Stochastic Modeling Techniques and Data Analysis international conference (SMTDA '10), Chania, Greece, June 8-11, 2010.

Long memory based approximation of filtering in non linear switching systems

Noufel Abbassi and Wojciech Pieczynski

CITI Department, CNRS UMR 5157, Telecom SudParis,
Evry, France
e-mail : Wojciech.Pieczynskii@it-sudparis.eu
e-mail : Noufel.Abbassi@it-sudparis.eu

Abstract: In this paper we consider conditionally Gaussian state space models with Markovian switches and we propose a new method of approximating the optimal solution by the use of Markov chains hidden with long memory noise. We show through experiments that our method can be more efficient than the classical particle filter based approximation.
Keywords: Conditionally Gaussian state space model, Markov switching, Markov chains hidden with long memory noise, Expectation-Maximization, Iterative conditional estimation.

1 Introduction

Let $X_1^N = (X_1, ..., X_N)$, $R_1^N = (R_1, ..., R_N)$ and $Y_1^N = (Y_1, ..., Y_N)$ be three sequences of random variables. Each X_n and Y_n take their values from $\mathbb{R}$, while each R_n takes its values from a finite set of switches $\Omega = \{\omega_1, ..., \omega_K\}$. The sequences X_1^N and R_1^N are hidden and the sequence Y_1^N is observed. For each $n = 1$, …, N we will denote by $p(r_n | y_1^n)$ the distribution of R_n conditional on $Y_1^n = y_1^n$, and we will denote by $E[X_n | r_n, y_1^n)$ the expectation of X_n conditional on $(R_n, Y_1^n) = (r_n, y_1^n)$. We deal with the classical problem of filtering, which consists of computation of $p(r_{n+1} | y_1^{n+1})$ and $E[X_{n+1} | r_{n+1}, y_1^{n+1})$ with a reasonable complexity. We consider the following classical partly non linear model:

$$R_1^N \text{ is a Markov chain;} \quad (1.1)$$

$$X_{n+1} = F_n(R_{n+1}) X_n + H_n(R_{n+1}) W_{n+1} \text{ ; and} \quad (1.2)$$

$$Y_n = G_n(R_n, X_n) + K(R_n) Z_n . \quad (1.3)$$

where X_1, W_2, …, W_N , V_1, …, V_N are independent Gaussian variables, and for each $n = 1$, …, N, $F_n(R_n)$, $H_n(R_n)$, $G_n(R_n)$, $K_n(R_n)$ are real numbers depending on switches. Such models are of interest in numerous situations (Cappe et al. (2005), Costa et al. (2005), Zoeter and Heskes (2006)), among others. However, it has been well known since (Tugnait (1982)) that exact filtering and smoothing are not feasible with linear - or even polynomial - complexity in time in such models, and different approximations must be used. Many papers deal with this approximation problem and a rich bibliography can be seen in (Andrieu et al. (2003), Cappe et al. (2005), Costa et al. (2005), Zoeter and Heskes (2006)), among others.

The problem lies in the fact that the distribution of (R_1^N, Y_1^N) is not easy to manage and, in particular, does not allow one to compute the conditional probabilities $p(r_n | y_1^n)$ with a reasonable complexity. In fact, we have

$$p(x_{n+1}, r_{n+1} | y_1^{n+1}) = p(r_{n+1} | y_1^{n+1}) p(x_{n+1} | r_{n+1}, y_1^{n+1}), \quad (1.4)$$

where the probabilities $p(x_{n+1} | r_{n+1}, y_1^{n+1})$ can be recursively computed with a Kalman filter once they have been classically approximated by Gaussian densities, and the probabilities $p(r_{n+1} | y_1^{n+1})$ have to be approximated. Then they are approximated with different methods among which particle filtering is widely used (Andrieu et al. (2003)). The aim of this paper is to propose an alternate approximation of the distribution of (R_1^N, Y_1^N) based on the two following models. The first one is the very classical hidden Markov chain (HMM) model, in which $p(r_n | y_1^n)$ are computable with the classical recursive "forward" procedure. The second one is the recent Markov chain hidden with the long-memory noise (HMM-LMN) (Lanchantin et al. (2008)), in which $p(r_n | y_1^n)$ are computable too. In these two models we propose an original recursive parameter estimation method based on the general "Iterative Conditional Estimation" (ICE) procedure, which implies the possibility of partially unsupervised filtering.

Finally, we will consider the model (1.1)-(1.3) in which the distribution of (R_1^N, Y_1^N) will be approximated either by an HMM or an HMM-LMN distribution.

2 Filtering switches in hidden Markov models

Let (R_1^N, Y_1^N) be the classical HMM, whose distribution is of the form

$$p(r_1^N, y_1^N) = p(r_1) p(y_1 | r_1) \prod_{n=1}^{N-1} p(r_{n+1} | r_n) p(y_{n+1} | r_{n+1}). \quad (2.1)$$

The conditional probabilities $p(r_{n+1} | y_1^{n+1})$ are then computed from $p(r_n | y_1^n)$ by

$$p(r_{n+1}|y_1^{n+1}) = \frac{p(y_{n+1}|r_{n+1})\sum_{r_t\in\Omega} p(r_{n+1}|r_n)p(r_n|y_1^n)}{\sum_{r_{t+1}\in\Omega} p(y_{n+1}|r_{n+1})\sum_{r_t\in\Omega} p(r_{n+1}|r_n)p(r_n|y_1^n)} \qquad (2.2)$$

Let us consider the following distribution, called a « partially » Markov distribution (Lanchantin et al. (2008)):

$$p(r,y) = p(r_1)p(y_1|r_1)\prod_{n=1}^{N-1} p(r_{n+1}|r_n)p(y_{n+1}|r_{n+1},y_1^n) \qquad (2.3)$$

One can see how the model (2.3) extends the classical HMM (2.1) : the former is obtained setting $p(y_{n+1}|r_{n+1},y_1^n) = p(y_{n+1}|r_{n+1})$ in the latter. However, similarly to (2.2) one has

$$p(r_{n+1}|y_1^{n+1}) = \frac{p(y_{n+1}|r_{n+1},y_1^n)\sum_{r_t\in\Omega} p(r_{n+1}|r_n)p(r_n|y_1^n)}{\sum_{r_{t+1}\in\Omega} p(y_{n+1}|r_{n+1},y_1^n)[\sum_{r_t\in\Omega} p(r_{n+1}|r_n)p(r_n|y_1^n)]} \qquad (2.4)$$

3 Parameter estimation

In the classical HMM with Gaussian noise the model parameters are usually estimated with the classical "Expectation-Maximization" (EM) algorithm. They can also be estimated by the "Iterative Conditional Estimation" (ICE) algorithm, which is another iterative parameter estimation method. For fixed N both EM and ICE provide a sequence of parameters θ^0, θ^1, $\ldots, \theta^m$, where the vector of parameters θ contains the distribution of (R_1, R_2), which defines the distribution of the stationary Markov chain R_1^N, K means and K variances defining the K Gaussian distributions $p(y_n|r_n=\omega_1)$, …, $p(y_n|r_n=\omega_K)$, which are identical for each $n=1$, …, N.

ICE provides the next θ^{q+1} from the current θ^q and $y_1^N = (y_1,...,y_N)$. For $j,k=1$, …, K, let $p_{jk} = p(r_1=\omega_j, r_2=\omega_k)$ and let μ_j, σ_j^2 be the common mean and the variance of the Gaussian distributions $p(y_1|r_1=\omega_j)$, …, $p(y_N|r_N=\omega_j)$. The next values p_{jk}^{q+1}, μ_j^{q+1}, $(\sigma_j^{q+1})^2$ are obtained from the current θ^q and $y_1^N = (y_1,...,y_N)$ in the following way. The parameters p_{jk}^{q+1} are given by

$$p_{jk}^{q+1} = \frac{1}{n}\sum_{i=1}^{n-1} p(r_i=\omega_j, r_{i+1}=\omega_k|y_1^n,\theta^q), \qquad (3.1)$$

where $p(r_i = \omega_j, r_{i+1} = \omega_k | y_1^N, \theta^q)$ are classically computed with the "forward" and "backward" recursions. To obtain μ_j^{q+1}, $(\sigma_j^{q+1})^2$, one samples $r_1^{N,q} = (r_1^q, ..., r_N^q)$ according to $p(r_1^N | y_1^N, \theta^q)$ and one sets

$$\mu_j^{q+1} = \frac{\sum_{i=1}^{N} y_i 1_{[r_i = \omega_j]}}{\sum_{i=1}^{N} 1_{[r_i = \omega_j]}} \qquad (3.2)$$

$$(\sigma_j^{q+1})^2 = \frac{\sum_{i=1}^{N} (y_i - \mu_j^{q+1})^2 1_{[r_i = \omega_j]}}{\sum_{i=1}^{N} 1_{[r_i = \omega_j]}} \qquad (3.3)$$

EM provides the next θ^{q+1} from the current θ^q and $y_1^N = (y_1, ..., y_N)$ in the following way. The next parameters p_{jk}^{q+1} are given by (3.1), exactly as in the case of ICE. The next μ_j^{q+1}, $(\sigma_j^{q+1})^2$ are computed with

$$\mu_j^{q+1} = \frac{\sum_{i=1}^{N} p(r_i = \omega_j | y_1^N, \theta^q) y_i}{\sum_{i=1}^{N} p(r_i = \omega_j | y_1^N, \theta^q)} \qquad (3.4)$$

$$(\sigma_j^{q+1})^2 = \frac{\sum_{i=1}^{N} p(r_i = \omega_j | y_1^N, \theta^q)(y_i - \mu_j^{q+1})^2}{\sum_{i=1}^{N} p(r_i = \omega_j | y_1^N, \theta^q)} \qquad (3.5)$$

Let us notice that there is a stochastic aspect in the sequence produced by ICE, while the sequence related to EM is deterministic. This can make EM more sensitive to the initialization θ^0 than ICE. However, numerous comparisons between EM and ICE have been performed and, on the whole, in the case of classical HMM with Gaussian noise they provide similar results (Benmiloud and Pieczynski (1995)).

We propose the following adaptive parameter estimation method, based on EM or ICE. Let $\hat{\theta}^n$ be the parameter obtained from y_1^n. Then $\hat{\theta}^{n+1}$ is obtained from y_1^{n+1} by applying EM (or ICE) and using $\hat{\theta}^n$ as the initial value. Thus one obtains a sequence $\hat{\theta}^0$, $\hat{\theta}^1$, ..., $\hat{\theta}^N$, each $\hat{\theta}^n$ being estimated from y_1^n. The computation of $p(r_{n+1}|y_1^{n+1})$ from $p(r_n|y_1^n)$ using (2.2) is then unsupervised and is performed in two steps:

(i) compute $\hat{\theta}^{n+1}$ with EM (or ICE) from y_1^{n+1} using $\hat{\theta}^n$ as initialization;

(ii) compute $p(r_{n+1}|y_1^{n+1})$ from $p(r_n|y_1^n)$ using (2.2) and $\hat{\theta}^{n+1}$.

Thus for each $n=1, \ldots, N-1$ one has to perform a finite number, defined in some way, of EM or ICE iterations.

In the following section these methods will be called "adaptive EM (AEM) and adaptive ICE (AICE).

Let us return to the « partially » Markov distribution defined by (2.3) and let us consider the following particular case. One considers K Gaussian distributions $p^1(y_1^N)$, …, $p^K(y_1^N)$ which are used to define the distributions $p(y_{n+1}|r_{n+1}, y_1^n)$ in (2.3): for each $n=1, \ldots, N-1$ and $k=1, \ldots, K$, $p(y_{n+1}|r_{n+1}=\omega_k, y_1^n)$ is the conditional Gaussian distribution given by the Gaussian distribution $p^k(y_1^{n+1})$, which is the marginal distribution of the Gaussian distribution $p^k(y_1^N)$. Besides, for each $k=1$, …, K, $p^k(y_1^N)$ is defined by the mean vector $M^k=(m^k,...,m^k)$ and a variance-covariance matrix $\Gamma^k=[\gamma_{ij}^k]_{1\le i,j\le N}$, with

$$\gamma_{ij}^k=\sigma_k^2(1+|i-j|)^{-a_k} \qquad (3.5)$$

The distribution of such a model, which will be called in the following HMM with "long memory noise" (HMM-LMN), is then defined by the parameters p_{ij}, which give the distribution of the Markov chain X, and K triplets (m^1,σ_1^2,a_1), …, (m^K,σ_K^2,a_K). HMM-LMN has been recently proposed in (Lanchantin et al. (2008)) and an extension of ICE, which is not trivial, to the HMM-LMN context has been described and successfully tested. As above, we propose using this ICE to estimate the parameters in an "adaptive" manner. One has the same two steps as above:

(i) compute $\hat{\theta}^{n+1}$ with ICE from y_1^{n+1} using $\hat{\theta}^n$ as initialization;

(ii) compute $p(r_{n+1}|y_1^{n+1})$ from $p(r_n|y_1^n)$ using (2.4) and $\hat{\theta}^{n+1}$.

4 Experiments

Let us consider an HMM with two classes $\Omega=\{\omega_1,\omega_2\}$. The distribution of R_1^N is defined by $p(r_1=\omega_1)=p(r_1=\omega_2)=0.5$ and the transitions $p(r_{n+1}=\omega_1|r_n=\omega_2)=p(r_{n+1}=\omega_2|r_n=\omega_1)=\rho$. The two Gaussian noise distributions are $N(\mu_1,\sigma_1^2)=N(-1/2,1/2)$, $N(\mu_2,\sigma_2^2)=N(1/2,1/3)$. We consider $N=200$ for the sample size and the results for four different values of ρ are presented for each case in Table 1. An example of the evolution with respect to n of the estimation with AEM and AICE of ρ (for the true $\rho = 0.2$) is presented in Figure 1.

It is difficult to compare AEM and AICE as they are very sensitive to the initialization and give very stochastic results for small n. However, in all experiments performed they are of similar efficiency when n increases.

ρ	0.1	0.2	0.3	0.4
EM $\hat{\rho}$	0.10	0.19	0.36	0.51
ICE $\hat{\rho}$	0.09	0.18	0.39	0.52
EM τ	8.5%	6.8%	9.0%	17.1%
ICE τ	13.0%	6.0%	11.3%	22.0%

Tab. 1. Estimates of ρ and error ratio τ of unsupervised adaptive filtering.

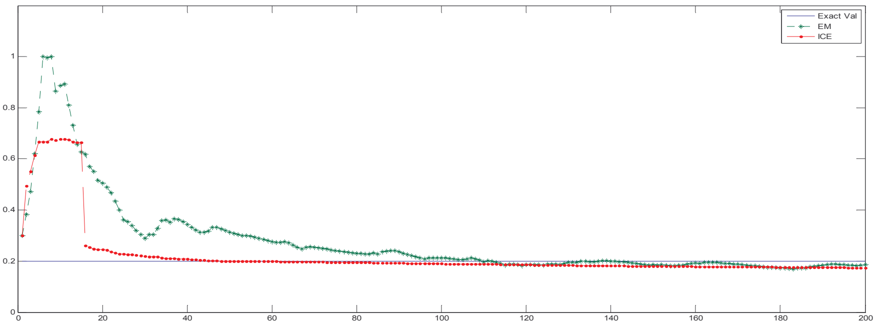

Fig. 1. Evolution with n of the estimation with AEM and AICE of ρ.

Let us now consider the case of data simulated with the model (1.1)-(1.3) and filtered by three methods. The first one is the method based on the particle filter (Andrieu et al. (2003), Doucet et al. 2001)). The second one is based on HMM, and the third one on HMM-LMN. As ρ has to be known in the particle filter based method, we also assume it to be known in the other two methods; however, let us underline the fact that it could be estimated which is an advantage of the HMM and HMM-LMN based methods over the particle filter based one.

True parameters are $N(\mu_1, \sigma_1^2) = N(-1/2, 1/2)$, $N(\mu_2, \sigma_2^2) = N(1/2, 1/3)$, $F(\omega_1) = -0.25$, $F(\omega_2) = 0.25$, $G(\omega_1) = -2$, $G(\omega_2) = 2$, $H(\omega_1) = 0.1$, $H(\omega_2) = 0.5$, $K(\omega_1) = 0.5$, $K(\omega_2) = 1$. The squared error is given by

$$\varepsilon = \sqrt{\frac{1}{200}\sum_{i=1}^{200}(x_i - \hat{x}_i)^2} \qquad (4.1)$$

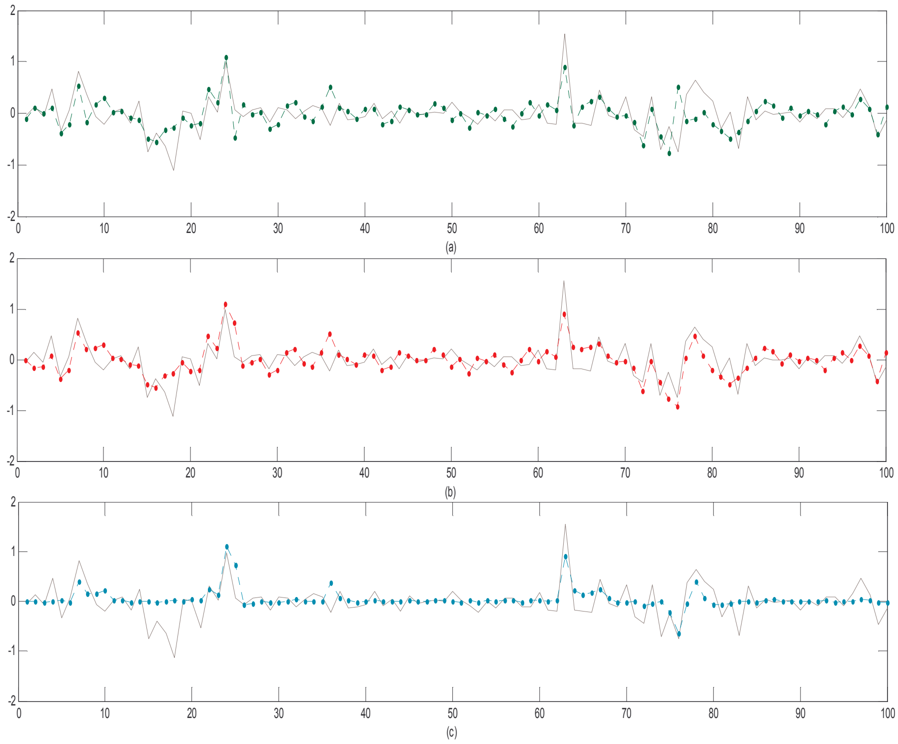

Fig. 2. Data simulated according to the model (1.1)-(1.3) with $\rho = 0.1$ and the true parameters related to Tab. 2.(continuous line) (a) : particle filter (500 particles sampled) based filtering ; (b) HMM-LMN based filtering, and (c) : HMM based filtering.

ρ	0.1	0.40	0.80
	Error ratio τ		
Particle Filter	17.0%	31.7%	37.0%
HMM-LMN	13.3%	27.3%	27.5%
HMM	29.0%	32.0%	37.3%
	Squared error ε		
Particle Filter	0.0273	0.16	0.19
HMM-LMN	0.0211	0.09	0.11
HMM	0.0301	0.12	0.13

Tab. 2. Error ratio τ of unsupervised adaptive segmentation and squared error ε (4.1) in the case of data simulated with the model (1.1)-(1.3). $\rho = 0.1$ is given for the three methods, means and variances in HMM and HMM-LMN are estimated with AICE.

According to the results presented in Table 2 and other similar results obtained, one can say that, on the whole, an HMM-LMN based method is the most efficient, while the Particle Filter based method works better than the HMM based one.

5 Conclusion

In this paper we considered the problem of optimal filtering in the conditionally Gaussian state space models with Markovian switches given by (1.1)-(1.3). We presented two original methods of approximation of the distribution of the couple (R_1^N, Y_1^N), where $R_1^N = (R_1, ..., R_N)$ is the random chain of switches and $Y_1^N = (Y_1, ..., Y_N)$ is the random chain of observations. In the first one this distribution is approximated by the classical hidden Markov model (HMM) distribution, and in the second one it is approximated by a recent Markov model hidden with a "long memory noise" model (HMM-LMN) (Lanchantin et al. 2008). In these two models the parameters can be estimated by the "adaptive" methods also proposed in the paper. Using these two approximations makes the exact filtering possible with a reasonable complexity. The two related filtering methods have then been compared to the classical Particle Filter based approximation. Different experiments showed that the HMM-LMN based methods take the upper hand over the Particle Filter based ones, while the efficiency of the latter is, roughly speaking, similar to the efficiency of the HMM based methods.

As perspective, let us mention further comparisons of our methods with some recent models based exact filtering, as the method proposed in (Pieczynski (2008)).

6 References

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

[2] Benmiloud, B. and Pieczynski, W., Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images, Traitement du Signal, 12, 5, 433-454 (1995).

[3] Cappé, O., Moulines, E., and Ryden, T., *Inference in hidden Markov models*, Springer, Series in Statistics, Springer (2005).

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

[5] Derrode, S., and Pieczynski, W., Signal and image segmentation using pairwise Markov chains, *IEEE Trans. on Signal Processing*, 52, 9, 2477-2489 (2004).

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

[7] Doucet, A., Gordon, N. J., and Krishnamurthy, V., Particle filters for state estimation of Jump Markov Linear Systems, *IEEE Trans. on Signal Processing*, 49, 3, 613-624 (2001).

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

[9] Pieczynski, W., Exact calculation of optimal filter in semi-Markov switching model, (IASC 2008), December 5-8, Yokohama, Japan (2008).

Stochastic Modeling Techniques and Data Analysis international conference (SMTDA '10), Chania, Greece, June 8-11, 2010.

[10] Ristic, B., Arulampalam, S., and Gordon, N., *Beyond the Kalman Filter - Particle filters for tracking applications*, Artech House, Boston, MA (2004).

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

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