Particle Markov chain Monte Carlo

Fredrik Lindsten
Division of Automatic Control
Linköping University, Sweden
Particle Markov chain Monte Carlo

Particle MCMC (PMCMC) introduced in the seminal paper,


More on backward simulation in PMCMC,


Bayesian system identification

Consider a nonlinear, discrete-time state-space model,

\[ x_{t+1} = f_t(x_t, u_t; \theta) + v_t(\theta), \]
\[ y_t = h_t(x_t, u_t; \theta) + e_t(\theta). \]

We observe

\[ D_T = \{ u_t, y_t \}_{t=1}^T. \]

**Bayesian model:** \( \theta \) random variable with prior density \( \pi(\theta) \).

**Aim:** Find \( p(\theta \mid D_T) \).
**Aim:** Find $p(\theta \mid D_T)$. 
**Aim:** Find $p(\theta, x_{1:T} \mid D_T)$. 
Gibbs sampler for SSMs

**Aim:** Find \( p(\theta, x_{1:T} \mid D_T) \).

Alternate between updating \( \theta \) and updating \( x_{1:T} \).
Gibbs sampler for SSMs

**Aim:** Find $p(\theta, x_{1:T} \mid D_T)$.

Alternate between updating $\theta$ and updating $x_{1:T}$.

**MCMC:** Gibbs sampling for state-space models. Iterate, 
- Draw $\theta[r] \sim p(\theta \mid x_{1:T}[r-1], D_T)$;
- Draw $x_{1:T}[r] \sim p(x_{1:T} \mid \theta[r], D_T)$.

The above procedure results in a Markov chain, 

$$\{\theta[r], x_{1:T}[r]\}_{r \geq 1}$$

with stationary distribution $p(\theta, x_{1:T} \mid D_T)$. 
Gibbs sampler

\[ \text{ex) Sample from,} \]

\[ \mathcal{N} \left( \begin{pmatrix} x \\ y \end{pmatrix} ; \begin{pmatrix} 10 \\ 10 \end{pmatrix} , \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \right) . \]

Gibbs sampler

- Draw \( x' \sim p(x \mid y); \)
- Draw \( y' \sim p(y \mid x'). \)
Gibbs sampling for linear system identification.

\[
\begin{bmatrix}
  x_{t+1} \\
  y_t
\end{bmatrix} = \begin{bmatrix}
  A & B \\
  C & D
\end{bmatrix}
\begin{bmatrix}
  x_t \\
  u_t
\end{bmatrix} + \begin{bmatrix}
  v_t \\
  e_t
\end{bmatrix}.
\]

Iterate,

- Draw \( \theta' \sim p(\theta \mid x_{1:T}, D_T) \);
- Draw \( x'_{1:T} \sim p(x_{1:T} \mid \theta', D_T) \).

Fredrik Lindsten
Particle Markov chain Monte Carlo

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET
What about the general nonlinear/non-Gaussian case?

- Draw $\theta' \sim p(\theta | x_{1:T}, D_T)$;
- Draw $x'_{1:T} \sim p(x_{1:T} | \theta', D_T)$.
What about the general nonlinear/non-Gaussian case?

- Draw $\theta' \sim p(\theta \mid x_{1:T}, D_T)$; \textbf{OK!}
- Draw $x'_{1:T} \sim p(x_{1:T} \mid \theta', D_T)$. \textbf{Hard!}

\textbf{Problem:} $p(x_{1:T} \mid \theta, D_T)$ not available!
What about the general nonlinear/non-Gaussian case?

- Draw $\theta' \sim p(\theta | x_{1:T}, D_T)$; \hspace{1cm} OK!
- Draw $x'_{1:T} \sim p(x_{1:T} | \theta', D_T)$. \hspace{1cm} Hard!

**Problem:** $p(x_{1:T} | \theta, D_T)$ not available!

**Idea:** Approximate $p(x_{1:T} | \theta, D_T)$ using particle smoother.
Sampling strategy:

- Run a particle filter

Backward simulator

Sampling strategy:

- Run a particle filter

Backward simulator

Sampling strategy:

- Run a particle filter

Sampling strategy:

- Run a particle filter

Sampling strategy:

- Run a particle filter
- Sample a trajectory

\[ x'_{1:T} \overset{\text{approx.}}{\sim} p(x_{1:T} \mid \theta, D_T) \]

Problems

Problems with this approach,

- Based on particle filter (PF) \( \Rightarrow \) approximate sample.
- Relies on large \( N \) to be successful.
- A lot of wasted computations.
Problems with this approach,

- Based on particle filter (PF) ⇒ approximate sample.
- Relies on large $N$ to be successful.
- A lot of wasted computations.

To get around these problems,

Analyze PF + MCMC together ⇒ PMCMC
Particle Markov chain Monte Carlo

• Combines PF and MCMC in a systematic manner.
• “Exact approximation” of MCMC samplers.
• Family of Bayesian inference methods,
  • Particle Metropolis-Hastings (PMH)
  • Particle Gibbs (PG)
Particle Markov chain Monte Carlo

- Combines PF and MCMC in a systematic manner.
- “Exact approximation” of MCMC samplers.
- Family of Bayesian inference methods,
  - Particle Metropolis-Hastings (PMH)
  - Particle Gibbs (PG) – with backward simulation
The particle filter

- **Resampling**: \( \{ x_{1:t-1}^i, w_{t-1}^i \}_{i=1}^N \rightarrow \{ \tilde{x}_{1:t-1}^i, 1/N \}_{i=1}^N \).

- **Propagation**: \( x_t^i \sim R^\theta_t(dx_t | \tilde{x}_{1:t-1}^i) \) and \( x_{1:t}^i = \{ \tilde{x}_{1:t-1}^i, x_t^i \} \).

- **Weighting**: \( w_t^i = W^\theta_t(x_{1:t}^i) \).

\[ \Rightarrow \{ x_{1:t}^i, w_t^i \}_{i=1}^N \]
The particle filter

- Resampling + Propagation:

\[(a_t^i, x_t^i) \sim M_t^\theta(a_t, x_t) = \frac{w_{t-1}^{a_t}}{\sum_l w_{t-1}^l} R_t^\theta(x_t | x_{1:t-1}^{a_t}).\]

- Weighting: \(w_t^i = W_t^\theta(x_{1:t}^i).\)

\[\Rightarrow \{x_{1:t}^i, w_t^i\}_{i=1}^N\]
Random variables generated by the PF. Let,

\[ x_t = \{x_1^t, \ldots, x_N^t\}, \quad a_t = \{a_1^t, \ldots, a_N^t\} \]

The PF generates a **single sample** on \( X^{NT} \times \{1, \ldots, N\}^{N(T-1)} \) with density,

\[
\psi^\theta(x_{1:T}, a_{2:T}) \triangleq \prod_{i=1}^N R_1^\theta(x_1^i) \prod_{t=2}^T \prod_{i=1}^N M_t^\theta(a_t^i, x_t^i).
\]
Extended target density

What is the target density?

- Must admit $p(x_{1:T}, \theta \mid D_T)$ as a marginal.
- As close as possible to $\psi$. 
Extended target density

What is the target density?

- Must admit $p(x_{1:T}, \theta \mid D_T)$ as a marginal.
- As close as possible to $\psi$.

Let $x^k_{1:T} = x^{b_1}_{1:T} = \{x^{b_1}_1, \ldots, x^{b_T}_T\}$ be a specific path.
Extended target density

What is the target density?

- Must admit $p(x_{1:T}, \theta \mid D_T)$ as a marginal.
- As close as possible to $\psi$.

Let $x^{k}_{1:T} = x^{b_1:T}_{1:T} = \{x^{b_1}_1, \ldots, x^{b_T}_T\}$ be a specific path.

Introduce extended target,

$$
\phi(\theta, x_{1:T}, a_{2:T}, k) = \phi(\theta, x^{b_1:T}_{1:T}, b_{1:T}) \phi(x^{-b_1:T}_{1:T}, a^{-b_2:T}_{2:T} \mid \theta, x^{b_1:T}_{1:T}, b_{1:T})
$$
Extended target density

What is the target density?

• Must admit $p(x_{1:T}, \theta \mid D_T)$ as a marginal.
• As close as possible to $\psi$.

Let $x_{1:T}^k = x_{1:T}^{b_1} = \{x_1^{b_1}, \ldots, x_T^{b_T}\}$ be a specific path.

Introduce extended target,

$$
\phi(\theta, x_{1:T}, a_{2:T}, k) = \phi(\theta, x_{1:T}^{b_1}, b_{1:T}) \phi(x_{1:T}^{-b_1}, a_{2:T}^{-b_2} \mid \theta, x_{1:T}^{b_1}, b_{1:T})
$$

$$
\triangleq \frac{p(x_{1:T}^{b_1}, \theta \mid D_T)}{N^T} \prod_{i=1 \atop i \neq b_1}^N R_i^{\theta}(x_1^i) \prod_{t=2}^T \prod_{i=1 \atop i \neq b_t}^N M_t^{\theta}(a_t^i, x_t^i).
$$

marginal conditional
Particle Gibbs with backward simulation (PG-BS)

Multi-stage Gibbs sampler, targeting $\phi$,

i) Draw $\theta' \sim \phi(\theta | x_{1:T}^{b_1:T}, b_{1:T})$;

ii) Draw $\{x_{1:T}^{t',-b_1:T}, a_{2:T}^{t',-b_2:T}\} \sim \phi(x_{1:T}^{-b_1:T}, a_{2:T}^{-b_2:T} | \theta', x_{1:T}^{b_1:T}, b_{1:T})$;

iii) Draw, for $t = T, \ldots, 1$,

$$b_t' \sim \phi(b_t | \theta', x_{1:t}^{t',-b_1:t}, a_{2:t}^{t',-b_2:t}, x_{1:T}^{b_1:T}, b_{t+1:T}).$$
Particle Gibbs with backward simulation (PG-BS)

Multi-stage Gibbs sampler, targeting $\phi$,

i) Draw $\theta' \sim \phi(\theta \mid x^{b_1:T}_{1:T}, b_{1:T})$;

ii) Draw $\{x'_{1:T}^{b_1:T}, a'_{2:T}^{b_2:T}\} \sim \phi(x_{1:T}^{b_1:T}, a_{2:T}^{b_2:T} \mid \theta', x^{b_1:T}_{1:T}, b_{1:T})$;

iii) Draw, for $t = T, \ldots, 1$,

$$b'_{t} \sim \phi(b_{t} \mid \theta', x'_{1:t}^{b_1:T}, a'_{2:t}^{b_2:T}, x^{b_1:T}_{1:T}, b'_{t+1:T}).$$

---

Step i) By construction,

$$\phi(\theta \mid x^{b_1:T}_{1:T}, b_{1:T}) = p(\theta \mid x^{b_1:T}_{1:T}, D_T).$$

Sampling is assumed to be feasible.
Step ii) By construction,

\[ \phi(x_{1:T}^{-b_1:T}, a_{2:T}^{-b_2:T} \mid \theta, x_{1:T}^{b_1:T}, b_{1:T}) = \prod_{i=1}^{N} R_{1}^{\theta}(x_{i}^{1}) \prod_{t=2}^{T} \prod_{i=1}^{N} M_{t}^{\theta}(a_{i}, x_{i}). \]
Step ii) By construction,

$$
\phi(x^{-b_1:T}_1, a^{-b_2:T}_2 \mid \theta, x^{b_1:T}_1, b_1:T) = \prod_{i=1}^{N} R^\theta_1(x^i_1) \prod_{t=2}^{T} \prod_{i=1}^{N} M^\theta_t(a^i_t, x^i_t).
$$

Conditional PF (conditioned on \{x^t_1:T, b_1:T\}),

1. **Initialize** \((t = 1)\):
   
   (a) Draw \(x^i_1 \sim R^\theta_1(x_1)\) for \(i \neq b_1\) and set \(x^{b_1}_1 = x'_1\).
   
   (b) Set \(w^i_1 = W^\theta_1(x^i_1)\) for \(i = 1, \ldots, N\).

2. **for** \(t = 2, \ldots, T\):
   
   (a) Draw \((a^i_t, x^i_t) \sim M^\theta_t(a_t, x_t)\) for \(i \neq b_t\).
   
   (b) Set \(x^{b_t}_t = x'_t\) and \(a^{b_t}_t = b_{t-1}\).
   
   (c) Set \(x^i_{1:t} = \{x^{a^i_{t-1}}_{1:t-1}, x^i_t\}\) and \(w^i_t = W^\theta_t(x^i_{1:t})\) for \(i = 1, \ldots, N\).
Step iii) Sequence of Gibbs steps. For $t = T, \ldots, 1$, draw,

$$b_t \sim \phi(b_t \mid x_{1:t}, a_{2:t}, x_{t+1:T}, b_{t+1:T})$$

(⋆)
Step iii) Sequence of Gibbs steps. For \( t = T, \ldots, 1 \), draw,

\[
b_t \sim \phi(b_t \mid x_{1:t}, a_{2:t}, x_{t+1:T}, b_{t+1:T}) \quad (\star)
\]

By expanding

\[
p(x_{1:t} \mid \theta, D_t) \propto W_t^\theta(x_{1:t}) R_t^\theta(x_t \mid x_{1:t-1}) p(x_{1:t-1} \mid \theta, D_{t-1}),
\]

we can show that \((\star)\) corresponds to

\[
P(b_t = i) \propto w_t^i p(x_{i+1}^{b_t+1} \mid \theta, x_i^i).
\]

Sampling \( b_{1:T} \) corresponds exactly to a run of a backward simulator!
Algorithm 1 PG-BS: Particle Gibbs with backward simulation

1. **Initialize:** Set $\theta[0], x_{1:T}[0]$ and $b_{1:T}[0]$ arbitrarily.

2. **For** $r \geq 1$, **iterate:**
   
   (a) Draw $\theta[r] \sim p(\theta | x_{1:T}[r-1], D_T)$.

   (b) Run a conditional PF, targeting $p(x_{1:T} | \theta[r], D_T)$, conditioned on $\{x_{1:T}[r-1], b_{1:T}[r-1]\}$.

   (c) Run a backward simulator to generate $b_{1:T}[r]$ and set $x_{1:T}[r]$ to the corresponding particle trajectory.

\[
\{\theta[r], x_{1:T}[r]\}_{r \geq 1} \text{ has stationary distribution } p(\theta, x_{1:T} | D_T).
\]
Stochastic volatility model,

\[ x_{t+1} = \theta_1 x_t + v_t, \]
\[ y_t = e_t \exp \left( \frac{1}{2} x_t \right), \]
\[ v_t \sim \mathcal{N}(0, \theta_2), \]
\[ e_t \sim \mathcal{N}(0, 1). \]
Stochastic volatility model,

\[ x_{t+1} = \theta_1 x_t + v_t, \]
\[ y_t = e_t \exp \left( \frac{1}{2} x_t \right), \]

\[ v_t \sim \mathcal{N}(0, \theta_2), \]
\[ e_t \sim \mathcal{N}(0, 1). \]
ex) Wiener system identification

- Find $\theta = \{G, h(\cdot)\}$.
  - Parametric (state-space) model for $G$.
  - Nonparametric model for $h$, based on Gaussian process.
- Example system
  - 4th order linear system, $T = 1000$.
  - Blind identification ($u_t = 0$).
- PG-BS with
  - $N = 5$ particles.
  - 15000 iterations of the Gibbs sampler.
ex) Wiener system identification, cont’d.

Bode diagram

Nonlinear mapping

Fredrik Lindsten
Particle Markov chain Monte Carlo
Summary

Particle Gibbs with backward simulation

- Combines PF and MCMC in a systematic manner.
- Provably convergent for any $N \geq 2$ – and it works in practice!
- Makes efficient use of the available particles.
- How does it scale with the state dimension?
- Models with strong dependencies between state and parameter?

PG-BS only one member of the PMCMC family – there are other methods with different properties.

MATLAB code available at:
http://www.control.isy.liu.se/~lindsten/code/
Particle Markov chain Monte Carlo

Particle MCMC (PMCMC) introduced in the seminal paper,


More on backward simulation in PMCMC,


Stochastic volatility example

- $\theta_1 = 0.9$, $\theta_2 = 0.5^2$.
- $T = 5000$. 
LGSS example

\[
A = \begin{pmatrix}
-0.5107 & 1 & 0 & 0 & 0 \\
-1.0705 & 0 & 1 & 0 & 0 \\
-0.4268 & 0 & 0 & 1 & 0 \\
-0.1080 & 0 & 0 & 0 & 1 \\
-0.0005 & 0 & 0 & 0 & 0 \\
\end{pmatrix}, \\
B = \begin{pmatrix}
-1.6599 \\
-0.9034 \\
-2.3697 \\
-0.8543 \\
-0.2029 \\
\end{pmatrix},
\]

\[
C = (1 \ 0 \ 0 \ 0 \ 0).
\]

- \( Q = 0.05I_5, \ R = 0.01 \).
- \( u_t \sim \mathcal{N}(0, 0.01) \).
- \( T = 1000 \).
- MNIW prior with subspace initialization for \( A \) and \( B \).