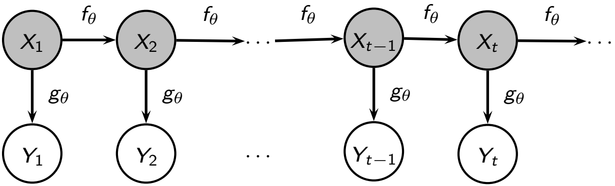
Inference in state-space models with multiple paths from conditional SMC

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 $\{Y_t, t \geq 1\}$

$$Y_t | (X_t = x_t) \sim g_\theta(\cdot | x_t).$$

The Metropolis-Hastings algorithm

We are interested in sampling from the posterior distribution of θ 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 θ.

Metropolis-Hastings (MH)

Given θ,

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

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

otherwise reject and keep θ.

Intractability

The MH acceptance ratio:

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

Intractable likelihood:

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

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 an 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 θ 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

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

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

Alternatives:

- Correlated pseudo-marginal algorithm of Deligiannidis et al. (2018).
- 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 θ and θ', 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

$$\frac{p_{\theta'}(y)}{p_\theta(y)} = \frac{p_{\tilde{\theta}}(y)}{p_\theta(y)} \frac{p_{\theta'}(y)}{p_{\tilde{\theta}}(y)} = \int \int \frac{p_{\theta'}(x', y)}{p_{\tilde{\theta}}(x', y)} p_\theta(x|y) R_{\tilde{\theta}}(x, x') dx dx'$$

  Variance reduction (Crooks, 1998; Jarzynski, 1997; Neal, 2001, 1996).

- 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

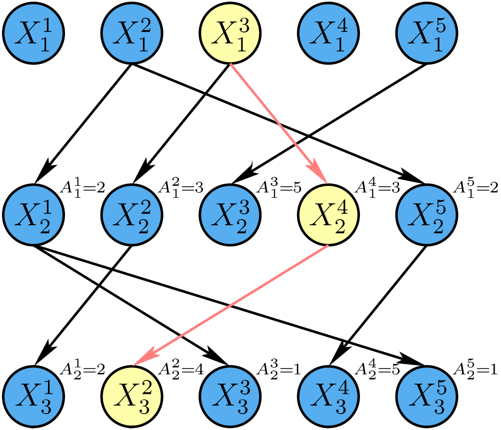
AIS MCMC for SSM

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

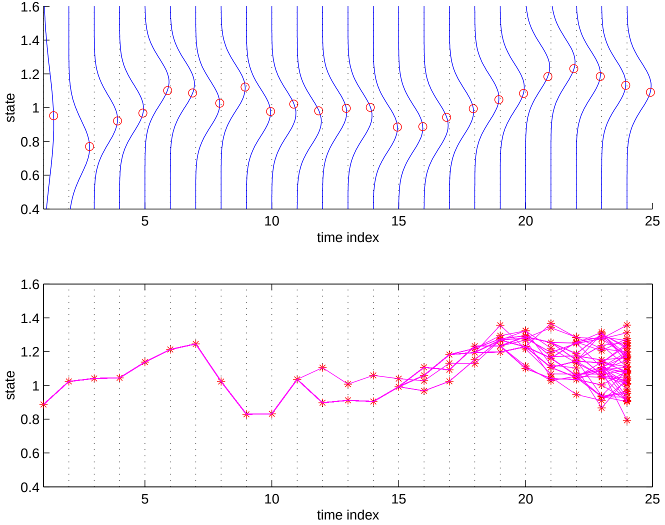
$$\min\left\{1, \frac{q(\theta', \theta)}{q(\theta, \theta')} \frac{\eta(\theta')}{\eta(\theta)} \frac{p_{\tilde{\theta}}(x, y)}{p_{\theta}(x, y)} \frac{p_{\theta'}(x', y)}{p_{\tilde{\theta}}(x', y)}\right\};$$

otherwise reject and keep (θ, x).

Reversible kernel: conditional SMC (Andrieu et al., 2010)



cSMC and path degeneracy



Picture created by Olivier Cappé.

cSMC with backward sampling (Whiteley, 2010)

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

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

- $R_\theta^M(x, \cdot)$ is a $p_\theta(x|y)$ invariant Markov kernel (Andrieu et al., 2010).
- More importantly, R_θ^M 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

$$\begin{aligned} X_t &= X_{t-1}/2 + 25X_{t-1}/(1+X_{t-1}^2) + 8\cos(1.2t) + V_t, \quad t \geq 2, \\ Y_t &= X_t^2/20 + W_t, \quad t \geq 1, \end{aligned}$$

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

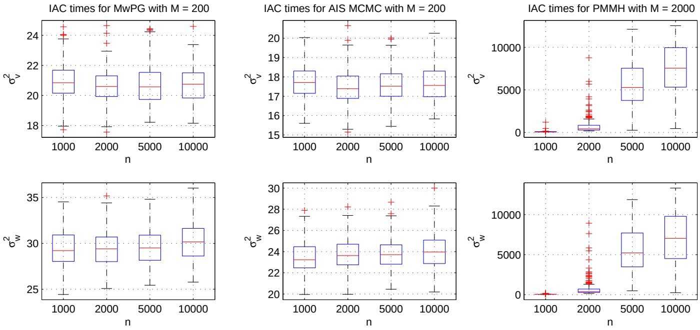
$$\theta = (\sigma_v^2, \sigma_w^2).$$

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:

	AIS MCMC cSMC+BS		MwPG		PMMH	
	σ_v^2	σ_w^2	σ_v^2	σ_w^2	σ_v^2	σ_w^2
$n = 1000$	17.7	23.5	20.9	29.4	71.3	59.2
$n = 2000$	17.5	23.7	20.6	29.4	759.0	757.9
$n = 5000$	17.6	23.7	20.7	29.6	5808.6	5663.5
$n = 10000$	17.6	24.0	20.7	30.2	7368.1	7170.9

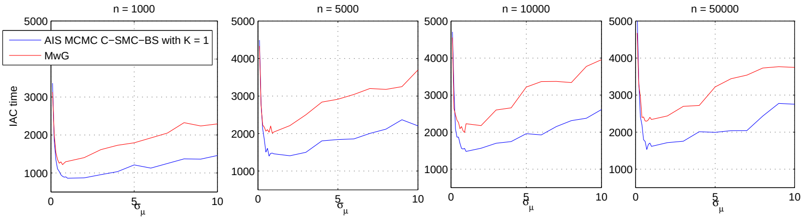


Comparison on a 'sticky' model (Pitt and Shephard, 1999)

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

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 σ_μ of the symmetric RW proposal.



Preliminary theoretical results

Considered the situation

$$p_\theta(y_{1:n}) := \prod_{t=1}^{n} p_\theta(y_t) = \prod_{t=1}^{n} \int_{\mathsf{X}} p_\theta(x_t, y_t) dx_t$$

We make some assumptions including

1. $\Theta \subset \mathbb{R}$ is compact,
2. For any $x, y \in \mathsf{X} \times \mathsf{Y}$, $\theta \mapsto \log p_\theta(x, y)$ is three times differentiable with derivatives uniformly bounded in θ, 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

$$\lim_{M \to \infty} \sup_{(\theta,x,y) \in \Theta \times \mathsf{X} \times \mathsf{Y}} \| R_{\theta,y}^{M}\left(x, \cdot\right) - p_\theta(\cdot \mid y) \|_{tv} = 0.$$

Some asymptotics

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

Some asymptotics

- $\xi = \{X_t, X'_t\}_{t\geq 1}$, $\omega = \{Y_t\}_{t\geq 1}$,
- $\mathbb{E}_n^\omega$: conditional expectations given observations ω
- $\tilde{r}_n(\theta, \theta'; \xi)$: 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 $\left\{M_n\right\} \in \mathbb{N}^{\mathbb{N}}$ satisfying $M_n \geq M_0$ for $n \geq n_0$,

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

where

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

$\sigma^2(\theta, \theta') := \frac{-(\theta' - \theta)^2}{2} \mathbb{E}_{\tilde{\theta}} \left[\partial^2_{\tilde{\theta}} \log p_{\tilde{\theta}}\left(X|Y\right)\right]$, and $\tilde{\theta} = (\theta + \theta')/2$.

More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider
$$\mathscr{P}_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0, \ldots, K+1\}$$
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
$$\mathscr{P}_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0, \dots, K+1\}$$
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
$$\mathscr{R}_{\theta,\theta',K} := \left\{R_{\theta,\theta',k} : \mathsf{X}^n \times \mathcal{X}^{\otimes n} \to [0,1], \quad k = 1, \dots, K\right\}.$$

More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider
$$\mathscr{P}_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0,\ldots,K+1\}$$
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
$$\mathscr{R}_{\theta,\theta',K} := \left\{R_{\theta,\theta',k} : \mathsf{X}^n \times \mathcal{X}^{\otimes n} \to [0,1], \quad k = 1,\ldots,K\right\}.$$

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

More general AIS MCMC with $K \geq 1$ intermediate steps

- Consider
$$\mathscr{P}_{\theta,\theta',K} := \{\pi_{\theta,\theta',k}, \quad k = 0, \ldots, K+1\}$$
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
$$\mathscr{R}_{\theta,\theta',K} := \left\{R_{\theta,\theta',k} : \mathsf{X}^n \times \mathcal{X}^{\otimes n} \to [0,1], \quad k = 1, \ldots, K\right\}.$$

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

With θ', 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.

	AIS MCMC cSMC+BS		MwPG		PMMH	
	σ_v^2	σ_w^2	σ_v^2	σ_w^2	σ_v^2	σ_w^2
$K = 1$	17.7	20.9	22.9	29.8	161.9	309.3
$K = 2$	14.5	15.7	22.1	28.4	41.8	43.5
$K = 3$	13.9	15.6	22.8	28.1	22.6	21.6
$K = 4$	15.0	15.9	20.0	31.1	19.0	19.3
$K = 5$	13.4	14.9	20.4	25.8	18.9	17.5
$K = 6$	13.0	13.1	20.8	26.3	16.9	16.0
$K = 7$	13.7	12.4	18.3	26.5	16.6	14.1
$K = 8$	13.7	12.6	22.7	27.6	14.3	13.7
$K = 9$	12.0	12.2	21.9	29.8	16.3	14.0
$K = 10$	13.5	13.7	22.7	26.7	14.9	14.0

Using all paths of cSMC

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

$$\phi_{\tilde{\theta}}(\mathbf{k} \mid \zeta)$$

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

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

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

Theorem: Unbiased estimator of acceptance ratio

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

$$\sum_{\mathbf{k} \in [M]^n} \phi_{\tilde{\theta}}(\mathbf{k}|\zeta) \mathring{r}_{x,\zeta^{(\mathbf{k})}}(\theta, \theta'; \tilde{\theta})$$

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

Sample $\theta' \sim \boldsymbol{q}(\theta, \cdot)$ and $v \sim \mathcal{U}(0,1)$

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

if $v \leq 1/2$ **then**

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

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

  Accept (θ', x') with probability

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

  otherwise reject.

Sample $\theta' \sim \boldsymbol{q}(\theta, \cdot)$ and $v \sim \mathcal{U}(0,1)$

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

if $v \leq 1/2$ **then**

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

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

Accept (θ', x') with probability

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

otherwise reject.

else

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

Set $x' = \zeta^{(\mathbf{k})}$ with probability $\phi_{\tilde{\theta}}(\mathbf{k}|\zeta)$

Accept (θ', x') with probability

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

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

Let x' be drawn with backward sampling conditional upon ζ.

Then,

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

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

Computation load

The computations needed can be performed with a complexity of $\mathcal{O}(M^2 n)$

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

However, $\mathcal{O}(M^2 n)$ can still be overwhelming, especially when M is large.

Reduced computation via subsampling

- Let $u^{(1)}, \ldots, u^{(N)}$ be independently sampled paths via backward sampling conditional on ζ.
- We can still target $p(\theta, x|y)$ using

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

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

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

Moreover, sampling N paths can be parallelised.

Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0,1)$.
Set $\tilde{\theta} = (\theta + \theta')/2$.
if $v \leq 1/2$ **then**
    Run a $\mathrm{cSMC}(M, \tilde{\theta}, x)$ to obtain particles ζ.
    Draw N paths with backward sampling, $u^{(1)}, \ldots, u^{(N)}$.
    Set $x' = u^{(k)}$ w.p. $\propto \mathring{r}_{x,u^{(k)}}(\theta, \theta'; \tilde{\theta})$
    Accept (θ', x') with probability

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

    otherwise reject, and keep (θ, x).

Sample $\theta' \sim q(\theta, \cdot)$ and $v \sim \mathcal{U}(0,1)$.
Set $\tilde{\theta} = (\theta + \theta')/2$.
if $v \leq 1/2$ **then**
    Run a $\mathrm{cSMC}(M, \tilde{\theta}, x)$ to obtain particles ζ.
    Draw N paths with backward sampling, $u^{(1)}, \ldots, u^{(N)}$.
    Set $x' = u^{(k)}$ w.p. $\propto \mathring{r}_{x,u^{(k)}}(\theta, \theta'; \tilde{\theta})$
    Accept (θ', x') with probability

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

    otherwise reject, and keep (θ, x).
else
    Run a $\mathrm{cSMC}(M, \tilde{\theta}, x)$ to obtain particles ζ.
    Set $u^{(1)} = x$
    Draw N paths with backward sampling $x', u^{(2)}, u^{(3)}, \ldots, u^{(N)}$.
    Accept (θ', x') with probability

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

    otherwise reject, and keep (θ, x).

Theorem: Exactness

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

Corollary (to the exactness of the algorithm)

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

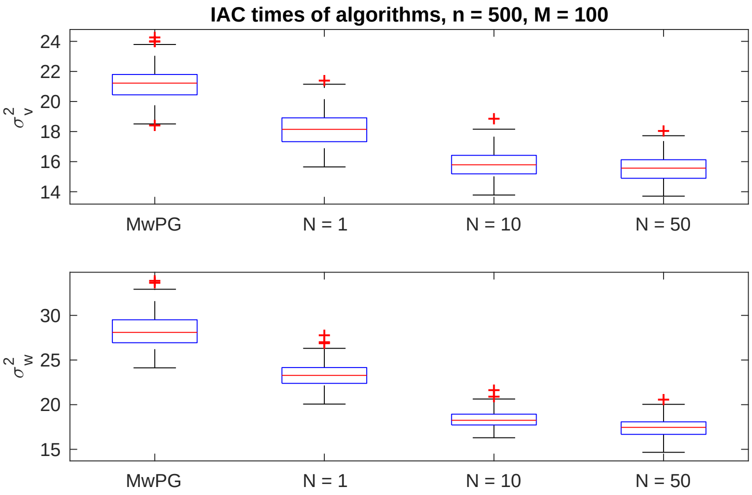
Let $u^{(1)} = x$ and $x', u^{(2)}, u^{(3)}, \ldots, u^{(N)}$ be N independent paths sampled with backward sampling conditioned on ζ.

Then

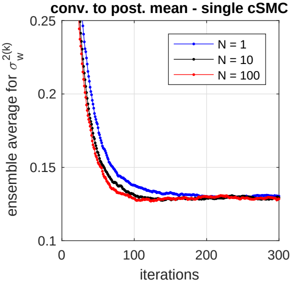
$$\left[\frac{1}{N}\sum_{i=1}^{N} \mathring{r}_{x', u^{(i)}}(\theta', \theta; \tilde{\theta})\right]^{-1}$$

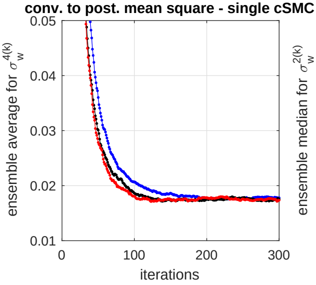
is an unbiased estimator of $r(\theta, \theta')$.

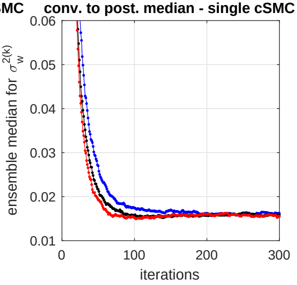
Numerical example: IAC time

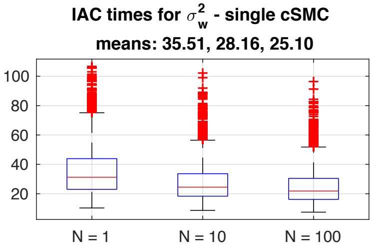


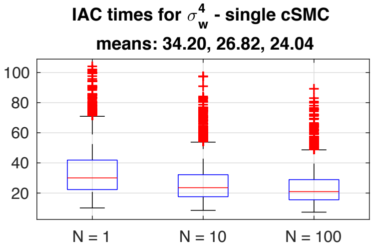
Numerical example: convergence vs IAC time











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

"On the utility of Metropolis-Hastings with asymmetric acceptance ratio" , arXiv:1803.09527

Thank you!

References:

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:269–342.

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, 37(2):569–1078.

Andrieu, C. and Vihola, M. (2014). Establishing some order amongst exact approximations of MCMCs. *arXiv:1404.6909*.

Andrieu, C. and Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability*, 25(2):1030–1077.

Chopin, N. and Singh, S. (2015). On particle Gibbs sampling. *Bernoulli*, 21(3):1855–1883.

Crooks, G. (1998). Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5-6):1481–1487.

Deligiannidis, G., Doucet, A., and Pitt, M. K. (2018). The correlated pseudomarginal method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) (forthcoming)*.

Jarzynski, C. (1997). Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018?5035.

Lindsten, F., Douc, R., , and Moulines, E. (2015). Uniform ergodicity of the particle Gibbs sampler. *Scandinavian Journal of Statistics*, 42(3):775?797.

Neal, R. (1996). Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6(4):353–366.

Neal, R. (2001). Annealed importance sampling. *Statistics and Computing*, 11:125–139.

Neal, R. M. (2004). Taking bigger Metropolis steps by dragging fast variables. Technical report, University of Toronto.

Pitt, M. K. and Shephard, N. (1999). Analytic convergence rates and parametrization issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis*, 20(1):63–85.

Whiteley, N. (2010). Discussion of 'Particle Markov chain Monte Carlo methods' by Andrieu et al. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):306–307.