The XIII International Conference
"Applied Stochastic Models and Data Analysis"
(ASMDA-2009)
June 30-July 3, 2009, Vilnius, LITHUANIA

ISBN 978-9955-28-463-5
L. Sakalauskas, C. Skiadas and
E. K. Zavadskas (Eds.): ASMDA-2009
Selected papers. Vilnius, 2009, pp. 199–202

EXACT BAYESIAN PREDICTION IN NON-GAUSSIAN MARKOV-SWITCHING MODEL

Noémie Bardel[1], François Desbouvries[2]

Telecom Sudparis / CITI department & CNRS UMR 5157,
9 rue Charles Fourier, 91011 Evry, France
E-mail: Noemie.Bardel, Francois.Desbouvries@it-sudparis.eu

Abstract: In this paper we consider a class of recently introduced jump-Markov switching models, involving a hidden process *X*, an observed process *Y* and a latent process *R* which models the switches or changes of regimes in (*X*,*Y*). We address the Bayesian prediction problem, and we show that the *p*-step ahead a posteriori conditional expectation (and associated conditional covariance matrix) can be computed exactly linearly in time.

Keywords: Bayesian restoration, jump-Markov models, non-Gaussian models.

1. Introduction

Let $X_{1:N} = (X_1,...,X_N)$ be a hidden random sequence with values in R^q, $Y_{1:N} = (Y_1,...,Y_N)$ an observed random sequence with values in R^m, and $R_{1:N} = (R_1,...,R_N)$ a discrete random sequence with values in a finite set $S = \{1,...,s\}$, which usually models the random changes of regime, or switches of the distribution of (X_n,Y_n). The three chains are linked via some probability distribution $p(x_{1:N}, r_{1:N}, y_{1:N})$. Bayesian restoration consists in efficiently computing a posterior probability density function (pdf) of interest, namely $p(x_k \mid y_{1:n})$ for some value of *k* and *n*.

As is well known, the exact (recursive) computation of $p(x_k \mid y_{1:n})$ is not possible in many commonly used stochastic models and one needs to resort to approximations. Let us consider for instance the classical conditionally linear Gaussian model, also called jump-Markov state-space system, which consists in considering that R is a Markov chain and, roughly speaking, that conditionally on R, the couple (X,Y) is the classical Gaussian dynamic linear system. This is summarized in the following :

$$R \text{ is a Markov chain;} \qquad (1)$$

$$X_{n+1} = F_n(R_n)X_n + W_n \; ; \qquad (2)$$

$$Y_n = H_n(R_n)X_n + Z_n , \qquad (3)$$

where matrices $F_n(R_n)$ and $H_n(R_n)$ depend on R_n, W_1, …, W_N are Gaussian vectors in R^q, Z_1, …, Z_N are Gaussian vectors in R^m, and X_1, W_1, …, W_N, Z_1, …, Z_N are independent (see the oriented dependence graph in Figure 1, (a)). For fixed $R_1 = r_1$, …, $R_n = r_n$, … the computation of $E[X_n | y_{1:n}]$, say, is obtained by classical Kalman-like methods. However, it has been well known since Tugnait, 1982 that exact computation is no longer possible with random Markov R and different approximations must be used, including particle filtering methods, see e.g. Tugnait, 1982, Andrieu *et al.*, 2003, Ristic *et al.*, 2004, Cappé *et al.*, 2005, Costa *et al.*, 2005, Zoeter *et al.*, 2006, or Giordani *et al.*, 2007.

On the other hand, in most situations we are indeed more interested by some moment $E(g(x_k) \mid y_{1:n})$ than by pdf $p(x_k \mid y_{1:n})$ itself. In particular, the conditional expectation $E(x_k \mid y_{1:n})$ is of particular interest since it is the solution to the Bayesian estimation problem with quadratic loss.

The Bayesian prediction problem which we address in this paper consists in computing efficiently the conditional expectation $E[X_{n+p} | y_{1:n}]$ and associated conditional covariance matrix $Cov[X_{n+p} | y_{1:n}]$ in a particular class of stochastic dynamical models with Markov regime. More precisely, the contribution of this paper consists in showing that $E[X_{n+p}|y_{1:n}]$ and $Cov[X_{n+p}|y_{1:n}]$ can be computed exactly, with complexity linear in time, in a recent jump-Markov model proposed in Pieczynski, 2008.

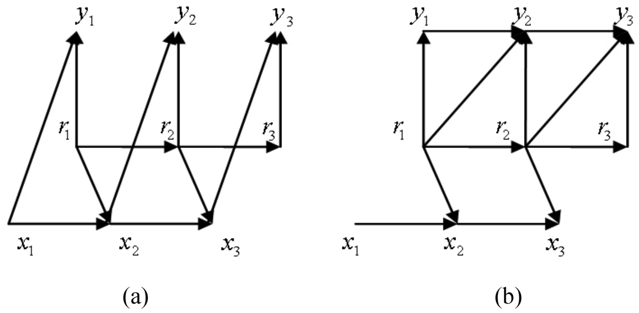

Fig. 1: Dependence oriented graphs of: (a) the classical Markov switching model; (b) the Markov switching model considered in this paper.

In this paper we thus consider the following Markov-switching model (see figure 1, (b)) :

$$R_n \text{ is a Markov Chain;} \quad (4)$$

$$(R_n, Y_n) \text{ is a Markov Chain;} \quad (5)$$

$$X_{n+1} = F_n(R_n)X_n + U_n, \quad (6)$$

where $\{U_n\}_{n\in\{1,...,N\}}$ are independent zero-mean random vectors, such that for each $n \in \{1,...,N\}$, U_n is independent from $(R_{1:N}, Y_{1:N})$. Note that in (6) (as compared to (2)) vectors U_n are not necessarily Gaussian; nevertheless, exact computation of the conditional posterior mean $E[X_{n+p}|y_{1:n}]$ will be feasible, as we shall see in section 2. From (4)-(6) we see that conditionally on $R_{1:n}$, $X_{1:n}$ and $Y_{1:n}$ are independent; but of course $X_{1:n}$ and $Y_{1:n}$ are actually dependent, and are linked through the Markov chain $R_{1:n}$.

2. Exact Bayesian prediction

Notation

For each integer *k* and for each $n \in \{1,...,N\}$, let us set :

$$M_{n+k}(r_{n+k}, y_{1:n}) = \int_{R^q} x_{n+k}\, p(x_{n+k}, r_{n+k} \mid y_{1:n}) dx_{n+k} \ . \quad (7)$$

If the covariance matrix Σ_n of U_n exists for all *n*, let us set :

$$V_{n+k}(r_{n+k}, y_{1:n}) = \int_{R^q} x_{n+k} x_{n+k}^T\, p(x_{n+k}, r_{n+k} \mid y_{1:n}) dx_{n+k} \ . \quad (8)$$

Of course, $E(X_{n+p} \mid Y_{1:n} = y_{1:n})$ and $Cov(x_{n+p} \mid y_{1:n})$ can be computed from $M_{n+p}(r_{n+p}, y_{1:n})$ and $V_{n+p}(r_{n+p}, y_{1:n})$ as:

$$E(X_{n+p} \mid Y_{1:n} = y_{1:n}) = \sum_{r_{n+p}} M_{n+p}(r_{n+p}, y_{1:n}) \text{ and}$$

$$Cov(x_{n+p} \mid y_{1:n}) = \sum_{r_{n+p}} V_{n+p}(r_{n+p}, y_{1:n}) - (\sum_{r_{n+p}} M_{n+p}(r_{n+p}, y_{1:n}))(\sum_{r_{n+p}} M_{n+p}(r_{n+p}, y_{1:n}))^T \ .$$

In the following we thus focus on the computation of $M_{n+p}(r_{n+p}, y_{1:n})$ and $V_{n+p}(r_{n+p}, y_{1:n})$.

Proposition

Let $(X_{1:N}, R_{1:N}, Y_{1:N})$ satisfy (4)-(6), with given transitions $p(r_{n+1}, y_{n+1} \mid r_n, y_n)$ and $p(r_{n+1} \mid r_n)$. Then $M_{n+p}(r_{n+p}, y_{1:n})$ can be recursively computed with linear complexity in time index by the following scheme:

- compute $M_n(r_n, y_{1:n})$ with the algorithm presented in Pieczynski, 2008;
- for each integer $p \geq 0$, compute

$$M_{n+p+1}(r_{n+p+1}, y_{1:n}) = \sum_{r_{n+p}} F_{n+p}(r_{n+p}) M_{n+p}(r_{n+p}, y_{1:n}) p(r_{n+p+1} \mid r_{n+p}) . \tag{9}$$

Furthermore, if the covariance matrix Σ_n of U_n exists for all *n*, it is possible to compute $V_{n+p}(r_{n+p}, y_{1:n})$ as follows:

- compute $V_n(r_n, y_{1:n})$ with the algorithm presented in Pieczynski, 2008;
- for each integer $p \geq 0$, compute:

$$\begin{aligned} V_{n+p+1}(r_{n+p+1}, y_{1:n}) = \sum_{r_{n+p}} & p(r_{n+p+1} \mid r_{n+p}) \\ & \times \left[F_{n+p}(r_{n+p}) V_{n+p}(r_{n+p}, y_{1:n}) F_{n+p}(r_{n+p})^T + \Sigma_{n+p} \right] \end{aligned} \tag{10}$$

Proof

We have:

$$\begin{aligned} p(x_{n+p+1}, r_{n+p+1} \mid y_{1:n}) &= \int_{R^q} \sum_{r_{n+p}} p(x_{n+p+1}, r_{n+p+1}, x_{n+p}, r_{n+p} \mid y_{1:n}) dx_{n+p} \\ &= \int_{R^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p} \mid y_{1:n}) p(x_{n+p+1}, r_{n+p+1} \mid x_{n+p}, r_{n+p}, y_{1:n}) dx_{n+p} . \end{aligned} \tag{11}$$

On the other hand, by the Bayes formula:

$$\begin{aligned} & p(x_{n+p+1}, r_{n+p+1} \mid x_{n+p}, r_{n+p}, y_{1:n}) \\ & = p(x_{n+p+1} \mid x_{n+p}, r_{n+p}, r_{n+p+1}, y_{1:n}) p(r_{n+p+1} \mid x_{n+p}, r_{n+p}, y_{1:n}) . \end{aligned}$$

Then, from (4) and (6), $p(x_{n+p+1} \mid x_{n+p}, r_{n+p}, r_{n+p+1}, y_{1:n})$ reduces to $p(x_{n+p+1} \mid x_{n+p}, r_{n+p})$ and $p(r_{n+p+1} \mid x_{n+p}, r_{n+p}, y_{1:n})$ reduces to $p(r_{n+p+1} \mid r_{n+p})$.

We next multiply (11) by x_{n+p+1} and integrate with respect to x_{n+p+1} to get:

$$\begin{aligned} & M_{n+p+1}(r_{n+p+1}, y_{1:n}) \\ & = \int_{R^q} x_{n+p+1} p(x_{n+p+1}, r_{n+p+1} \mid y_{1:n}) dx_{n+p+1} \\ & = \int_{R^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p} \mid y_{1:n}) \\ & \times \left[\int_{R^q} x_{n+p+1} p(x_{n+p+1} \mid x_{n+p}, r_{n+p}) dx_{n+p+1} \right] p(r_{n+p+1} \mid r_{n+p}) dx_{n+p} \end{aligned}$$

Since the $\{U_n\}_{n \in \{1,\dots,N\}}$ are independent, zero-mean and independent from $(R_{1:N}, Y_{1:N})$, we have :

$$\int_{R^q} x_{n+p+1} p(x_{n+p+1} \mid x_{n+p}, r_{n+p}) dx_{n+p+1} = F_{n+p}(r_{n+p}) x_{n+p} .$$

Finally:

$$
\begin{aligned}
& M_{n+p+1}(r_{n+p+1}, y_{1:n}) \\
& = \int_{R^q} \sum_{r_{n+p}} p(x_{n+p}, r_{n+p} \mid y_{1:n}) F_{n+p}(r_{n+p}) x_{n+p} p(r_{n+p+1} \mid r_{n+p}) dx_{n+p} \\
& = \sum_{r_{n+p}} F_{n+p}(r_{n+p}) \left[\int_{R^q} x_{n+p} p(x_{n+p}, r_{n+p} | y_{1:n}) dx_{n+p} \right] p(r_{n+p+1} \mid r_{n+p}) \\
& = \sum_{r_{n+p}} F_{n+p}(r_{n+p}) M_{n+p}(r_{n+p}, y_{1:n}) p(r_{n+p+1} \mid r_{n+p})
\end{aligned}
$$

which completes the proof of (9).

(10) is obtained similarly, by multiplying (11) by $x_{n+p+1} x_{n+p+1}{}^T$ and integrating with respect to x_{n+p+1}.

References

Tugnait, J. K. 1982. Adaptive estimation and identification for discrete systems with Markov jump parameters, *IEEE Transactions on Automatic Control* 25, 1054–1065.

Andrieu, C., Davy, M., and Doucet, A. 2003. Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, *IEEE Trans. on Signal Processing*, 51(7), 1762–1770.

Ristic, B., Arulampalam, S., and Gordon, N. 2004. *Beyond the Kalman Filter – Particle filters for tracking applications*, Artech House, Boston, MA.

Cappé, O., Moulines, E., and Ryden, T. 2005. *Inference in hidden Markov models*, Springer.

Costa O. L. V., Fragoso M. D., and Marques, R. P. 2005. *Discrete time Markov jump linear systems*, New York, Springer-Verlag.

Zoeter, O., and Heskes, T. 2006. Deterministic approximate inference techniques for conditionally Gaussian state space models, *Statistical Computation*, 16, 279–292.

Giordani, P., Kohn, R., and van Dijk, D. 2007. A unified approach to nonlinearity, structural change, and outliers, *Journal of Econometrics*, 137, 112–133.

Pieczynski, W. 2008. Exact calculation of optimal filter in semi-Markov switching model, *Fourth World Conference of the International Association for Statistical Computing (IASC 2008)*, December 5-8, Yokohama, Japan.