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

Exact Smoothing in Hidden Conditionally Markov Switching Chains

Wojciech Pieczynski [1]

[1] *Institut Telecom, Telecom SudParis,*
Dept. CITI, CNRS UMR 5157
8, rue Charles Fourier, 91 000 Evry, France
E-mail: Wojciech.Pieczynski@it-sudparis.eu

Abstract: The problem considered in this paper is the problem of the exact calculation of smoothing in hidden switching state-space systems. There is a hidden state-space chain X, the switching Markov chain R, and the observed chain Y. In the classical, widely used "conditionally Gaussian state-space linear model" (CGSSLM) the exact calculation with complexity linear in time is not feasible and different approximations have to be made. Different alternative models, in which the exact calculations are feasible, have been recently proposed (2008). The core difference between these models and the classical ones is that the couple (R, Y) is a Markov one in the recent models, while it is not in the classical ones. Here we propose a further extension of these recent models by relaxing the hypothesis of the Markovianity of X conditionally on (R, Y). In fact, in all classical models and as well as in the recent ones, the hidden chain X is always a Markov one conditionally on (R, Y). In the proposed model it can be of any form. In particular, different "long memory" processes can be considered. In spite of this larger generality, we show that the smoothing formulae are still calculable exactly with the complexity polynomial in time.
Keywords: exact smoothing, Markov switches, partially Markov chains,

1 Introduction

Let us consider $X_1^N = (X_1, ..., X_N)$ and $Y_1^N = (Y_1, ..., Y_N)$ two sequences of random vectors, and let $R_1^N = (R_1, ..., R_N)$ be a finite-values Markov chain. Each X_n takes its values from R^q, while each Y_n takes its values from R^m. The sequences X_1^N and R_1^N are hidden and the sequence Y_1^N is observed. We deal with the problem of smoothing, which consists of a computation, for each $n = 1$, …, N, of the conditional expectation $E[X_n | R_n = r_n, Y_1^N = y_1^N]$. To simplify notations, this expectation, as well as other similar quantities, will be denoted by $E[X_n | r_n, y_1^N]$. To fix ideas, let us consider the classical widely used "conditionally Gaussian state-space linear model" (CGSSLM), which consists of considering that R_1^N is a Markov chain and, roughly speaking, (X_1^N, Y_1^N) is the classical linear system conditionally on R_1^N. 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 X_1, W_1, …, W_N are independent (conditionally on R_1^N) Gaussian vectors in R^q, Z_1, …, Z_N are independent (conditionally on R_1^N) Gaussian vectors in R^m, $F_1(R_1)$, …, $F_N(R_N)$ are matrices of size $q \times q$ depending on switches, and $H_1(R_1)$, …, $H_N(R_N)$ are matrices of size $q \times m$ also depending on switches R_1, …, R_N. It has been well known since the publication of (Tugnait, 1982) that the exact computation of $E[X_n | r_n, y_1^N]$ is not feasible with linear - or even polynomial - complexity in time in such models, and different approximations must be used. There are dozens of papers dealing with this approximation problem and a rich bibliography can be seen in recent books (Cappe *et al.* 2005, Costa *et al.* 2005). Roughly speaking, there are two families of approximating methods: the stochastic ones, based on the Monte Carlo Markov Chains (MCMC) principle (Andrieu *et al.* 2003, Cappe *et al.* 2005, Giordani *et al.* 2007, among others), and deterministic ones (Costa *et al.* 2005, Zoeter *et al.* 2006, among others). To remedy this impossibility of exact computation, two different models have been recently proposed in (Abbassi and Pieczynski 2008, Pieczynski 2008). Based on ideas issued from the general triplet Markov chains considerations (Pieczynski and Desbouvries 2005), they make the exact computation of optimal Kalman-like filters possible. The exact calculation of smoothing is also possible in these models, as shown in (Bardel *et al.* 2009). The general idea leading to these models is to consider the independence of the X_1^N and Y_1^N conditionally on R_1^N. Of course, this does not mean that X_1^N and Y_1^N are independent. Once this hypothesis assumed, there is a wide range of different models in which exact smoothing and exact filtering can be performed with complexity linear in time. In addition, Gaussian distributions are not required, neither for $p(y_1^N | r_1^N, x_1^N) = p(y_1^N | r_1^N)$ nor for $p(x_1^N | r_1^N, y_1^N) = p(x_1^N | r_1^N)$. Moreover, the distribution of R_1^N is no longer necessarily a Markov distribution and can be extended to others ones, such as a semi-Markov distribution.

Afterward these early models were extended to more general ones, in which the independence of X_1^N and Y_1^N conditionally on R_1^N is no longer required (Pieczynski 2009, Pieczynski and Desbouvries 2009). Called “Markov marginal switching hidden models” (MMSHMs), these models verify:

$$(R_1^N, Y_1^N) \text{ is a Markov chain;} \tag{4}$$

$$X_{n+1} = F_n(R_n, Y_n) X_n + W_n, \tag{5}$$

where X_1 is given, W_1, …, W_N are independent random centered vectors in R^q with finite covariance matrices Σ_1, …, Σ_N, and $F_1(R_1)$, …, $F_N(R_N)$ are matrices of size $q \times q$ depending on the switches and on the observations.

The important difference between the classical model (1)-(3) and the recent model (4)-(5) is the following. In the classical model (1)-(3) the couple (X_1^N, R_1^N) is Markovian, the couple (R_1^N, Y_1^N) is not, and neither filtering nor smoothing is possible with complexity linear - or even polynomial - in time. In the models (4)-(5) the couple (R_1^N, Y_1^N) is Markovian (note that R_1^N is not necessarily Markovian (Pieczynski 2007)), the couple (X_1^N, R_1^N) is not in the general case, and both filtering and smoothing are calculable with complexity linear in time. From a modeling point of view, it does not seem to appear clearly why one model should fit real situations better than the other; however, from the computational point of view the possibility of exact calculations is a clear advantage of the MMSHM model over the classical one.

The aim of this paper is to extend the models (4)-(5) to more general ones, in which the triplet (X_1^N, R_1^N, Y_1^N) in no longer necessarily a Markov one. In fact, let us focus on the distribution $p(x_1^N | r_1^N, y_1^N)$ of the hidden chain X_1^N conditional on the couple (R_1^N, Y_1^N). In both models (1)-(3) and (4)-(5) this distribution is a Markov one, and thus it is a "short memory" distribution. In practice, many phenomena must be modeled by a "long- memory" - thus non Markovian – distribution (Beran and Taqqu 1994, Doukhan *et al.* 2003). Our aim is to show that exact smoothing remains feasible for a wide range of distributions $p(x_1^N | r_1^N, y_1^N)$, including "long memory" distributions.

Finally, the problem we deal with is the following. Let us assume that we have a random chain X_1^N with stochastic switches modeled by a Markov chain R_1^N. Neither X_1^N nor R_1^N are observable and we observe a "noisy" version Y_1^N. What can be done to estimate (X_1^N, R_1^N)? We propose a new model, which extends the model (4)-(5) and in which both $p(r_n | y_1^N)$ and $E[X_n | Y_1^N = y_1^N]$ are computable with complexity polynomial in time. This makes an estimation of (X_1^N, R_1^N) for very large N feasible. The general idea is to use the recent results presented in (Lanchantin *et al.* 2007), where one considers a long memory chain with switches (X_1^N, R_1^N) in which it is possible to compute $E[X_n | r_1^N]$. Roughly speaking, here we add a noise Y_1^N to the model studied in (Lanchantin *et al.* 2007), and we show the calculability of $p(r_n | y_1^N)$ and $E[X_n | y_1^N]$.

The new model is proposed and discussed in the next section, and the exact computation of smoothing is described in the third one.

2. Hidden conditionally Markov switching chains

Let us consider the triplet (X_1^N, R_1^N, Y_1^N) as above. The starting point of the model we propose is to consider that, conditionally on Y_1^N, the distribution of the couple (X_1^N, R_1^N) is the distribution of the "partially Markov chain", which is an extension of the "partially Markov Gaussian chain" (PMGC) introduced in (Lanchantin *et al.* 2007). Although the situation here is different from the situation in this paper where X_1^N was observed and R_1^N was being searched for, the extension of the PMGC used here will here be a basic brick to build a model for the triplet (X_1^N, R_1^N, Y_1^N). A PMGC verifies

$$p(x_{n+1}, r_{n+1} | x_1^n, r_1^n) = p(r_{n+1} | r_n) p(x_{n+1} | r_{n+1}, x_1^n) \quad (6)$$

where $p(x_{n+1} | r_{n+1}, x_1^n)$ are assumed Gaussian. We see that (X_1^N, R_1^N) is Markovian with respect to R_1^N, but is not necessarily Markovian with respect to X_1^N, which is why it is named "partially" Markov.
Finally, the model we propose is the following

Definition
A triplet (X_1^N, R_1^N, Y_1^N) will be said to be a "hidden conditionally Markov switching chain" (HCMSC) if it verifies

(R_1^N, Y_1^N) is a Markov chain ; (7)

for $n = 2, \ldots, N$, $X_n = F^n(R_n, Y_n) X_1^{n-1} + W_n$, (8)

with $F^n(r_n, y_n) = [F_1(r_n, y_n), F_2(r_n, y_n), ..., F_{n-1}(r_n, y_n)]$, where each $F_i(r_n, y_n)$ is a matrix of size $q \times q$ depending on (r_n, y_n), and X_1, W_1, …, W_N are independent centred vectors in R^q such that each W_n is independent from (R_1^N, Y_1^N).
The oriented dependence graphs of the classical CGSSLM and the new HCMSC are presented in Figure1, (a) and (b), respectively.

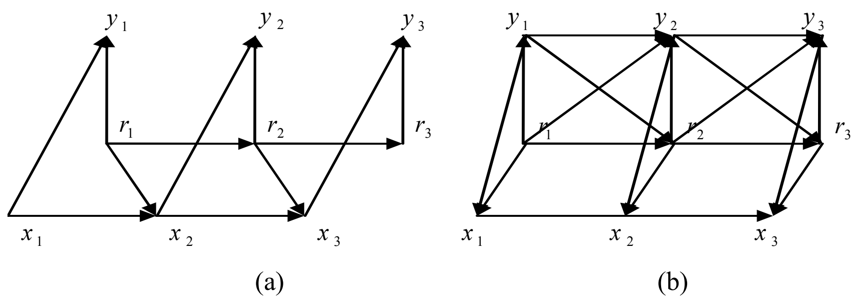

Fig. 1 : (a) classical model (1)-(3); (b) new model (7)-(8).

Let us underline the following points:
(i) The hypotheses on the sequence W_1, …, W_N are very weak, as only the existence of their means is requested. If the chain W_1, …, W_N is Gaussian then (X_1^N, R_1^N) is a PMGC conditionally on Y_1^N;
(ii) the variables W_1, …, W_N do not necessarily have a covariance matrix;
(iii) the chain R_1^N is not necessarily Markovian, which is the reason why we call the model a "conditionally Markov switching" model and not a "Markov switching" model. However, R_1^N is Markovian conditionally on Y_1^N, which is of core importance in the computation of the smoothing.

3. Exact smoothing in HCMSC

Let us consider an HCMSC (X_1^N, R_1^N, Y_1^N). As (R_1^N, Y_1^N) is a Markov chain, we have $p(r_n | r_{n-1}, y_1^N) = p(r_n | r_{n-1}, y_{n-1}^N)$. Then these transitions and $p(r_1 | y_1^N)$ are classically given by $p(r_n | r_{n-1}, y_{n-1}^N) = \frac{\beta_n(r_n)}{\beta_{n-1}(r_{n-1})}$ and $p(r_1 | y_1^N) = \frac{\beta_1(r_1)}{\sum_{r_1} \beta_1(r_1)}$, where $\beta_n(r_n)$ are classically computable by the backward recursions $\beta_N(r_N) = 1$, $\beta_{n-1}(r_{n-1}) = \sum_{r_n} p(r_n, y_n | r_{n-1}, y_{n-1}) \beta_n(r_n)$: see (Derrode and Pieczynski 2004). Finally, $p(r_n | y_1^N)$ are classically computed from $p(r_1 | y_1^N)$ and the transitions $p(r_n | r_{n-1}, y_1^N)$ by the forward recursions: $p(r_1 | y_1^N)$ given above, $p(r_n | y_1^N) = \sum_{r_{n-1}} p(r_{n-1} | y_1^N) p(r_n | r_{n-1}, y_1^N)$. Knowing the transitions $p(r_n | r_{n-1}, y_1^N)$ and the marginal distributions $p(r_n | y_1^N)$ makes feasible to compute, for $1 \le k < n \le N$, all $p(r_n | r_k, y_1^N)$ with complexity polynomial in time. Consequently, for $1 \le k < n \le N$ all $p(r_k | r_n, y_1^N)$ are also computable with complexity polynomial in time.
We can state the following result

Proposition

Let (X_1^N, R_1^N, Y_1^N) be an HCMSC such that $p(r_1, y_1)$ and the transitions $p(r_{n+1}, y_{n+1} | r_n, y_n)$ of the Markov chain (R_1^N, Y_1^N) are given.
Then $p(x_n | y_1^N)$ and $E[X_n | y_1^N]$ can be computed with polynomial complexity in time in the following way:
(i) compute $p(r_n | y_1^N)$ for each $1 \le n \le N$, and $p(r_k | r_n, y_1^N)$ for each $1 \le k < n \le N$ as specified above;
(ii) for known $E[X_1 | r_1, y_1^N]$, $E[X_2 | r_2, y_1^N]$, …, $E[X_{n-1} | r_{n-1}, y_1^N]$, compute $E[X_n | r_n, y_1^N]$ with

$$E[X_n | r_n, y_1^N] = F^n(r_n, y_n) E[X_1^{n-1} | r_n, y_1^N] = \sum_{1 \le k \le n-1} F_k^n(r_n, y_n) E[X_k | r_n, y_1^N], \quad (9)$$

where $E[X_k | r_n, y_1^N]$ are given by

$$E[X_k | r_n, y_1^N] = \sum_{r_k} E[X_k | r_k, y_1^N] p(r_k | r_n, y_1^N); \quad (10)$$

(iii) compute $E[X_n | y_1^N] = \sum_{r_n} E[X_n | r_n, y_1^N] p(r_n | y_1^N)$.

Proof.
Taking the expectation of (8) conditional to $(R_n, Y_1^N) = (r_n, y_1^N)$ leads to (9). To show (10) let us notice that according to the model, the variables R_n and X_k are independent conditionally on (R_k, Y_1^N). This leads to $E[X_k | r_n, y_1^N] =$

$\sum_{r_k} E[X_k, r_k | r_n, y_1^N] = \sum_{r_k} E[X_k | r_k, r_n, y_1^N] p(r_k | r_n, y_1^N) = \sum_{r_k} E[X_k | r_k, y_1^N] p(r_k | r_n, y_1^N)$, which is (10) and which ends the proof.

References

Abbassi, N. and Pieczynski, W. 2008. Exact Filtering In Semi-Markov Jumping System, Sixth International Conference of Computational Methods in Sciences and Engineering, September 25-30, Hersonissos, Crete, Greece.

Andrieu, C., Davy, C. 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.

Bardel N., Desbouvries F., Pieczynski, W., and Barbaresco, F. 2009. An exact non-Gaussian Markov switching Bayesian smoothing algorithm with application to Track-Before-Detect, submitted to *Radar 2009*.

Beran J. and Taqqu M. S. 1994. *Statistics for Long-Memory processes.* Monographs on Statistics and Applied Probability, Chapman and Hall, New York.

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.

Derrode S. and Pieczynski W. 2004. Signal and Image Segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, 52(9): 2477-2489.

Doukhan P., Oppenheim G., and Taqqu M. S. 2003. *Long-Range Dependence*, Birkhauser.

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

Lanchantin, P., Lapuyade-Lahorgue, J., and Pieczynski, W. 2008. Unsupervised segmentation of triplet Markov chains hidden with long-memory noise, *Signal Processing*, 88(5): 1134-1151.

Pieczynski W., and Desbouvries, F. 2005. On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA), Brest, France.

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 Mathématique*.

Pieczynski, W. and Desbouvries, F. 2009. Exact Bayesian smoothing in triplet switching Markov chains, Complex data modeling and computationally intensive statistical methods for estimation and prediction (S. Co 2009), September 14-16, Milan, Italy, 2009.

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

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