A data augmentation scheme embedding a sequential Monte Carlo method for Bayesian parameter inference in state space models

Irene Votsi [1] & Paul-Henry Cournède [1]

[1] *Ecole CentraleSupélec, Laboratory of Mathematics in Interaction with Computer Science, Grande Voie des Vignes, 92290 Châtenay-Malabry, France*
eirini.votsi@ecp.fr, paul-henry.cournede@ecp.fr

Résumé. Les modèles à espace d'état (SSMs) sont utilisés dans de nombreux domaines scientifiques et techniques pour représenter des séries temporelles et/ou des systèmes dynamiques. Dans ce travail, on revisite l'algorithme d'augmentation de données de Tanner and Wong (1987) pour l'estimation bayésienne de paramètres pour les SSMs. On propose l'emploi de méthodes de Monte-Carlo séquentielles et adaptatives pour améliorer les performances de l'algorithme original. La nouvelle approche est testée sur un example académique dans l'optique de l 'étendre à des modèles plus riches (non-linéaires/non-Gaussiens) issus du domaine de la modélisation de la croissance des plantes.

Mots-clés. Augmentation de données, modèles à espace d'état, méthodes de Monte Carlo séquentielles, méthodes de Monte-Carlo

Abstract. State space models (SSMs) are successfully used in many areas of science to describe time series and/or dynamical systems. In this work, we revisit the data augmentation algorithm introduced by Tanner and Wong (1987) for bayesian parameter estimation in SSMs. We propose to employ sequential Monte-Carlo and adaptive Monte-Carlo Markov chain methods to improve the performance of the algorithm. We provide a first numerical example that allows us to evaluate the convergence of the posterior estimate to the true posterior distribution. Our objective is to evaluate the performance of the proposed method to nonlinear/non-Gaussian models, with a special interest to plant growth models.

Keywords. Data augmentation, state-space models, sequential Monte Carlo methods, Monte-Carlo methods

1 Introduction

State space models are one of the most successful statistical modeling ideas that have came up in the last forty years: the use of latent states makes the model generic enough to handle a variety of complicated real-world problems, whereas the relatively simple prior dependence structure (the "Markov" property) still allows for the use of efficient computational procedures. A state space model consists of two equations: the transition equation and the observation (or measurement) equation which are respectively given by

$$x_{t+1} = f_t(x_t, \eta_t), \tag{1}$$
$$y_t = h_t(x_t, \epsilon_t). \tag{2}$$

It is assumed that the distributions of the state variable and observations admit density functions with respect to appropriate dominating measures dx_t and dy_t, respectively. These densities $p(x_{t+1}|x_t, \theta)$ and $p(y_t|x_t, \theta)$ corresponding to (1) and (2) respectively are called the observation (or measurement) and transition densities. The densities will typically depend upon a vector of unknown parameters $\theta \in \Theta$ that need to be estimated from the observations. In a Bayesian context θ is treated as a random variable with prior $p(\theta)$.

The state variables x_t and observations y_t may be continuous-valued, discrete-valued, or a combination of the two. The functions η_t and ϵ_t are possibly nonlinear but of known form. Time is denoted by the subscripts t. Given a batch of observations $y_{1:T}$, we wish to learn θ and infer the latent states $x_{1:T}$. For a presentation of SSMs with different focuses and applications, see, e.g., Cappé et al. (2005).

Here we revisit the *data augmentation* or *imputation-posterior algorithm* (IP) introduced by Tanner and Wong (1987) for parameter learning in problems with missing data. This algorithm provides an appealing, iterative scheme to estimate the posterior distribution of the parameters of interest, $p(\theta|y_{1:T})$. We present an academic example that allows us to evaluate the convergence of the posterior estimate to the true posterior distribution. For our purposes we employ Markov chain Monte Carlo (MCMC) and sequential Monte-Carlo (SMC) methods. Our objective is to evaluate the performance of this new approach to nonlinear/non-Gaussian models, with a special focus on plant growth models (Trevezas & Cournède, 2013), to improve its convergence rate and eventually to compare it with the state-of-art methods in terms of computational cost and predictability.

2 Methods

The term *data augmentation* refers to methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables. In a Bayesian context, Tanner and Wong (1987) introduced the idea of data augmentation to learn model parameters, θ, given a sequence of observations $y_{1:T}$, for missing data problems. The authors observed that in incomplete data problems the posterior distribution of θ is the fixed point solution of an integral equation and introduced the IP algorithm (Algorithm 1). If we apply the fixed point principle to an SSM, we obtain that the posterior $p(\theta|y_{1:T})$ of the unknown parameter θ is a fixed point solution of the following integral equation:

$$p(\theta|y_{1:T}) = \int\int p(\theta|x_{1:T}, y_{1:T})p(x_{1:T}|y_{1:T}, \theta')p(\theta'|y_{1:T})dx_{1:T}d\theta'. \tag{3}$$

Given some approximation $g_i(\theta)$ of the posterior $p(\theta|y_{1:T})$ one may use Eq. (3) to improve it:

$$g_{i+1}(\theta) = \int K(\theta,\theta')g_i(\theta')d\theta', \quad K(\theta,\theta') = \int p(\theta|x_{1:T}, y_{1:T})p(x_{1:T}|y_{1:T}, \theta')dx_{1:T}.$$

for $i \leftarrow 1$ **to** K **do**
- *Imputation step* ;
- **for** $j \leftarrow 1$ **to** M **do**
  - Draw $\theta^{(j)} \sim g_i(\theta)$;
  - Draw $x_{1:T}^{(j)} \sim p(x_{1:T}|y_{1:T}, \theta)$;
- **end**
- *Posterior step* ;
- Update the current approximation of $p(\theta|y_{1:T})$, $g_i(\theta)$, by the mixture of conditional densities of θ given the augmented data patterns generated in the imputation step

$$g_{i+1}(\theta) = \frac{1}{M}\sum_{j=1}^{M} p(\theta|y_{1:T}, x_{1:T}^{(j)}).$$

end

Algorithm 1: Imputation Posterior Algorithm

In order to implement the IP algorithm, we have to discuss how to sample the parameters $\theta^{(j)}$ from the current posterior estimate $g_i(\theta)$. For this purpose we employ an MCMC algorithm, the Adaptive Metropolis Algorithm introduced by Haario et al. (2001). We further need to sample the states $x_{1:T}^{(j)}$ from the distribution $p(x_{1:T}|y_{1:T},\theta^{(j)})$. To achieve this goal, we propose the use of particle filtering or sequential Monte-Carlo methods (Doucet et al. 2001).

For a given value of the parameter vector θ, N particles are drawn from a proposal distribution $q_t(\cdot)$ and sequentially propagated over time. The particles that best fit the data $y_{1:T}$ are given more weight through resampling. Finally, under mild conditions, we obtain unbiased estimates of $x_{1:T}^{(j)}$ from the target distribution $p(x_{1:T}|y_{1:T},\theta)$. Different particle filtering algorithms are obtained by different choices of the proposal distribution $q_t(\cdot)$. In our context, q_0 is the initial law of the Markov chain. The proposal density $q_t(\cdot)$ $(t \geq 1)$ is the transition density $p(\cdot|x_t,\theta)$ as proposed by Gordon et al. (1993) in their so-called SIR filter (Algorithm 2).

Initialization ;

At $t = 0$ draw $x_0^{(i)} \sim q_0(x_0)$ and set $w_0^{(i)} = \frac{p(x_0^{(i)})}{q_0(x_0^{(i)})}$;

for $t \leftarrow 1$ **to** T **do**

Iteration i ;

Draw $x_t^{(i)} \sim q_t(x_t|x_{t-1}^{(i)})$ and compute the importance weights by

$$w_t^{(i)} \sim w_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)},\theta)p(x_t^{(i)}|x_{t-1}^{(i)},\theta)}{q_t(x_t^{(i)}|x_{t-1}^{(i)})}$$

Normalize the importance weights $\widehat{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N}\widehat{w}_t^{(j)}}$;

Resample N particles with probabilities $\{\widehat{w}_t^{(i)}\}_{i=1}^{N}$ and set $w_t^{(i)} = \frac{1}{N}$.

end

Algorithm 2: SIR Filter (each iteration is for $i = 1, \ldots, N$)

3 A numerical example

Consider the linear Gaussian SSM taken from the summer school presentation of (Doucet, 2012) :

$$x_{t+1} = \theta x_t + \sigma_V V_t, \quad V_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{4}$$

$$y_t = x_t + \sigma_W W_t, \quad W_t \overset{iid}{\sim} \mathcal{N}(0,1). \tag{5}$$

θ is treated as an unknown parameter with uniform prior in $(-1, 1)$. In this case

$$p(\theta|x_{0:T}, y_{1:T}) \propto N(\theta; m_t, \sigma_t^2) 1_{(-1,1)}(\theta),$$

where

$$\sigma_t^2 = \Big(\sum_{k=2}^{t} x_{k-1}^2\Big)^{-1}, \quad m_t = \Big(\sum_{k=2}^{t} x_{k-1}^2\Big)^{-1}\Big(\sum_{k=2}^{t} x_{k-1}x_k\Big).$$

The true posterior distribution is calculated by using the classical Bayes' formula

$$p(\theta|y_{1:T}) \quad = \quad \frac{p(\theta)p(y_{1:T}|\theta)}{\int_\Theta p(\theta)p(y_{1:T}|\theta)d\theta},$$

where the likelihood $p(y_{1:T}|\theta)$ is computed by means of Kalman filter. A batch of $T = 100$ observations are simulated for true parameter value equal to $\theta = -0.1$ with $\sigma_V = 1$ and $\sigma_W = 0.1$. The total number of adaptive MCMC iterations performed at each IP iteration is equal to 3000 with a burnin period of 100. A number of $N = 2000$ particles are propagated at each run of the SIR filter over time. Figure 1 presents the true posterior distribution of parameter θ and its estimate resulting from our approach. We further present the estimates that result from the state-of-the-art algorithms PMMH (Andrieu et al. 2010) and PGAS (Lindsten et al. 2014), with $N = 5000$ and $N = 1000$ particles, respectively.

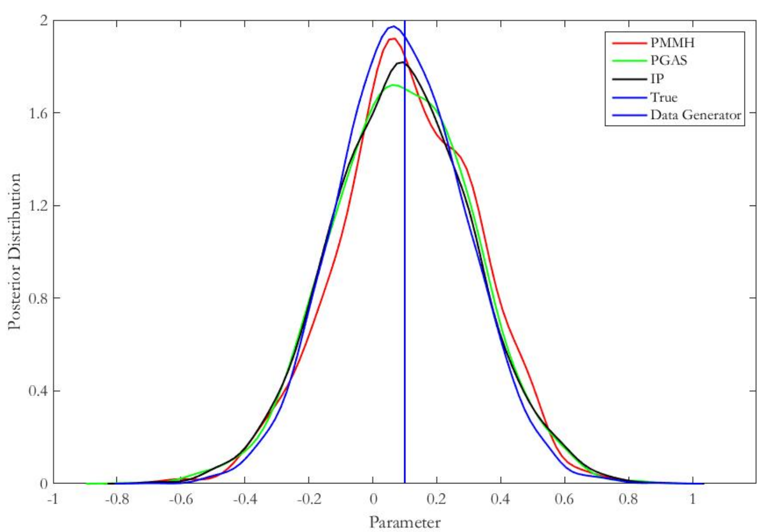

Figure 1: Posterior distribution for the parameter of model (4)–(5) for $K = 3000$ IP iterations ($M = 10$ terms for the approximation of the current posterior).

4 Conclusion and perspectives

In this work we present a new approach to perform bayesian parameter estimation for state space models. This approach is based on data augmentation techniques and sequential Monte-Carlo methods and is evaluated numerically by an academic example. Further research includes the application of the method to non-linear/ non-Gaussian models, in particular plant growth models and its development to make state inference.

Acknowledgements

We would like to thank F.Lindsten (Senior Lecturer, Uppsala University) and M.Hamdaoui (Associate Professor, Université de Lorraine) for fruitful discussions.

Bibliography

[1] Tanner, M.A. & Wong, W.H. (1987), The calculation of posterior distribution by data augmentation, *Journal of the American Statistical Association*, 82(398), 528-540.
[2] Cappé, O., Moulines, E. & Rydén, T. (2005), *Inference in hidden Markov models*, Springer Press, New-York.
[3] Haario, H., Saksman, E. & Tamminen, J. (2001), An adaptive Metropolis algorithm, *Bernoulli*, 7, 223-242.
[4] Doucet, A., de Freitas, N. & Gordon, N. (2001), *Sequential Monte Carlo methods in practice. Information Science and Statistics*, Springer-Verlag, New-York.
[5] Trevezas, S. & Cournède, P-H. (2013), A sequential Monte Carlo approach for MLE in a plant growth model, *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2), 250-270.
[6] Gordon, N.J, Salmond, D.J. & Smith, A.F.M. (1993), A novel approach to nonlinear and non-Gaussian Bayesian state estimation, *IEE Proceedings. Part F: Radar and Sonar Navigation*, 140(2), 107-113.
[7] Andrieu, C., Doucet, A. & Holenstein, R. (2010), Particle Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society (B)*, 72(3), 269-342.
[8] Lindsten, F., Jordan, M.I. & Schon, T.B. (2014), Particle Gibbs with ancestor sampling, *Journal of Machine Learning Research*, 15, 2145-2184.
[9] Arnaud Doucet, Sequential Monte Carlo methods for Bayesian computation, 2012 Machine Learning Summer School in Kyoto, Japan (27/08/2012–7/09/2012).