UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1987:2

ISSN 0349-8034

ESTIMATION OF TRANSITION MATRICES IN MULTIDIMENSIONAL MARKOVIAN CHAINS AND DYNAMIC SYSTEMS

by

Sture Holm,

Irene Isaksson and

Vladimir Khokhlov

Statistiska institutionen

Göteborgs Universitet
Viktoriagatan 13
S 411 25 Göteborg
Sweden

CONTENTS

1 Introduction

In this paper we will study estimation of the transition matrix A in a m-dimensional random sequence $\{X(n), n=0,1,...\}$ generated by

$$X(n+1) = AX(n) + Z(n+1) \quad n=0,1,2,...$$

where
$Z(n)$ $n=1,2,...$ are i.i.d. m-dimensional random variables. A is an (unknown) m×m matrix and X(0) is an m-dimensional random variable independent of $\{Z(n) : n=1,2,...\}$.

If $||A|| < 1$ the process $\{X(n) : n=0,1,2,...\}$ is weakly stationary for suitable starting variable X(0). This process is irregular for some A:s but has quite a periodic character for other A:s. Processes with such A:s might be good models e.g. for interacting predator-pray-populations in biology (Isaksson (1987)).

If $||A|| > 1$ the process $\{X(n) : n=0,1,2,...\}$ is an "exploding" nonstationary dynamic system. This type of model might have applications e.g. in biology and economy.

Both in the stationary and nonstationary case, the character of the process depends strongly on A, and accurate estimates are needed.

The simplest estimates of A are based on observations of consecutive transitions. If the components of Z(n) are independent with expection 0 and X(n) is observed for n=0,1,..., **N**

a natural simple estimate is obtained by minimizing

$$\sum_{n=1}^{N} (X(n) - AX(n-1))^T (X(n) - AX(n-1))$$

In the stationary case this is essentially the Yule-Walker estimate determined by the equation

$$\hat{R}(a) = A\hat{R}(0)$$

where $\hat{R}(m)$ is the estimate of the covariance function $R(m)$ of $\{X(n) : n=0,1,...\}$. The only difference is the drop of some terms at the ends, which are usually included in the estimates in $\hat{R}(m)$.

In a simplified onedimensional situation we demonstrate that it might be better to base the estimate on transition in several steps. This however means that we have to determine the matrix A from A^m for some $m \geq 2$. Generally A^m (and A) are not symmetric. We propose an iteration method and we show that it converges. The solution of $A^m=B$ (for some $m \geq 2$) is not unique and the result of the iteration depends on the starting point. For reasonably small m the difficulty of non-uniqueness can however quite easily be overcome by using side information on the approximate values of the elements in A.

2 A simple motivation

Suppose that $\{X(n) : n=0,\pm1,\ldots\}$ is an autoregressive stationary Gaussian random process with expectation 0. In the model formulation

$$X(n+1) = aX(n) + Z(n+1)$$

the innovations $Z(n+1)$ are independent Gaussian with expectation 0 and some variance σ^2.

Suppose further that the parameters a and σ^2 of the process are to be estimated by a reasonably long series of observation $\{X(n) :=1,2,\ldots,N\}$.

A very simple estimate of a is obtained by the Yule-Walker equation

$$\hat{a} = \hat{R}(1) \;/\; \hat{R}(0)$$

where $\hat{R}(m)$ is the (empirical covariance) estimate of the covariance function $R(m) = \mathrm{Cov}(X(n), X(n+m))$.

Since the covariance function equals

$$R(m) = \sigma^2 \cdot a^{|m|}$$

it is also possible to estimate a by

$$\hat{a}_m = (\hat{R}(m) \;/\; \hat{R}(0))^{1/m}$$

or

$$\hat{a}_m = -(\hat{R}(m) \;/\; \hat{R}(0))^{1/m}$$

if we can determine the sign of $\hat{a}$ on other grounds. Suppose in the following discussion that we know that $a > 0$ and thus consider only $\hat{a} > 0$.

What are the variances of the different estimates $\hat{a}_m$?

From the well-known formulas of the asymptotic variances and covariances of the estimates $\hat{R}(m)$ it easily follows that

$$\text{Var } \frac{\hat{R}(m)}{\hat{R}(0)} \simeq \frac{R(m)}{N \cdot R(0)} (R(0)-R(m))$$

This means that

$$\text{Var } \hat{a}_m \simeq \frac{R(0)-R(m)}{N \cdot m^2 \cdot (R(m)/R(0))^{1-2/m}}$$

which has a minimum when $R(m)/R(0) \approx 0.2$ (or $a^m \approx 0.2$). The minimum is not very distinct, but the variance can be considerably larger than the minimal value for m far away from its optimal value.

If e.g. $a=0.9$ the variance for $m=1$ is 4-5 times the minimal variance obtained for $m=7$.

Even if this is a simplified one dimensional example, it indicates that it might be a good idea to estimate the matrix A in the model $X(n+1) = AX(n) + Z(n+1)$ $n=0,1,\ldots$ by using an intermediate estimate of A^m for an m-step transformation of the model.

3 Estimation of transition matrix for one step and for several steps

Let us consider first the estimation of the transition matrix A in the model

$$X(n+1) = AX(n) + Z(n+1)$$

when we have N+1 observations $\{X(n) : n=0,1,\ldots,N\}$. Since the $Z(n)$: s are independent a natural estimate is obtained by minimizing the "sum of squares"

$$\sum_{n=0}^{N-1} (X(n+1) - AX(n))^T (X(n+1) - AX(n))$$

The minimum is easily seen to be obtained when

$$A \cdot \left(\sum_{n=0}^{N-1} X(n)X^T(n)\right) = \sum_{n=0}^{N-1} X(n+1)X^T(n)$$

In the stationary case this gives only a small modification of the Yule-Walker equation. But the equation works in the non-stationary case too.

If the model

$$X(n+1) = AX(n) + Z(n+1)$$

is rewritten into a model for m steps we get

$$X(n+m) = A^m X(n) + \sum_{k=0}^{m-1} A^k Z(n+m-k)$$

Although the random vectors

$$\sum_{k=0}^{m-1} A^k Z(n+m-k) \qquad n=m,m+1,\ldots$$

are not independent, a reasonable estimate of A^m is obtained by minimizing

$$\sum_{n=0}^{N-m} (X(n+m) - A^m X(n))^T (X(n+m) - A^m X(n))$$

Like in the case m=1 the minimum is obtained when

$$A^m \left(\sum_{n=0}^{N-m} X(n) X^T(n) \right) = \sum_{n=0}^{n-m} X(n+m) X^T(n)$$

Thus an estimate of A^m is easily obtained, the only problem remaining is to get A from A^m. This will be treated in the next section.

4 Existence and uniqueness

Generally there is no unique solution to the equation

$$A^m = B$$

for given B. For some B:s there is no real solution and for some B:s there are several real solutions. If det $B \neq 0$ we must have det $A \neq 0$ for a possible solution. Any matrix A with det $A \neq 0$ can be written in the form

$$A = G^{-1}DG$$

where G is a matrix with det $G \neq 0$ and D is diagonal with the diagonal elements equal to the eigenvalues of A. The columns (rows) of G are determined by the eigenvectors of A. Further A and $B=A^m$ have the same eigenvectors and if the eigenvalues of A are λ_k, then the eigenvalues of $B=A^m$ are λ_k^m. The eigenvalues of A and B may be complex, but for real matrices they must be real or appear in complex conjugated pairs.

Suppose now for instance that B is a 2x2 matrix with two different eigenvalues λ_1^* and λ_2^*, at least one negative. Then the equation

$$A^2 = B$$

has no real solution, since there does not exist real or complex conjugated λ_1 and λ_2 such that $\lambda_1^2 = \lambda_1^*$ and $\lambda_2^2 = \lambda_2^*$. On the other hand if λ_1^* and λ_2^* are complex conjugated there exist two pairs λ_1, λ_2 of complex conjugated numbers such that

$$\lambda_1^2 = \lambda_1^* \quad \text{and} \quad \lambda_2^2 = \lambda_2^* ,$$

and corresponding real matrices

$$A = G^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} G$$

where G is determined by the eigenvectors of B.

This simple example illustrates the principle problem in this context. It is to be observed that lack of solution does not occur in isolated points only but in open sets. In the simple case above the condition for getting at least one negative eigenvalue, of two noncoinciding, is that

$$b_{11} + b_{22} < 0$$

and

$$b_{11}^2 + b_{22}^2 - 4b_{12}\, b_{21} > 0$$

In estimation problems we must take care of the possibility that A^m is estimated in a point where we have no solution. This means that we happen to get an estimate outside the possibility parameter region, which is a problem appearing also in other branches of statistics.

5 Numerical methods in estimation procedure

In this section we will consider iterative methods for solving the equation

$$A^m = B$$

appearing in the estimation problem. We will use a variant of the Newton-Kantorowich method.

Consider an operator P(x). A matrix A which makes P(A) a zero operation can be found by an iteration procedure

$$x_{k+1} = x_k - (P'(x_k))^{-1} P(x_k)$$

The convergence of this type of iteration procedure is treated in Kantorowich and Akilov (1982), p. 532. The general result is the following.

Kantorowich theorem. Suppose we have an operator P defined in an open subset Ω of a Banach space and the operator is mapping the subset into another Banach space and has a continuous second derivative in Ω_0, $\Omega_0 \subset \Omega$. Suppose, in addition, that

1) $\Gamma_0 = |P'(\kappa_0)|^{-1}$ exists

2) $\|\Gamma_0 P(\kappa_0)\| \leq \eta$

3) $\|\Gamma_0 P''(\kappa)\| \leq K \quad (\kappa \varepsilon \Omega_0)$

Then, provided

$$h = K\eta \leq \frac{1}{2}$$

and

$$\tau \geq \tau_0 = \frac{1-\sqrt{1-2h}}{h}\,\eta$$

equation $P(\kappa)$ has a solution κ_*, and Newton's process converges to this solution. Furthermore

$$\|\kappa_* - \kappa_0\| \leq \tau_0$$

Also, if for $h < \frac{1}{2}$ we have

$$\tau < \tau_0 = \frac{1+\sqrt{1-2h}}{h}\,\eta$$

while for $h = \frac{1}{2}$ we have

$$\tau \leq \tau_0$$

then κ_* is the unique solution in Ω_0.

The rate of convergence is given by

$$\|\kappa_* - \kappa_n\| \leq \frac{1}{2^n}\,(2h)^{2^n}\,\frac{\eta}{h}\,, \quad (n=0,1,\ldots)$$

In our case the operator $P(A)$ equals

$$P(A) = \text{vec}\,\{A^m - B\}$$

where vec is an operator which transforms a matrix in a column vector.

The first derivative of $P(A)$ is $P'(A) = I\otimes A^{m-1} + A\otimes A^{m-2} + \ldots + A^{m-1}\otimes I$ but in, our procedure we will use an approximation

$$P'(A) \approx m\,A^{m-1}\otimes I$$

using only one of the Cartesian products.

Hence our iteration equation is

$$\text{vec}\{A_{k+1}\} = \text{vec}\{A_k\} -$$

$$- |n' A_k^{m-1} \otimes I|^{-1} \text{vec}\{A_k^m - B\} \qquad k=0,1,...$$

which by using some rules of matrix algebra (McDonald Swaminathan (1973)) is the same as the iteration equation

$$A_{k+1} = A_k - \frac{1}{m} A_k^{1-m} (A_k^m - B) \qquad k=0,1,...$$

or

$$A_{k+1} = \frac{m-1}{m} A_k + \frac{1}{m} A_k^{1-m} B \qquad k=0,1,...$$

As we have pointed out in section 4 the matrix B and a solution A to the equation $A^m=B$, have the same eigenvectors. The following lemma shows that with a special start of the iteration, the same holds for the iteration approximations.

Lemma. Suppose that the iteration

$$A_{k+1} = \frac{m-1}{m} A_k + \frac{1}{m} A_k^{1-m} B \qquad k=0,1,...$$

is started with

$$A_0 = a_0 I$$

where I is the unit matrix. Then all A_k for k=1,2,... have the same eigenvectors as the matrix B.

If λ denotes the eigenvalue of B corresponding to a certain eigenvector and λ_k denotes the eigenvalue of A_k corresponding to the same eigenvector then these eigenvalues are determined by

$$\lambda_0 = a_0$$

$$\lambda_{k+1} = \frac{m-1}{m}\,\lambda_k + \frac{\lambda}{m\,\lambda_k^{m-1}} \qquad k=1,2,\ldots$$

Proof. Any vector g is an eigenvector of $A_0=I$ since $a_0\, I_g=a_0 g$. Further if g is an eigenvector of B with eigenvalue λ and of A_k with eigenvalue λ_λ then

$$A_{k+1}\, g = \frac{m-1}{m}\, A_k\, g + \frac{1}{m}\, A_k^{1-m}\, B\, g =$$

$$= \frac{m-1}{m}\, \lambda_k\, g + \frac{1}{m}\, A_k^{1-m}\, \lambda\, g =$$

$$= \frac{m-1}{m}\, \lambda_k\, g + \frac{1}{m}\, \lambda_k^{1-m}\, \lambda\, g$$

Thus the lemma follows by induction.

Let us now consider the convergence of the procedure. This can be obtained by the general Kantorowich theory. But it is also possible to get a convergence condition and convergence rate directly from the eigenvalue property in the lemma.

Let λ_* be a solution of

$$\lambda_*^m = \lambda$$

and consider

$$\lambda_{k+1} - \lambda_* = \frac{m-1}{m}\, \lambda_k + \frac{(\lambda_*)^m}{m\,\lambda_k^{m-1}} - \lambda^*$$

or

$$\frac{\lambda_{k+1}-\lambda_*}{\lambda_*} = \frac{m-1}{m}\, \frac{\lambda_k}{\lambda_*} + \frac{\lambda_*^{m-1}}{m\,\lambda_k^{m-1}} - 1$$

Straight-forward calculations now show that if

$$\left|\frac{\lambda_k - \lambda_*}{\lambda_*}\right| < \Delta = \frac{2}{m-1} \quad ,$$

then

$$\left|\frac{\lambda_{k+1} - \lambda_*}{\lambda_*}\right| < \left|\frac{\lambda_k - \lambda_*}{\lambda_*}\right|$$

Thus if the starting point λ_0 satisfies

$$\left|\frac{\lambda_0 - \lambda_*}{\lambda_*}\right| < \frac{2}{m-1} \quad ,$$

the procedure converges to λ_*. This also means that if the procedure starts with a matrix, whose eigenvalues are close enough to the eigenvalues of a solution matrix the procedure will give convergence to this matrix.

A reasonable starting point is

$$A_0 = (\text{Det } B)^{1/m\nu} \cdot I$$

if B is of type $\nu x \nu$. The reason is that the product of the eigenvalues of A_0 are then equal to $(\text{Det } B)^{1/m}$, which equals the product of the eigenvalues of the unknown solution A. Thus A_0 is of the "correct size".

REFERENCES

Kantorowich, L.V. and Akilov, G.P. (1982): Functional Analysis. Pegamon Press Inc., Oxford.

McDonald, R. and Swaminathan (1973): A simple matrix calculus with application to multivariate analysis. General Systems, 18 p. 37-54.