci donnent lieu à une matrice de transition homogènes de premier ordre idempotente.

□

PROPOSITION 6. *Le modèle (6) permet une modélisation autorégressive correcte d'un plus grand nombre de séquences de données que le modèle (2).*

DÉMONSTRATION DE LA PROPOSITION 6

La démonstration de cette proposition découle directement de celle de la proposition 5 : le modèle (6), en utilisant la matrice Q^g pour le retard d'ordre g, $g = 1, \ldots, l$, permet aussi la modélisation autorégressive correcte de séquences de données dont la matrice de transition de premier ordre n'est pas idempotente.

□

PROPOSITION 7. *Si la modélisation autorégressive des données fournie par le modèle (6) est correcte, alors celle fournie par le modèle (7) l'est aussi.*

DÉMONSTRATION DE LA PROPOSITION 7

Si la séquence de données étudiée est parfaitement homogène d'ordre 1, alors Q représente exactement, $\forall t$, les probabilités de transition entre $t-1$ et t, Q^2 représente exactement, $\forall t$, les probabilités de transition entre $t-2$ et t, etc... De plus, comme nous supposons que les données sont parfaitement homogènes d'ordre 1, la relation suivante est vraie :

$$Q^g = Q_g \quad , \quad g = 1, \ldots, l$$

Ainsi, dans ce cas, les modèles (6) et (7) sont strictement identiques et donnent par conséquent les mêmes résultats.

□

La proposition 7 ci-dessus n'est valable que pour des données strictement homogènes. Par contre, la proposition ci-dessous a un plus large champ d'application puisqu'elle s'applique au cas où les données ne sont pas homogènes.

PROPOSITION 8. *Le modèle (7) permet une modélisation autorégressive correcte d'un plus grand nombre de séquences de données que les modèles (2) et (6).*

DÉMONSTRATION DE LA PROPOSITION 8

Nous venons de voir que, si la séquence de données est parfaitement homogène d'ordre 1, les modèles (6) et (7) sont identiques. Par contre, si les données ne sont pas parfaitement homogènes d'ordre 1, ces deux modèles seront différents. En effet, si les données ne sont pas homogènes d'ordre 1, la matrice Q ne contient qu'une *approximation moyenne* des vraies probabilités de transition entre $t-1$ et t, $\forall t$. Partant de là, les matrices Q^g, $g = 2, \ldots, l$, non seulement ne représentent pas les vraies probabilités de transition entre $t-g$ et t, mais de plus ne sont pas forcément les meilleures approximations homogènes de ces probabilités de transition. Il en découle que le modèle (7) représente mieux les données que le (6), puisque les matrices Q_g sont les meilleures approximations homogènes des probabilités de transition entre les époques $t-g$ et t. De façon similaire, Q n'étant pas la meilleure approximation des probabilités de transition entre $t-2$ et $t, \ldots, t-l$ et t, le modèle (7) représente mieux les données que le modèle (2).

□

L'exemple suivant est une illustration de la proposition 8. Soit la séquence de données :

$$\ldots, 1, 1, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2, \ldots$$

La matrice de transition homogène d'ordre 1 de ces données s'écrit :

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

ce qui implique :

$$Q^2 = Q^3 = \ldots = Q^l = Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Or, d'après les données, les matrices de transition homogènes d'ordre $(1, g)$ sont :

$$Q_{2k+1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad , \quad k \in \mathbb{N}$$

$$Q_{4k+2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad k \in \mathbb{N}$$

$$Q_{4k+4} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad k \in \mathbb{N}$$

Les matrices Q et Q^g ne sont donc pas les meilleures approximations homogènes des probabilités de transition entre $t - g$ et t, et, sur cet exemple, le modèle (7) est plus performant que les modèles (2) et (6).

THÉORÈME 3. *Choix d'un modèle*

Parmi les trois modèles proposés, le modèle (7), utilisant les matrices de transition d'ordre $(1, g)$, $g = 1, \ldots, l$, *est le plus adapté à la représentation autorégressive de toute séquence de données, homogène ou non. Dans le cas où les données sont strictement homogènes d'ordre 1, les modèles (6) et (7) sont identiques.*

DÉMONSTRATION DU THÉORÈME 3

Le théorème 3 découle directement des propositions 5 à 8.

□

Au vu du théorème 3, le modèle (6) ne semble pas se justifier puisqu'il n'est jamais meilleur que le (7), mais en pratique, si les données utilisées sont strictement homogènes d'ordre 1, l'utilisation du modèle (6) se justifie, car il est beaucoup plus facile à estimer que le (7) (une seule matrice à estimer au lieu de l).

4. Traitement des données manquantes

Dans ce chapitre, nous adaptons le traitement des données manquantes proposé par Mehran (1989) au cas des modèles (6) et (7).

PROPOSITION 9. *Soit la séquence* $\ldots, X_{t-2}, X_{t-1}, X_t, \ldots$, *dans laquelle la donnée* X_{t-k}, $k \leqslant l$, *est manquante. Dans ce cas, le modèle autorégressif utilisant les matrices d'ordre* $(1, g)$ *est identique au modèle (7), excepté que le terme* $q_{(k)i_k i_0}$ *est remplacé par :*

$$\begin{aligned} q_{(k)i_k i_0} &= \varphi_1 q_{(k+1)i_{k+1} i_0} + \ldots + \varphi_l q_{(k+l)i_{k+l} i_0} \\ &= \sum_{r=1}^{l} \varphi_r q_{(k+r)i_{k+r} i_0} \end{aligned} \tag{11}$$

où $q_{(k+r)i_{k+r}i_0}$ *est la probabilité de passer de l'état* i_{k+r} *à l'état* i_0 *dans la matrice* Q_{k+r}.

DÉMONSTRATION DE LA PROPOSITION 9

Sachant que, dans le cadre du modèle (7), nous avons :

$$\sum_{i_k \in V} q_{(r)i_{k+r}i_k} = 1 \quad , \quad r > 0$$

$$\sum_{r=1}^{l} \varphi_r = 1$$

$$\sum_{i_k \in V} q_{(r)i_{k+r}i_k} q_{(k)i_k i_0} = q_{(k+r)i_{k+r}i_0}$$

nous pouvons écrire :

$$\begin{aligned}
&P(X_t = i_0 | X_{t-1} = i_1, \ldots, X_{t-k+1} = i_{k-1}, X_{t-k-1} = i_{k+1}, \ldots, X_{t-l} = i_l) \\
&\quad = \sum_{i_k \in V} P(X_t = i_0 | X_{t-1} = i_1, \ldots, X_{t-k} = i_k, \ldots, X_{t-l} = i_l) \\
&\qquad\qquad P(X_{t-k} = i_k | X_{t-k-1} = i_{k+1}, \ldots, X_{t-k-l} = i_{k+l}) \\
&\quad \cong \sum_{i_k \in V} \Big(\sum_{g=1}^{l} \varphi_g q_{(g)i_g i_0}\Big)\Big(\sum_{r=1}^{l} \varphi_r q_{(r)i_{k+r}i_k}\Big) \\
&\quad = \sum_{\substack{g=1 \\ g \neq k}}^{l} \varphi_g q_{(g)i_g i_0} \Big(\sum_{r=1}^{l} \varphi_r \Big(\sum_{i_k \in V} q_{(r)i_{k+r}i_k}\Big)\Big) + \varphi_k \sum_{r=1}^{l} \varphi_r \Big(\sum_{i_k \in V} q_{(r)i_{k+r}i_k} q_{(k)i_k i_0}\Big) \\
&\quad = \sum_{\substack{g=1 \\ g \neq k}}^{l} \varphi_g q_{(g)i_g i_0} + \varphi_k \sum_{r=1}^{l} \varphi_r q_{(k+r)i_{k+r}i_0}
\end{aligned}$$

ce qui vérifie la proposition.

□

Pour l'ensemble des valeurs, $\chi'_{t-k} Q_k$ est remplacé par :

$$\begin{aligned}
\chi'_{t-k} Q_k &= \varphi_1 \chi'_{t-k-1} Q_{k+1} + \ldots + \varphi_l \chi'_{t-k-l} Q_{k+l} \\
&= \sum_{r=1}^{l} \varphi_r \chi'_{t-k-r} Q_{k+r} \qquad (12)
\end{aligned}$$

Un argument similaire est utilisé lorsqu'il y a plus d'une donnée manquante. Par exemple, si les données X_{t-k} et X_{t-h}, $h < k$, $h \leqslant l$, $k - h \leqslant l$, manquent,

les termes $q_{(h)i_h i_0}$ et $q_{(k)i_k i_0}$ doivent être remplacés dans le modèle, conformément à l'équation (11), par :

$$q_{(h)i_h i_0} = \varphi_1 q_{(h+1)i_{h+1} i_0} + \ldots + \varphi_{k-h} q_{(k)i_k i_0} + \ldots + \varphi_l q_{(h+l)i_{h+l} i_0}$$

$$= \sum_{r=1}^{l} \varphi_r q_{(h+r)i_{h+r} i_0} \tag{13}$$

$$q_{(k)i_k i_0} = \sum_{r=1}^{l} \varphi_r q_{(k+r)i_{k+r} i_0} \tag{14}$$

De plus, le terme $q_{(k)i_k i_0}$ apparaissant dans l'équation (13) doit être à son tour remplacé par (14).

De façon générale, une donnée manquante X_{t-z} est remplacée par :

$$q_{(z)i_z i_0} = \sum_{r=1}^{l} \varphi_r q_{(z)i_{z+r} i_0} \tag{15}$$

et pour l'ensemble des valeurs par :

$$\chi'_{t-z} Q_z = \sum_{r=1}^{l} \varphi_r \chi'_{t-z-r} Q_{z+r} \tag{16}$$

Il peut arriver que certaines matrices Q_{z+r} soient impossibles à estimer à l'aide des données à disposition. Dans ce cas, Q_{z+r} peut être approximée par le produit matriciel $Q_r Q_z$.

Dans le cas du modèle (6), il suffit de remplacer chaque matrice d'ordre $(1, g)$ par la matrice Q^g correspondante pour que les résultats énoncés ci-dessus soient valables.

5. Illustration numérique

Dans ce paragraphe, nous nous proposons de montrer par le biais d'une simulation que le modèle utilisant les matrices de transition d'ordre $(1, g)$ pour le retard d'ordre g est le plus performant parmi les trois proposés. A cet effet, nous avons généré des séries de données à partir d'une matrice de transition d'ordre 2, puis nous avons comparé cette matrice avec les matrices d'ordre 2 construites à partir des modèles autorégressifs.

5.1 Les données

Nous avons repris ici les données sur la mobilité professionnelle des physiciens américains, déjà utilisées par Logan (1981). De 1960 à 1964, des physiciens

américains ont été observés du point de vue de l'évolution, d'année en année, de leur activité principale. Celle-ci peut être de trois types : management (M), recherche (R) et enseignement (E). Nous donnons ci-dessous la matrice d'ordre 2 de ces données :

$$Q2 = \begin{array}{cc|ccccccccc} & & t & M & M & M & R & R & R & E & E & E \\ & & t-1 & M & R & E & M & R & E & M & R & E \\ t-2 & t-1 & & & & & & & & & & \\ \hline M & M & & 0.90 & 0 & 0 & 0.07 & 0 & 0 & 0.03 & 0 & 0 \\ R & M & & 0.70 & 0 & 0 & 0.26 & 0 & 0 & 0.04 & 0 & 0 \\ E & M & & 0.63 & 0 & 0 & 0.07 & 0 & 0 & 0.30 & 0 & 0 \\ M & R & & 0 & 0.35 & 0 & 0 & 0.60 & 0 & 0 & 0.05 & 0 \\ R & R & & 0 & 0.11 & 0 & 0 & 0.83 & 0 & 0 & 0.06 & 0 \\ E & R & & 0 & 0.07 & 0 & 0 & 0.59 & 0 & 0 & 0.34 & 0 \\ M & E & & 0 & 0 & 0.27 & 0 & 0 & 0.09 & 0 & 0 & 0.64 \\ R & E & & 0 & 0 & 0.05 & 0 & 0 & 0.31 & 0 & 0 & 0.64 \\ E & E & & 0 & 0 & 0.08 & 0 & 0 & 0.08 & 0 & 0 & 0.84 \end{array}$$

A partir de cette matrice, nous avons généré 15 séries de données : 5 séries de taille 100 (notées d100a, ..., d100e), 5 séries de taille 1000 (notées d1000a, ..., d1000e) et 5 séries de taille 10000 (notées d10000a, ..., d10000e). Pour chacune de ces 15 séries, nous avons estimé à l'ordre 2 les trois modèles autorégressifs envisagés dans cet article, les matrices de transition étant estimées par la méthode du maximum de vraisemblance. Nous avons ensuite construit à partir de chaque modèle la matrice de transition d'ordre $l = 2$, notée $ARQ2$, ayant pour éléments (ici, dans le cas du modèle (7)) :

$$P(X_t = i_0, X_{t-1} = i_1, \ldots, X_{t-l+1} = i_{l-1} | X_{t-1} = j_1, \ldots, X_{t-l} = j_l) = \begin{cases} \sum_{g=1}^{2} \varphi_g q_{(g)j_g i_0} & , \text{ si } i_1 = j_1 \\ 0 & , \text{ sinon} \end{cases}$$

Finalement, nous avons comparé ces matrices à la matrice ayant servi à générer les données, à l'aide de l'indice de qualité de la modélisation décrit plus loin. Nous avons également comparé les matrices de transition d'ordre 2 (notées $\hat{Q}2$) calculées directement sur chacune des 15 séries de données avec la matrice ayant servi à les générer.

5.2 Indice de qualité de la modélisation (I_{qm})

Pour avoir une idée de la qualité (c'est-à-dire de la ressemblance) d'une matrice par rapport à une autre, nous avons construit un indice représentant l'écart moyen entre les probabilités des deux matrices. Formellement, pour un modèle d'ordre l sur m valeurs, cet indice s'écrit :

$$I_{qm}\left(\frac{ARQl}{Ql}\right) = \frac{1}{m^{(l+1)}} \sum_{i=1}^{m^l} \sum_{j=1}^{m^l} |Ql(i,j) - ARQl(i,j)| \qquad (17)$$

où $Ql(i,j)$ est la case d'indices (i,j) de la vraie matrice de transition d'ordre l, $ARQl(i,j)$ est la case d'indices (i,j) de la modélisation de la matrice d'ordre l,

| | représente la valeur absolue, et où $m^{(l+1)}$ est le nombre de probabilités pouvant être non-nulles dans les matrices Ql et $ARQl$. Parmi un ensemble de modèles à disposition, le meilleur sera celui dont l'indice de qualité de la modélisation est le plus petit.

5.3 Les calculs

Tous les calculs ont été effectués sur des ordinateurs DEC 3000-800S et PC486. Les estimations des paramètres des modèles autorégressifs ont été réalisées à l'aide du logiciel TSP. Les autres calculs ont été effectués sur le logiciel MATLAB. En particulier, les matrices de transition et les indices de qualité de la modélisation ont été calculés à l'aide d'un programme en langage MATLAB, disponible sur demande auprès de l'auteur.

5.4 Les résultats

Les trois tableaux ci-après donnent pour les séries de 100, 1000 et 10000 données les différents résultats obtenus, soit : les paramètres autorégressifs des modèles d'ordre 2 (φ_1, φ_2), l'indice de qualité de la modélisation de $ARQ2$ par rapport à la matrice ayant servi à générer les données ($I_{qm}(\frac{ARQ2}{Q2})$), et l'indice de qualité de la modélisation de la matrice $\hat{Q}2$ de chacune des séries de données par rapport à la matrice les ayant générées ($I_{qm}(\frac{\hat{Q}2}{Q2})$).

Remarque : le modèle (2) est noté ici «R», le modèle (6) «RP» et le modèle (7) «B».

5.4.1 Matrices de transition et sous-identification

Le faible nombre de données des séries de taille 100 (et dans une moindre mesure de celles de taille 1000) influence fortement la qualité des modélisations issues de ces séries. Les matrices d'ordres (1,1) ou (1,2) sont sous-identifiées (ie : certaines transition n'apparaissant jamais ou pratiquement jamais dans les données utilisées, les probabilités de transition correspondantes ne peuvent pas être calculées) et diffèrent beaucoup d'une série à l'autre, ce qui est reflété par les indices de qualité de la modélisation et par les paramètres φ_1 et φ_2. En revanche, pour les séries de 10000 données, les résultats sont assez homogènes, conséquence d'une bonne estimation de toutes les matrices de transition.

Afin de justifier l'utilisation des modèles «B», nous avons vérifié empiriquement l'égalité des distributions limites des matrices Q_1 et Q_2 de chaque série. La plus grande différence trouvée entre deux probabilités est égale à 0.03 pour les séries de 100 données, à 0.007 pour les séries de 1000 données et à 0.0027 pour les séries de 10000 données, ce qui nous semble suffisamment petit pour rendre l'utilisation de «B» raisonnable.

TABLEAU 1
Séries de 100 données

Données	Modèles	φ_1	φ_2	$I_{qm}\left(\frac{ARQ2}{Q2}\right)$	$I_{qm}\left(\frac{\hat{Q}2}{Q2}\right)$
d100a	R	0.27671	0.72329	0.1411	0.1407
	RP	0.54988	0.45012	0.2022	
	B	0.23191	0.76809	0.1083	
d100b	R	0.20820	0.79180	0.2557	0.2207
	RP	0.54669	0.45331	0.2952	
	B	0.16154	0.83846	0.2437	
d100c	R	0.05759	0.94241	0.1255	0.0988
	RP	0.43350	0.56650	0.2126	
	B	0.15109	0.84891	0.0781	
d100d	R	0.07272	0.92728	0.1874	0.1342
	RP	0.53825	0.46175	0.2464	
	B	0.12773	0.87227	0.0982	
d100e	R	0.33932	0.66068	0.2138	0.2142
	RP	0.55587	0.44413	0.2577	
	B	0.30729	0.69271	0.2050	

TABLEAU 2
Séries de 1000 données

Données	Modèles	φ_1	φ_2	$I_{qm}\left(\frac{ARQ2}{Q2}\right)$	$I_{qm}\left(\frac{\hat{Q}2}{Q2}\right)$
d1000a	R	0.22546	0.77454	0.0883	0.0343
	RP	0.48134	0.51866	0.1855	
	B	0.24191	0.75809	0.0576	
d1000b	R	0.22708	0.77292	0.0738	0.0268
	RP	0.47991	0.52009	0.1768	
	B	0.23564	0.76436	0.0466	
d1000c	R	0.10125	0.89875	0.1382	0.0461
	RP	0.43710	0.56290	0.2098	
	B	0.17235	0.82765	0.0783	
d1000d	R	0.11358	0.88642	0.1044	0.0322
	RP	0.43564	0.56436	0.1937	
	B	0.17789	0.82211	0.0675	
d1000e	R	0.24358	0.75642	0.0782	0.0333
	RP	0.49796	0.50204	0.1823	
	B	0.25181	0.74819	0.0496	

TABLEAU 3
Séries de 10000 données

Données	Modèles	φ_1	φ_2	$I_{qm}\left(\frac{ARQ2}{Q2}\right)$	$I_{qm}\left(\frac{\hat{Q}2}{Q2}\right)$
d10000a	R	0.19595	0.80405	0.0890	0.0099
	RP	0.47290	0.52710	0.1885	
	B	0.22372	0.77628	0.0559	
d10000b	R	0.22541	0.77459	0.0802	0.0106
	RP	0.48369	0.51631	0.1817	
	B	0.24022	0.75978	0.0496	
d10000c	R	0.21810	0.78190	0.0808	0.0128
	RP	0.47729	0.52271	0.1825	
	B	0.23511	0.76489	0.0493	
d10000d	R	0.20174	0.79826	0.0882	0.0142
	RP	0.47452	0.52548	0.1875	
	B	0.22563	0.77437	0.0573	
d10000e	R	0.19426	0.80574	0.0938	0.0148
	RP	0.46790	0.53210	0.1895	
	B	0.22212	0.77788	0.0576	

5.4.2 Les paramètres autorégressifs

Tous les paramètres φ_1 et φ_2 étant positifs, il n'a pas été nécessaire de vérifier la non-négativité des résultats fournis par les modèles.

Dans le cas des séries de tailles 100 et 1000, les estimations de φ_1 et φ_2 varient considérablement pour un même modèle d'une série à l'autre, mais, comme relevé précédemment, c'est une conséquence directe de la sous-identification des matrices de transition de ces modèles.

De façon générale, les modèles de type «RP» tendent à attribuer une importance égale aux périodes $t-1$ et $t-2$, φ_1 étant en moyenne légèrement supérieur à 0.5 et φ_2 légèrement inférieur. Au contraire, les modèles de types «R» et «B» attribuent une très grande importance à la période $t-2$, au détriment de $t-1$ (environ 0.2 pour φ_1 contre 0.8 à φ_2). Il est à remarquer que, excepté dans le cas des séries de 100 données manifestement sous-identifiées, le φ_1 du modèle «B» est systématiquement supérieur de quelques centièmes à celui de «R», et inversement pour φ_2.

5.4.3 Les indices de qualité de la modélisation

Nous pouvons tout d'abord remarquer que, pour chacune des séries de données, le plus mauvais résultat parmi les indices $I_{qm}\left(\frac{ARQ2}{Q2}\right)$ est systématiquement celui

du modèle «RP», ce qui n'est pas contraire à la théorie puisque les données utilisées ne sont pas homogènes d'ordre 1. Nous pouvons aussi constater que, pour chacune des 15 séries de données, le meilleur indice est systématiquement celui du modèle «B», ce qui confirme pleinement la théorie.

En comparant les indices des modèles «B» $\left(I_{qm}\left(\frac{ARQ2}{Q2}\right)\right)$ avec ceux des matrices d'ordre 2 calculées directement sur chacune des séries $\left(I_{qm}\left(\frac{\hat{Q}2}{Q2}\right)\right)$, nous constatons ceci :

- Dans le cas des séries de 100 données, les résultats des modèles «B» sont tout-à-fait comparables, et même meilleurs dans trois cas sur cinq, à ceux des matrices d'ordre 2.
- Dans le cas des séries de 1000 données, les modèles «B» présentent des résultats systématiquement moins bons à ceux des matrices d'ordre 2, mais dans un rapport moyen assez faible de 1.74 (ie : l'écart moyen de «B» est en moyenne 1.74 fois plus grand que celui de $\hat{Q}2$).
- Dans le cas des séries de 10000 données, les modèles «B» présentent des résultats systématiquement moins bons à ceux des matrices d'ordre 2, dans un rapport moyen de 4.42.

5.4.4 Conclusion de l'illustration numérique

Conformément à la théorie développée dans les chapitres précédents, notre illustration montre que, dans le cas de données hétérogènes, la modélisation autorégressive utilisant la matrice d'ordre $(1, g)$ pour le retard d'ordre g est la meilleure parmi les trois envisagées. En particulier, lorsque le nombre de données à disposition est restreint, une telle modélisation peut remplacer avantageusement une matrice d'ordre 2 traditionnelle.

6. Conclusion

Les deux nouvelles modélisations autorégressives présentées dans cet article (utilisation de la matrice Q^g pour le retard d'ordre g dans le cas de données homogènes d'ordre 1, et surtout, utilisation de la matrice Q_g pour le retard d'ordre g dans le cas de données hétérogènes), permettent une nette amélioration des résultats par rapport à ceux fournis par le modèle de Raftery, tout en n'étant pas significativement plus compliquées à estimer. Ces modèles sont une alternative valable aux chaînes de Markov traditionnelles d'ordre supérieur à 1, particulièrement lorsque le nombre de données à disposition est restreint et conduit à une sous-identification des matrices de transition.

Bibliographie

Bellman, R. (1960) *Introduction to Matrix Analysis*. Ed. McGraw-Hill.

Berchtold, A. (1994) *Modélisation autorégressive des chaînes de Markov d'ordre l*. Cahiers du Département d'Econométrie, Université de Genève.

Cox, D.R., Miller, H.R. (1965) *The Theory of Stochastic Processes*. Ed. Methuen & Co Limited.

Doob, J.L. (1990) *Stochastic Processes*. Ed. John Wiley & Sons.

Feller, W. (3e édition) *An Introduction to Probability Theory and its Applications, Vol. 1*. Ed. John Wiley & Sons.

Gourieroux, C., Monfort, A. (1990) *Séries temporelles et modèles dynamiques*. Ed. Economica.

Grimmett, G.R., Stirzaker, D.R. (1992) *Probability and Random Processes*. Ed. Clarendon Press.

Jacobs, P.A., Lewis, P.A.W. (1978) *Discrete Time Series Generated by Mixtures I : Correlational and Runs Properties*. Journal of the Royal Statistical Society B, Vol. 40, 94-105.

Jacobs, P.A., Lewis, P.A.W. (1978) *Discrete Time Series Generated by Mixtures II : Asymptotic properties*. Journal of the Royal Statistical Society B, Vol. 40, 222-228.

Jacobs, P.A., Lewis, P.A.W. (1978) *Discrete Time Series Generated by Mixtures III : Autoregressive Processes*. Naval Postgraduate School Technical Report NPS 55-78-022.

Jacobs, P.A., Lewis, P.A.W. (1983) *Stationary Discrete Autoregressive-Moving Average Time Series Generated by Mixtures*. Journal of Time Series Analysis, Vol. 4, 18-36.

Katz, R.W. (1981) *On Some Criteria for Estimating the Order of a Markov Chain*. Technometrics, Vol. 23, No 3, 243-249.

Kemeny, J.G., Snell, J.L. (1976) *Finite Markov Chains*. Ed. Springer-Verlag.

Kendall, M., Stuart, A., Ord, J.K. (1983) *The Advanced Theory of Statistics, Vol. 3*. Ed. Charles Griffin & Company Limited.

Logan, J.A (1981) *A Structural Model of the Higher-Order Markov Process Incorporating Reversion Effects*. The Journal of Mathematical Sociology, Vol. 8, 75-89.

Mehran, F. (1989) *Analysis of Discrete Longitudinal Data : Infinite-Lag Markov Models*. Statistical Data Analysis and Inference 533-541. Ed. Dodge.

PC-Matlab (1987) *User's Guide*. Ed. MathWorks Inc.

Pegram, G.G.S. (1975) *A Multinomial Model for Transitions Probability Matrices*. Journal of Applied Probability, Vol. 12, 498-506.

Pegram, G.G.S. (1980) *An Autoregressive Model for Multilag Markov Chains*. Journal of Applied Probability, Vol. 17, 350-362.

Raftery, A.E. (1985) *A Model for High-Order Markov Chains*. Journal of the Royal Statistical Society B, Vol. 47, No 3, 528-539.

Raftery, A.E., Tavaré, S. (1994) *Estimation and Modelling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model*. Applied Statistics, Vol. 43, No 1, 179-199.

Seneta, E. (1973) *Non-Negatives Matrices*. Ed. George Allen & Unwin Ltd.

Stirzaker, D.R. (1994) *Elementary Probability*. Ed. Cambridge University Press.

Tricot, C. *Notes de statistique VI : Ergodicité*. Université de Genève.

TSP (1988) *Reference Manual, version 4.1*. Ed. TSP International.

TSP (1988) *User's Guide, version 4.1*. Ed. TSP International.