Approximation particulaire par FFBS de la loi de lissage pour des HMM dans des espaces d'états généraux

Randal Douc, Aurélien Garivier, Eric Moulines, Jimmy Olsson

ON THE FORWARD FILTERING BACKWARD SMOOTHING PARTICLE APPROXIMATIONS OF THE SMOOTHING DISTRIBUTION IN GENERAL STATE SPACES MODELS

Randal Douc & Aurélien Garivier & Eric Moulines & Jimmy Olsson

randal.douc@it-sudparis.eu
Institut des Télécoms/Télécom SudParis 9 rue Charles Fourier, 91011 Evry-Cedex, France
aurelien.garivier@telecom-paristech.fr, eric.moulines@telecom-paristech.fr
Institut des Télécoms/Télécom ParisTech, CNRS UMR 5141 46 rue Barrault, 75634 Paris, Cédex 13, France
jimmy@maths.lth.se
Center of Mathematical Sciences Lund University, Box 118 SE-22100 Lund, SWEDEN

Mots-clés : chaînes de Markov cachées, filtrage particulaire, lissage, échantillonnage
Keywords : hidden Markov models, particle filtering, smoothing, sampling

Résumé :

Dans les espaces d'état généraux, il est souvent complexe de construire une bonne approximation de la loi de lissage d'un état, ou d'une suite d'états, conditionnellement aux observations passées, présentes et futures. L'objet de cette contribution est de fournir un cadre rigoureux pour le calcul ou l'approximation de telles distributions de lissage, et d'analyser dans une cadre commun différent schémas qui y parviennent. Après un exposé générique des différents résultats développés jusqu'ici, nous proposons de nouvelles extensions permettant d'approcher les distributions jointes de lissage dans des cadres assez généraux, et dont la complexité reste souvent linéaire en le nombre de particules.

Abstract :

A prevalent problem in general state-space models is the approximation of the smoothing distribution of a state, or a sequence of states, conditional on the observations from the past, the present, and the future. The aim of this paper is to provide a rigorous foundation for the calculation, or approximation, of such smoothed distributions, and to analyse in a common unifying framework different schemes to reach this goal. Through a cohesive and generic exposition of the scientific literature we offer several novel extensions allowing to approximate joint smoothing distribution in the most general case with a cost growing linearly with the number of particles.

Introduction :

Consider the nonlinear *state space model*, where the *state process* $\{X_t\}_{t\geq 0}$ is a Markov chain on some general state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ having initial distribution χ and transition kernel Q. The state process is hidden but partially observed through the *observations*

$\{Y_t\}_{t \geq 0}$, which are Y-valued random variables being independent conditionally on the latent state sequence $\{X_t\}_{t \geq 0}$; in addition, there exists a σ-finite measure λ on $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$, and a transition density function $x \mapsto g(x, y)$, referred to as the *likelihood*, such that $\mathbb{P}(Y_t \in A \mid X_t) = \int_A g(X_t, y) \lambda(dy)$ for all $A \in \mathcal{B}(\mathsf{Y})$. The kernel Q and the likelihood function $x \mapsto g(x, y)$ are assumed to be known. We shall consider the case in which the observations have arbitrary but fixed values $y_{0:T}$, where we have, for any quantities $a_m, \ldots, a_n$, introduced the vector notation $a_{m:n} \stackrel{\text{def}}{=} (a_m, \ldots, a_n)$. We shall also denote $[a_{m:n}, b_{p:q}] \stackrel{\text{def}}{=} (a_m, \ldots, a_n, b_p, \ldots, b_q)$. Nonlinear state space models of this kind arise frequently in many scientific and engineering disciplines such as target tracking, finance, environmental sciences, to list but a few.

Statistical inference in general state space models involves computing the *posterior distribution* of a batch of state variables $X_{s:s'}$ conditioned on a batch of observations $Y_{t:T}$, which we denote by $\phi_{s:s'|t:T}$ (here the dependence on the observations $Y_{t:T}$ is implicit). The posterior distribution can be computed in closed form only in very specific cases, principally, when the state space model is linear and Gaussian or when the state space X is a finite set. In the vast majority of cases, nonlinearity or non-Gaussianity render analytic solutions intractable [1, 9, 13, 2].

These limitations have stimulated the interest in alternative strategies being able to handle more general state and measurement equations without putting strong a priori constraints on the behaviour of the posterior distributions. Among these, *Sequential Monte Carlo* (SMC) *methods* play a central role. SMC methods refer to a class of algorithms designed for approximating a *sequence of probability distributions* over a *sequence of probability spaces* by updating recursively in time a set of random *particles* with associated nonnegative weights. These algorithms are all based on *selection* and *mutation* operations and can thus be seen as combinations of the *sequential importance sampling* and *sampling importance resampling* methods introduced in [7] and [14], respectively. SMC methods have emerged as a key tool for approximating state posterior distributions in general state space models; see, for instance, [11, 12, 4, 13] and the references therein.

The recursive formulas generating the filter and joint smoothing distributions $\phi_{T|0:T}$ and $\phi_{0:T|0:T}$, respectively, are closely related and follow the same dynamics. Thus, using the basic *filtering* version of the particle filter actually provides as a by-product an approximation of the joint smoothing distribution in the sense that the particle *paths* and their associated weights can be considered as a weighted sample approximating $\phi_{0:T|0:T}$. From these joint draws one may readily obtain fixed lag or fixed interval smoothed samples by simply extracting the required components from the sampled particle paths and retaining the same weights.

This appealingly simple scheme can be used successfully for estimating the smoothing joint smoothing distribution for small values of T or any marginal smoothing distribution $\phi_{s|0:T}$, with $s \leq T$, when s and T are close; however, when T is large or when s and T are remote, the associated particle approximations are depleted and inaccurate. This observation has triggered a wealth of publications proposing particle methods to approximate the fixed-lag smoothing distribution. In this article, we consider the so-called forward-filtering backward smoothing (FFBS) algorithm, which shares some similarities with the so-called forward-backward algorithm for discrete state-space HMM. This algorithm consists in reweighting, in a backward pass, the weighted sample approximating the filtering distribution (see [10], [8], [5] and [6]); this algorithm is very challenging to analyze and, up to now, only a consistency result is available in [6] (the proof of this result is plagued by a difficult to correct mistake). As we will see below, the analysis of this algorithm is very challenging, because the FFBS smoothing estimate at a given time instant explicitly depends upon all the particles and weights drawn before and after this time instant. It is therefore impossible to analyze directly the convergence of this approximation using the standard techniques developed to study the interacting particle approximations of the Feyman-Kac flows (see the monograph [3]).

The forward filtering-backward smoothing approach requires $O(NT)$ operations to sample one path approximately distributed according to $\phi_{\chi,0:T|0:T}$ where N is the number of particles and T the time horizon. The FFBS algorithm requires $O(N^2T)$ operations to approximate the same distribution. However, in many situations, this computational cost can considerably reduced, making the algorithm manageable in many interesting scenarios (such as the smoothing of very large discrete state-space Markov chains with sparse transition matrix).

Our contribution is the following. First, the forward filtering backward smoothing algorithm for estimating the fixed-interval smoothing distribution is introduced. An exponential deviation inequality follows. A joint Central Limit Theorem (CLT) for the weighted samples approximating the filtering and the smoothing distribution is obtained. Finally, time-uniform exponential bounds and CLT are then obtained, under some additional mixing conditions on the kernel Q.

Bibliographie

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, 1979.

[2] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, 2005.

[3] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications.* Springer, 2004.

[4] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice.* Springer, New York, 2001.

[5] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208, 2000.

[6] S. J. Godsill, A. Doucet, and M. West. Monte carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.

[7] J. Handschin and D. Mayne. Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering. In *Int. J. Control*, volume 9, pages 547–559, 1969.

[8] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7:175–193, 1998.

[9] T. Kailath, A. Sayed, and B. Hassibi. *Linear Estimation.* Prentice-Hall, 2000.

[10] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.

[11] J. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. *J. Am. Statist. Assoc.*, 93(443):1032–1044, 1998.

[12] J.S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, New York, 2001.

[13] B. Ristic, M. Arulampalam, and A. Gordon. *Beyond Kalman Filters: Particle Filters for Target Tracking.* Artech House, 2004.

[14] D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong). *J. Am. Statist. Assoc.*, 82:543–546, 1987.