MCMC for state space models

August 23, 2012

We consider the following class of Gaussian linear state space models: For $t = 1, \ldots, T$

$$\begin{aligned} Y_t &= F_t\theta_t + \epsilon_t, \ \ \epsilon_t \sim N_m(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + \eta_t, \ \ \eta_t \sim N_p(0, W_t), \end{aligned} \tag{1}$$

together with

$$\theta_0 \sim N_p(m_0, C_0),$$

where F_t and G_t are known matrices of dimension $m \times p$ and $p \times p$ respectively and $\{\epsilon_t\}_{t\geq 1}$ and $\{\eta_t\}_{t\geq 1}$ are independent. In this article we propose some efficient MCMC algorithms that can be used to explore the posterior densities associated with the above dynamic linear models.

We first consider one simple example of the above class of models known as the steady state models given by

$$\begin{aligned} Y_t &= \theta_t + \epsilon_t, \epsilon_t \sim N(0, V), \\ \theta_t &= \theta_{t-1} + \eta_t, \eta_t \sim N(0, W), \end{aligned} \tag{2}$$

for $t = 1, \ldots, T$ with

$$\theta_0 \sim N(m_0, C_0).$$

Obviously, the above is a state space model with $m = 1 = p$, $F_t = G_t = 1$. For a Bayesian analysis of the above model, we use the following prior distributions on V and W; $V \sim$ IG(α_1, β_1) and $W \sim$ IG(α_2, β_2) for known $\alpha_i, \beta_i, i = 1, 2$. (Here $X \sim$ IG (α, β) means that the pdf of X is $p(x) = \beta^\alpha/\Gamma(\alpha)x^{-(\alpha+1)}\exp(-\beta/x)$ for $x > 0$.) Let the posterior density of the parameters (V, W) be denoted by $\pi(V, W|y)$. There are Markov chain Monte Carlo (MCMC) algorithms that can be used to explore the posterior density $\pi(V, W|y)$. Here we present a well-known data augmentation (DA) algorithm. Note that

$$\pi(V, W|y) = \int \pi((V, W), \theta_{0:T}|y) d\theta_{0:T},$$

where $\theta_{0:T} = (\theta_0, \theta_1, \ldots, \theta_T)$ and $\pi((V,W), \theta_{0:T}|y)$ is the joint posterior density of the parameters (V, W) and the states $\theta_{0:T}$ given by

$$\begin{aligned}\pi((V,W),\theta_{0:T}|y) \quad \propto \quad & V^{-T/2}\exp\{-\sum_{t=1}^{T}(y_t-\theta_t)^2/2V\}W^{-T/2}\exp\{-\sum_{t=1}^{T}(\theta_t-\theta_{t-1})^2/2W\} \\ & \exp\{-(\theta_0-m_0)^2/2C_0\}V^{-(\alpha_1+1)}\exp(-\beta_1/V)W^{-(\alpha_2+1)}\exp(-\beta_2/W).\end{aligned}$$

It is easy to see that conditional on $\theta_{0:T}, y$, V and W are conditionally independent, that is,

$$\pi((V,W)|\theta_{0:T},y) = \pi(V|\theta_{0:T},y)\pi(W|\theta_{0:T},y),$$

where $\pi(V|\theta_{0:T}, y)$ is the density of IG$(\alpha_1 + T/2, \beta_1 + \sum_{t=1}^{T}(y_t - \theta_t)^2/2)$, and $W|\theta_{0:T}, y$ follows IG$(\alpha_2 + T/2, \beta_2 + \sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2)$. The conditional density $\pi(\theta_{0:T}|(V, W), y)$ can be written as

$$\pi(\theta_{0:T}|(V,W),y) = \pi(\theta_T|(V,W),y)\pi(\theta_{T-1}|\theta_T,(V,W),y)\ldots\pi(\theta_0|\theta_1,\ldots,\theta_T,(V,W),y).$$

The so-called forward filtering backward sampling (FFBS) algorithm for sampling from $\pi(\theta_{0:T}|(V, W), y)$ is to use the above identity where Kalman filter and Kalman smoother can be used to make draws from each of the densities in the right-hand side. From the above discussion we obtain the following DA algorithm, which we call Algorithm I, that can be used to move from the current state $(V^{(i)}, W^{(i)})$ to the new value $(V^{(i+1)}, W^{(i+1)})$:

Algorithm I

1. use the FFBS algorithm to draw $\theta_{0:T}$ from the conditional density $\pi(\theta_{0:T}|(V^{(i)}, W^{(i)}), y)$
2. draw $V^{(i+1)}$ from IG$(\alpha_1 + T/2, \beta_1 + \sum_{t=1}^{T}(y_t - \theta_t)^2/2)$ and independently draw $W^{(i+1)}$ from IG$(\alpha_2 + T/2, \beta_2 + \sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2)$.

Standard Markov chain theory implies that the above DA algorithm is reversible with respect to $\pi(V, W|y)$ and hence $\pi(V, W|y)$ is its stationary density.

We now propose an MCMC algorithm based on the above DA algorithm which may have faster convergence to the stationary distribution than the DA algorithm. Consider the transformation $(\theta_{0:T}, W) \rightarrow (\gamma_{0:T}, W)$, where $\gamma_{0:T} = (\gamma_0, \gamma_1, \ldots, \gamma_T)$, $\gamma_0 = \theta_0$, $\gamma_1 = (\theta_1 - \theta_0)/W$, $\gamma_2 = (\theta_2 - \theta_1)/W$, $\ldots$, and $\gamma_T = (\theta_T - \theta_{T-1})/W$, that is, $\gamma_{0:T} = A\theta_{0:T}$, where

$$A = \frac{1}{W}\begin{pmatrix} W & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}.$$

The steady-state model in (2) can be written as

$$Y_t = \gamma_0 + W \sum_{j=1}^{t} \gamma_j + \epsilon_t, \epsilon_t \sim N(0, V), \tag{3}$$

$\gamma_t \sim N(0,1)$, for $t = 1, \ldots, T, \gamma_0 \sim N(m_0, C_0)$. The priors on V and W are as mentioned before. Since the jacobian of the transformation is W^T, the joint posterior density of (V, W) and $\gamma_{0:T}$ is given by

$$\begin{aligned} \pi((V,W),\gamma_{0:T}|y) \quad \propto \quad & V^{-T/2} \exp(-\sum_{t=1}^{T}(y_t - \gamma_0 - W\sum_{j=1}^{t}\gamma_j)^2/2V) W^{T/2} \exp(-\sum_{t=1}^{T}\gamma_t^2/2) \\ & \exp(-(\gamma_0 - m_0)^2/2C_0) V^{-(\alpha_1+1)} \exp(-\beta_1/V) W^{-(\alpha_2+1)} \exp(-\beta_2/W) \end{aligned} \tag{4}$$

Note that V and W are no more conditionally independent. The conditional density of V given $W, \gamma_{0:T}, y$ is IG $(\alpha_1 + T/2, \beta_1 + 0.5\sum_{t=1}^{T}(y_t - \gamma_0 - W\sum_{j=1}^{t}\gamma_j)^2)$, whereas the conditional density of W given $V, \gamma_{0:T}, y$ is not a standard density. The conditional pdf of W given $V, \gamma_{0:T}, y$ is

$$\pi(W|V,\gamma_{0:T},y) \propto \exp\Big[\{-\sum_{t=1}^{T}\big(\sum_{j=1}^{t}\gamma_j\big)^2 W^2 + 2\sum_{t=1}^{T}(y_t - \gamma_0)\big(\sum_{j=1}^{t}\gamma_j\big)W\}/2V\Big] W^{T/2-(\alpha_2+1)} \exp(-\beta_2/W). \tag{5}$$

So

$$\log \pi(W|V, \gamma_{0:T}, y) = -aW^2 + bW + (T/2 - \alpha_2 - 1)\log W - \beta_2/W + C,$$

where C is a constant, and $a = \sum_{t=1}^{T}\big(\sum_{j=1}^{t}\gamma_j\big)^2/2V$, and $b = \sum_{t=1}^{T}(y_t - \gamma_0)\big(\sum_{j=1}^{t}\gamma_j\big)/V$. Let

$$g(W) = -aW^2 + bW + (T/2 - \alpha_2 - 1)\log W - \beta_2/W.$$

Since $g''(W) = -2a - (T/2 - \alpha_2 - 1)/W^2 - 2\beta_2/W^3 < 0$, the density $\pi(W|V, \gamma_{0:T}, y)$ is log concave at least when $T/2 > \alpha_2 + 1$. So we can use Gilks and Wild's (1992) adaptive rejection sampler to sample from the conditional density $\pi(W|V, \gamma_{0:T}, y)$. We can construct a DA algorithm using the joint posterior density $\pi((V, W), \gamma_{0:T}|y)$ in (4). But, we have something else in mind. Following Yu and Meng (2011), we consider the following two algorithms. Suppose the current state of the chain is $(V^{(i)}, W^{(i)})$.

Algorithm II

1. use the FFBS algorithm, as in the step I of Algorithm I, to draw $\theta_{0:T}$ from the conditional density $\pi(\theta_{0:T}|(V^{(i)}, W^{(i)}), y)$
2. draw $V' \sim$ IG$(\alpha_1 + T/2, \beta_1 + \sum_{t=1}^{T}(y_t - \theta_t)^2/2)$ and independently draw $W' \sim$ IG$(\alpha_2 + T/2, \beta_2 + \sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2)$

3. draw $\gamma_{0:T}$ from $\pi(\gamma_{0:T}|(V', W'), y)$ by first drawing $\theta'_{0:T} \sim \pi(\theta_{0:T}|(V^{(i)}, W^{(i)}), y)$ as in Step 1 and then setting $\gamma_{0:T} = A\theta'_{0:T}$.

4. draw $W^{(i+1)}$ from $\pi(W|V', \gamma_{0:T}, y)$ given in (5) and then draw $V^{(i+1)}$ from IG $(\alpha_1 + T/2, \beta_1 + 0.5\sum_{t=1}^{T}(y_t - \gamma_0 - W^{(i+1)}\sum_{j=1}^{t}\gamma_j)^2)$.

Algorithm III

1. use the FFBS algorithm, as in the step I of Algorithm I, to draw $\theta_{0:T}$ from the conditional density $\pi(\theta_{0:T}|(V^{(i)}, W^{(i)}), y)$

2. draw $V' \sim \text{IG}(\alpha_1 + T/2, \beta_1 + \sum_{t=1}^{T}(y_t - \theta_t)^2/2)$ and independently draw $W' \sim \text{IG}(\alpha_2 + T/2, \beta_2 + \sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2)$

3. draw $W^{(i+1)}$ from $\pi(W|V', \gamma_{0:T}, y)$ given in (5) where $\gamma_{0:T} = A\theta_{0:T}$ and then draw $V^{(i+1)}$ from IG $(\alpha_1 + T/2, \beta_1 + 0.5\sum_{t=1}^{T}(y_t - \gamma_0 - W^{(i+1)}\sum_{j=1}^{t}\gamma_j)^2)$.

If in the Step 3 of Algorithm III, $W^{(i+1)}$ is drawn from the marginal density $\pi(W|\gamma_{0:T}, y)$ instead of the full conditional density $\pi(W|V', \gamma_{0:T}, y)$, then Algorithm III becomes a GIS algorithm given in Yu and Meng (2011). The marginal density $\pi(W|\gamma_{0:T}, y)$ is given by

$$\pi(W|\gamma_{0:T}, y) \propto \frac{W^{T/2-\alpha_2-1}\exp(-\beta_2/W)}{(\beta_1 + 0.5\sum_{t=1}^{T}(y_t - \gamma_0 - W\sum_{j=1}^{t}\gamma_j)^2)^{T/2+\alpha_1}}.$$

It may be difficult to draw directly e.g. using a rejection sampler from the density $\pi(W|\gamma_{0:T}, y)$. In that case, we can use a Metropolis Hastings step here.

It is easy to see that $\pi(V, W|y)$ is the invariant density of the Markov chain produced by Algorithm II. Below we prove that Algorithm III has $\pi(V, W|y)$ has its invariant density.

Proof: Let $k((V', W')|(V, W))$ be the Markov transition density (Mtd) of the Markov chain associated with the Algorithm III. We need to show that

$$\int\int k((V', W')|(V, W))\pi(V, W|y)dVdW = \pi(V', W'|y).$$

The Mtd $k((V', W')|(V, W))$ is given by

$$\begin{aligned} k((V', W')|(V, W)) \quad = \quad & \int\int\int\int \pi(V'|W', \gamma_{0:T}, y)\pi(W'|V'', \gamma_{0:T}|y)\pi(\gamma_{0:T}|V'', W'', \theta_{0:T}, y) \\ & \pi((V'', W'')|\theta_{0:T}, y)\pi(\theta_{0:T}|V, W, y)d\theta_{0:T}dV''dW''d\gamma_{0:T}. \end{aligned}$$

Then

$$
\begin{aligned}
& \int\int k((V',W')|(V,W))\pi((V,W)|y)dVdW \\
= & \int\int\int \pi(V'|W',\gamma_{0:T},y)\pi(W'|V'',\gamma_{0:T}|y)\pi(\gamma_{0:T}|V'',W'',\theta_{0:T},y)\pi(V'',W'',\theta_{0:T}|y)d\theta_{0:T}d\gamma_{0:T}dV''dW'' \\
= & \int\int\int \pi(V'|W',\gamma_{0:T},y)\pi(W'|V'',\gamma_{0:T}|y)\pi(V'',\gamma_{0:T}|y)d\gamma_{0:T}dV'' \\
= & \int\int\int \pi(V'|W',\gamma_{0:T},y)\pi(W',\gamma_{0:T}|y)d\gamma_{0:T} \\
= & \pi(V',W'|y).
\end{aligned}
$$

We now present another algorithm that can be used to analyze the random walk plus noise model given in (2).

Algorithm IV

1. use the FFBS algorithm, as in the step I of Algorithm I, to draw $\theta_{0:T}^{(i+1)}$ from the conditional density $\pi(\theta_{0:T}|(V^{(i)},W^{(i)}),y)$
2. draw $V' \sim \text{IG}(\alpha_1 + T/2, \beta_1 + \sum_{t=1}^{T}(y_t - \theta_t)^2/2)$ and independently draw $W' \sim \text{IG}(\alpha_2 + T/2, \beta_2 + \sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2)$
3. draw $\theta'_{0:T}$ from $\pi(\theta_{0:T}|(V',W'),y)$ using the simulation smoother as in (Durbin and Koopman, 2002, p. 609)
4. draw $V^{(i+1)} \sim \text{IG}(\alpha_1+T/2, \beta_1+\sum_{t=1}^{T}(y_t-\theta'_t)^2/2)$ and independently draw $W^{(i+1)} \sim \text{IG}(\alpha_2+T/2, \beta_2 + \sum_{t=1}^{T}(\theta'_t - \theta_{t-1})^2/2)$

In step 3 of Algorithm IV, we can also use the simulation smoother given in Jong and Shephard (1995). We can construct a similar algorithm for the general state space model given in (1). Let $\pi(V_t)$ and $\pi(W_t)$ be the prior densities of V_t and W_t respectively. Suppose ψ be the vector of unknown parameters. The form of the conditional density $\pi(\psi|\theta_{0:T},y)$ will depend on the priors $\pi(V_t)$ and $\pi(W_t)$. Let $\psi^{(i)}$ be the current state of ψ. The following steps are used to move to the new state $\psi^{(i+1)}$.

1. use the FFBS algorithm, to draw $\theta_{0:T}^{(i+1)}$ from the conditional density $\pi(\theta_{0:T}|\psi^{(i)},y)$
2. draw $\psi' \sim \pi(\psi|\theta_{0:T},y)$
3. draw $\theta'_{0:T}$ from $\pi(\theta_{0:T}|\psi',y)$ using the simulation smoother presented in Durbin and Koopman (2002)

4. draw $\psi^{(i+1)} \sim \pi(\psi|\theta'_{0:T}, y)$

The above algorithms require more computations than the data augmentation algorithm of Frühwirth-Schnatter (1994) or the simulation smoother algorithms in Jong and Shephard (1995) and Durbin and Koopman (2002). But, it may happen that the gain in efficiency by adding the above steps to each iteration of Frühwirth-Schnatter's (1994) DA algorithm may outweigh the extra computational burden. For example, we should consider some simulation comparisons of Algorithm I and Algorithm II- IV for several cases, e.g., V known, W unknown, both V and W unknown with different values of the hyperparameters.

References

DURBIN, J. and KOOPMAN, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, **89** 603–615.

FRÜHWIRTH-SCHNATTER, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, **15** 183–202.

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41** 337–348.

JONG, P. D. and SHEPHARD, N. (1995). The simulation smoother for time series models. *Biometrika*, **82** 339–350.

YU, Y. and MENG, X.-L. (2011). To center or not to center: that is not the question - An ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, **20** 531–570.