EXACT BAYESIAN SMOOTHING IN TRIPLET SWITCHING MARKOV CHAINS

Wojciech Pieczynski and François Desbouvries

Telecom SudParis / CITI department and CNRS UMR 5157
9, rue Charles Fourier, 91011 Evry, France
(e-mail: `wojciech.pieczynski,francois.desbouvries@it-sudparis.eu`)

ABSTRACT. Bayesian smoothing in conditionally linear Gaussian models, also called jump-Markov state-space systems, is an NP-hard problem. As a result, a number of approximate methods - either deterministic or Monte Carlo based- have been developed. In this paper we address the Bayesian smoothing problem in another triplet Markov chain model, in which the switching process R is not necessarily Markovian and the additive noises do not need to be Gaussian. We show that in this model the smoothing posterior mean and covariance matrix can be computed exactly with complexity linear in time.

1 INTRODUCTION

Let $X_{1:N} = (X_1, \cdots, X_N)$ be a hidden random sequence with values in $\mathbb{R}^q$, $Y_{1:N} = (Y_1, \cdots, Y_N)$ an observed random sequence with values in $\mathbb{R}^m$. One often says that "$X_{1:N}$ is only observed through $Y_{1:N}$" or that "$Y_{1:N}$ is a noisy version of $X_{1:N}$. According to the latter viewpoint, the distribution $p(y_{1:N}|x_{1:N})$ of $y_{1:N}$ conditional on $x_{1:N}$ is sometimes called the "noise distribution". Let also $R_{1:N} = (R_1, \cdots, R_N)$ be an unobserved discrete random sequence with values in a finite set $S = \{1, \cdots, s\}$ which 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 $p(x_{1:N}, r_{1:N}, y_{1:N})$, which should be designed in such a way that physical situations of interest are rather well fitted, and on the other hand estimating the couple $(x_{1:N}, r_{1:N})$ (or only the sequence $x_{1:N}$) from the observed sequence $y_{1:N}$ is computationally feasible. More precisely, the particular Bayesian smoothing problem which we address here consists in computing, for each $N = 1, 2, \cdots$ and each $n, 1 \leq n \leq N$, the conditional expectation $E(X_n|Y_{1:N} = y_{1:N})$ (or, in short, $E(X_n|y_{1:N})$) and associated conditional covariance matrix $Cov(X_n|Y_{1:N} = y_{1:N})$ (also denoted by $Cov(X_n|y_{1:N})$). The contribution of this paper consists in showing that $E(X_n|y_{1:N})$ and $Cov(X_n|y_{1:N})$ can be computed exactly, with complexity linear in time, in the recent model proposed in Pieczynski (2009).

2 MARKOV MARGINAL SWITCHING HIDDEN MODEL (MMSHM)

Let us first consider the classical conditionally linear Gaussian model, also called jump-Markov state-space system, which consists in considering that $R_{1:N}$ is a Markov chain and, roughly speaking, that conditionally on $R_{1:N}$, the couple $(X_{1:N}, Y_{1:N})$ is the classical Gaussian dynamic linear system. This is summarized in the following :

$$R_{1:N} \text{ is a Markov chain;} \tag{1}$$
$$X_{n+1} = F_n(R_n)X_n + W_n; \tag{2}$$
$$Y_n = H_n(R_n)X_n + Z_n, \tag{3}$$

where matrices $F_n(R_n)$ and $H_n(R_n)$ depend on R_n, $W_1, \cdots, W_N$ are Gaussian vectors in $\mathbb{R}^q$, $Z_1, \cdots, Z_N$ are Gaussian vectors in $\mathbb{R}^m$, and $X_1, W_1, \cdots, W_N, Z_1, \cdots, Z_N$ are independent. Let us notice that variables W_n and Z_n can also depend on R_n; however, we will keep model (1)-(3) for sake of simplicity. For fixed $R_1 = r_1, \cdots, R_n = r_n$, $E(X_n|y_{1:N})$ can be computed by classical Kalman-like smoothing methods (see e.g. Ait-el-Fquih and Desbouvries (2008) and references therein). However, it has been well known since Tugnait (1982) that exact computation is no longer possible with random Markov $R_{1:N}$ and different approximations must be used, see e.g. Andrieu *et al.* (2003), Cappé *et al.* (2005), Costa *et al.* (2005), Giordani *et al.* (2007), Ristic *et al.* (2004), Zoeter and Heskes (2006). To remedy this, different models have been recently proposed in Pieczynski (2009). The core novelty of these models with respect to the classical one (1)-(3) is that the couple $(R_{1:N}, Y_{1:N})$ is Markovian, which in turn ensures that one can compute $p(y_{n+1}|y_{1:n})$. This is a key computational point because the impossibility of filtering and smoothing in model (1)-(3) comes from the fact that $p(y_{n+1}|y_{1:n})$ cannot be computed exactly (see Andrieu *et al.* (2003), Tugnait (1982)). Let us thus consider the following model, first introduced in Pieczynski (2009) :

Definition

The triplet $(X_{1:N}, R_{1:N}, Y_{1:N})$ will be called a "Markov marginal switching hidden model" (MMSHM) if:

$$(R_{1:N}, Y_{1:N}) \text{ is a Markov chain;} \tag{4}$$
$$X_{n+1} = F_n(R_n, Y_n)X_n + W_n, \tag{5}$$

where matrix $F_n(R_n, Y_n)$ depends on (R_n, Y_n), $W_1, \cdots, W_n$ are independent zero-mean random vectors in $\mathbb{R}^q$ such that W_n is independent from $(R_{1:N}, Y_{1:N})$ for each n, $1 \leq n \leq N$. Note that variables $W_1, \cdots, W_n$ are not necessarily Gaussian and do not necessarily have a covariance matrix. This model is an extension of the early model proposed in Pieczynski (2008).

In (4) the chain $R_{1:N}$ does not need to be Markovian, which is the reason why we call (4)-(5) a "Markov marginal switching" model and not a "Markov switching" model. As studied in Derrode and Pieczynski (2004), such a Markov chain $(R_{1:N}, Y_{1:N})$ can be much more complex, and much more efficient in unsupervised hidden discrete data segmentation, than the classical hidden Markov chain (HMC), in which $R_{1:N}$ is Markovian. Theoretical results specifying under which conditions on a Markov chain $(R_{1:N}, Y_{1:N})$ the chain $R_{1:N}$ is Markovian can be found in Pieczynski (2007).

Finally, in the classical Markov switching model (1)-(3) $R_{1:N}$ is Markovian and $(R_{1:N}, Y_{1:N})$ is not, while in the MMSHM (4)-(5) $(X_{1:N}, R_{1:N})$ is not necessarily Markovian but $(R_{1:N}, Y_{1:N})$ is. From a modeling point of view, it does not seem to appear clearly why any of these properties should fit real situations better than the other; however, from a computational point of view the possibility of exact calculations is a clear advantage of the MMSHM model over the classical Markov-switching model : as we now see, in (4)-(5) $E(X_n|y_{1:N})$ and $Cov(X_n|y_{1:N})$ can be computed exactly with a computational cost linear in the time index N.

3 Exact Bayesian smoothing in MMSHM

Let us consider an MMSHM $(X_{1:N}, Y_{1:N}, R_{1:N})$. Let us set, for $1 \leq n \leq N-1$:

$$E(X_{n+1}, r_n|y_{1:N}) = E(X_{n+1}|r_n, y_{1:N})p(r_n|y_{1:N}); \tag{6}$$

$$E(X_{n+1}X_{n+1}^T, r_n|y_{1:N}) = E(X_{n+1}X_{n+1}^T|r_n, y_{1:N})p(r_n|y_{1:N}). \tag{7}$$

Of course, we have

$$E(X_{n+1}|y_{1:N}) = \sum_{r_n} E(X_{n+1}, r_n|y_{1:N}), \tag{8}$$

$$Cov(X_{n+1}|y_{1:N}) = \sum_{r_n} E(X_{n+1}X_{n+1}^T, r_n|y_{1:N}) - E(X_{n+1}|y_{1:N})E(X_{n+1}|y_{1:N})^T, \tag{9}$$

and thus it is sufficient to compute $E(X_{n+1}, r_n|y_{1:N})$ and $E(X_{n+1}X_{n+1}^T, r_n|y_{1:N})$. In the following proposition we show that $E(X_{n+1}, r_n|y_{1:N})$ and $E(X_{n+1}X_{n+1}^T, r_n|y_{1:N})$ can be computed with complexity linear in time :

Proposition 1.

Let $(X_{1:N}, Y_{1:N}, R_{1:N})$ be an MMSHM with given transitions $p(r_{n+1}, y_{n+1}|r_n, y_n)$. Then $E(X_{n+1}, r_n|y_{1:N})$ can be computed with linear complexity in time index N in the following way :

- Compute $p(r_1|y_{1:N})$ and $p(r_n|r_{n-1}, y_{1:N})$ for $2 \leq n \leq N$ as

$$p(r_1|y_{1:N}) = \frac{\beta_1(r_1)}{\sum_{r_1}\beta_1(r_1)}, p(r_n|r_{n-1}, y_{1:N}) = \frac{\beta_n(r_n)}{\beta_{n-1}(r_{n-1})}, \tag{10}$$

  where $\beta_n(r_n)$ are computed by the backward recursions

$$\beta_N(r_N) = 1, \beta_{n-1}(r_{n-1}) = \sum_{r_n} p(r_n, y_n|r_{n-1}, y_{n-1})\beta_n(r_n); \tag{11}$$

- For $2 \leq n \leq N$, compute $p(r_n|y_{1:N})$ by the forward recursion

$$p(r_{n+1}|y_{1:N}) = \sum_{r_n} p(r_n|y_{1:N})p(r_{n+1}|r_n, y_{1:N}); \tag{12}$$

- Compute $E(X_{n+1}, r_n|y_{1:N})$ from $E(X_n, r_{n-1}|y_{1:N})$ by the recursion :

$$E(X_{n+1}, r_n|y_{1:N}) = F_n(r_n, y_n) \sum_{r_{n-1}} E(X_n, r_{n-1}|y_{1:N})p(r_n|r_{n-1}, y_{n-1:N}). \tag{13}$$

If, in addition, the covariance matrices $\Sigma_1, \cdots, \Sigma_N$ of $W_1, \cdots, W_N$ exist, then $E(X_{n+1}X_{n+1}^T, r_n|y_{1:N})$ satisfies

$$E(X_{n+1}X_{n+1}^T, r_n|y_{1:N}) = F_n(r_n,y_n)[\sum_{r_{n-1}} E(X_nX_n^T, r_{n-1}|y_{1:N})p(r_n|r_{n-1},y_{n-1:N})]F_n^T(r_n,y_n) + \Sigma_n p(r_n|y_{1:N}), \quad (14)$$

and thus $Cov(X_{n+1}|y_{1:N})$ can also be computed with linear complexity in time index N.

Proof.

(10)-(12) extend from hidden to Pairwise Markov chains $(R_{1:N}, Y_{1:N})$ (Derrode and Pieczynski (2004), Pieczynski (2007)) the classical calculations (Baum and Petrie (1966), Baum and Eagon (1967)). We now address (13). By assumption, $X_{n+1} = F_n(R_n,Y_n)X_n + W_n$. Since W_n and $(R_n, Y_{1:N})$ are independent, and W_n is zero-mean, we have

$$\begin{aligned} E(X_{n+1}|r_n,y_{1:N}) &= F_n(r_n,y_n)E(X_n|r_n,y_{1:N}) \\ &= F_n(r_n,y_n)\sum_{r_{n-1}} E(X_n|r_{n-1},r_n,y_{1:N})p(r_{n-1}|r_n,y_{1:N}). \end{aligned} \quad (15)$$

On the other hand, from model (4)-(5) $E(X_n|r_{n-1},r_n,y_{1:N}) = E(X_n|r_{n-1},y_{1:N})$, and thus

$$E(X_{n+1}|r_n,y_{1:N}) = F_n(r_n,y_n)\sum_{r_{n-1}} E(X_n|r_{n-1},y_{1:N})p(r_{n-1}|r_n,y_{1:N}). \quad (16)$$

Multiplying both sides by $p(r_n|y_{1:N})$ gives (13). Equation (14) is shown similarly : the independence of $W_1, \cdots, W_N$ implies that X_n and W_n are independent conditionally on $(R_{1:N}, Y_{1:N})$, so (5) gives

$$\begin{aligned} E(X_{n+1}X_{n+1}^T|r_n,y_{1:N}) &= F_n(r_n,y_n)E(X_nX_n^T|r_n,y_{1:N})F_n^T(r_n,y_n) + E(W_nW_n^T|r_n,y_{1:N}) \\ &= F_n(r_n,y_n)E(X_nX_n^T|r_n,y_{1:N})F_n^T(r_n,y_n) + \Sigma_n. \end{aligned} \quad (17)$$

On the other hand,

$$\begin{aligned} E(X_nX_n^T|r_n,y_{1:N}) &= \sum_{r_{n-1}} E(X_nX_n^T, r_{n-1}|r_n,y_{1:N}) \\ &= \sum_{r_{n-1}} E(X_nX_n^T|r_{n-1},y_{1:N})p(r_{n-1}|r_n,y_{1:N}). \end{aligned}$$

Injecting into (17) and multiplying by $p(r_n|y_{1:N})$ gives (14), which ends the proof.

Remarks.

- As far as estimating X_{n+1} (and not R_{n+1}) is concerned, our method enables us to compute $E(X_{n+1}|y_{1:N})$ and $Cov(X_{n+1}|y_{1:N})$, but not the distribution $p(x_{n+1}|y_{1:N})$, which indeed is a very rich mixture distribution. We thus solve the Bayesian smoothing problem for the loss function $L(x^1,x^2) = ||x^1 - x^2||^2$ only. Note however that this problem remains of interest, and indeed the quadratic loss function is used in many applications;

- Let now the problem consist in estimating simultaneously X_{n+1} and R_{n+1}. Then our method enables us to compute the exact Bayesian solution associated to the family of loss functions

$$L((x^1,r^1),(x^2,r^2)) = ||x^1 - x^2||^2 L'(r^1,r^2) \tag{18}$$

in which L' is arbitrary. To see this, let us notice that for a given $y_{1:N}$, the Bayesian estimator $(\hat{x}_n,\hat{r}_n)$ associates to $y_{1:N}$ the couple $(\hat{x}_n,\hat{r}_n)$ which minimizes the function :

$$(x_n,r_n) \mapsto \sum_{r'_n} \int_{\mathbb{R}^q} L((x_n,r_n),(x'_n,r'_n)) p(x'_n,r'_n|y_{1:N}) dx'_n.$$

Given (18), the couple $(\hat{x}_n,\hat{r}_n)$ minimizes

$$(x_n,r_n) \mapsto \int_{\mathbb{R}^q} ||x_n - x'_n||^2 [\sum_{r'_n} L'(r_n,r'_n) p(r'_n|y_{1:N}) p(x'_n|r'_n,y_{1:N})] dx'_n.$$

For fixed r_n, the minimum of this function is reached for

$$\hat{x}_n(r_n) = \sum_{r'_n} \frac{L'(r_n,r'_n) p(r'_n|y_{1:N})}{\sum_{r'_n} L'(r_n,r'_n) p(r'_n|y_{1:N})} E(X_n|r'_n,y_{1:N}), \tag{19}$$

which can be computed because $p(r_n|y_{1:N})$ and $E(X_n|r'_n,y_{1:N})$ can both be computed in (4)-(5). Thus we can first search $\hat{r}_n$ which minimizes $r_n \mapsto \hat{x}_n(r_n)$ in (19), and finally set $\hat{x}_n = E(X_n|\hat{r}_n,y_{1:N})$.
Let us finally remark that in case we are only interested in the Bayesian estimation of R_{n+1}, then the arbitrary loss function $L'(r^1,r^2)$ in (19) leads to the solution $\hat{r}_n^*$ which minimizes the function $r_n \mapsto \sum_{r'_n} L'(r_n,r'_n) p(r'_n|y_{1:N})$, and thus $\hat{r}_n^*$ differs from $\hat{r}_n$.

REFERENCES

AIT-EL-FQUIH, B., DESBOUVRIES, F. (2008): On Bayesian Fixed-Interval Smoothing Algorithms. *IEEE Transactions on Automatic Control*, *53-10*, 2437–42.

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

BAUM, L. E., PETRIE, T. (1966): Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, *37*, 1554-63.

BAUM, L. E., EAGON, J. A. (1967): An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bull. Amer. Meteorol. Soc.*, *73*, 360–63.

CAPPÉ, O., MOULINES, E., RYDEN, T. (2005): *Inference in hidden Markov models*. Springer, New York.

COSTA, O. L. V., FRAGOSO, M. D., MARQUES, R. P. (2005) *Discrete time Markov jump linear systems.* Springer-Verlag, New York.

DERRODE, S., PIECZYNSKI, W. (2004) Signal and Image Segmentation using Pairwise Markov Chains. *IEEE Transactions on Signal Processing*, *52(9)*, 2477–2489.

GIRDANI., P., KOHN, R., Van DIJK, D. (2007): A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, *137*, 112–133.

PIECZYNSKI, W. (2007) Multisensor triplet Markov chains and theory of evidence. *International Journal of Approximate Reasoning*, *45(1)*, 1–16.

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.

PIECZYNSKI, W. (2009). Exact filtering in Markov marginal switching hidden models. submitted to *Comptes Rendus de l'Académie des Sciences - Mathematiques*.

PIECZYNSKI, W., DESBOUVRIES, F. (2003): Kalman filtering using pairwise Gaussian models *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003)*, Hong-Kong, April.

RISTIC, B, ARULAMPALAM, S., GORDON, N. (2004): *Beyond the Kalman Filter - Particle filters for tracking applications.* Artech House, Boston, MA.

TUGNAIT, J. K. (1982) : Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control*, *AC-25*, 1054–1065.

ZOETER, O., HESKES, T. (2006): Deterministic approximate inference techniques for conditionally Gaussian state space models. *Statistical Computation*, *16*, 279–292.