Bayesian Computations Using the Sequential MCMC Filter

Nicholas K. K. Chia (1st author) Dominic S. Lee
Signal Processing Lab, DSO National Laboratories
20 Science Park Drive, Singapore 118230
ckangkia@dso.org.sg ldominic@dso.org.sg

Abstract: In Bayesian computations, we are interested in the posterior distribution $p(\Theta_k \mid X_k)$ where $X_k \equiv \{x_1, x_2, \Lambda\ , x_k\}$ is the set of measurements up to time-step k and $\Theta_k \equiv \{y, q_1, q_2, \Lambda\ , q_k\}$ is the cumulative set of static parameters y and time-evolving states q. Using finite samples to represent the posterior distributions of interest, which evolves over time, the main challenge is to ensure that the filter remains stable. In this paper, we present the sequential MCMC filter, which combines both *Sequential Importance Sampling* (SIS) and *Markov Chain Monte Carlo* (MCMC) to achieve computational efficiency and stability.

1. Motivations for the Sequential MCMC filter

We are given a series of measurement data $X_N \equiv \{x_1, x_2, \Lambda\ , x_N\}$, which can be adequately modelled using $\Theta_N \equiv \{y, q_1, q_2, \Lambda\ , q_N\}$. Here, y is the time-invariant component or *static parameter(s)* and q_j is the time-evolving component or *state*. The complete probabilistic solution at each time-step k is contained in the joint posterior distribution given by $p(\Theta_k \mid X_k)$. Using Bayes' rule, we can express the posterior distribution at time-step k, $p(\Theta_k \mid X_k)$, in terms of the distribution at the previous time-step k -1, as follows:

$$p(\Theta_k | X_k) \propto p(x_k | \Theta_k, X_{k-1})\, p(\Theta_k | X_{k-1}) = p(x_k | \Theta_k, X_{k-1})\, p(q_k | \Theta_{k-1}, X_{k-1})\, p(\Theta_{k-1} | X_{k-1}). \tag{1}$$

We start with a weighted sample of size n, $\Theta_{k-1,1}, \ldots, \Theta_{k-1,n}$, with weights $w_{k-1,1}, \ldots, w_{k-1,n}$, representing $p(\Theta_{k-1} | X_{k-1})$. We denote such a weighted sample by $(\Theta_{k-1,1}, w_{k-1,1}), \ldots, (\Theta_{k-1,n}, w_{k-1,n})$. From equation (1), one way to get a weighted sample, $(\Theta_{k,1}, w_{k,1}), \ldots, (\Theta_{k,n}, w_{k,n})$, that represents $p(\Theta_k | X_k)$ is to generate $q_{k,j}$ from $p(q_k | \Theta_{k-1,j}, X_{k-1})$, augment it to $\Theta_{k-1,j}$ to form $\Theta_{k,j}$, i.e.

$$q_{k,j} \sim p(q_k | \Theta_{k-1,j}, X_{k-1}),\ \Theta_{k,j} = (\Theta_{k-1,j}, q_{k,j}), \tag{2}$$

and then compute its updated weight by

$$w_{k,j} \propto p(x_k | \Theta_{k,j}, X_{k-1})\, w_{k-1,j}. \tag{3}$$

We shall refer to this method of obtaining the desired Monte Carlo sample as simple *sequential importance sampling* (SIS) (Liu and Chen, 1998).

A problem that arises with SIS is that with a finite sample, the weights become increasingly skewed over time, adversely affecting the sample's ability to adequately represent the distribution. This phenomenon is known as *sample degeneration*. One approach is to perform a check on the *skewness* of the weights of the sample at each time-step. If the weights are not too skewed, SIS is performed. Otherwise, SIS together with re-sampling is performed to counter sample degeneration. With a finite sample size, this algorithm has been shown to delay degeneration but the problem is not entirely resolved. When the measurement sequence is long, the filter does not remain stable.

In our sequential MCMC filter which was formulated in (Lee and Chia), we first check whether the resulting Monte Carlo sample provides an adequate representation of the distribution of interest. If it does, we proceed with simple SIS for the next time-step; otherwise, we perform MCMC with the desired distribution as target distribution and with a proposal distribution that is constructed from the Monte Carlo sample produced by the simple SIS. In this way, we avoid the drawbacks of re-sampling such as increase in random variation in the resulting sample and decrease in diversity of the sample points. Performing a full MCMC from time to time "refreshes" the Monte

Carlo sample and removes any approximation errors that may have accumulated from the SIS steps due to the finite sample size. Consequently, the sequential MCMC filter is guaranteed to be stable as long as there are enough resources to perform the MCMC properly. This has been demonstrated in many of our signal-processing experiments with long sequences of low SNR signals.

2. Algorithm

The outline of our sequential MCMC filter is given below:

1. Start of time-step k + 1: we have $(\Theta_{k,1}, w_{k,1}), \ldots, (\Theta_{k,n_k}, w_{k,n_k})$ from $p(\Theta_k | X_k)$.
   Set m = 1.
2. Simple SIS: Obtain $(\Theta_{k+m,1}, w_{k+m,1}), \ldots, (\Theta_{k+m,n_k}, w_{k+m,n_k})$ by
$$q_{k+m,j} \sim p(q_{k+m} | \Theta_{k+m-1,j}, X_k),$$
$$\Theta_{k+m,j} = (\Theta_{k+m-1,j}, q_{k+m,j}),$$
$$w_{k+m,j} \propto p(x_{k+1}, \ldots, x_{k+m} | \Theta_{k+m,j}, X_k) w_{k,j},$$
   for $j = 1, \ldots, n_k$.
3. If $(\Theta_{k+m,1}, w_{k+m,1}), \ldots, (\Theta_{k+m,n_k}, w_{k+m,n_k})$ adequately represent $p(\Theta_{k+m} | X_{k+m})$, then
   Set $m = m + 1$. Go to Step 2.
   Else
   (a) Construct $\hat{p}(\Theta_{k+m} | X_{k+m})$ using $(\Theta_{k+m,1}, w_{k+m,1}), \ldots, (\Theta_{k+m,n_k}, w_{k+m,n_k})$.
   (b) MCMC: Obtain $(\Theta_{k+m,1}, w_{k+m,1}), \ldots, (\Theta_{k+m,n_{k+m}}, w_{k+m,n_{k+m}})$ by MCMC with $p(\Theta_{k+m} | X_{k+m})$ as target density and $\hat{p}(\Theta_{k+m} | X_{k+m})$ as proposal density.
   (c) Set $k = k + m$. Go to Step 1.

To check if $(\Theta_{k+m,1}, w_{k+m,1}), \ldots, (\Theta_{k+m,n_k}, w_{k+m,n_k})$ adequately represents $p(\Theta_{k+m} | X_{k+m})$ in Step 3, we measure how well $p(\Theta_{k+m} | X_k)$ "predicts" $p(\Theta_{k+m} | X_{k+m})$ by computing the Kullback-Leibler distance between their two respective representative samples, $(\Theta_{k+m,1}, w_{k,1})$, …, $(\Theta_{k+m,n_k}, w_{k,n_k})$ and $(\Theta_{k+m,1}, w_{k+m,1})$, …, $(\Theta_{k+m,n_k}, w_{k+m,n_k})$. This is shown as follows:

$$k(w_{k+m}, w_k) = \sum_{j=1}^{n_k} w_{k+m,j} (\log w_{k+m,j} - \log w_{k,j}). \tag{4}$$

We refer to the number m when MCMC is required as the *batch size*. It is determined *adaptively* by specifying a threshold for $k(w_{k+m}, w_k)$. Bigger batches mean that MCMC is performed less frequently. Thus, MCMC is performed sequentially for batches of measurements, hence the name sequential MCMC filter. Within a batch, simple SIS is performed.

In experiments with parameters only, we observe that the batch sizes increase over time, demonstrating the feasibility of the sequential MCMC filter for real signal-processing applications.

Reference

J. S. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems," Journal of the American Statistical Association, vol. 93, pp.1032-1044, 1998.

Lee, D. S. and Chia, N. K., "Bayesian Signal Processing with the Sequential MCMC Filter," submitted for publication.