A Variation on the Particle-Gibbs with Backward-Simulation Algorithm

Pete Bunch

October 17, 2012

1 Outline

Particle-Gibbs (PG) is a particle-MCMC algorithm for joint estimation of states and parameters in a hidden Markov model (HMM). In each iteration, a particle filter is run, from which a state trajectory is sampled. Parameters are then sampled conditional on this trajectory. The basic PG algorithm suffers from poor mixing when the number of filter particles, N, is too small for the number of observations, T, resulting in a degenerate representation of the smoothing distribution. Particle-Gibbs with backward-simulation (PG-BS) is an improved version of PG, in which backward simulation is used to sample a trajectory from the particle filter output. This improves mixing, because the trajectory is not restricted to those appearing in the particle filter output. It is, however, required to use the individual state values which appear in the particle filter. Here we propose a further extension to particle-Gibbs by using a new version of backward-simulation in which the state values used when sampling a trajectory are not restricted to those appearing in the particle filter output. This might improve mixing further.

2 Particle Filtering

3 Basic Particle-Gibbs

4 Particle-Gibbs with Backward Simulation

5 Improved Backward Simulation

6 Improved Backward Simulation for Particle-Gibbs

The improved backward simulation algorithm samples from,

$$\begin{aligned}
&\psi_\theta(x'_{1:T}, b_{1:T}|\mathbf{x}_{1:T}, \mathbf{a}_{2:T}) \\
&= \frac{w_T^{(b_T)}}{\sum_j w_T^{(j)}} \times \frac{w_{T-1}^{(b_{T-1})} p_\theta(x'_T, y_T|x_{T-1}^{(b_{T-1})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T|x_{T-1}^{(j)})} \times \prod_{t=2}^{T-1} \frac{w_{t-1}^{(b_{t-1})} p_\theta(x'_t, x'_{t+1}, y_t|x_{t-1}^{(b_{t-1})})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t|x_{t-1}^{(j)})} \times p_\theta(x'_1|x'_2, y_1).
\end{aligned} \tag{1}$$

The extended target distribution is,

$$\begin{aligned}
&\phi(x'_{1:T}, b_{1:T}, \mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta) \\
&= \frac{p(x'_{1:T}, \theta|y_{1:T})}{N^T} \\
&\quad \times \left[\prod_{i \neq b_1} M(x_1^{(i)})\right] \prod_{t=2}^{T} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})})\right] \\
&\quad \times \frac{w_{T-1}^{(a_T^{(b_T)})} p_\theta(x_T^{(b_T)}, y_T|x_{T-1}^{(a_T^{(b_T)})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T|x_{T-1}^{(j)})} \prod_{t=2}^{T-1} \frac{w_{t-1}^{(a_t^{(b_t)})} p_\theta(x_t^{(b_t)}, x'_{t+1}, y_t|x_{t-1}^{(a_t^{(b_t)})})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t|x_{t-1}^{(j)})} p_\theta(x_1^{(b_1)}|x'_2, y_1)
\end{aligned} \tag{2}$$

$$\begin{aligned}
&= \frac{p(x'_{1:T}, \theta|y_{1:T})}{N^T} \\
&\quad \times \left[\prod_{i \neq b_1} M(x_1^{(i)})\right] p_\theta(x_1^{(b_1)}|x'_2, y_1) \\
&\quad \times \prod_{t=2}^{T-1} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})})\right] \frac{w_{t-1}^{(a_t^{(b_t)})} p_\theta(x_t^{(b_t)}, x'_{t+1}, y_t|x_{t-1}^{(a_t^{(b_t)})})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t|x_{t-1}^{(j)})} \\
&\quad \times \left[\prod_{i \neq b_T} \frac{w_{T-1}^{(a_T^{(i)})} \nu_{T-1}^{(a_T^{(i)})}}{\sum_j w_{T-1}^{(j)} \nu_{T-1}^{(j)}} M(x_T^{(i)}|x_{T-1}^{(a_T^{(i)})})\right] \frac{w_{T-1}^{(a_T^{(b_T)})} p_\theta(x_T^{(b_T)}, y_T|x_{T-1}^{(a_T^{(b_T)})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T|x_{T-1}^{(j)})}.
\end{aligned} \tag{3}$$

To use this in a Gibbs sampler, we need to know the following things:

- That the marginal distribution $\phi(x'_{1:T}, b_{1:T}, \theta)$ is the desired posterior.
- That the conditional $\phi(x'_{1:T}, b_{1:T}|\mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta)$ is equal to $\psi_\theta(x'_{1:T}, b_{1:T}|\mathbf{x}_{1:T}, \mathbf{a}_{2:T})$.
- The conditional, $\phi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}|x'_{1:T}, b_{1:T}, \theta)$ and how to sample from it using a conditional particle filter.

These are all very similar to the standard PG-BS.

6.1 Marginal

Here we show that the extended target distribution admits the posterior as a marginal

$$\begin{aligned}
&\phi(x'_{1:T}, b_{1:T}, \theta) \\
&= \sum_{\mathbf{a}_{2:T}} \int \phi(x'_{1:T}, b_{1:T}, \mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta) d\mathbf{x}_{1:T} \\
&= \frac{p(x'_{1:T}, \theta|y_{1:T})}{N^T} \qquad (4)
\end{aligned}$$

I can't be bothered to write this all out, but its pretty obvious. We do the integrals and summations alternately backwards through time. Each step eliminates a term from the time-product in the extended distribution expression. The N^{-T} term is a uniform distribution over $b_{1:T}$.

6.2 Trajectory Sampling Conditional

Here we show that the improved backward-simulation algorithm samples from the conditional distribution, $\phi(x'_{1:T}, b_{1:T}|\mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta)$. This requires a repeated application of the following identity, which follows from the weight definition,

$$p_\theta(y_t|x_t^{(i)}) p_\theta(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}) = w_t^{(i)} M(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}) \nu_{t-1}^{(a_t^{(i)})}. \qquad (5)$$

Here goes...

$$\begin{aligned}
&\phi(x'_{1:T}, b_{1:T}|\mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta) \\
&\propto \phi(x'_{1:T}, b_{1:T}, \mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta) \\
&= \frac{p(x'_{1:T}, \theta|y_{1:T})}{N^T} \\
&\quad \times \left[\prod_{i \neq b_1} M(x_1^{(i)}) \right] p_\theta(x_1^{(b_1)}|x'_2, y_1) \\
&\quad \times \prod_{t=2}^{T-1} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)}|x_{t-1}^{(a_t^{(i)})}) \right] \frac{w_{t-1}^{(a_t^{(b_t)})} p_\theta(x_t^{(b_t)}, x'_{t+1}, y_t|x_{t-1}^{(a_t^{(b_t)})})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t|x_{t-1}^{(j)})} \\
&\quad \times \left[\prod_{i \neq b_T} \frac{w_{T-1}^{(a_T^{(i)})} \nu_{T-1}^{(a_T^{(i)})}}{\sum_j w_{T-1}^{(j)} \nu_{T-1}^{(j)}} M(x_T^{(i)}|x_{T-1}^{(a_T^{(i)})}) \right] \frac{w_{T-1}^{(a_T^{(b_T)})} p_\theta(x_T^{(b_T)}, y_T|x_{T-1}^{(a_T^{(b_T)})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T|x_{T-1}^{(j)})}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{p(x'_{1:T}, \theta | y_{1:T})}{N^T} \\
&\quad \times \left[\prod_{i \neq b_1} M(x_1^{(i)}) \right] \frac{p_\theta(x'_2 | x_1^{(b_1)}) w_1^{(b_1)} M(x_1^{(b_1)})}{p_\theta(x'_2, y_1)} \\
&\quad \times \prod_{t=2}^{T-1} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)} | x_{t-1}^{(a_t^{(i)})}) \right] \frac{w_{t-1}^{(a_t^{(b_t)})} p_\theta(x'_{t+1} | x_t^{(b_t)}) w_t^{(b_t)} M(x_t^{(b_t)} | x_{t-1}^{(a_t^{(b_t)})}) \nu_{t-1}^{(a_t^{(b_t)})}}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t | x_{t-1}^{(j)})} \\
&\quad \times \left[\prod_{i \neq b_T} \frac{w_{T-1}^{(a_T^{(i)})} \nu_{T-1}^{(a_T^{(i)})}}{\sum_j w_{T-1}^{(j)} \nu_{T-1}^{(j)}} M(x_T^{(i)} | x_{T-1}^{(a_T^{(i)})}) \right] \frac{w_{T-1}^{(a_T^{(b_T)})} w_T^{(b_T)} M(x_T^{(b_T)} | x_{T-1}^{(a_T^{(b_T)})}) \nu_{T-1}^{(a_T^{(b_T)})}}{\sum_j w_{T-1}^{(j)} p_\theta(y_T | x_{T-1}^{(j)})} \\
&\propto \frac{p(x'_{1:T}, \theta | y_{1:T})}{N^T} \\
&\quad \times \left[\prod_{i} M(x_1^{(i)}) \right] \frac{p_\theta(x'_2 | x_1^{(b_1)}) w_1^{(b_1)}}{p_\theta(x'_2, y_1)} \\
&\quad \times \prod_{t=2}^{T-1} \left[\prod_{i} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)} | x_{t-1}^{(a_t^{(i)})}) \right] \frac{p_\theta(x'_{t+1} | x_t^{(b_t)}) w_t^{(b_t)}}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t | x_{t-1}^{(j)})} \\
&\quad \times \left[\prod_{i} \frac{w_{T-1}^{(a_T^{(i)})} \nu_{T-1}^{(a_T^{(i)})}}{\sum_j w_{T-1}^{(j)} \nu_{T-1}^{(j)}} M(x_T^{(i)} | x_{T-1}^{(a_T^{(i)})}) \right] \frac{w_T^{(b_T)}}{\sum_j w_{T-1}^{(j)} p_\theta(y_T | x_{T-1}^{(j)})} \\
&\propto \frac{1}{N^T} \psi_\theta(\mathbf{x}_{1:T}, \mathbf{a}_{2:T}) \\
&\quad \times \frac{w_1^{(b_1)} p_\theta(x'_2 | x_1^{(b_1)}) p_\theta(y_1 | x'_1) p_\theta(x'_1)}{p_\theta(x'_2, y_1)} \\
&\quad \times \prod_{t=2}^{T-1} \frac{w_t^{(b_t)} p_\theta(x'_{t+1} | x_t^{(b_t)}) p_\theta(y_t | x'_t) p_\theta(x'_t | x'_{t-1})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t | x_{t-1}^{(j)})} \\
&\quad \times \frac{w_T^{(b_T)} p_\theta(y_T | x'_T) p_\theta(x'_T | x'_{T-1})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T | x_{T-1}^{(j)})} \\
&\propto \frac{p_\theta(x'_2 | x'_1) p_\theta(y_1 | x'_1) p_\theta(x'_1)}{p_\theta(x'_2, y_1)} \\
&\quad \times \prod_{t=2}^{T-1} \frac{w_{t-1}^{(b_{t-1})} p_\theta(x'_t | x_{t-1}^{(b_t)}) p_\theta(y_t | x'_t) p_\theta(x'_{t+1} | x'_t)}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t | x_{t-1}^{(j)})} \\
&\quad \times \frac{w_T^{(b_T)}}{\sum_j w_T^{(j)}} \times \frac{w_{T-1}^{(b_{T-1})} p_\theta(y_T | x'_T) p_\theta(x'_T | x_{T-1}^{(b_{T-1})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T | x_{T-1}^{(j)})} \\
&= \psi_\theta(x'_{1:T}, b_{1:T} | \mathbf{x}_{1:T}, \mathbf{a}_{2:T}). \qquad (6)
\end{aligned}
$$

6.3 Conditional Particle Filter

The other conditional from we need to sample is,

$$
\begin{aligned}
&\phi(\mathbf{x}_{1:T}, \mathbf{a}_{2:T} | x'_{1:T}, b_{1:T}, \theta) \\
&= \frac{\phi(x'_{1:T}, b_{1:T}, \mathbf{x}_{1:T}, \mathbf{a}_{2:T}, \theta)}{\phi(x'_{1:T}, b_{1:T}, \theta)} \\
&= \left[\prod_{i \neq b_1} M(x_1^{(i)})\right] p_\theta(x_1^{(b_1)} | x'_2, y_1) \\
&\quad \times \prod_{t=2}^{T-1} \left[\prod_{i \neq b_t} \frac{w_{t-1}^{(a_t^{(i)})} \nu_{t-1}^{(a_t^{(i)})}}{\sum_j w_{t-1}^{(j)} \nu_{t-1}^{(j)}} M(x_t^{(i)} | x_{t-1}^{(a_t^{(i)})})\right] \frac{w_{t-1}^{(a_t^{(b_t)})} p_\theta(x_t^{(b_t)}, x'_{t+1}, y_t | x_{t-1}^{(a_t^{(b_t)})})}{\sum_j w_{t-1}^{(j)} p_\theta(x'_{t+1}, y_t | x_{t-1}^{(j)})} \\
&\quad \times \left[\prod_{i \neq b_T} \frac{w_{T-1}^{(a_T^{(i)})} \nu_{T-1}^{(a_T^{(i)})}}{\sum_j w_{T-1}^{(j)} \nu_{T-1}^{(j)}} M(x_T^{(i)} | x_{T-1}^{(a_T^{(i)})})\right] \frac{w_{T-1}^{(a_T^{(b_T)})} p_\theta(x_T^{(b_T)}, y_T | x_{T-1}^{(a_T^{(b_T)})})}{\sum_j w_{T-1}^{(j)} p_\theta(y_T | x_{T-1}^{(j)})}.
\end{aligned} \tag{7}
$$

This is a (new form of) conditional particle filter. The intractable distributions can be sampled with MH in the same manner as the improved backward simulation algorithm.