The Conditional Particle Filter

Sumeetpal S. Singh
Cambridge University Engineering Department

jointly with A. Lee, M. Vihola
older work with F. Lindsten, E. Moulines
older still with N. Chopin, B. Kuhlenschimdt

The Institute for Mathematical Sciences, NUS, 28 Aug. 2018

State-space Model

Running example (Yu & Meng 2011):

$$X_{t+1} = \rho X_t + \sigma W_{t+1}, \qquad W_{t+1} \sim^{\text{i.i.d.}} N(0,1)$$
$$Y_t | X_t = x_t \sim \text{Poisson}(e^{x_t + \mu})$$

State-space Model

Running example (Yu & Meng 2011):

$$X_{t+1} = \rho X_t + \sigma W_{t+1}, \qquad W_{t+1} \overset{\text{i.i.d.}}{\sim} N(0,1)$$
$$Y_t | X_t = x_t \sim \text{Poisson}(e^{x_t + \mu})$$

In general:

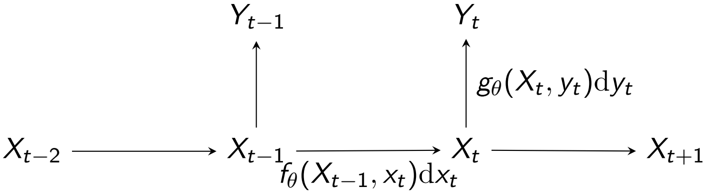


Figure : Evolution of the random variables of a HMM.

Inference Objective

The posterior: $p(\theta, x_{0:T} | y_{0:T}), \qquad \theta = (\mu, \rho, \sigma)$

Inference Objective

The posterior: $p(\theta, x_{0:T} | y_{0:T}), \quad \theta = (\mu, \rho, \sigma)$

Gibbs sampler (one cycle): $(\theta, x_{0:T}) \rightarrow (\theta', x'_{0:T})$

Inference Objective

The posterior: $p(\theta, x_{0:T} | y_{0:T}), \qquad \theta = (\mu, \rho, \sigma)$

Gibbs sampler (one cycle): $(\theta, x_{0:T}) \to (\theta', x'_{0:T})$

$$
\begin{aligned}
\sigma' | (x_{0:T}, \mu, \rho) \quad &\sim \quad Gamma\,(\cdots) \\
\rho' | (x_{0:T}, \mu, \sigma') \quad &\sim \quad Normal\,(\cdots) \\
\mu' | (x_{0:T}, \sigma', \rho') \quad &\sim \quad Normal\,(\cdots) \\
x'_{0:T} | (\sigma', \mu', \rho') \quad &\sim \quad p(x_{0:T} | \theta', y_{0:T})
\end{aligned}
$$

Inference Objective

The posterior: $p(\theta, x_{0:T} | y_{0:T}), \qquad \theta = (\mu, \rho, \sigma)$

Gibbs sampler (one cycle): $(\theta, x_{0:T}) \rightarrow (\theta', x'_{0:T})$

$$\begin{aligned}
\sigma' | (x_{0:T}, \mu, \rho) &\sim \mathit{Gamma}(\cdots) \\
\rho' | (x_{0:T}, \mu, \sigma') &\sim \mathit{Normal}(\cdots) \\
\mu' | (x_{0:T}, \sigma', \rho') &\sim \mathit{Normal}(\cdots) \\
x'_{0:T} | (\sigma', \mu', \rho') &\sim p(x_{0:T} | \theta', y_{0:T})
\end{aligned}$$

In general cannot sample from $p(x_{0:T} | \theta', y_{0:T})$

Inference Objective

The posterior: $p(\theta, x_{0:T} | y_{0:T})$, $\theta = (\mu, \rho, \sigma)$

Gibbs sampler (one cycle): $(\theta, x_{0:T}) \rightarrow (\theta', x'_{0:T})$

$$
\begin{aligned}
\sigma' | (x_{0:T}, \mu, \rho) &\sim \mathit{Gamma}(\cdots) \\
\rho' | (x_{0:T}, \mu, \sigma') &\sim \mathit{Normal}(\cdots) \\
\mu' | (x_{0:T}, \sigma', \rho') &\sim \mathit{Normal}(\cdots) \\
x'_{0:T} | (\sigma', \mu', \rho') &\sim p(x_{0:T} | \theta', y_{0:T})
\end{aligned}
$$

In general cannot sample from $p(x_{0:T} | \theta', y_{0:T})$

An old remedy is one at a time: $x_i | (x'_{0:i-1}, x_{i+1:T}, \theta')$

Particle Filtering

- Popularised by Gordon, Salmond and Smith (1993)

- Popularised by Gordon, Salmond and Smith (1993)
- (Sequential) Importance sampling method to approximate

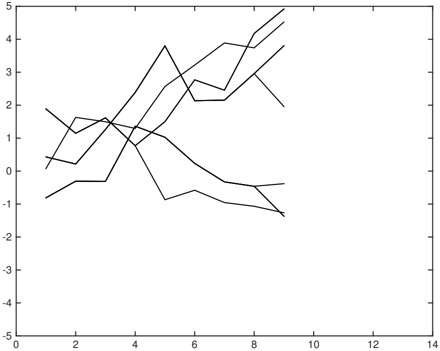
$$p(x_{0:T}|\theta, y_{0:T})$$

using non-iid samples:

$$\mathbb{E}(h(X_{0:T})|\theta, y_{0:T}) \approx \sum_{i=1}^{N} h(X_{0:T}^{(i)})W_T^{(i)}$$

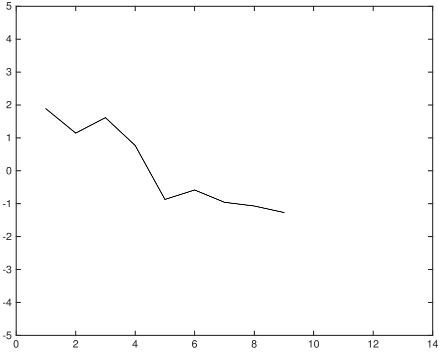
Particle Filter execution for sampling $p(x_{0:T}|y_{0:T})$

Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$

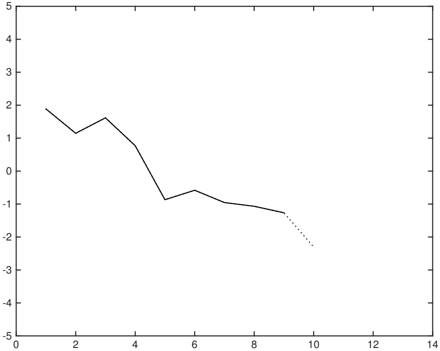


Particle Filter execution for sampling $p(x_{0:T}|y_{0:T})$

Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$



Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$

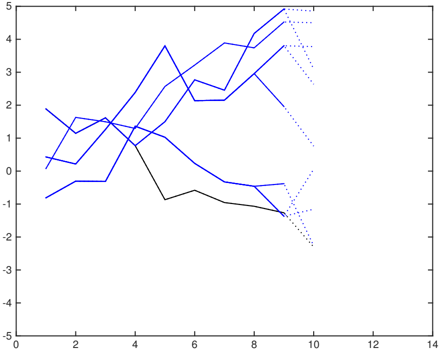


Sample: $X_{t+1}^{(i)} \sim f(X_t^{(i)}, x_{t+1})$

Weight: $w_{t+1}^{(i)} = g(X_{t+1}^{(i)}, y_{t+1})$

Particle Filter execution for sampling $p(x_{0:T}|y_{0:T})$

Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$



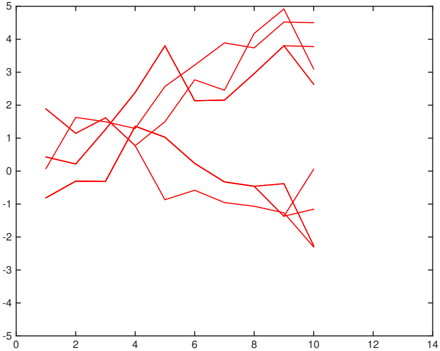
Sample: $X_{t+1}^{(i)} \sim f(X_t^{(i)}, x_{t+1})$

Weight: $w_{t+1}^{(i)} = g(X_{t+1}^{(i)}, y_{t+1})$

$$p(x_{0:t+1}|y_{0:t+1}) \approx \sum_{i=1}^{N} W_{t+1}^{(i)} \delta_{X_{0:t+1}^{(i)}}$$

Particle Filter execution for sampling $p(x_{0:T}|y_{0:T})$

Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$



Sample: $X_{t+1}^{(i)} \sim f(X_t^{(i)}, x_{t+1})$

Weight: $w_{t+1}^{(i)} = g(X_{t+1}^{(i)}, y_{t+1})$

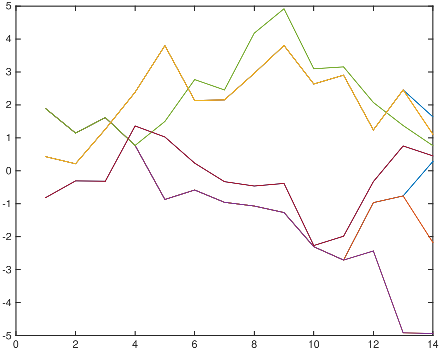
$$p(x_{0:t+1}|y_{0:t+1}) \approx \sum_{i=1}^{N} W_{t+1}^{(i)} \delta_{X_{0:t+1}^{(i)}}$$

Resample:

$$p(x_{0:t+1}|y_{0:t+1}) \approx \sum_{i=1}^{N} \delta_{X_{0:t+1}^{(i)}}$$

Particle Filter execution for sampling $p(x_{0:T}|y_{0:T})$

Given $\sum_{i=1}^{N} \delta_{X_{0:t}^{(i)}} \approx p(x_{0:t}|y_{0:t})$, approximate $p(x_{0:t+1}|y_{0:t+1})$



Sample: $X_{t+1}^{(i)} \sim f(X_t^{(i)}, x_{t+1})$

Weight: $w_{t+1}^{(i)} = g(X_{t+1}^{(i)}, y_{t+1})$

$$p(x_{0:t+1}|y_{0:t+1}) \approx \sum_{i=1}^{N} W_{t+1}^{(i)} \delta_{X_{0:t+1}^{(i)}}$$

Resample:

$$p(x_{0:t+1}|y_{0:t+1}) \approx \sum_{i=1}^{N} \delta_{X_{0:t+1}^{(i)}}$$

Particle Filtering (cont'd)

- The final (Gibbs) step for states, $(\theta', x_{0:T}) \rightarrow (\theta', x'_{0:T})$ with

$$x'_{0:T}|(\sigma', \mu', \rho') \quad \sim \quad \text{PF approx. of } p(x_{0:T}|\theta', y_{0:T})$$

Particle Filtering (cont'd)

- The final (Gibbs) step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T}|(\sigma', \mu', \rho') \quad \sim \quad \text{PF approx. of } p(x_{0:T}|\theta', y_{0:T})$$

- Why not, since used extensively in EM and gradient methods to learn θ

$$Q(\theta, \theta') = \mathbb{E}\left\{ \log p(x_{0:T}, y_{0:T}|\theta') \middle|\, \theta, y_{0:T} \right\}$$

Particle Filtering (cont'd)

- The final (Gibbs) step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T} | (\sigma', \mu', \rho') \quad \sim \qquad \text{PF approx. of } p(x_{0:T} | \theta', y_{0:T})$$

- Why not, since used extensively in EM and gradient methods to learn θ

$$Q(\theta, \theta') = \mathbb{E}\left\{ \log p(x_{0:T}, y_{0:T} | \theta') \middle| \, \theta, y_{0:T} \right\}$$

- In practise particle number N must grow linearly with T (Many results on the error of Particle filter estimates, e.g. Del Moral's book 2004, ...)

Particle Filtering (cont'd)

- The final (Gibbs) step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T}|(\sigma', \mu', \rho') \quad \sim \quad \text{PF approx. of } p(x_{0:T}|\theta', y_{0:T})$$

- Why not, since used extensively in EM and gradient methods to learn θ

$$Q(\theta, \theta') = \mathbb{E}\left\{ \log p(x_{0:T}, y_{0:T}|\theta') \middle| \theta, y_{0:T} \right\}$$

- In practise particle number N must grow linearly with T (Many results on the error of Particle filter estimates, e.g. Del Moral's book 2004, ...)
- The bias free (mathematically correct) way (ADH2010) is to use the *conditional* Particle Filter

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \rightarrow (\theta', x'_{0:T})$ with

$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

- A CPF simulates a PF with N particles for T time steps as "usual" but with one particle set to $X^{(1)}_{0:T} = x_{0:T}$

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \rightarrow (\theta', x'_{0:T})$ with

$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

- A CPF simulates a PF with N particles for T time steps as "usual" but with one particle set to $X^{(1)}_{0:T} = x_{0:T}$
  – Then choose one particle randomly according to its weight

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

- A CPF simulates a PF with N particles for T time steps as "usual" but with one particle set to $X_{0:T}^{(1)} = x_{0:T}$
  – Then choose one particle randomly according to its weight
- In effect, the CPF is a Markov kernel:

$$X'_{0:T} \sim P_{\theta',N}(x_{0:T}, dx'_{0:T})$$

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

- A CPF simulates a PF with N particles for T time steps as "usual" but with one particle set to $X_{0:T}^{(1)} = x_{0:T}$
  – Then choose one particle randomly according to its weight
- In effect, the CPF is a Markov kernel:

$$X'_{0:T} \sim P_{\theta',N}(x_{0:T}, dx'_{0:T})$$

- Invariant measure is $p(x_{0:T}|\theta', y_{0:T})$ for any $N \geq 2$

The Conditional Particle Filter (Andrieu, Doucet & Holenstein, 2010)

- The final step for states, $(\theta', x_{0:T}) \to (\theta', x'_{0:T})$ with

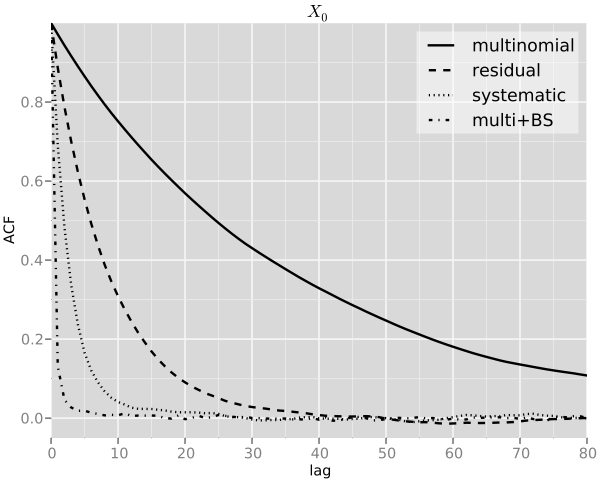
$$x'_{0:T} \sim \text{conditional Particle Filter (CPF)}$$

- A CPF simulates a PF with N particles for T time steps as "usual" but with one particle set to $X^{(1)}_{0:T} = x_{0:T}$
  – Then choose one particle randomly according to its weight
- In effect, the CPF is a Markov kernel:

$$X'_{0:T} \sim P_{\theta',N}(x_{0:T}, dx'_{0:T})$$

- Invariant measure is $p(x_{0:T}|\theta', y_{0:T})$ for any $N \geq 2$
- Effective sampler? How should N grow with T?

Sampling $p(x_{0:399}|y_{0:399})$ with 200 particles (Chopin & S., 2013)



Statistic: ACF X_0

Coupling the CPF (Chopin & S., 2013)

Coupling the CPF (Chopin & S., 2013)

- The outputs of two CPFs

$$P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

  with different inputs $x_{0:T}$ and $\tilde{x}_{0:T}$

  but implemented with common random numbers

  can be the same with (high) probability

Coupling the CPF (Chopin & S., 2013)

- The outputs of two CPFs

$$P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

with different inputs $x_{0:T}$ and $\tilde{x}_{0:T}$

but implemented with common random numbers

can be the same with (high) probability

- Thus if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$X_{0:T} \overset{d}{=} P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad \tilde{X}_{0:T} \overset{d}{=} P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

Coupling the CPF (Chopin & S., 2013)

- The outputs of two CPFs

$$P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

with different inputs $x_{0:T}$ and $\tilde{x}_{0:T}$

but implemented with common random numbers

can be the same with (high) probability

- Thus if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$X_{0:T} \stackrel{d}{=} P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad \tilde{X}_{0:T} \stackrel{d}{=} P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

- We have

$$\mathbb{P}(X_{0:T} \neq \tilde{X}_{0:T}) \leq \epsilon$$

Uniform Ergodicity

Chopin+S. (2013)

Can construct a coupling$(P_N(x_{0:T}, \cdot), P_N(\tilde{x}_{0:T}, \cdot))$ with probability at least $1 - \epsilon$.

$$\|P_N^k(x_{0:T}, dx'_{0:T}) - p(dx'_{0:T}|y_{0:T})\|_{\mathrm{tv}} \leq \epsilon^k$$

Uniform Ergodicity

Chopin+S. (2013)

Can construct a coupling$(P_N(x_{0:T}, \cdot), P_N(\tilde{x}_{0:T}, \cdot))$ with probability at least $1 - \epsilon$.

$$\| P_N^k(x_{0:T}, dx'_{0:T}) - p(dx'_{0:T} | y_{0:T}) \|_{\mathrm{tv}} \leq \epsilon^k$$

Kuhlenschimdt + S. (2014)

$$\| P_N^k(x_{0:T}, dx'_{0:T}) - p(dx'_{0:T} | y_{0:T}) \|_{\mathrm{tv}} \leq \text{Const.} \times \left(\frac{T}{N}\right)^k$$

Uniform Ergodicity

Chopin+S. (2013)

Can construct a coupling$(P_N(x_{0:T},\cdot), P_N(\tilde{x}_{0:T},\cdot))$ with probability at least $1-\epsilon$.

$$\|P_N^k(x_{0:T}, dx'_{0:T}) - p(dx'_{0:T}|y_{0:T})\|_{\mathrm{tv}} \leq \epsilon^k$$

Kuhlenschimdt + S. (2014)

$$\|P_N^k(x_{0:T}, dx'_{0:T}) - p(dx'_{0:T}|y_{0:T})\|_{\mathrm{tv}} \leq \text{Const.} \times \left(\frac{T}{N}\right)^k$$

Andrieu, Lee, Vihola (2014); Douc, Lindsten, Moulines (2014)

$$P_N(x_{0:T}, dx'_{0:T}) \geq \epsilon(N,T) p(x'_{0:T}|y_{0:T})$$

and

$$\liminf_T \epsilon(N,T) > 0 \quad \text{provided } N \propto T$$

Cost per iteration

- These results say particles must increase linearly with T costing T^2 per application of $P_{T,N}$

Cost per iteration

- These results say particles must increase linearly with T costing T^2 per application of $P_{T,N}$
- Could CPF work with a fixed number of particles? Costing NT per application of CPF or $P_{T,N}$.

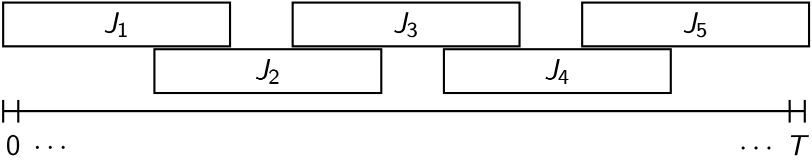
Cost per iteration

- These results say particles must increase linearly with T costing T^2 per application of $P_{T,N}$
- Could CPF work with a fixed number of particles? Costing NT per application of CPF or $P_{T,N}$.
- N. Whiteley (RSS discussion of PMCMC, 2010) suggested an extra *backward* step that tries to modify (recursively, backward in time) the ancestry of the selected trajectory.

Cost per iteration

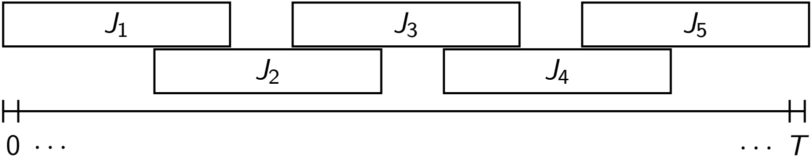
- These results say particles must increase linearly with T costing T^2 per application of $P_{T,N}$
- Could CPF work with a fixed number of particles? Costing NT per application of CPF or $P_{T,N}$.
- N. Whiteley (RSS discussion of PMCMC, 2010) suggested an extra *backward* step that tries to modify (recursively, backward in time) the ancestry of the selected trajectory.
- Highly successful in practise but no theoretical verification.

Blocked Gibbs sampler for $p(x_{0:T}|y_{0:T})$



- Group states $x_{0:T}$ into m overlapping blocks

Blocked Gibbs sampler for $p(x_{0:T}|y_{0:T})$

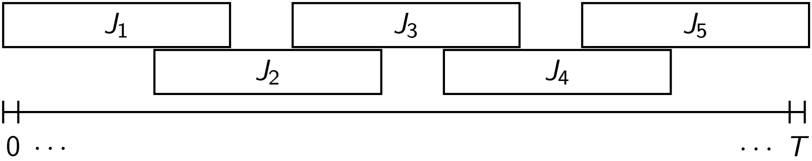


- Group states $x_{0:T}$ into m overlapping blocks
- When sampling block $J_i = r : s$, sample from

$$p(x_{r:s}|x_{r-1}, y_{r:s}, x_{s+1})$$

while holding remaining states unchanged.

Blocked Gibbs sampler for $p(x_{0:T}|y_{0:T})$

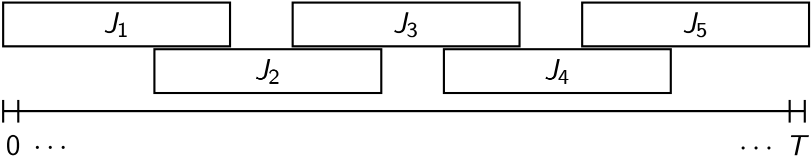


- Group states $x_{0:T}$ into m overlapping blocks
- When sampling block $J_i = r : s$, sample from

$$p(x_{r:s}|x_{r-1}, y_{r:s}, x_{s+1})$$

  while holding remaining states unchanged.
- Cycle through the blocks in any order, sequentially, odd-even etc.

Blocked Gibbs sampler for $p(x_{0:T}|y_{0:T})$

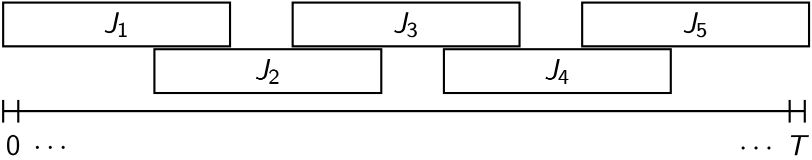


- Group states $x_{0:T}$ into m overlapping blocks
- When sampling block $J_i = r : s$, sample from

$$p(x_{r:s}|x_{r-1}, y_{r:s}, x_{s+1})$$

  while holding remaining states unchanged.
- Cycle through the blocks in any order, sequentially, odd-even etc.
- Effectively sampling $p(x_{0:T}|y_{0:T})$ using the Markov kernel $\mathcal{P}(x_{0:T}, dx'_{0:T})$

Blocked Gibbs sampler for $p(x_{0:T}|y_{0:T})$



- Group states $x_{0:T}$ into m overlapping blocks
- When sampling block $J_i = r : s$, sample from

$$p(x_{r:s}|x_{r-1}, y_{r:s}, x_{s+1})$$

  while holding remaining states unchanged.
- Cycle through the blocks in any order, sequentially, odd-even etc.
- Effectively sampling $p(x_{0:T}|y_{0:T})$ using the Markov kernel $\mathcal{P}(x_{0:T}, dx'_{0:T})$

$$\mathcal{P} = \mathcal{P}_o \mathcal{P}_e \qquad \text{where} \qquad \begin{cases} \mathcal{P}_o = P_{J_1} P_{J_3} \cdots P_{J_m} \\ \mathcal{P}_e = P_{J_2} P_{J_4} \cdots P_{J_{m-1}} \end{cases}$$

Uniform ergodicity: ideal blocked sampling

- If $\{J_1, \ldots, J_m\}$ be an arbitrary cover of $\{1, \ldots, T\}$

If $\mathcal{P}$ is the blocked Gibbs kernel of one complete sweep then

$$||p(dx_{0:T}|y_{0:T}) - \mu\mathcal{P}^k||_{\mathrm{tv}} \leq (T+1)\lambda^k$$

where λ is T-independent (S., Lindsten and Moulines, 2015)

- If $\{J_1, \ldots, J_m\}$ be an arbitrary cover of $\{1, \ldots, T\}$

If $\mathcal{P}$ is the blocked Gibbs kernel of one complete sweep then

$$||p(dx_{0:T}|y_{0:T}) - \mu\mathcal{P}^k||_{\mathrm{tv}} \leq (T+1)\lambda^k$$

where λ is T-independent (S., Lindsten and Moulines, 2015)

- Once you decide on a block size and overlap proportion, works for any time-series length T
- Rate quickens, $\lambda \to 0$, as block overlap increases.

Uniform ergodicity: ideal blocked sampling

- If $\{J_1, \ldots, J_m\}$ be an arbitrary cover of $\{1, \ldots, T\}$

If $\mathcal{P}$ is the blocked Gibbs kernel of one complete sweep then

$$||p(dx_{0:T}|y_{0:T}) - \mu\mathcal{P}^k||_{\mathrm{tv}} \leq (T+1)\lambda^k$$

where λ is T-independent (S., Lindsten and Moulines, 2015)

- Once you decide on a block size and overlap proportion, works for any time-series length T
- Rate quickens, $\lambda \to 0$, as block overlap increases.
- Recall $\mathcal{P} = P_{J_1} P_{J_2} \cdots P_{J_m}$. Idea is to approximate each P_{J_i} with CPF.

Uniform ergodicity: fixed N and any T!

- Approximate each block kernel P_{J_i} with CPF $P_{J_i,N}$:

  (ideal) $\mathcal{P} = P_{J_1} P_{J_2} \cdots P_{J_m}$ (CPF) $\mathcal{P}_N = P_{J_1,N} P_{J_2,N} \cdots P_{J_m,N}$

Uniform ergodicity: fixed N and any T!

- Approximate each block kernel P_{J_i} with CPF $P_{J_i,N}$:

(ideal) $\mathcal{P} = P_{J_1} P_{J_2} \cdots P_{J_m}$ (CPF) $\mathcal{P}_N = P_{J_1,N} P_{J_2,N} \cdots P_{J_m,N}$

If $\mathcal{P}_N$ is the blocked pGibbs kernel of one complete sweep then

$$||p(dx_{0:T}|y_{0:T}) - \mu\mathcal{P}_N^k||_{\text{tv}} \leq (T+1)\lambda_N^k$$

(S., Lindsten and Moulines, 2015)

Uniform ergodicity: fixed N and any T!

- Approximate each block kernel P_{J_i} with CPF $P_{J_i,N}$:

(ideal) $\mathcal{P} = P_{J_1} P_{J_2} \cdots P_{J_m}$ (CPF) $\mathcal{P}_N = P_{J_1,N} P_{J_2,N} \cdots P_{J_m,N}$

If $\mathcal{P}_N$ is the blocked pGibbs kernel of one complete sweep then

$$||p(dx_{0:T}|y_{0:T}) - \mu \mathcal{P}_N^k||_{\text{tv}} \leq (T+1)\lambda_N^k$$

(S., Lindsten and Moulines, 2015)

- Rate is

$$\lambda_N = \sqrt{\lambda} + \text{Const.} \times \text{max block size} \times \frac{1}{N}$$

Coupling the CPF (Chopin & S., 2013)

- The outputs of two CPFs

$$P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

with different inputs $x_{0:T}$ and $\tilde{x}_{0:T}$
but implemented with common random numbers
can be the same with (high) probability

- Thus if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$X_{0:T} \stackrel{d}{=} P_N(x_{0:T}, \mathrm{d}x'_{0:T}) \qquad \tilde{X}_{0:T} \stackrel{d}{=} P_N(\tilde{x}_{0:T}, \mathrm{d}x'_{0:T})$$

- We have

$$\mathbb{P}(X_{0:T} \neq \tilde{X}_{0:T}) \leq \epsilon$$

Coupling for Unbiased estimation

Set $X_{0:T}[1] \leftarrow \mathrm{CPF}(x_{0:T})$, $\tilde{X}_{0:T}[1] = x_{0:T}$, $x_{0:T}$ arbitrary.
for $n = 2, 3, \ldots$ **do**
    $(X_{0:T}[n], \tilde{X}_{0:T}[n]) \leftarrow \mathrm{CCPF}(X_{0:T}[n-1], \tilde{X}_{0:T}[n-1])$
    if $X_{0:T}[n] = \tilde{X}_{0:T}[n]$ **then output**

$$Z = h(X_{0:T}[1]) + \sum_{k=2}^{n} h(X_{0:T}[k]) - h(\tilde{X}_{0:T}[k])$$

end for

- Unbiased estimation:

$$\mathbb{E}(h(Z)) = \int h(x_{0:T}) p(x_{0:T}|y_{0:T}) \mathrm{d}x_{0:T}$$

- Jacob, Lindsten, Schon (2017) use the CCPF within the scheme of Glynn & Rhee (2014),

Coupling for Unbiased estimation

- Works because (i) $\tilde{X}_{0:T}[k] \stackrel{d}{=} X_{0:T}[k-1]$,
  (ii) coupling time is finite and
  (iii) the coupling CCPF is for ergodic kernels

$$\|P_N^k - p(x_{0:T} \mid y_{0:T})\|_{\mathrm{tv}} \xrightarrow{k\to\infty} 0$$

Coupling for Unbiased estimation

- Works because (i) $\tilde{X}_{0:T}[k] \stackrel{d}{=} X_{0:T}[k-1]$,
  (ii) coupling time is finite and
  (iii) the coupling CCPF is for ergodic kernels

$$\|P_N^k - p(x_{0:T} \mid y_{0:T})\|_{\mathrm{tv}} \stackrel{k\to\infty}{\longrightarrow} 0$$

- Under weak assumptions

Lee, S., Vihola (2018)

There exists a constant c such that for any $N \geq 2$ the coupling time

$$\mathbb{P}(\tau \geq k) \leq \left(\frac{c}{c+N}\right)^k$$

Coupling for Unbiased estimation

- Works because (i) $\tilde{X}_{0:T}[k] \overset{d}{=} X_{0:T}[k-1]$,
  (ii) coupling time is finite and
  (iii) the coupling CCPF is for ergodic kernels

$$\|P_N^k - p(x_{0:T} \mid y_{0:T})\|_{\mathrm{tv}} \overset{k\to\infty}{\longrightarrow} 0$$

- Under weak assumptions

Lee, S., Vihola (2018)

There exists a constant c such that for any $N \geq 2$ the coupling time

$$\mathbb{P}(\tau \geq k) \leq \left(\frac{c}{c+N}\right)^k$$

Under stronger assumptions, the coupling time is stable provided

$$N \propto 2^T$$

The Coupled Conditional Backward Particle Filter, or CCBPF (Lee, S., Vihola, 2018)

- The problem here is we rely one *one-shot* coupling: if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$\mathbb{P}(X_{0:T} \neq \tilde{X}_{0:T}) \leq \frac{c}{N + c}$$

The Coupled Conditional Backward Particle Filter, or CCBPF (Lee, S., Vihola, 2018)

- The problem here is we rely one *one-shot* coupling: if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$\mathbb{P}(X_{0:T} \neq \tilde{X}_{0:T}) \leq \frac{c}{N+c}$$

- Is there a version which will work with a fix number of particles N irrespective of T?

The Coupled Conditional Backward Particle Filter, or CCBPF (Lee, S., Vihola, 2018)

- The problem here is we rely one *one-shot* coupling: if $(X_{0:T}, \tilde{X}_{0:T}) \sim \mathrm{CCPF}(x_{0:T}, \tilde{x}_{0:T})$ then

$$\mathbb{P}(X_{0:T} \neq \tilde{X}_{0:T}) \leq \frac{c}{N+c}$$

- Is there a version which will work with a fix number of particles N irrespective of T?
- The idea is rely to coupling progressively

$$\kappa_n = \max\left\{0 \leq t \leq T \ : \ X_{0:t}[n] = \tilde{X}_{0:t}[n]\right\}$$

- With CCPF implemented with *backward sampling* the coupling boundary κ_n drifts to the right!

The CCBPF (Lee, S., Vihola, 2018)

Let τ = first time n s.t. $X_{0:T}[n] = \tilde{X}_{0:T}[n]$ then for any positive constants $\alpha > 1$ and $\beta < 1/\alpha$

$$\mathbb{P}(\tau \geq n) \leq \alpha^T \beta^n, \qquad \text{for all } n, T$$

if particle number N is large enough

The CCBPF (Lee, S., Vihola, 2018)

Let τ = first time n s.t. $X_{0:T}[n] = \tilde{X}_{0:T}[n]$ then for any positive constants $\alpha > 1$ and $\beta < 1/\alpha$

$$\mathbb{P}(\tau \geq n) \leq \alpha^T \beta^n, \qquad \text{for all } n, T$$

if particle number N is large enough

- Among the corollaries, an important one is coupling for unbiased simulation is assured in time proportional to time series length T .

$$\mathbb{P}\left(\text{Coupling time exceeds } T\right) \xrightarrow{T\to\infty} 0$$

The CCBPF (Lee, S., Vihola, 2018)

Let τ = first time n s.t. $X_{0:T}[n] = \tilde{X}_{0:T}[n]$ then for any positive constants $\alpha > 1$ and $\beta < 1/\alpha$

$$\mathbb{P}(\tau \geq n) \leq \alpha^T \beta^n, \qquad \text{for all } n, T$$

if particle number N is large enough

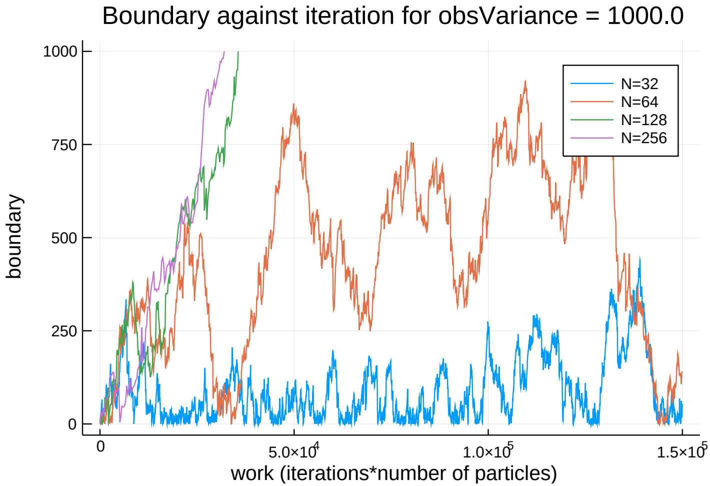
- Among the corollaries, an important one is coupling for unbiased simulation is assured in time proportional to time series length T .

$$\mathbb{P}\left(\text{Coupling time exceeds } T\right) \overset{T\to\infty}{\longrightarrow} 0$$

- The conjecture that Whiteley's backward sampling version of Andrieu et al's CPH is stable for a fix N and any T is true

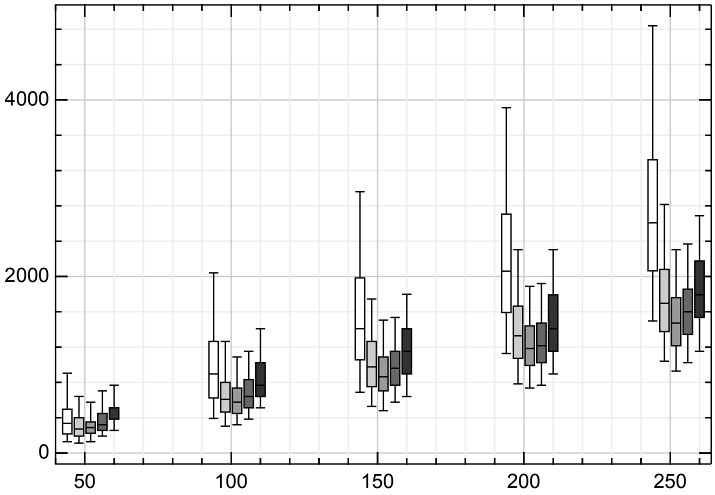
$$\|P_N^n - p(x_{0:T}|y_{0:T})\|_{\mathrm{tv}} \leq \alpha^T \beta^n, \qquad (\forall N > N_0, T, n)$$

Particle number cut-off behaviour of CCBPF



Statistic: Cost of coupling (first time n s.t.) $X_{0:999}[n] = \tilde{X}_{0:999}[n]$

Optimal particle number behaviour of CCBPF



Statistic: Best N for coupling $X_{0:T}[n] = \tilde{X}_{0:T}[n]$

References

1 N. Chopin and S.S. Singh, "On particle Gibbs sampling," *Bernoulli*, 2014.

2 N. Chopin and S.S. Singh, "Stability of Conditional Sequential Monte Carlo," 2014, arXiv:1806.06520

3 S.S. Singh, F. Lindsten and E. Moulines, "Blocking strategies and stability of particle Gibbs samplers," *Biometrika*, 2017.

4 A. Lee, S.S. Singh and M. Vihola, "The Conditional Backward Particle Filter," *ArXiv preprint,* 2018.

5 A. Andrieu, A. Lee and M. Vihola, *Bernoulli*, 2018

6 F. Lindsten, R. Douc and E. Moulines, *Scand. J. Stat.*, 2015.

7 P. W. Glynn and C.H. Rhee., *J. Appl. Probab.*, 2014.

8 P. E. Jacob, F. Lindsten, and T. B. Schon. *Preprint arXiv:1701.02002,* 2017.