Particle Markov Chain Monte Carlo Methods

Arnaud Doucet
University of British Columbia, Vancouver, Canada

Kyoto, 15th June 2011

General State-Space Models

- State-space models also known as Hidden Markov models are ubiquitous time series models in ecology, econometrics, engineering, statistics etc.

General State-Space Models

- State-space models also known as Hidden Markov models are ubiquitous time series models in ecology, econometrics, engineering, statistics etc.
- Let $\{X_n\}_{n \geq 1}$ be a latent/hidden Markov process defined by

$$X_1 \sim \mu_\theta(\cdot) \text{ and } X_n | (X_{n-1} = x_{n-1}) \sim f_\theta(\cdot | x_{n-1}) \ .$$

General State-Space Models

- State-space models also known as Hidden Markov models are ubiquitous time series models in ecology, econometrics, engineering, statistics etc.
- Let $\{X_n\}_{n\geq 1}$ be a latent/hidden Markov process defined by

$$X_1 \sim \mu_\theta(\cdot) \text{ and } X_n | (X_{n-1} = x_{n-1}) \sim f_\theta(\cdot | x_{n-1}) .$$

- We only have access to a process $\{Y_n\}_{n\geq 1}$ such that, conditional upon $\{X_n\}_{n\geq 1}$, the observations are statistically independent and

$$Y_n | (X_n = x_n) \sim g_\theta(\cdot | x_n) .$$

General State-Space Models

- State-space models also known as Hidden Markov models are ubiquitous time series models in ecology, econometrics, engineering, statistics etc.
- Let $\{X_n\}_{n\geq 1}$ be a latent/hidden Markov process defined by

$$X_1 \sim \mu_\theta(\cdot) \text{ and } X_n | (X_{n-1} = x_{n-1}) \sim f_\theta(\cdot | x_{n-1}) .$$

- We only have access to a process $\{Y_n\}_{n\geq 1}$ such that, conditional upon $\{X_n\}_{n\geq 1}$, the observations are statistically independent and

$$Y_n | (X_n = x_n) \sim g_\theta(\cdot | x_n) .$$

- θ is an *unknown* parameter of prior density $p(\theta)$.

Examples of State-Space Models

- *Canonical univariate SV model* (Ghysels et al., 1996)

$$X_n = \alpha + \phi \left(X_{n-1} - \alpha \right) + \sigma V_n,$$
$$Y_n = \exp \left(X_n / 2 \right) W_n,$$

where $X_1 \sim \mathcal{N} \left(\alpha, \sigma^2 / \left(1 - \phi^2 \right) \right)$, $V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N} \left(0, 1 \right)$ and $W_m \overset{\text{i.i.d.}}{\sim} \mathcal{N} \left(0, 1 \right)$ and $\theta = \left(\alpha, \phi, \sigma \right)$.

Examples of State-Space Models

- *Canonical univariate SV model* (Ghysels et al., 1996)

$$X_n = \alpha + \phi \left(X_{n-1} - \alpha \right) + \sigma V_n,$$
$$Y_n = \exp \left(X_n / 2 \right) W_n,$$

where $X_1 \sim \mathcal{N} \left(\alpha, \sigma^2 / \left(1 - \phi^2 \right) \right)$, $V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $W_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\theta = (\alpha, \phi, \sigma)$.

- *Wishart processes for multivariate SV* (Gourieroux et al., 2009)

$$X_n^m = M X_{n-1}^m + V_n^m, \; V_n^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Xi), \; m = 1, ..., K$$
$$\Sigma_n = \sum_{m=1}^K X_n^m \left(X_n^m \right)^{\mathsf{T}},$$
$$Y_n | \Sigma_n \sim \mathcal{N}(0, \Sigma_n).$$

where $\theta = (M, \Xi)$.

Examples of State-Space Models

- *U.S./U.K. exchange rate model* (Engle & Kim, 1999). Log exchange rate values Y_n are modeled through

$$
\begin{aligned}
Y_n &= \alpha_n + \eta_n, \\
\alpha_n &= \alpha_{n-1} + \sigma_\alpha V_{n,1}, \\
\eta_n &= a_1 \eta_{n-1} + a_2 \eta_{n-2} + \sigma_{\eta, Z_n} V_{n,2}
\end{aligned}
$$

where $V_{n,1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $V_{n,2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $Z_n \in \{1,2,3,4\}$ is an unobserved Markov chain of unknown transition matrix.

Examples of State-Space Models

- *U.S./U.K. exchange rate model* (Engle & Kim, 1999). Log exchange rate values Y_n are modeled through

$$\begin{aligned} Y_n &= \alpha_n + \eta_n, \\ \alpha_n &= \alpha_{n-1} + \sigma_\alpha V_{n,1}, \\ \eta_n &= a_1 \eta_{n-1} + a_2 \eta_{n-2} + \sigma_{\eta, Z_n} V_{n,2} \end{aligned}$$

  where $V_{n,1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $V_{n,2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $Z_n \in \{1,2,3,4\}$ is an unobserved Markov chain of unknown transition matrix.
- This can be reformulated as a state-space by selecting $X_n = \begin{bmatrix} \alpha_n & \eta_n & \eta_{n-1} & Z_n \end{bmatrix}^{\mathsf{T}}$ and $\theta = (a_1, a_2, \sigma_\alpha, \sigma_{1:4}, P)$.

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).
- **Econometrics**: stochastic volatility models, nonlinear term structures (Li, *JBES*, 2011; Giordani, Kohn & Pitt, *JCGS*, 2011; Andreasen 2011)

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).
- **Econometrics**: stochastic volatility models, nonlinear term structures (Li, *JBES*, 2011; Giordani, Kohn & Pitt, *JCGS*, 2011; Andreasen 2011)
- **Epidemiology**: disease dynamic models (Ionides et al., *JASA*, 2011).

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).
- **Econometrics**: stochastic volatility models, nonlinear term structures (Li, *JBES*, 2011; Giordani, Kohn & Pitt, *JCGS*, 2011; Andreasen 2011)
- **Epidemiology**: disease dynamic models (Ionides et al., *JASA*, 2011).
- **Ecology**: population dynamic (Thomas et al., 2009; Peters et al., 2011).

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).
- **Econometrics**: stochastic volatility models, nonlinear term structures (Li, *JBES*, 2011; Giordani, Kohn & Pitt, *JCGS*, 2011; Andreasen 2011)
- **Epidemiology**: disease dynamic models (Ionides et al., *JASA*, 2011).
- **Ecology**: population dynamic (Thomas et al., 2009; Peters et al., 2011).
- **Environmentrics**: Phytoplankton-Zooplankton model (Parslow et al., 2009), Paleoclimate reconstruction (Rougier, 2010).

Other Applications

- **Macroeconomics**: dynamic generalized stochastic equilibrium (Flury & Shephard, *Econometrics Review*, 2011; Smith, *J. Econometrics*, 2012).
- **Econometrics**: stochastic volatility models, nonlinear term structures (Li, *JBES*, 2011; Giordani, Kohn & Pitt, *JCGS*, 2011; Andreasen 2011)
- **Epidemiology**: disease dynamic models (Ionides et al., *JASA*, 2011).
- **Ecology**: population dynamic (Thomas et al., 2009; Peters et al., 2011).
- **Environmentrics**: Phytoplankton-Zooplankton model (Parslow et al., 2009), Paleoclimate reconstruction (Rougier, 2010).
- **Biochemical Systems**: stochastic kinetic models (Wilkinson & Golightly, 2010).

Bayesian Inference in General State-Space Models

- Given a collection of observations $y_{1:T} := (y_1, ..., y_T)$, we are interested in carrying out inference about θ and $X_{1:T} := (X_1, ..., X_T)$.

Bayesian Inference in General State-Space Models

- Given a collection of observations $y_{1:T} := (y_1, \dots, y_T)$, we are interested in carrying out inference about θ and $X_{1:T} := (X_1, \dots, X_T)$.
- Inference relies on the posterior density

$$
\begin{aligned}
p(\theta, x_{1:T} | y_{1:T}) &= p(\theta | y_{1:T}) \, p_\theta(x_{1:T} | y_{1:T}) \\
&\propto p(\theta, x_{1:T}, y_{1:T})
\end{aligned}
$$

where

$$
p(\theta, x_{1:T}, y_{1:T}) \propto p(\theta) \, \mu_\theta(x_1) \prod_{n=2}^{T} f_\theta(x_n | x_{n-1}) \prod_{n=1}^{T} g_\theta(y_n | x_n) \quad .
$$

Bayesian Inference in General State-Space Models

- Given a collection of observations $y_{1:T} := (y_1, ..., y_T)$, we are interested in carrying out inference about θ and $X_{1:T} := (X_1, ..., X_T)$.
- Inference relies on the posterior density

$$\begin{aligned} p(\theta, x_{1:T} | y_{1:T}) &= p(\theta | y_{1:T}) p_\theta(x_{1:T} | y_{1:T}) \\ &\propto p(\theta, x_{1:T}, y_{1:T}) \end{aligned}$$

where

$$p(\theta, x_{1:T}, y_{1:T}) \propto p(\theta) \mu_\theta(x_1) \prod_{n=2}^{T} f_\theta(x_n | x_{n-1}) \prod_{n=1}^{T} g_\theta(y_n | x_n) .$$

- No closed-form expression for $p(\theta, x_{1:T} | y_{1:T})$, numerical approximations are required.

Common MCMC Approaches and Limitations

- **MCMC Idea**: Simulate an ergodic Markov chain $\{\theta(i), X_{1:T}(i)\}_{i \geq 0}$ of invariant distribution $p(\theta, x_{1:T} | y_{1:T})$... infinite number of possibilities.

Common MCMC Approaches and Limitations

- **MCMC Idea**: Simulate an ergodic Markov chain $\{\theta(i), X_{1:T}(i)\}_{i\geq 0}$ of invariant distribution $p(\theta, x_{1:T} | y_{1:T})$... infinite number of possibilities.
- Typical strategies consists of updating iteratively $X_{1:T}$ conditional upon θ then θ conditional upon $X_{1:T}$.

Common MCMC Approaches and Limitations

- **MCMC Idea**: Simulate an ergodic Markov chain $\{\theta(i), X_{1:T}(i)\}_{i\geq 0}$ of invariant distribution $p(\theta, x_{1:T} | y_{1:T})$... infinite number of possibilities.
- Typical strategies consists of updating iteratively $X_{1:T}$ conditional upon θ then θ conditional upon $X_{1:T}$.
- To update $X_{1:T}$ conditional upon θ, use MCMC kernels updating subblocks according to $p_\theta(x_{n:n+K-1} | y_{n:n+K-1}, x_{n-1}, x_{n+K})$.

Common MCMC Approaches and Limitations

- **MCMC Idea**: Simulate an ergodic Markov chain $\{\theta(i), X_{1:T}(i)\}_{i\geq 0}$ of invariant distribution $p(\theta, x_{1:T} | y_{1:T})$... infinite number of possibilities.
- Typical strategies consists of updating iteratively $X_{1:T}$ conditional upon θ then θ conditional upon $X_{1:T}$.
- To update $X_{1:T}$ conditional upon θ, use MCMC kernels updating subblocks according to $p_\theta(x_{n:n+K-1} | y_{n:n+K-1}, x_{n-1}, x_{n+K})$.
- Standard MCMC algorithms are inefficient if θ and $X_{1:T}$ are strongly correlated.

Common MCMC Approaches and Limitations

- **MCMC Idea**: Simulate an ergodic Markov chain $\{\theta(i), X_{1:T}(i)\}_{i\geq 0}$ of invariant distribution $p(\theta, x_{1:T} | y_{1:T})$... infinite number of possibilities.
- Typical strategies consists of updating iteratively $X_{1:T}$ conditional upon θ then θ conditional upon $X_{1:T}$.
- To update $X_{1:T}$ conditional upon θ, use MCMC kernels updating subblocks according to $p_\theta(x_{n:n+K-1} | y_{n:n+K-1}, x_{n-1}, x_{n+K})$.
- Standard MCMC algorithms are inefficient if θ and $X_{1:T}$ are strongly correlated.
- Strategy impossible to implement when it is only possible to sample from the prior but impossible to evaluate it pointwise.

Metropolis-Hastings (MH) Sampling

- To bypass these problems, we want to update jointly θ and $X_{1:T}$.

Metropolis-Hastings (MH) Sampling

- To bypass these problems, we want to update jointly θ and $X_{1:T}$.
- Assume that the current state of our Markov chain is $(\theta, x_{1:T})$, we propose to update simultaneously the parameter and the states using a proposal

$$q\left(\left(\theta^*, x_{1:T}^*\right) \middle| \left(\theta, x_{1:T}\right)\right) = q\left(\theta^* \middle| \theta\right) \; q_{\theta^*}\left(x_{1:T}^* \middle| y_{1:T}\right).$$

Metropolis-Hastings (MH) Sampling

- To bypass these problems, we want to update jointly θ and $X_{1:T}$.
- Assume that the current state of our Markov chain is $(\theta, x_{1:T})$, we propose to update simultaneously the parameter and the states using a proposal

$$q\left(\left(\theta^{*}, x_{1:T}^{*}\right) \middle| \left(\theta, x_{1:T}\right) \right) = q\left(\theta^{*} \middle| \theta\right) \; q_{\theta^{*}}\left(x_{1:T}^{*} \middle| y_{1:T} \right).$$

- The proposal $(\theta^{*}, x_{1:T}^{*})$ is accepted with MH acceptance probability

$$1 \wedge \frac{p\left(\theta^{*}, x_{1:T}^{*} \middle| y_{1:T}\right)}{p\left(\theta, x_{1:T} \middle| y_{1:T}\right)} \frac{q\left(\left(x_{1:T}, \theta\right) \middle| \left(x_{1:T}^{*}, \theta^{*}\right)\right)}{q\left(\left(x_{1:T}^{*}, \theta^{*}\right) \middle| \left(x_{1:T}, \theta\right)\right)}$$

Metropolis-Hastings (MH) Sampling

- To bypass these problems, we want to update jointly θ and $X_{1:T}$.
- Assume that the current state of our Markov chain is $(\theta, x_{1:T})$, we propose to update simultaneously the parameter and the states using a proposal

$$q\left(\left(\theta^*, x^*_{1:T}\right) \middle| \left(\theta, x_{1:T}\right)\right) = q\left(\theta^* \middle| \theta\right)\; q_{\theta^*}\left(x^*_{1:T} \middle| y_{1:T}\right).$$

- The proposal $(\theta^*, x^*_{1:T})$ is accepted with MH acceptance probability

$$1 \wedge \frac{p\left(\theta^*, x^*_{1:T} \middle| y_{1:T}\right)}{p\left(\theta, x_{1:T} \middle| y_{1:T}\right)} \frac{q\left(\left(x_{1:T}, \theta\right) \middle| \left(x^*_{1:T}, \theta^*\right)\right)}{q\left(\left(x^*_{1:T}, \theta^*\right) \middle| \left(x_{1:T}, \theta\right)\right)}$$

- **Problem**: Designing a proposal $q_{\theta^*}\left(x^*_{1:T} \middle| y_{1:T}\right)$ such that the acceptance probability is not extremely small is very difficult.

"Idealized" Marginal MH Sampler

- Consider the following so-called marginal Metropolis-Hastings (MH) algorithm which uses as a proposal

$$q\left(\left(x_{1:T}^{*},\theta^{*}\right)\middle|\left(x_{1:T},\theta\right)\right)=q\left(\theta^{*}\middle|\theta\right)p_{\theta^{*}}\left(x_{1:T}^{*}\middle|y_{1:T}\right).$$

"Idealized" Marginal MH Sampler

- Consider the following so-called marginal Metropolis-Hastings (MH) algorithm which uses as a proposal

$$q\left(\left(x_{1:T}^{*},\theta^{*}\right)\middle|\left(x_{1:T},\theta\right)\right)=q\left(\theta^{*}\middle|\theta\right)p_{\theta^{*}}\left(x_{1:T}^{*}\middle|y_{1:T}\right).$$

- The MH acceptance probability is

$$1\wedge\frac{p\left(\theta^{*},x_{1:T}^{*}\middle|y_{1:T}\right)}{p\left(\theta,x_{1:T}\middle|y_{1:T}\right)}\frac{q\left(\left(x_{1:T},\theta\right)\middle|\left(x_{1:T}^{*},\theta^{*}\right)\right)}{q\left(\left(x_{1:T}^{*},\theta^{*}\right)\middle|\left(x_{1:T},\theta\right)\right)}$$

$$=1\wedge\frac{p_{\theta^{*}}\left(y_{1:T}\right)p\left(\theta^{*}\right)}{p_{\theta}\left(y_{1:T}\right)p\left(\theta\right)}\frac{q\left(\theta\middle|\theta^{*}\right)}{q\left(\theta^{*}\middle|\theta\right)}$$

"Idealized" Marginal MH Sampler

- Consider the following so-called marginal Metropolis-Hastings (MH) algorithm which uses as a proposal

$$q\left(\left(x_{1:T}^{*},\theta^{*}\right)\middle|\left(x_{1:T},\theta\right)\right)=q\left(\theta^{*}\middle|\theta\right)p_{\theta^{*}}\left(x_{1:T}^{*}\middle|y_{1:T}\right).$$

- The MH acceptance probability is

$$1\wedge\frac{p\left(\theta^{*},x_{1:T}^{*}\middle|y_{1:T}\right)}{p\left(\theta,x_{1:T}\middle|y_{1:T}\right)}\frac{q\left(\left(x_{1:T},\theta\right)\middle|\left(x_{1:T}^{*},\theta^{*}\right)\right)}{q\left(\left(x_{1:T}^{*},\theta^{*}\right)\middle|\left(x_{1:T},\theta\right)\right)}$$

$$=1\wedge\frac{p_{\theta^{*}}\left(y_{1:T}\right)p\left(\theta^{*}\right)}{p_{\theta}\left(y_{1:T}\right)p\left(\theta\right)}\frac{q\left(\theta\middle|\theta^{*}\right)}{q\left(\theta^{*}\middle|\theta\right)}$$

- In this MH algorithm, $X_{1:T}$ has been essentially integrated out.

Implementation Issues

- **Problem 1**: We do not know $p_\theta(y_{1:T}) = \int p_\theta(x_{1:T}, y_{1:T})\, dx_{1:T}$ analytically.

Implementation Issues

- **Problem 1**: We do not know $p_\theta(y_{1:T}) = \int p_\theta(x_{1:T}, y_{1:T})\, dx_{1:T}$ analytically.
- **Problem 2:** We do not know how to sample from $p_\theta(x_{1:T} | y_{1:T})$.

Implementation Issues

- **Problem 1**: We do not know $p_\theta(y_{1:T}) = \int p_\theta(x_{1:T}, y_{1:T})\, dx_{1:T}$ analytically.
- **Problem 2:** We do not know how to sample from $p_\theta(x_{1:T} | y_{1:T})$.
- **"Idea"**: Use SMC approximations of $p_\theta(x_{1:T} | y_{1:T})$ and $p_\theta(y_{1:T})$.

Sequential Monte Carlo aka Particle Filters

- Given θ, SMC methods provide approximations of $p_\theta\left(x_{1:T} \middle| y_{1:T}\right)$ and $p_\theta\left(y_{1:T}\right)$.

Sequential Monte Carlo aka Particle Filters

- Given θ, SMC methods provide approximations of $p_\theta\left(x_{1:T} \middle| y_{1:T}\right)$ and $p_\theta\left(y_{1:T}\right)$.
- To sample from $p_\theta\left(x_{1:T} \middle| y_{1:T}\right)$, SMC proceed sequentially by first approximating $p_\theta\left(x_1 \middle| y_1\right)$ and $p_\theta\left(y_1\right)$ at time 1 then $p_\theta\left(x_{1:2} \middle| y_{1:2}\right)$ and $p_\theta\left(y_{1:2}\right)$ at time 2 and so on.

Sequential Monte Carlo aka Particle Filters

- Given θ, SMC methods provide approximations of $p_\theta(x_{1:T}|y_{1:T})$ and $p_\theta(y_{1:T})$.
- To sample from $p_\theta(x_{1:T}|y_{1:T})$, SMC proceed sequentially by first approximating $p_\theta(x_1|y_1)$ and $p_\theta(y_1)$ at time 1 then $p_\theta(x_{1:2}|y_{1:2})$ and $p_\theta(y_{1:2})$ at time 2 and so on.
- SMC methods approximate the distributions of interest via a cloud of N particles which are propagated using *Importance Sampling* and *Resampling* steps.

Importance Sampling

- Assume you have at time $n-1$

$$\widehat{p}_\theta\left(x_{1:n-1} \middle| y_{1:n-1} \right) = \frac{1}{N} \sum_{k=1}^{N} \delta_{X^k_{1:n-1}} \left(x_{1:n-1} \right).$$

Importance Sampling

- Assume you have at time $n-1$

$$\widehat{p}_{\theta}\left(x_{1:n-1} \middle| y_{1:n-1}\right) = \frac{1}{N}\sum_{k=1}^{N} \delta_{X_{1:n-1}^{k}}\left(x_{1:n-1}\right).$$

- By sampling $\overline{X}_{n}^{k} \sim f_{\theta}\left(\cdot \middle| X_{n-1}^{k}\right)$ and setting $\overline{X}_{1:n}^{k} = \left(X_{1:n-1}^{k}, \overline{X}_{n}^{k}\right)$ then

$$\widehat{p}_{\theta}\left(x_{1:n} \middle| y_{1:n-1}\right) = \frac{1}{N}\sum_{k=1}^{N} \delta_{\overline{X}_{1:n}^{k}}\left(x_{1:n}\right).$$

Importance Sampling

- Assume you have at time $n-1$

$$\widehat{p}_{\theta}\left(\left.x_{1:n-1}\right|y_{1:n-1}\right)=\frac{1}{N}\sum_{k=1}^{N}\delta_{X_{1:n-1}^{k}}\left(x_{1:n-1}\right).$$

- By sampling $\overline{X}_{n}^{k}\sim f_{\theta}\left(\left.\cdot\right|X_{n-1}^{k}\right)$ and setting $\overline{X}_{1:n}^{k}=\left(X_{1:n-1}^{k},\overline{X}_{n}^{k}\right)$ then

$$\widehat{p}_{\theta}\left(\left.x_{1:n}\right|y_{1:n-1}\right)=\frac{1}{N}\sum_{k=1}^{N}\delta_{\overline{X}_{1:n}^{k}}\left(x_{1:n}\right).$$

- Our target at time n is

$$p_{\theta}\left(\left.x_{1:n}\right|y_{1:n}\right)=\frac{g_{\theta}\left(\left.y_{n}\right|x_{n}\right)p_{\theta}\left(\left.x_{1:n}\right|y_{1:n-1}\right)}{\int g_{\theta}\left(\left.y_{n}\right|x_{n}\right)p_{\theta}\left(\left.x_{1:n}\right|y_{1:n-1}\right)dx_{1:n}}$$

so by substituting $\widehat{p}_{\theta}\left(\left.x_{1:n}\right|y_{1:n-1}\right)$ to $p_{\theta}\left(\left.x_{1:n}\right|y_{1:n-1}\right)$ we obtain

$$\overline{p}_{\theta}\left(\left.x_{1:n}\right|y_{1:n}\right)=\sum_{k=1}^{N}W_{n}^{k}\delta_{\overline{X}_{1:n}^{k}}\left(x_{1:n}\right),\quad W_{n}^{k}\propto g_{\theta}\left(\left.y_{n}\right|\overline{X}_{1:n}^{k}\right).$$

Resampling

- We have a "weighted" approximation $\overline{p}_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)$ of $p_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)$

$$\overline{p}_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)=\sum_{k=1}^{N}W_n^k\delta_{\overline{X}_{1:n}^k}\left(x_{1:n}\right).$$

Resampling

- We have a "weighted" approximation $\overline{p}_\theta\left(x_{1:n}\middle|\, y_{1:n}\right)$ of $p_\theta\left(x_{1:n}\middle|\, y_{1:n}\right)$

$$\overline{p}_\theta\left(x_{1:n}\middle|\, y_{1:n}\right) = \sum_{k=1}^{N} W_n^k \delta_{\overline{X}_{1:n}^k}\left(x_{1:n}\right).$$

- To obtain N samples $X_{1:n}^k$ approximately distributed according to $p_\theta\left(x_{1:n}\middle|\, y_{1:n}\right)$, we just resample

$$X_{1:n}^k \sim \overline{p}_\theta\left(\cdot\middle|\, y_{1:n}\right)$$

to obtain

$$\widehat{p}_\theta\left(x_{1:n}\middle|\, y_{1:n}\right) = \frac{1}{N}\sum_{k=1}^{N} \delta_{X_{1:n}^k}\left(x_{1:n}\right).$$

Resampling

- We have a "weighted" approximation $\overline{p}_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)$ of $p_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)$

$$\overline{p}_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)=\sum_{k=1}^{N}W_n^k\delta_{\overline{X}_{1:n}^k}\left(x_{1:n}\right).$$

- To obtain N samples $X_{1:n}^k$ approximately distributed according to $p_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)$, we just resample

$$X_{1:n}^k\sim\overline{p}_\theta\left(\left.\cdot\right|y_{1:n}\right)$$

to obtain

$$\widehat{p}_\theta\left(\left.x_{1:n}\right|y_{1:n}\right)=\frac{1}{N}\sum_{k=1}^{N}\delta_{X_{1:n}^k}\left(x_{1:n}\right).$$

- Particles with high weights are copied multiples times, particles with low weights die.

Bootstrap Filter (Gordon, Salmond & Smith, 1993)

At time $n = 1$

- Sample $\overline{X}_1^k \sim \mu_\theta(\cdot)$ then

$$\overline{p}_\theta(x_1 | y_1) = \sum_{k=1}^{N} W_1^k \delta_{\overline{X}_1^k}(x_1), \; W_1^k \propto g_\theta\left(y_1 | \overline{X}_1^k\right).$$

Bootstrap Filter (Gordon, Salmond & Smith, 1993)

At time $n = 1$

- Sample $\overline{X}_1^k \sim \mu_\theta(\cdot)$ then

$$\overline{p}_\theta(x_1 | y_1) = \sum_{k=1}^{N} W_1^k \delta_{\overline{X}_1^k}(x_1), \; W_1^k \propto g_\theta\left(y_1 | \overline{X}_1^k\right).$$

- Resample $X_1^k \sim \overline{p}_\theta(x_1 | y_1)$ to obtain $\widehat{p}_\theta(x_1 | y_1) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_1^k}(x_1)$.

Bootstrap Filter (Gordon, Salmond & Smith, 1993)

At time $n = 1$

- Sample $\overline{X}_1^k \sim \mu_\theta(\cdot)$ then

$$\overline{p}_\theta(x_1 | y_1) = \sum_{k=1}^{N} W_1^k \delta_{\overline{X}_1^k}(x_1), \; W_1^k \propto g_\theta\left(y_1 | \overline{X}_1^k\right).$$

- Resample $X_1^k \sim \overline{p}_\theta(x_1 | y_1)$ to obtain $\widehat{p}_\theta(x_1 | y_1) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_1^k}(x_1)$.

Bootstrap Filter (Gordon, Salmond & Smith, 1993)

At time $n = 1$

- Sample $\overline{X}_1^k \sim \mu_\theta(\cdot)$ then

$$\overline{p}_\theta(x_1 | y_1) = \sum_{k=1}^{N} W_1^k \delta_{\overline{X}_1^k}(x_1), \; W_1^k \propto g_\theta\left(y_1 | \overline{X}_1^k\right).$$

- Resample $X_1^k \sim \overline{p}_\theta(x_1 | y_1)$ to obtain $\widehat{p}_\theta(x_1 | y_1) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_1^k}(x_1)$.

At time $n \geq 2$

- Sample $\overline{X}_n^k \sim f_\theta\left(\cdot | X_{n-1}^k\right)$, set $\overline{X}_{1:n}^k = \left(X_{1:n-1}^k, \overline{X}_n^k\right)$ and

$$\overline{p}_\theta(x_{1:n} | y_{1:n}) = \sum_{k=1}^{N} W_n^k \delta_{\overline{X}_{1:n}^k}(x_{1:n}), \; W_n^k \propto g_\theta\left(y_n | \overline{X}_n^k\right).$$

Bootstrap Filter (Gordon, Salmond & Smith, 1993)

At time $n = 1$

- Sample $\overline{X}_1^k \sim \mu_\theta(\cdot)$ then

$$\overline{p}_\theta(x_1 | y_1) = \sum_{k=1}^{N} W_1^k \delta_{\overline{X}_1^k}(x_1), \quad W_1^k \propto g_\theta\left(y_1 | \overline{X}_1^k\right).$$

- Resample $X_1^k \sim \overline{p}_\theta(x_1 | y_1)$ to obtain $\widehat{p}_\theta(x_1 | y_1) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_1^k}(x_1)$.

At time $n \geq 2$

- Sample $\overline{X}_n^k \sim f_\theta\left(\cdot | X_{n-1}^k\right)$, set $\overline{X}_{1:n}^k = \left(X_{1:n-1}^k, \overline{X}_n^k\right)$ and

$$\overline{p}_\theta(x_{1:n} | y_{1:n}) = \sum_{k=1}^{N} W_n^k \delta_{\overline{X}_{1:n}^k}(x_{1:n}), \quad W_n^k \propto g_\theta\left(y_n | \overline{X}_n^k\right).$$

- Resample $X_{1:n}^k \sim \overline{p}_\theta(x_{1:n} | y_{1:n})$ to obtain $\widehat{p}_\theta(x_{1:n} | y_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{1:n}^k}(x_{1:n})$.

SMC Output

- At time T, we obtain the following approximation of the posterior of interest

$$\widehat{p}_\theta\left(x_{1:T}\middle| y_{1:T}\right) = \frac{1}{N}\sum_{k=1}^{N}\delta_{X_{1:T}^k}\left(dx_{1:T}\right)$$

and an approximation of $p_\theta\left(y_{1:T}\right)$ is given by

$$\widehat{p}_\theta\left(y_{1:T}\right) = \widehat{p}_\theta\left(y_1\right)\prod_{n=2}^{T}\widehat{p}_\theta\left(y_n\middle| y_{1:n-1}\right) = \prod_{n=1}^{T}\left(\frac{1}{N}\sum_{k=1}^{N} g_\theta\left(y_n\middle| X_n^k\right)\right).$$

SMC Output

- At time T, we obtain the following approximation of the posterior of interest

$$\widehat{p}_\theta\left(x_{1:T} \middle| y_{1:T}\right) = \frac{1}{N}\sum_{k=1}^{N}\delta_{X_{1:T}^k}\left(dx_{1:T}\right)$$

  and an approximation of $p_\theta\left(y_{1:T}\right)$ is given by

$$\widehat{p}_\theta\left(y_{1:T}\right) = \widehat{p}_\theta\left(y_1\right)\prod_{n=2}^{T}\widehat{p}_\theta\left(y_n \middle| y_{1:n-1}\right) = \prod_{n=1}^{T}\left(\frac{1}{N}\sum_{k=1}^{N} g_\theta\left(y_n \middle| X_n^k\right)\right).$$

- These approximations are asymptotically (i.e. $N \rightarrow \infty$) consistent under very weak assumptions.

Some Theoretical Results

- Under *mixing assumptions* (Del Moral, 2004), we have

$$\left\| \mathcal{L}\left(X_{1:T} \in \cdot\right) - p_\theta\left(\cdot \mid y_{1:T}\right) \right\|_{\text{tv}} \leq C_\theta \frac{T}{N}$$

where $X_{1:T} \sim \mathbb{E}\left[\widehat{p}_\theta\left(\cdot \mid y_{1:T}\right)\right]$.

Some Theoretical Results

- Under *mixing assumptions* (Del Moral, 2004), we have

$$\left\| \mathcal{L}\left(X_{1:T} \in \cdot\right) - p_\theta\left(\cdot \mid y_{1:T}\right)\right\|_{\text{tv}} \leq C_\theta \frac{T}{N}$$

  where $X_{1:T} \sim \mathbb{E}\left[\widehat{p}_\theta\left(\cdot \mid y_{1:T}\right)\right]$.

- Under *mixing assumptions* (Del Moral et al., 2010) we also have

$$\frac{\mathbb{V}\left[\widehat{p}_\theta\left(y_{1:T}\right)\right]}{p_\theta^2\left(y_{1:T}\right)} \leq D_\theta \frac{T}{N}.$$

Some Theoretical Results

- Under *mixing assumptions* (Del Moral, 2004), we have

$$\left\| \mathcal{L}\left(X_{1:T} \in \cdot\right) - p_\theta\left(\cdot \mid y_{1:T}\right) \right\|_{\text{tv}} \leq C_\theta \frac{T}{N}$$

  where $X_{1:T} \sim \mathbb{E}\left[\widehat{p}_\theta\left(\cdot \mid y_{1:T}\right)\right]$.

- Under *mixing assumptions* (Del Moral et al., 2010) we also have

$$\frac{\mathbb{V}\left[\widehat{p}_\theta\left(y_{1:T}\right)\right]}{p_\theta^2\left(y_{1:T}\right)} \leq D_\theta \frac{T}{N}.$$

- Loosely speaking, the performance of SMC only degrade linearly with time rather than exponentially for naive approaches.

Some Theoretical Results

- Under *mixing assumptions* (Del Moral, 2004), we have

$$\left\| \mathcal{L}\left(X_{1:T} \in \cdot\right) - p_\theta\left(\cdot \mid y_{1:T}\right) \right\|_{\text{tv}} \leq C_\theta \frac{T}{N}$$

  where $X_{1:T} \sim \mathbb{E}\left[\widehat{p}_\theta\left(\cdot \mid y_{1:T}\right)\right]$.
- Under *mixing assumptions* (Del Moral et al., 2010) we also have

$$\frac{\mathbb{V}\left[\widehat{p}_\theta\left(y_{1:T}\right)\right]}{p_\theta^2\left(y_{1:T}\right)} \leq D_\theta \frac{T}{N}.$$

- Loosely speaking, the performance of SMC only degrade linearly with time rather than exponentially for naive approaches.
- **Problem**: We cannot compute analytically the particle filter proposal $q_\theta\left(x_{1:T} \mid y_{1:T}\right) = \mathbb{E}\left[\widehat{p}_\theta\left(x_{1:T} \mid y_{1:T}\right)\right]$ as it involves an expectation w.r.t all the variables appearing in the particle algorithm...

"Idealized" Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q\left(\cdot \mid \theta\left(i-1\right)\right)$.

"Idealized" Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q\left(\cdot\middle|\,\theta\left(i-1\right)\right)$.
- Sample $X_{1:T}^* \sim p_{\theta^*}\left(\cdot\middle|\,y_{1:T}\right)$.

"Idealized" Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q\left(\cdot|\,\theta\left(i-1\right)\right)$.
- Sample $X_{1:T}^* \sim p_{\theta^*}\left(\cdot|\,y_{1:T}\right)$.
- With probability

$$1 \wedge \frac{p_{\theta^*}\left(y_{1:T}\right) p\left(\theta^*\right)}{p_{\theta(i-1)}\left(y_{1:T}\right) p\left(\theta\left(i-1\right)\right)} \frac{q\left(\theta\left(i-1\right)|\,\theta^*\right)}{q\left(\theta^*|\,\theta\left(i-1\right)\right)}$$

set $\theta\left(i\right)=\theta^*$, $X_{1:T}\left(i\right)=X_{1:T}^*$ otherwise set $\theta\left(i\right)=\theta\left(i-1\right)$, $X_{1:T}\left(i\right)=X_{1:T}\left(i-1\right)$.

Particle Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q\left(\cdot \mid \theta(i-1)\right)$ and run an SMC algorithm to obtain $\widehat{p}_{\theta^*}\left(x_{1:T} \mid y_{1:T}\right)$ and $\widehat{p}_{\theta^*}\left(y_{1:T}\right)$.

Particle Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q(\cdot|\, \theta(i-1))$ and run an SMC algorithm to obtain $\widehat{p}_{\theta^*}(x_{1:T}|\, y_{1:T})$ and $\widehat{p}_{\theta^*}(y_{1:T})$.
- Sample $X_{1:T}^* \sim \widehat{p}_{\theta^*}(\cdot|\, y_{1:T})$.

Particle Marginal MH Sampler

At iteration i

- Sample $\theta^* \sim q(\cdot|\,\theta(i-1))$ and run an SMC algorithm to obtain $\widehat{p}_{\theta^*}(x_{1:T}|\,y_{1:T})$ and $\widehat{p}_{\theta^*}(y_{1:T})$.
- Sample $X^*_{1:T} \sim \widehat{p}_{\theta^*}(\cdot|\,y_{1:T})$.
- With probability

$$1 \wedge \frac{\widehat{p}_{\theta^*}(y_{1:T})\,p(\theta^*)}{\widehat{p}_{\theta(i-1)}(y_{1:T})\,p(\theta(i-1))}\frac{q(\theta(i-1)|\,\theta^*)}{q(\theta^*|\,\theta(i-1))}$$

set $\theta(i) = \theta^*$, $X_{1:T}(i) = X^*_{1:T}$ otherwise set $\theta(i) = \theta(i-1)$, $X_{1:T}(i) = X_{1:T}(i-1)$.

Validity of the Particle Marginal MH Sampler

- Assume that the 'idealized' marginal MH sampler is irreducible and aperiodic then, under very weak assumptions, the PMMH sampler generates a sequence $\{\theta(i), X_{1:T}(i)\}$ whose marginal distributions $\left\{\mathcal{L}^N(\theta(i), X_{1:T}(i) \in \cdot)\right\}$ satisfy for any $N \geq 1$

$$\left\|\mathcal{L}^N(\theta(i), X_{1:T}(i) \in \cdot) - p(\cdot|\, y_{1:T})\right\|_{\text{TV}} \rightarrow 0 \text{ as } i \rightarrow \infty\ .$$

Validity of the Particle Marginal MH Sampler

- Assume that the 'idealized' marginal MH sampler is irreducible and aperiodic then, under very weak assumptions, the PMMH sampler generates a sequence $\{\theta(i), X_{1:T}(i)\}$ whose marginal distributions $\{\mathcal{L}^N(\theta(i), X_{1:T}(i) \in \cdot)\}$ satisfy for any $N \geq 1$

$$\left\| \mathcal{L}^N(\theta(i), X_{1:T}(i) \in \cdot) - p(\cdot | y_{1:T}) \right\|_{\text{TV}} \rightarrow 0 \text{ as } i \rightarrow \infty .$$

- Corollary of a more general result: the PMMH sampler is a standard MH sampler of target distribution $\tilde{\pi}^N$ and proposal $\widetilde{q}^N$ defined on an extended space associated to all the variables used to generate the proposal.

Explicit Structure of the Target Distribution

- For pedagogical reasons, we limit ourselves to the case where $T = 1$.

Explicit Structure of the Target Distribution

- For pedagogical reasons, we limit ourselves to the case where $T = 1$.
- The proposal is

$$\widetilde{q}^N\left(\left(\theta^*, k^*, x_1^{*1:N}\right) \middle| \left(\theta, k, x_1^{1:N}\right)\right) = q\left(\theta^* \middle| \theta\right) \prod_{m=1}^{N} \mu_{\theta^*}\left(x_1^{*m}\right) \; w_1^{k^*}$$

Explicit Structure of the Target Distribution

- For pedagogical reasons, we limit ourselves to the case where $T = 1$.
- The proposal is

$$\widetilde{q}^N \left(\left(\theta^*, k^*, x_1^{*1:N} \right) \middle| \left(\theta, k, x_1^{1:N} \right) \right) = q\left(\theta^* \middle| \theta \right) \prod_{m=1}^{N} \mu_{\theta^*}\left(x_1^{*m} \right) \; w_1^{k^*}$$

- The artificial target is

$$\begin{aligned} \tilde{\pi}^N \left(\theta, k, x_1^{1:N} \right) &= \frac{p\left(\theta, x_1^k \middle| y_1 \right)}{N} \prod_{m=1; m \neq k}^{N} \mu_\theta \left(x_1^m \right) \\ &= \frac{1}{N} \frac{p(\theta)\, g_\theta\left(y_1 \middle| x_1^k \right)}{p_\theta(y_1)} \prod_{m=1}^{N} \mu_\theta \left(x_1^m \right) \end{aligned}$$

Explicit Structure of the Target Distribution

- For pedagogical reasons, we limit ourselves to the case where $T = 1$.
- The proposal is

$$\widetilde{q}^N\left(\left(\theta^*, k^*, x_1^{*1:N}\right)\middle|\left(\theta, k, x_1^{1:N}\right)\right) = q\left(\theta^* \middle| \theta\right) \prod_{m=1}^{N} \mu_{\theta^*}\left(x_1^{*m}\right) \; w_1^{k^*}$$

- The artificial target is

$$\begin{aligned}
\widetilde{\pi}^N\left(\theta, k, x_1^{1:N}\right) &= \frac{p\left(\theta, x_1^k \middle| y_1\right)}{N} \prod_{m=1; m\neq k}^{N} \mu_\theta\left(x_1^m\right) \\
&= \frac{1}{N} \frac{p\left(\theta\right) g_\theta\left(y_1 \middle| x_1^k\right)}{p_\theta\left(y_1\right)} \prod_{m=1}^{N} \mu_\theta\left(x_1^m\right)
\end{aligned}$$

- We have indeed

$$\frac{\widetilde{\pi}\left(\theta^*, k^*, x_1^{*1:N}\right)}{\widetilde{q}^N\left(\left(\theta^*, k^*, x_1^{*1:N}\right)\middle|\left(\theta, k, x_1^{1:N}\right)\right)} = \frac{p\left(\theta^*\right)}{q\left(\theta^* \middle| \theta\right)} \frac{\frac{1}{N}\sum_{i=1}^{N} g_{\theta^*}\left(y_1 \middle| x_1^{*i}\right)}{p_\theta\left(y_1\right)}$$

"Idealized" Block Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot | y_{1:T}, X_{1:T}(i-1))$.

"Idealized" Block Gibbs Sampler

At iteration i

- Sample $\theta\left(i\right) \sim p\left(\cdot\middle|y_{1:T}, X_{1:T}\left(i-1\right)\right)$.
- Sample $X_{1:T}\left(i\right) \sim p\left(\cdot\middle|y_{1:T}, \theta\left(i\right)\right)$.

“Idealized” Block Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot | y_{1:T}, X_{1:T}(i-1))$.
- Sample $X_{1:T}(i) \sim p(\cdot | y_{1:T}, \theta(i))$.
- Naive particle approximation where $X_{1:T}(i) \sim \widehat{p}(\cdot | y_{1:T}, \theta(i))$ is substituted to $X_{1:T}(i) \sim p(\cdot | y_{1:T}, \theta(i))$ is obviously incorrect.

Particle Gibbs Sampler

- A (collapsed) Gibbs sampler to sample from $\widetilde{\pi}^N$ for $T = 1$ can be implemented using

$$\widetilde{\pi}^N \left(\theta, x_1^{-k} \middle| k, x_1^k \right) = p \left(\theta | \, y_1, x_1^k \right) \prod_{m=1; m \neq k}^{N} \mu_\theta \left(x_1^m \right),$$

$$\widetilde{\pi}^N \left(K = k | \, \theta, x_1^{1:N} \right) = \frac{g_\theta \left(y_1 | \, x_1^k \right)}{\sum_{i=1}^{N} g_\theta \left(y_1 | \, x_1^i \right)}.$$

Particle Gibbs Sampler

- A (collapsed) Gibbs sampler to sample from $\widetilde{\pi}^N$ for $T = 1$ can be implemented using

$$\widetilde{\pi}^N \left(\theta, x_1^{-k} \middle| k, x_1^k \right) = p\left(\theta | y_1, x_1^k \right) \prod_{m=1; m \neq k}^{N} \mu_\theta \left(x_1^m \right),$$

$$\widetilde{\pi}^N \left(K = k | \theta, x_1^{1:N} \right) = \frac{g_\theta \left(y_1 | x_1^k \right)}{\sum_{i=1}^N g_\theta \left(y_1 | x_1^i \right)}.$$

- Note that even for fixed θ, this is a non-standard MCMC update for $p_\theta \left(x_1 | y_1 \right)$. This generalizes Baker's acceptance rule (Baker, 1965).

Particle Gibbs Sampler

- A (collapsed) Gibbs sampler to sample from $\widetilde{\pi}^N$ for $T = 1$ can be implemented using

$$\widetilde{\pi}^N \left(\theta, x_1^{-k} \middle| k, x_1^k \right) = p \left(\theta | \, y_1, x_1^k \right) \prod_{m=1; m \neq k}^{N} \mu_\theta \left(x_1^m \right),$$

$$\widetilde{\pi}^N \left(K = k | \, \theta, x_1^{1:N} \right) = \frac{g_\theta \left(y_1 | \, x_1^k \right)}{\sum_{i=1}^N g_\theta \left(y_1 | \, x_1^i \right)}.$$

- Note that even for fixed θ, this is a non-standard MCMC update for $p_\theta \left(x_1 | \, y_1 \right)$. This generalizes Baker's acceptance rule (Baker, 1965).
- The target and associated Gibbs sampler can be generalized to $T > 1$.

Particle Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot | y_{1:T}, X_{1:T}(i-1))$.

Particle Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot|y_{1:T}, X_{1:T}(i-1))$.
- Run a conditional SMC algorithm for $\theta(i)$ consistent with $X_{1:T}(i-1)$ and its ancestral lineage.

Particle Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot | y_{1:T}, X_{1:T}(i-1))$.
- Run a conditional SMC algorithm for $\theta(i)$ consistent with $X_{1:T}(i-1)$ and its ancestral lineage.
- Sample $X_{1:T}(i) \sim \widehat{p}(\cdot | y_{1:T}, \theta(i))$ from the resulting approximation (hence its ancestral lineage too).

Particle Gibbs Sampler

At iteration i

- Sample $\theta(i) \sim p(\cdot | y_{1:T}, X_{1:T}(i-1))$.
- Run a conditional SMC algorithm for $\theta(i)$ consistent with $X_{1:T}(i-1)$ and its ancestral lineage.
- Sample $X_{1:T}(i) \sim \widehat{p}(\cdot | y_{1:T}, \theta(i))$ from the resulting approximation (hence its ancestral lineage too).

- **Proposition**. Assume that the 'ideal' Gibbs sampler is irreducible and aperiodic then under very weak assumptions the particle Gibbs sampler generates a sequence $\{\theta(i), X_{1:T}(i)\}$ such that for any $N \geq 2$

$$\|\mathcal{L}((\theta(i), X_{1:T}(i)) \in \cdot) - p(\cdot | y_{1:T})\| \to 0 \text{ as } i \to \infty.$$

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_1 | y_1) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{ \overline{X}_1^{-b_1^k} \right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

Conditional SMC Algorithm

<u>*At time* 1</u>

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_1 | y_1) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{ \overline{X}_1^{-b_1^k} \right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_1 | y_1) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{ \overline{X}_1^{-b_1^k} \right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

At time $n = 2, ..., T$

- For $m \neq b_n^k$, sample $X_n^m \sim f_\theta\left(\cdot | \overline{X}_{n-1}^m\right)$, set $X_{1:n}^m = \left(\overline{X}_{1:n-1}^m, X_n^m\right)$ and $W_n^m \propto g_\theta(y_n | X_n^m)$, $\sum_{m=1}^N W_n^m = 1$.

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_1 | y_1) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{ \overline{X}_1^{-b_1^k} \right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

At time $n = 2, ..., T$

- For $m \neq b_n^k$, sample $X_n^m \sim f_\theta\left(\cdot | \overline{X}_{n-1}^m\right)$, set $X_{1:n}^m = \left(\overline{X}_{1:n-1}^m, X_n^m\right)$ and $W_n^m \propto g_\theta(y_n | X_n^m)$, $\sum_{m=1}^N W_n^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_{1:n} | y_{1:n}) = \sum_{m=1}^N W_n^m \delta_{X_{1:n}^m}(x_{1:n})$ to obtain $\left\{ \overline{X}_{1:n}^{-b_n^k} \right\}$ and set $\overline{X}_{1:n}^{b_n^k} = X_{1:n}^{b_n^k}$.

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta\left(y_1 \middle| X_1^m,\right)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta\left(x_1 \middle| y_1\right) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{\overline{X}_1^{-b_1^k}\right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

At time $n = 2, ..., T$

- For $m \neq b_n^k$, sample $X_n^m \sim f_\theta\left(\cdot \middle| \overline{X}_{n-1}^m\right)$, set $X_{1:n}^m = \left(\overline{X}_{1:n-1}^m, X_n^m\right)$ and $W_n^m \propto g_\theta\left(y_n \middle| X_n^m\right)$, $\sum_{m=1}^N W_n^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta\left(x_{1:n} \middle| y_{1:n}\right) = \sum_{m=1}^N W_n^m \delta_{X_{1:n}^m}(x_{1:n})$ to obtain $\left\{\overline{X}_{1:n}^{-b_n^k}\right\}$ and set $\overline{X}_{1:n}^{b_n^k} = X_{1:n}^{b_n^k}$.

Conditional SMC Algorithm

At time 1

- For $m \neq b_1^k$, sample $X_1^m \sim \mu_\theta(\cdot)$ and set $W_1^m \propto g_\theta(y_1 | X_1^m,)$, $\sum_{m=1}^N W_1^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_1 | y_1) = \sum_{m=1}^N W_1^m \delta_{X_1^m}(x_1)$ to obtain $\left\{ \overline{X}_1^{-b_1^k} \right\}$ and set $\overline{X}_1^{b_1^k} = X_1^{b_1^k}$.

At time $n = 2, ..., T$

- For $m \neq b_n^k$, sample $X_n^m \sim f_\theta\left(\cdot | \overline{X}_{n-1}^m\right)$, set $X_{1:n}^m = \left(\overline{X}_{1:n-1}^m, X_n^m\right)$ and $W_n^m \propto g_\theta(y_n | X_n^m)$, $\sum_{m=1}^N W_n^m = 1$.
- Resample $N-1$ times from $\widehat{p}_\theta(x_{1:n} | y_{1:n}) = \sum_{m=1}^N W_n^m \delta_{X_{1:n}^m}(x_{1:n})$ to obtain $\left\{ \overline{X}_{1:n}^{-b_n^k} \right\}$ and set $\overline{X}_{1:n}^{b_n^k} = X_{1:n}^{b_n^k}$.

At time $n = T$

- Sample $X_{1:T} \sim \widehat{p}_\theta(\cdot | y_{1:T})$.

Nonlinear State-Space Model

- Consider the following model

$$
\begin{aligned}
X_n &= \frac{1}{2}X_{n-1} + 25\frac{X_{n-1}}{1+X_{n-1}^2} + 8\cos 1.2n + V_n, \\
Y_n &= \frac{X_n^2}{20} + W_n
\end{aligned}
$$

where $V_n \sim \mathcal{N}\left(0,\sigma_v^2\right)$, $W_n \sim \mathcal{N}\left(0,\sigma_w^2\right)$ and $X_1 \sim \mathcal{N}\left(0,5^2\right)$.

Nonlinear State-Space Model

- Consider the following model

$$X_n = \frac{1}{2}X_{n-1} + 25\frac{X_{n-1}}{1+X_{n-1}^2} + 8\cos 1.2n + V_n,$$

$$Y_n = \frac{X_n^2}{20} + W_n$$

where $V_n \sim \mathcal{N}\left(0,\sigma_v^2\right)$, $W_n \sim \mathcal{N}\left(0,\sigma_w^2\right)$ and $X_1 \sim \mathcal{N}\left(0,5^2\right)$.

- Use the prior for $\{X_n\}$ as proposal distribution.

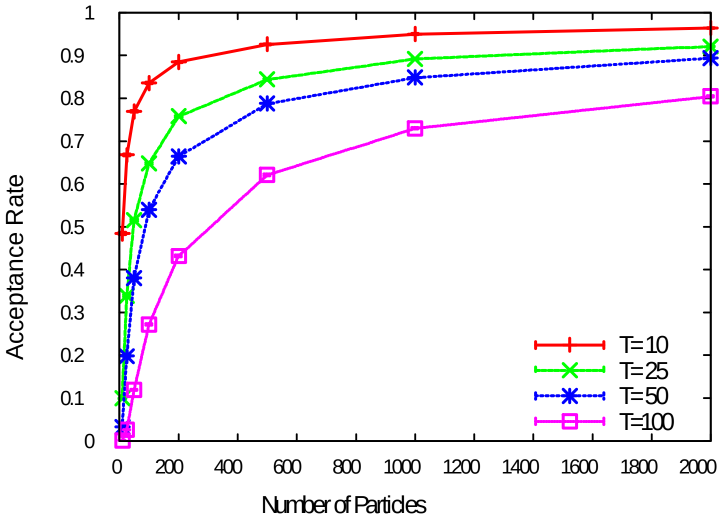
Nonlinear State-Space Model

- Consider the following model

$$
\begin{aligned}
X_n &= \frac{1}{2}X_{n-1} + 25\frac{X_{n-1}}{1+X_{n-1}^2} + 8\cos 1.2n + V_n, \\
Y_n &= \frac{X_n^2}{20} + W_n
\end{aligned}
$$

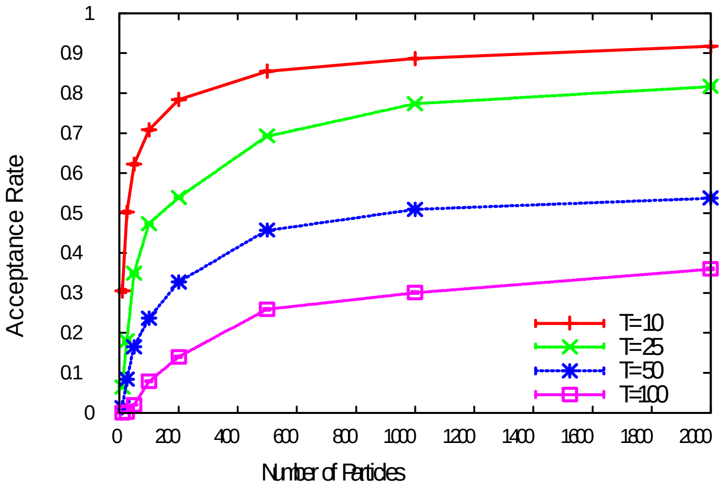
  where $V_n \sim \mathcal{N}\left(0, \sigma_v^2\right)$, $W_n \sim \mathcal{N}\left(0, \sigma_w^2\right)$ and $X_1 \sim \mathcal{N}\left(0, 5^2\right)$.
- Use the prior for $\{X_n\}$ as proposal distribution.
- For a fixed θ, we evaluate the expected acceptance probability as a function of N.

Average Acceptance Probability



Average acceptance probability when $\sigma_v^2 = \sigma_w^2 = 10$

Average Acceptance Probability



Average acceptance probability when $\sigma_v^2 = 10$, $\sigma_w^2 = 1$

Inference for Stochastic Kinetic Models

- Two species X_t^1 (prey) and X_t^2 (predator)

$$\begin{aligned}
\Pr\left(X^1_{t+dt}{=}x^1_t{+}1, X^2_{t+dt}{=}x^2_t \,\middle|\, x^1_t, x^2_t\right) &= \alpha\, x^1_t\, dt + o\left(dt\right),\\
\Pr\left(X^1_{t+dt}{=}x^1_t{-}1, X^2_{t+dt}{=}x^2_t{+}1 \,\middle|\, x^1_t, x^2_t\right) &= \beta\, x^1_t\, x^2_t\, dt + o\left(dt\right),\\
\Pr\left(X^1_{t+dt}{=}x^1_t, X^2_{t+dt}{=}x^2_t{-}1 \,\middle|\, x^1_t, x^2_t\right) &= \gamma\, x^2_t\, dt + o\left(dt\right),
\end{aligned}$$

observed at discrete times

$$Y_n = X^1_{n\Delta} + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

Inference for Stochastic Kinetic Models

- Two species X_t^1 (prey) and X_t^2 (predator)

$$\begin{aligned}
\Pr\left(X_{t+dt}^1{=}x_t^1{+}1, X_{t+dt}^2{=}x_t^2 \,\middle|\, x_t^1, x_t^2\right) &= \alpha\, x_t^1\, dt + o\left(dt\right), \\
\Pr\left(X_{t+dt}^1{=}x_t^1{-}1, X_{t+dt}^2{=}x_t^2{+}1 \,\middle|\, x_t^1, x_t^2\right) &= \beta\, x_t^1\, x_t^2\, dt + o\left(dt\right), \\
\Pr\left(X_{t+dt}^1{=}x_t^1, X_{t+dt}^2{=}x_t^2{-}1 \,\middle|\, x_t^1, x_t^2\right) &= \gamma\, x_t^2\, dt + o\left(dt\right),
\end{aligned}$$

  observed at discrete times

$$Y_n = X_{n\Delta}^1 + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

- We are interested in the kinetic rate constants $\theta = (\alpha, \beta, \gamma)$ a priori distributed as (Boys et al., 2008; Kunsch, 2011)

$$\alpha \sim \mathcal{G}(1, 10), \quad \beta \sim \mathcal{G}(1, 0.25), \quad \gamma \sim \mathcal{G}(1, 7.5).$$

Inference for Stochastic Kinetic Models

- Two species X_t^1 (prey) and X_t^2 (predator)

$$\Pr\left(X_{t+dt}^1 = x_t^1 + 1, X_{t+dt}^2 = x_t^2 \,\middle|\, x_t^1, x_t^2\right) = \alpha\, x_t^1\, dt + o\left(dt\right),$$
$$\Pr\left(X_{t+dt}^1 = x_t^1 - 1, X_{t+dt}^2 = x_t^2 + 1 \,\middle|\, x_t^1, x_t^2\right) = \beta\, x_t^1\, x_t^2\, dt + o\left(dt\right),$$
$$\Pr\left(X_{t+dt}^1 = x_t^1, X_{t+dt}^2 = x_t^2 - 1 \,\middle|\, x_t^1, x_t^2\right) = \gamma\, x_t^2\, dt + o\left(dt\right),$$

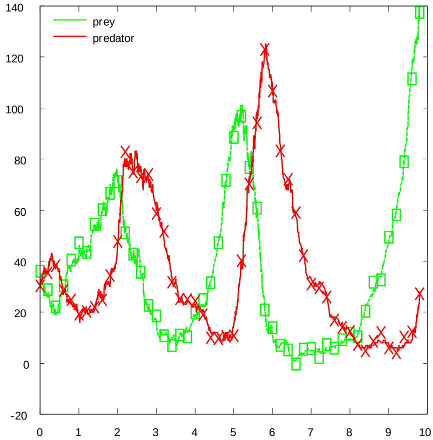
observed at discrete times

$$Y_n = X_{n\Delta}^1 + W_n \text{ with } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

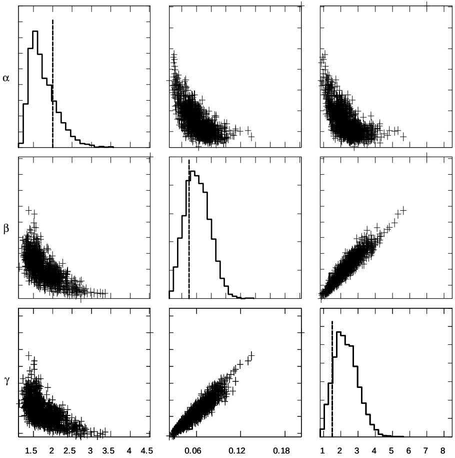
- We are interested in the kinetic rate constants $\theta = (\alpha, \beta, \gamma)$ a priori distributed as (Boys et al., 2008; Kunsch, 2011)

$$\alpha \sim \mathcal{G}(1, 10), \quad \beta \sim \mathcal{G}(1, 0.25), \quad \gamma \sim \mathcal{G}(1, 7.5).$$

- MCMC methods require reversible jumps, Particle MCMC requires only forward simulation.

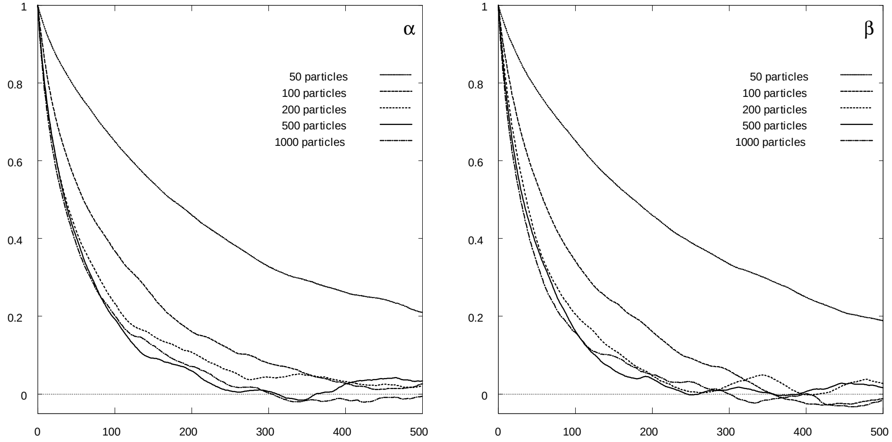


Simulated data



Estimated posteriors

Autocorrelation Functions



Autocorrelation of α (left) and β (right) for the PMMH sampler for various N.

Discussion

- PMCMC methods allow us to design 'good' high dimensional proposals based only on low dimensional (and potentially unsophisticated) proposals.

Discussion

- PMCMC methods allow us to design 'good' high dimensional proposals based only on low dimensional (and potentially unsophisticated) proposals.
- PMCMC allow us to perform Bayesian inference for dynamic models for which only forward simulation is possible.

Discussion

- PMCMC methods allow us to design ‘good’ high dimensional proposals based only on low dimensional (and potentially unsophisticated) proposals.
- PMCMC allow us to perform Bayesian inference for dynamic models for which only forward simulation is possible.
- Whenever an unbiased estimate of the likelihood function is available, “exact” Bayesian inference is possible.

Discussion

- PMCMC methods allow us to design 'good' high dimensional proposals based only on low dimensional (and potentially unsophisticated) proposals.
- PMCMC allow us to perform Bayesian inference for dynamic models for which only forward simulation is possible.
- Whenever an unbiased estimate of the likelihood function is available, "exact" Bayesian inference is possible.
- More precise quantitative convergence results need to be established.

References

- C. Andrieu, A.D. & R. Holenstein, Particle Markov chain Monte Carlo methods (with discussion), *J. Royal Statistical Society* B, 2010.

References

- C. Andrieu, A.D. & R. Holenstein, Particle Markov chain Monte Carlo methods (with discussion), *J. Royal Statistical Society* B, 2010.
- T. Flury & N. Shephard, Bayesian inference based only on simulated likelihood, *Econometrics Review*, 2011.