Inference in state-space models with multiple paths from conditional SMC

Sinan Yıldırım (Sabancı)  
joint work with  
Christophe Andrieu (Bristol), Arnaud Doucet (Oxford) and Nicolas Chopin (ENSAE)

September 20, 2018
State-space models

\{X_t, t \geq 1\} is a Markov chain

\[X_1 \sim \mu_\theta (\cdot) \text{ and } X_t | (X_{t-1} = x_{t-1}) \sim f_\theta (\cdot | x_{t-1}).\]

We only have access to a conditionally independent process \{Y_t, t \geq 1\}

\[Y_t | (X_t = x_t) \sim g_\theta (\cdot | x_t).\]
We are interested in sampling from the posterior distribution of $\theta$ given $y_{1:n}$:

$$p(\theta | y_{1:n}) \propto \eta(\theta) p_{\theta}(y_{1:n})$$

where $\eta(\theta)$ is the prior density for $\theta$.

### Metropolis-Hastings (MH)

**Given $\theta$,**

1. Propose $\theta' \sim q(\theta, \cdot)$.
2. Accept $\theta'$ with probability

$$\min \left\{ 1, \frac{q(\theta', \theta) \eta(\theta') p_{\theta'}(y_{1:n})}{q(\theta, \theta') \eta(\theta) p_{\theta}(y_{1:n})} \right\},$$

otherwise reject and keep $\theta$. 

Intractability

The MH acceptance ratio:

\[ r(\theta, \theta') := \frac{q(\theta', \theta) \eta(\theta') p_{\theta'}(y_{1:n})}{q(\theta, \theta') \eta(\theta) p_{\theta}(y_{1:n})} \]

Intractable likelihood:

\[ p_{\theta}(y_{1:n}) = \int_{\mathcal{X}^n} \mu_{\theta}(x_1) g_{\theta}(y_1|x_1) \prod_{i=2}^{n} g_{\theta}(y_i|x_i) f_{\theta}(x_i|x_{i-1}) dx_{1:n} \]

There are effective methods to estimate \( p_{\theta}(y_{1:n}) \) unbiasedly, based on sequential Monte Carlo (SMC), aka the particle filter.
Pseudo-marginal approach for state-space models

Replace $p_\theta(y_{1:n})$ with a non-negative unbiased estimator obtained from SMC.

$$\mathbb{E}[\hat{p}_\theta(y_{1:n})] = p_\theta(y_{1:n}), \quad \theta \in \Theta.$$ 

Pseudo-marginal MH (Andrieu and Roberts, 2009; Andrieu et al., 2010)

Given $\theta$ and $\hat{p}_\theta(y_{1:n})$,

1. Propose $\theta' \sim q(\theta, \cdot)$.
2. Run an SMC algorithm to obtain $\hat{p}_{\theta'}(y_{1:n})$.
3. Accept $(\theta', \hat{p}_{\theta'}(y_{1:n}))$ with probability

$$\min \left\{ 1, \frac{q(\theta', \theta)}{q(\theta, \theta')} \frac{\eta(\theta')}{\eta(\theta)} \frac{\hat{p}_{\theta'}(y_{1:n})}{\hat{p}_\theta(y_{1:n})} \right\}$$

otherwise reject and keep $(\theta, \hat{p}_\theta(y_{1:n}))$. 

Scalability issues with pseudo-marginal algorithms

In pseudo-marginal algorithms, one estimates \(p_\theta(y_{1:n})\) and \(p_{\theta'}(y_{1:n})\) independently.

- For fixed number of particles, the variability of \(\hat{p}_\theta(y_{1:n})\) increases with \(n\),
- As \(\theta' \to \theta\) the variability in \(\hat{p}_{\theta'}(y_{1:n})/\hat{p}_\theta(y_{1:n})\) does not vanish.

Variability of the acceptance ratio

\[
\frac{q(\theta', \theta) \eta(\theta') \hat{p}_{\theta'}(y_{1:n})}{q(\theta, \theta') \eta(\theta) \hat{p}_\theta(y_{1:n})}.
\]

has a big impact on the performance (Andrieu and Vihola, 2014, 2015).

Alternatives:

- Estimate \(p_{\theta'}(y_{1:n})/p_\theta(y_{1:n})\) directly.
Estimating the likelihood ratio

Shorthand notation: Let $x = x_{1:n}$ and $y = y_{1:n}$.

Given $\theta$ and $\theta'$, choose $\tilde{\theta} \in \Theta$, e.g. $\tilde{\theta} = (\theta + \theta')/2$.

- If $R_{\tilde{\theta}}$ is a Markov transition kernel leaving $p_{\tilde{\theta}}(\cdot|y)$ invariant, then
  \[
  \frac{p_{\theta'}(y)}{p_{\theta}(y)} = \frac{p_{\tilde{\theta}}(y)}{p_{\theta}(y)} \frac{p_{\theta'}(y)}{p_{\tilde{\theta}}(y)} = \int \int \frac{p_{\theta'}(x', y)}{p_{\tilde{\theta}}(x', y)} p_{\theta}(x|y) R_{\tilde{\theta}}(x, x') dx dx'
  \]


- If further we have reversibility
  \[
  p_{\tilde{\theta}}(x|y) R_{\tilde{\theta}}(x, x') = p_{\tilde{\theta}}(x'|y) R_{\tilde{\theta}}(x', x),
  \]
  then we have a valid algorithm for $p(\theta, x|y) \propto \eta(\theta) p_{\theta}(x, y)$ (Neal, 2004).
AIS based MCMC with one-step annealing

**AIS MCMC for SSM**

1. Given \((\theta, x)\), sample \(\theta' \sim q(\theta, \cdot)\).
2. Let \(\bar{\theta} = (\theta + \theta')/2\) and sample \(x' \sim R_{\bar{\theta}}(x, \cdot)\).
3. Accept \((\theta', x')\) with probability

\[
\min \left\{ 1, \frac{q(\theta', \theta)}{q(\theta, \theta')} \frac{\eta(\theta')}{\eta(\theta)} \frac{p_{\theta}(x, y)}{p_{\theta'}(x', y)} \right\} ;
\]

otherwise reject and keep \((\theta, x)\).
Reversible kernel: conditional SMC (Andrieu et al., 2010)
cSMC and path degeneracy

Picture created by Olivier Cappé.
cSMC with backward sampling (Whiteley, 2010)

cSMC($M, \theta, x$): cSMC at $\theta$ with $M$ particles conditional on the path $x$.

Let $R^M_{\theta}(x, \cdot)$ be the Markov kernel of cSMC($M, \theta, x$)+backward sampling.

- $R^M_{\theta}(x, \cdot)$ is a $p_{\theta}(x|y)$ invariant Markov kernel (Andrieu et al., 2010).

- More importantly, $R^M_{\theta}$ is reversible (Chopin and Singh, 2015).
Numerical experiments

**Non-linear benchmark model**

The latent variables and observations of the state-space model are defined as

\[
X_t = X_{t-1}/2 + 25X_{t-1}/(1 + X_{t-1}^2) + 8\cos(1.2t) + V_t, \quad t \geq 2, \\
Y_t = X_t^2/20 + W_t, \quad t \geq 1,
\]

where \(X_1 \sim \mathcal{N}(0, 10)\), \(V_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_v^2)\), \(W_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_w^2)\).

The static parameter of the model is then

\[
\theta = (\sigma_v^2, \sigma_w^2).
\]
Comparisons with fixed $N$ and varying $n$

$M = 200$ for MwPG (Lindsten et al., 2015) and AIS MCMC and $M = 2000$ for PMMH.

Integrated autocorrelation (IAC) times averaged over 200 runs:

<table>
<thead>
<tr>
<th></th>
<th>AIS MCMC</th>
<th>cSMC+BS</th>
<th>MwPG</th>
<th>PMMH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma^2_v$</td>
<td>$\sigma^2_w$</td>
<td>$\sigma^2_v$</td>
<td>$\sigma^2_w$</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>17.7</td>
<td>23.5</td>
<td>20.9</td>
<td>29.4</td>
</tr>
<tr>
<td>$n = 2000$</td>
<td>17.5</td>
<td>23.7</td>
<td>20.6</td>
<td>29.4</td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>17.6</td>
<td>23.7</td>
<td>20.7</td>
<td>29.6</td>
</tr>
<tr>
<td>$n = 10000$</td>
<td>17.6</td>
<td>24.0</td>
<td>20.7</td>
<td>30.2</td>
</tr>
</tbody>
</table>

IAC times for MwPG with $M = 200$

IAC times for AIS MCMC with $M = 200$

IAC times for PMMH with $M = 2000$
Comparison on a ‘sticky’ model (Pitt and Shephard, 1999)

\[ X_t = 0.98X_{t-1} + V_t, \quad V_t \sim \mathcal{N}(0, 0.02), \quad Y_t = X_t + \mu + W_t, \quad W_t \sim \mathcal{N}(0, 0.1), \]

For this comparison, \( R_{\theta,\theta',k}(x, \cdot) = \pi_{\theta,\theta',k} \) is used.

IAC times averaged over 100 runs for each for the standard deviation \( \sigma_\mu \) of the symmetric RW proposal.
Preliminary theoretical results

Considered the situation

\[
p_\theta(y_{1:n}) := \prod_{t=1}^{n} p_\theta(y_t) = \prod_{t=1}^{n} \int_X p_\theta(x_t, y_t) dx_t
\]

We make some assumptions including

1. \( \Theta \subset \mathbb{R} \) is compact,

2. For any \( x, y \in X \times Y \), \( \theta \mapsto \log p_\theta(x, y) \) is three times differentiable with derivatives uniformly bounded in \( \theta, x, y \).

3. \( \theta' = \theta + \frac{\epsilon}{\sqrt{n}} \) with \( \epsilon \sim q_0(\cdot) \), a symmetric distribution with density bounded away from 0.

4. \( \tilde{\theta} = \frac{\theta + \theta'}{2} \).

5. AIS MCMC that uses a product of \( p_\theta(\cdot | y) \)-invariant reversible Markov transition kernels \( R_{\theta,y}^{M_n} \) with

\[
\lim_{M \to \infty} \sup_{(\theta, x, y) \in \Theta \times X \times Y} \| R_{\theta,y}^M(x, \cdot) - p_\theta(\cdot | y) \|_{tv} = 0.
\]
Some asymptotics

- $\xi = \{X_t, X'_t\}_{t \geq 1}$, $\omega = \{Y_t\}_{t \geq 1}$
- $E^n_\omega$: conditional expectations given observations $\omega$
- $\tilde{r}_n(\theta, \theta'; \xi)$: estimated accepted ratio used in AIS MCMC.
- $r_n(\theta, \theta')$: The exact acceptance ratio of the MH.
Some asymptotics

- $\xi = \{X_t, X'_t\}_{t \geq 1}$, $\omega = \{Y_t\}_{t \geq 1}$
- $E^\omega$: conditional expectations given observations $\omega$
- $\tilde{r}_n(\theta, \theta'; \xi)$: estimated accepted ratio used in AIS MCMC.
- $r_n(\theta, \theta')$: The exact acceptance ratio of the MH.

Theorem

Under some assumptions... $P$–a.s., for any $\varepsilon_0 > 0$ there exist $n_0, M_0 \in \mathbb{N}$ such that for any sequence $\{M_n\} \in \mathbb{N}^\mathbb{N}$ satisfying $M_n \geq M_0$ for $n \geq n_0$,

$$\sup_{n \geq n_0} \left| E^\omega_n \left[ \min\{1, \tilde{r}_n(\theta, \theta', \xi)\} \right] - E^\omega_n \left[ \min\{1, r_n(\theta, \theta') \exp(Z)\} \right] \right| \leq \varepsilon_0,$$

where

$$Z|(\theta, \theta', \omega) \sim \mathcal{N}\left(-\frac{\sigma^2(\theta, \theta')}{2}, \sigma^2(\theta, \theta')\right),$$

$$\sigma^2(\theta, \theta') := \frac{-(\theta' - \theta)^2}{2} \mathbb{E}_{\tilde{\theta}} \left[ \partial_{\tilde{\theta}}^2 \log p_{\tilde{\theta}}(X|Y) \right], \text{ and } \tilde{\theta} = (\theta + \theta')/2.$$
More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider

$$\mathcal{P}_{\theta, \theta', K} := \{\pi_{\theta, \theta', k}, \quad k = 0, \ldots, K + 1\}$$

where

$$\pi_{\theta, \theta', k}(x) := p_{\theta_k}(x|y) \propto p_{\theta_k}(x, y) =: \gamma_{\theta, \theta', k}(x),$$

with annealing

$$\theta_k = \theta k/(K + 1) + \theta'(1 - k/(K + 1))$$
More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider

  \[ \mathcal{P}_{\theta, \theta', K} := \{ \pi_{\theta, \theta', k}, \quad k = 0, \ldots, K + 1 \} \]

  where

  \[ \pi_{\theta, \theta', k}(x) := p_{\theta_k}(x|y) \propto p_{\theta_k}(x, y) =: \gamma_{\theta, \theta', k}(x), \]

  with annealing

  \[ \theta_k = \theta k/(K + 1) + \theta'(1 - k/(K + 1)) \]

- Also, consider the associated Markov kernels

  \[ \mathcal{R}_{\theta, \theta', K} := \{ R_{\theta, \theta', k} : X^n \times X^{\otimes n} \to [0, 1], \quad k = 1, \ldots, K \} \].
More general AIS MCMC with $K \geq 1$ intermediate steps

Consider

$$\mathcal{P}_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0, \ldots, K + 1\}$$

where

$$\pi_{\theta,\theta',k}(x) := p_{\theta_k}(x|y) \propto p_{\theta_k}(x, y) =: \gamma_{\theta,\theta',k}(x),$$

with annealing

$$\theta_k = \theta k/(K + 1) + \theta'(1 - k/(K + 1))$$

Also, consider the associated Markov kernels

$$\mathcal{R}_{\theta,\theta',K} := \{R_{\theta,\theta',k} : X^n \times X^\otimes n \to [0, 1], \quad k = 1, \ldots, K\}.$$

In this case the estimator is of the form

$$\prod_{k=0}^{K} \frac{\gamma_{\theta,\theta',k+1}(x_k)}{\gamma_{\theta,\theta',k}(x_k)}$$

where $x_1 \sim R_{\theta,\theta',k}(x_0, \cdot)$ and $x_k \sim R_{\theta,\theta',k}(x_{k-1}, \cdot), \quad k = 2, \ldots, K$. 
More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider

$$P_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0, \ldots, K + 1\}$$

where

$$\pi_{\theta,\theta',k}(x) := p_{\theta_k}(x|y) \propto p_{\theta_k}(x, y) =: \gamma_{\theta,\theta',k}(x),$$

with annealing

$$\theta_k = \theta k/(K + 1) + \theta'(1 - k/(K + 1))$$

- Also, consider the associated Markov kernels

$$R_{\theta,\theta',K} := \{R_{\theta,\theta',k} : X^n \times X^k \rightarrow [0, 1], \quad k = 1, \ldots, K\}.$$

- In this case the estimator is of the form

$$\prod_{k=0}^{K} \frac{\gamma_{\theta,\theta',k+1}(x_k)}{\gamma_{\theta,\theta',k}(x_k)}$$

where $x_1 \sim R_{\theta,\theta',k}(x_0, \cdot)$ and $x_k \sim R_{\theta,\theta',k}(x_{k-1}, \cdot), \quad k = 2, \ldots, K$.

With $\theta', x_K$ being the proposal, we can design a correct algorithm.
Response to increasing budget: fixed $n$ and varying $M, K$

We fixed $n = 500$, run each combination 200 times. On each row:

- MCMC AIS with $M_0 = 500$ particles $K$ intermediate steps,
- MwPG and PMMH algorithms for $M = KM_0$ particles.

<table>
<thead>
<tr>
<th>$K$</th>
<th>AIS</th>
<th>MCMC</th>
<th>cSMC+BS</th>
<th>MwPG</th>
<th>PMMH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma^2_v$</td>
<td>$\sigma^2_w$</td>
<td>$\sigma^2_v$</td>
<td>$\sigma^2_w$</td>
<td>$\sigma^2_v$</td>
</tr>
<tr>
<td>$K = 1$</td>
<td>17.7</td>
<td>20.9</td>
<td>22.9</td>
<td>29.8</td>
<td>161.9</td>
</tr>
<tr>
<td>$K = 2$</td>
<td>14.5</td>
<td>15.7</td>
<td>22.1</td>
<td>28.4</td>
<td>41.8</td>
</tr>
<tr>
<td>$K = 3$</td>
<td>13.9</td>
<td>15.6</td>
<td>22.8</td>
<td>28.1</td>
<td>22.6</td>
</tr>
<tr>
<td>$K = 4$</td>
<td>15.0</td>
<td>15.9</td>
<td>20.0</td>
<td>31.1</td>
<td>19.0</td>
</tr>
<tr>
<td>$K = 5$</td>
<td>13.4</td>
<td>14.9</td>
<td>20.4</td>
<td>25.8</td>
<td>18.9</td>
</tr>
<tr>
<td>$K = 6$</td>
<td>13.0</td>
<td>13.1</td>
<td>20.8</td>
<td>26.3</td>
<td>16.9</td>
</tr>
<tr>
<td>$K = 7$</td>
<td>13.7</td>
<td>12.4</td>
<td>18.3</td>
<td>26.5</td>
<td>16.6</td>
</tr>
<tr>
<td>$K = 8$</td>
<td>13.7</td>
<td>12.6</td>
<td>22.7</td>
<td>27.6</td>
<td>14.3</td>
</tr>
<tr>
<td>$K = 9$</td>
<td>12.0</td>
<td>12.2</td>
<td>21.9</td>
<td>29.8</td>
<td>16.3</td>
</tr>
<tr>
<td>$K = 10$</td>
<td>13.5</td>
<td>13.7</td>
<td>22.7</td>
<td>26.7</td>
<td>14.9</td>
</tr>
</tbody>
</table>
Using all paths of cSMC

- Law of \( k := (k_1, \ldots, k_n) \) in backward sampling conditional on \( \zeta = x_{1:n}^{(1:M)} \):

\[
\phi_{\tilde{\theta}}(k \mid \zeta)
\]

Resulting path: \( \zeta^{(k)} \).

- For any \( \theta, \theta', \tilde{\theta} \in \Theta \), and paths \( x, x' \in X^n \), define

\[
\hat{r}_{x,x'}(\theta, \theta'; \tilde{\theta}) = \frac{q(\theta', \theta) \eta(\theta') p_{\theta'}(x', y) p_{\tilde{\theta}}(x, y)}{q(\theta, \theta') \eta(\theta) p_{\tilde{\theta}}(x', y) p_{\theta}(x, y)}.
\]

**Theorem: Unbiased estimator of acceptance ratio**

Let \( x \sim p_{\theta}(\cdot | y) \), \( \zeta | x \sim \text{cSMC}(M, \tilde{\theta}, x) \).

The expectation

\[
\sum_{k \in [M]^n} \phi_{\tilde{\theta}}(k \mid \zeta) \hat{r}_{x,\zeta^{(k)}}(\theta, \theta'; \tilde{\theta})
\]

is an unbiased estimator of \( r(\theta, \theta') \).
Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0, 1)$

2 Set $\tilde{\theta} = (\theta + \theta')/2$.

3 if $v \leq 1/2$ then

4 Run a cSMC($M, \tilde{\theta}, x$) to obtain $\zeta$.

5 Set $x' = \zeta^{(k)}$ w.p. $\phi_{\tilde{\theta}}(k|\zeta)\hat{r}_{x,\zeta^{(k)}}(\theta, \theta'; \tilde{\theta})$.

6 Accept $(\theta', x')$ with probability

$$\min \left\{ 1, \sum_{k \in [M]^n} \phi_{\tilde{\theta}}(k|\zeta)\hat{r}_{x,\zeta^{(k)}}(\theta, \theta'; \tilde{\theta}) \right\}$$

otherwise reject.
Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0, 1)$
Set $\bar{\theta} = (\theta + \theta')/2$.

\begin{enumerate}
\item[3] if $v \leq 1/2$ then
\begin{enumerate}
\item Run a cSMC($M, \bar{\theta}, x$) to obtain $\zeta$.
\item Set $x' = \zeta^{(k)}$ w.p. $\propto \phi_{\bar{\theta}}(k | \zeta) \hat{r}_{x, \zeta^{(k)}}(\theta, \theta'; \bar{\theta})$.
\item Accept $(\theta', x')$ with probability $\min \left\{ 1, \sum_{k \in [M]^n} \phi_{\bar{\theta}}(k | \zeta) \hat{r}_{x, \zeta^{(k)}}(\theta, \theta'; \bar{\theta}) \right\}$
\end{enumerate}

\item[7] else
\begin{enumerate}
\item Run a cSMC($M, \bar{\theta}, x$) to obtain $\zeta$.
\item Set $x' = \zeta^{(k)}$ with probability $\phi_{\bar{\theta}}(k | \zeta)$
\item Accept $(\theta', x')$ with probability $\min \left\{ 1, \left[ \sum_{k \in [M]^n} \phi_{\bar{\theta}}(k | \zeta) \hat{r}_{x', \zeta^{(k)}}(\theta', \theta; \bar{\theta}) \right]^{-1} \right\}$,
\end{enumerate}

\end{enumerate}
otherwise reject.
Theoretical results

Theorem: Exactness of the algorithm

The presented algorithm targets $p(\theta, x|y)$.

Corollary (to the exactness of the algorithm)

Let $x \sim p_\theta(\cdot|y)$, $\zeta|x \sim \text{cSMC}(M, \tilde{\theta}, x)$.

Let $x'$ be drawn with backward sampling conditional upon $\zeta$.

Then,

$$
\left[ \sum_{k \in [M]^n} \phi_{\tilde{\theta}}(k|\zeta) \hat{r}_{x',\zeta(k)}(\theta', \theta; \tilde{\theta}) \right]^{-1}
$$

is an unbiased estimator of $r(\theta, \theta')$. 
The computations needed can be performed with a complexity of $O(M^2 n)$

- Acceptance ratios can be performed by a sum-product algorithm.
- Sampling a path $\zeta^{(k)}$ can be performed with a forward-filtering backward-sampling algorithm.

However, $O(M^2 n)$ can still be overwhelming, especially when $M$ is large.
Let \( u^{(1)}, \ldots, u^{(N)} \) be independently sampled paths via backward sampling conditional on \( \zeta \).

We can still target \( p(\theta, x|y) \) using

\[
\frac{1}{N} \sum_{i=1}^{N} \hat{r}_{x,u(i)}(\theta, \theta'; \tilde{\theta}),
\]

which is an unbiased estimator of \( r(\theta, \theta') \).

Computational complexity: \( O(NMn) \) per iteration instead of \( O(M^2n) \);

Moreover, sampling \( N \) paths can be parallelised.
1. Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0, 1)$.
2. Set $\tilde{\theta} = (\theta + \theta')/2$.
3. \textbf{if} $v \leq 1/2$ \textbf{then}
   \begin{enumerate}[noitemsep,nolistsep]
   
   \item Run a cSMC($M, \tilde{\theta}, x$) to obtain particles $\zeta$.
   \item Draw $N$ paths with backward sampling, $u(1), \ldots, u(N)$.
   \item Set $x' = u^{(k)}$ w.p. $\propto \hat{r}_{x, u^{(k)}}(\theta, \theta'; \tilde{\theta})$
   \item Accept $(\theta', x')$ with probability
     \[
     \min \left\{ 1, \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{x, u^{(i)}}(\theta, \theta'; \tilde{\theta}) \right\};
     \]
   \end{enumerate}
4. \textbf{else}
   \begin{enumerate}[noitemsep,nolistsep]
   
   \item Run a cSMC($M, \tilde{\theta}, x$) to obtain particles $\zeta$.
   \item Set $u^{(1)} = x$
   \item Draw $N$ paths with backward sampling $x', u^{(2)}, u^{(3)}, \ldots, u^{(N)}$.
   \item Accept $(\theta', x')$ with probability
     \[
     \min \left\{ 1, \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{x', u^{(i)}}(\theta', \theta'; \tilde{\theta}) \right\};
     \]
   \end{enumerate}
5. \textbf{otherwise reject, and keep $(\theta, x)$}.
1. Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0, 1)$.

2. Set $\tilde{\theta} = (\theta + \theta')/2$.

3. **if** $v \leq 1/2$ **then**
   - Run a cSMC($M, \tilde{\theta}, x$) to obtain particles $\zeta$.
   - Draw $N$ paths with backward sampling, $u^{(1)}, \ldots, u^{(N)}$.
   - Set $x' = u^{(k)}$ w.p. $\propto r_{x, u^{(k)}}(\theta, \theta'; \tilde{\theta})$
   - Accept $(\theta', x')$ with probability
     \[
     \min \left\{ 1, \frac{1}{N} \sum_{i=1}^{N} r_{x, u^{(i)}}(\theta, \theta'; \tilde{\theta}) \right\} ;
     \]
   - **else** reject, and keep $(\theta, x)$.

4. **else**
   - Run a cSMC($M, \tilde{\theta}, x$) to obtain particles $\zeta$.
   - Set $u^{(1)} = x$.
   - Draw $N$ paths with backward sampling $x', u^{(2)}, u^{(3)}, \ldots, u^{(N)}$.
   - Accept $(\theta', x')$ with probability
     \[
     \min \left\{ 1, \left[ \frac{1}{N} \sum_{i=1}^{N} r_{x', u^{(i)}}(\theta', \theta'; \tilde{\theta}) \right]^{-1} \right\} ;
     \]
   - **else** reject, and keep $(\theta, x)$. 
Theorem: Exactness

The presented algorithm targets $p(\theta, x|y)$.

Corollary (to the exactness of the algorithm)

Let $x \sim p_{\theta}(\cdot|y)$, $\zeta|x \sim \text{cSMC}(M, \tilde{\theta}, x)$.

Let $u^{(1)} = x$ and $x', u^{(2)}, u^{(3)}, \ldots, u^{(N)}$ be $N$ independent paths sampled with backward sampling conditioned on $\zeta$.

Then

$$
\left[ \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{x', u^{(i)}}(\theta', \theta; \tilde{\theta}) \right]^{-1}
$$

is an unbiased estimator of $r(\theta, \theta')$. 
Numerical example: IAC time

<table>
<thead>
<tr>
<th></th>
<th>MwPG</th>
<th>N = 1</th>
<th>N = 10</th>
<th>N = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_v^2$</td>
<td>14</td>
<td>16</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>$\sigma_w^2$</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

IAC times of algorithms, $n = 500$, $M = 100$
Numerical example: convergence vs IAC time

**conv. to post. mean - single cSMC**

- **ensemble average for $\sigma_w^2(k)$**
- **conv. to post. mean square - single cSMC**
- **conv. to post. median - single cSMC**

IAC times for $\sigma_w^2$ - single cSMC
- means: 35.51, 28.16, 25.10

IAC times for $\sigma_w^4$ - single cSMC
- means: 34.20, 26.82, 24.04
Discussion

- Discussed AIS based MCMC for state-space models.
  “Scalable Monte Carlo inference for state-space models”, arXiv:1809.02527
- Discussed the use of multiple, (or all possible) paths from cSMC for AIS based MCMC for state-space models.
- If done in parallel, using multiple paths from cSMC can be beneficial in terms of
  ▶ convergence time
  ▶ IAC time (?
- The methodology presented for state-space models are a special case in a more general framework:
  ▶ Designing MH algorithms with asymmetric acceptance ratios can be useful in other models such as
    ▶ general latent variable models
    ▶ trans-dimensional models
    ▶ doubly intractable models

Thank you!
References:


