Introduction to Sequential Monte Carlo Methods

SMC for Static Problems
Online Bayesian Parameter Estimation

Prof. Nicholas Zabaras
Materials Process Design and Control Laboratory
Sibley School of Mechanical and Aerospace Engineering
101 Frank H. T. Rhodes Hall
Cornell University
Ithaca, NY 14853-3801

Email: nzabaras@gmail.com
URL: http://mpdc.mae.cornell.edu/

April 27, 2013

Contents

References

- C.P. Robert & G. Casella, Monte Carlo Statistical Methods, Chapter 11
- J.S. Liu, Monte Carlo Strategies in Scientific Computing, Chapter 3, Springer-Verlag, New York.
- A. Doucet, N. De Freitas & N. Gordon (eds), Sequential Monte Carlo in Practice, Springer-Verlag: 2001
- A. Doucet, N. De Freitas, N.J. Gordon, An introduction to Sequential Monte Carlo, in SMC in Practice, 2001
- D. Wilkison, Stochastic Modelling for Systems Biology, Second Edition, 2006
- E. Ionides, Inference for Nonlinear Dynamical Systems, PNAS, 2006
- J.S. Liu and R. Chen, Sequential Monte Carlo methods for dynamic systems, JASA, 1998
- A. Doucet, Sequential Monte Carlo Methods, Short Course at SAMSI
- A. Doucet, Sequential Monte Carlo Methods & Particle Filters Resources
- Pierre Del Moral, Feynman-Kac models and interacting particle systems (SMC resources)
- A. Doucet, Sequential Monte Carlo Methods, Video Lectures, 2007
- N. de Freitas and A. Doucet, Sequential MC Methods, N. de Freitas and A. Doucet, Video Lectures, 2010

References

- M.K. Pitt and N. Shephard, Filtering via Simulation: Auxiliary Particle Filter, JASA, 1999
- A. Doucet, S.J. Godsill and C. Andrieu, On Sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comp., 2000
- J. Carpenter, P. Clifford and P. Fearnhead, An Improved Particle Filter for Non-linear Problems, IEE 1999.
- A. Kong, J.S. Liu & W.H. Wong, Sequential Imputations and Bayesian Missing Data Problems, JASA, 1994
- O. Cappe, E. Moulines & T. Ryden, Inference in Hidden Markov Models, Springer-Verlag, 2005
- W Gilks and C. Berzuini, Following a moving target: MC inference for dynamic Bayesian Models, JRSS B, 2001
- G. Poyadjis, A. Doucet and S.S. Singh, Maximum Likelihood Parameter Estimation using Particle Methods, Joint Statistical Meeting, 2005
- N Gordon, D J Salmond, AFM Smith, Novel Approach to nonlinear non Gaussian Bayesian state estimation, IEE, 1993
- Particle Filters, S. Godsill, 2009 (Video Lectures)
- R. Chen and J.S. Liu, Predictive Updating Methods with Application to Bayesian Classification, JRSS B, 1996

References

- C. Andrieu and A. Doucet, Particle Filtering for Partially Observed Gaussian State-Space Models, JRSS B, 2002
- R Chen and J Liu, Mixture Kalman Filters, JRSSB, 2000
- A Doucet, S J Godsill, C Andrieu, On SMC sampling methods for Bayesian Filtering, Stat. Comp. 2000
- N. Kantas, A.D., S.S. Singh and J.M. Maciejowski, An overview of sequential Monte Carlo methods for parameter estimation in general state-space models, in Proceedings IFAC System Identification (SySid) Meeting, 2009
- C. Andrieu, A.Doucet & R. Holenstein, Particle Markov chain Monte Carlo methods, JRSS B, 2010
- C. Andrieu, N. De Freitas and A. Doucet, Sequential MCMC for Bayesian Model Selection, Proc. IEEE Workshop HOS, 1999
- P. Fearnhead, MCMC, sufficient statistics and particle filters, JCGS, 2002
- G. Storvik, Particle filters for state-space models with the presence of unknown static parameters, IEEE Trans. Signal Processing, 2002

References

- C. Andrieu, A. Doucet and V.B. Tadic, Online EM for parameter estimation in nonlinear-non Gaussian state-space models, Proc. IEEE CDC, 2005
- G. Poyadjis, A. Doucet and S.S. Singh, Particle Approximations of the Score and Observed Information Matrix in State-Space Models with Application to Parameter Estimation, *Biometrika*, 2011
- C. Caron, R. Gottardo and A. Doucet, On-line Changepoint Detection and Parameter Estimation for Genome Wide Transcript Analysis, Technical report 2008
- R. Martinez-Cantin, J. Castellanos and N. de Freitas. Analysis of Particle Methods for Simultaneous Robot Localization and Mapping and a New Algorithm: Marginal-SLAM. International Conference on Robotics and Automation
- C. Andrieu, A.D. & R. Holenstein, Particle Markov chain Monte Carlo methods (with discussion), JRSS B, 2010
- A Doucet, Sequential Monte Carlo Methods and Particle Filters, List of Papers, Codes, and Viedo lectures on SMC and particle filters
- Pierre Del Moral, Feynman-Kac models and interacting particle systems

References

- P. Del Moral, A. Doucet and A. Jasra, Sequential Monte Carlo samplers, JRSSB, 2006
- P. Del Moral, A. Doucet and A. Jasra, Sequential Monte Carlo for Bayesian Computation, Bayesian Statistics, 2006
- P. Del Moral, A. Doucet & S.S. Singh, Forward Smoothing using Sequential Monte Carlo, technical report, Cambridge University, 2009
- A. Doucet, Short Courses Lecture Notes (A, B, C)
- P. Del Moral, Feynman-Kac Formulae, Springer-Verlag, 2004
- Sequential MC Methods, M. Davy, 2007
- A Doucet, A Johansen, Particle Filtering and Smoothing: Fifteen years later, in Handbook of Nonlinear Filtering (edts D Crisan and B. Rozovsky), Oxford Univ. Press, 2011
- A. Johansen and A. Doucet, A Note on Auxiliary Particle Filters, Stat. Proba. Letters, 2008.
- A. Doucet et al., Efficient Block Sampling Strategies for Sequential Monte Carlo, (with M. Briers & S. Senecal), JCGS, 2006.
- C. Caron, R. Gottardo and A. Doucet, On-line Changepoint Detection and Parameter Estimation for Genome Wide Transcript Analysis, Stat Comput. 2011.

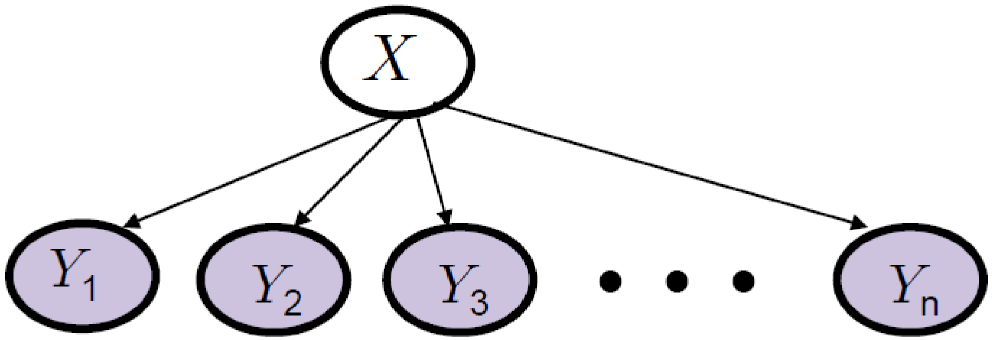
SEQUENTIAL MONTE CARLO METHODS FOR STATIC PROBLEMS

What about if all distributions π_n, $n \geq 1$ are defined on $\mathcal{X}$?

- This case can appear often, for example:
  - Sequential Bayesian Estimation: $\pi_n(x) = p(x \mid \boldsymbol{y}_{1:n})$
  - Classical Bayesian Inference: $\pi_n(x) = \pi(x)$
  - Global Optimization: $\pi_n(x) \propto \left[\pi(x)\right]^{\gamma_n}, \gamma_n \to \infty$
- It is not clear how SMC can be related to this problem since $\mathcal{X}_n = \mathcal{X}$ instead of $\mathcal{X}_n = \mathcal{X}^n$.

What about if all distributions π_n, $n \geq 1$ are defined on $\mathcal{X}$?

- This case can appear often, for example:
  - Sequential Bayesian Estimation: $\pi_n(x) = p(x \mid \boldsymbol{y}_{1:n})$



  - Global Optimization: $\pi_n(x) \propto \left[\pi(x)\right]^{\eta_n}, \eta_n \to \infty$
  - Sampling from a fixed distribution: $\pi_n(x) \propto \left(\mu(x)\right)^{\eta_n} \left(\pi(x)\right)^{1-\eta_n}$ where $\mu(x)$ is an easy to sample from distribution. Use a sequence

$$\eta_1 = 1 > \eta_2 > \ldots > \eta_{Final} = 0, \, i.e. \, \pi_1(x) \propto \mu(x), \, \pi_{Final}(x) \propto \pi(x)$$

What about if all distributions π_n, $n \geq 1$ are defined on $\mathcal{X}$?

- This case can appear often, for example:
  - Rare Event Simulation:

    $$\pi(A) << 1 : \pi_n(x) \propto \pi_n(x) 1_{E_n}(x), \; Normalizing\ Factor\ Z_1 = known$$

    $$Simulate\ with\ sequence : E_1 = \mathcal{X} \supset E_2 \supset \ldots \supset E_{Final} = A$$

    The required probability is then: $Z_{Final} = \pi(A)$

  - The Boolean Satisfiability Problem
  - Computing Matrix Permanents
  - Computing volumes in high dimensions

What about if all distributions π_n, $n \geq 1$ are defined on $\mathcal{X}$?

- Apply SMC to an augmented sequence of distributions $\left\{\widetilde{\pi}_n\right\}_{n\geq 1}$ on $\mathcal{X}^n$:

  For example:

$$\int \widetilde{\pi}_n\left(\boldsymbol{x}_{1:n-1}, x_n\right) d\boldsymbol{x}_{1:n-1} = \pi_n\left(x_n\right)$$

$$\widetilde{\pi}_n\left(\boldsymbol{x}_{1:n-1}, x_n\right) = \pi_n\left(x_n\right) \underbrace{\widetilde{\pi}_n\left(\boldsymbol{x}_{1:n-1} \mid x_n\right)}_{\substack{\text{Use any conditional} \\ \text{distribution on } \mathcal{X}^{n-1}}}$$

- C Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. 78, 2690–2693 (1997)
- Gavin E. Crooks, Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems, Journal of Statistical Physics March 1998, Volume 90, Issue 5-6, pp 1481-1487
- W Gilks and C. Berzuini, Following a moving target: MC inference for dynamic Bayesian Models,. JRSS B, 2001
- Neal, R. M. (2001) ``Annealed importance sampling'', Statistics and Computing, vol. 11, pp. 125-139
- N. Chopin, A sequential partile filter method for static problems, Biometrika, 2002, 89, 3, pp. 539-551.
- P. Del Moral, A. Doucet and A. Jasra, Sequential Monte Carlo for Bayesian Computation, Bayesian
- Statistics, 2006

SMC For Static Models

- Let $\{\pi_n(x)\}_{n\geq 1}$ be a sequence of probability distributions defined on $\mathcal{X}$ s.t. each $\pi_n(x)$ is known up to a normalizing constant:

$$\pi_n(x) = \underbrace{1/Z_n}_{Unknown} \underbrace{\gamma_n(x)}_{Known}$$

- We are interested to sample from $\pi_n(x)$ and compute Z_n sequentially.

- This is not the same as the standard SMC discussed earlier $\pi_n(\boldsymbol{x}_{1:n}) = p(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$ which was defined on $\mathcal{X}^n$.

SMC For Static Models

- We will construct an artificial distribution that is the product of the target distribution from which we want to sample and a backward kernel L as follows:

$$\tilde{\pi}_n(\boldsymbol{x}_{1:n}) = Z_n^{-1}\gamma_n(\boldsymbol{x}_{1:n}),\ where:$$

$$\gamma_n(\boldsymbol{x}_{1:n}) = \underbrace{\gamma_n(x_n)}_{Target}\underbrace{\prod_{k=1}^{n-1} L_k(x_k \mid x_{k+1})}_{\substack{Backward\\ Transitions}}$$

such that

$$\pi_n(x_n) = \int \tilde{\pi}_n(\boldsymbol{x}_{1:n})\, d\boldsymbol{x}_{1:n-1}$$

- The importance weights now become:

$$W_n = \frac{\gamma_n(\boldsymbol{x}_{1:n})}{q_n(\boldsymbol{x}_{1:n})} = W_{n-1}\frac{q_{n-1}(\boldsymbol{x}_{1:n-1})\gamma_n(\boldsymbol{x}_{1:n})}{\gamma_{n-1}(\boldsymbol{x}_{1:n-1})q_n(\boldsymbol{x}_{1:n})} = W_{n-1}\frac{\gamma_n(x_n)\prod_{k=1}^{n-1} L_k(x_k \mid x_{k+1})}{\gamma_{n-1}(x_{n-1})\prod_{k=1}^{n-2} L_k(x_k \mid x_{k+1}) q_n(x_n \mid x_{n-1})}$$

$$= W_{n-1}\frac{\gamma_n(x_n) L_{n-1}(x_{n-1} \mid x_n)}{\gamma_{n-1}(x_{n-1}) q_n(x_n \mid x_{n-1})}$$

SMC For Static Models

$$W_n = W_{n-1} \frac{\gamma_n(x_n) L_{n-1}(x_{n-1} \mid x_n)}{\gamma_{n-1}(x_{n-1}) q_n(x_n \mid x_{n-1})}$$

- Any MCMC kernel can be used for the proposal q(.|.).
- Since our interest is in computing only

$$\pi_n(x_n) = \int \tilde{\pi}_n(\boldsymbol{x}_{1:n}) d\boldsymbol{x}_{1:n-1} = \frac{1}{Z}\gamma_n(x_n)$$

there is no degeneracy problem.

P. Del Moral, A. Doucet and A. Jasra, Sequential Monte Carlo for Bayesian Computation, Bayesian Statistics, 2006

Algorithm: SMC For Static Problems

(1) Initialize at time n=1:

(2) At time n≥2:

- Sample $\widetilde{X}_n^{(i)} \sim q_n\left(x_n \mid X_{n-1}^{(i)}\right)$, and augment $\widetilde{\boldsymbol{X}}_{n-1:n}^{(i)} \sim \left(X_{n-1}^{(i)}, \widetilde{X}_n^{(i)}\right)$

- Compute the importance weights

$$W_n^{(i)} = W_{n-1}^{(i)} \frac{\gamma_n\left(\widetilde{X}_n^{(i)}\right) L_{n-1}\left(\widetilde{X}_{n-1}^{(i)} \mid \widetilde{X}_n^{(i)}\right)}{\gamma_{n-1}\left(\widetilde{X}_{n-1}^{(i)}\right) q_n\left(\widetilde{X}_n^{(i)} \mid \widetilde{X}_{n-1}^{(i)}\right)}$$

Then the weighted approximation is

$$\widetilde{\pi}_n\left(x_n\right) = \sum_{i=1}^{N} W_n^{(i)} \delta_{\widetilde{X}_n^{(i)}}\left(x_n\right)$$

- We finally resample from $X_n^{(i)} \sim \widetilde{\pi}_n\left(x_n\right)$ to obtain:

$$\pi_n\left(x_n\right) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_n^{(i)}}\left(x_n\right)$$

SMC For Static Models: Choice of L

- A default choice is first using a π_n-invariant MCMC kernel q_n and then the corresponding reversed kernel L_{n-1}:

$$L_{n-1}\left(x_{n-1} \mid x_n\right) = \frac{\pi_n\left(x_{n-1}\right) q_n\left(x_n \mid x_{n-1}\right)}{\pi_n\left(x_n\right)}$$

- Using this easy choice, we can simplify the expression for the weights:

$$W_n = W_{n-1} \frac{\gamma_n\left(x_n\right) L_{n-1}\left(x_{n-1} \mid x_n\right)}{\gamma_{n-1}\left(x_{n-1}\right) q_n\left(x_n \mid x_{n-1}\right)} = W_{n-1} \frac{\gamma_n\left(x_n\right)}{\gamma_{n-1}\left(x_{n-1}\right) q_n\left(x_n \mid x_{n-1}\right)} \frac{\pi_n\left(x_{n-1}\right) q_n\left(x_n \mid x_{n-1}\right)}{\pi_n\left(x_n\right)} \Rightarrow$$

$$W_n = W_{n-1} \frac{\gamma_n\left(X_{n-1}^{(i)}\right)}{\gamma_{n-1}\left(X_{n-1}^{(i)}\right)}$$

- This is known as *annealed importance sampling*. The particular choice has been used in physics and statistics.

W Gilks and C. Berzuini, Following a moving target: MC inference for dynamic Bayesian Models, JRSS B, 2001

ONLINE PARAMETER ESTIMATION

Online Bayesian Parameter Estimation

- Assume that our state model is defined with some unknown static parameter θ with some prior $p(\theta)$:

$$X_1 \sim \mu(.) \; and \; X_n \mid \left(X_{n-1} = x_{n-1}\right) \sim f_\theta\left(x_n \mid x_{n-1}\right)$$

$$Y_n \mid \left(X_n = x_n\right) \sim g_\theta\left(y_n \mid x_n\right)$$

- Given data $\mathbf{y}_{1:n}$, inference now is based on:

$$p\left(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right) = p\left(\theta \mid \boldsymbol{y}_{1:n}\right) p_\theta\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right),$$

$$where$$

$$p\left(\theta \mid \boldsymbol{y}_{1:n}\right) \propto p_\theta\left(\boldsymbol{y}_{1:n}\right) p\left(\theta\right)$$

- We can use standard SMC but on the extended space $Z_n=(X_n,\theta_n)$.

$$f\left(z_n \mid z_{n-1}\right) = \delta_{\theta_{n-1}}\left(\theta_n\right) f_\theta\left(x_n \mid x_{n-1}\right), \; g\left(y_n \mid z_n\right) = g_\theta\left(y_n \mid x_n\right)$$

- Note that θ is a static parameter –does not involve with n.

Online Bayesian Parameter Estimation

- For fixed θ, using our earlier error estimates

$$Var\left[\log \widehat{p}_{\theta}(\boldsymbol{y}_{1:n})\right] \leq \frac{Cn}{N}$$

- In a Bayesian context, the problem is even more severe as

$$p\left(\theta \mid \boldsymbol{y}_{1:n}\right) \propto p_{\theta}\left(\boldsymbol{y}_{1:n}\right) p\left(\theta\right)$$

- Exponential stability assumption cannot hold as $\theta_n = \theta_1$.

- To mitigate this problem, introduce MCMC steps on θ.

- C. Andrieu, N. De Freitas and A. Doucet, Sequential MCMC for Bayesian Model Selection, Proc. IEEE Workshop HOS, 1999
- P. Fearnhead, MCMC, sufficient statistics and particle filters, JCGS, 2002
- W Gilks and C. Berzuini, Following a moving target: MC inference for dynamic Bayesian Models, JRSS B, 2001

Online Bayesian Parameter Estimation

- When

$$p(\theta \mid \boldsymbol{y}_{1:n}, \boldsymbol{x}_{1:n}) = p\left(\theta \mid \underbrace{s_n(\boldsymbol{x}_{1:n}, \boldsymbol{y}_{1:n})}_{\substack{Fixed \\ Dimensional}}\right)$$

- This becomes an elegant algorithm that however still has the degeneracy problem since it uses $\widehat{p}(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$

- As dim(Z_n) = dim(X_n) + dim(θ), such methods are not recommended for high-dimensional θ, especially with vague priors.

Artificial Dynamics for θ

- A solution consists of perturbing the location of the particles $\left\{\theta^{(i)}\right\}$ in a way that does not modify their distributions; i.e. if at time n

$$\theta_n^{(i)} \sim p\left(\theta \mid \boldsymbol{y}_{1:n}\right)$$

then we would like a transition kernel such that if

$$\theta_n^{'(i)} \mid \theta_n^{(i)} \sim M_n\left(\theta_n^{(i)}, .\right)$$

Then:

$$\theta_n^{'(i)} \sim p\left(\theta \mid \boldsymbol{y}_{1:n}\right)$$

- In Markov chain language, we want $M_n\left(\theta, \theta'\right)$ to be $p\left(\theta \mid \boldsymbol{y}_{1:n}\right)$ invariant.

Using MCMC

- There is a whole literature on the design of such kernels known as Markov chain Monte Carlo e.g. the Metropolis-Hastings algorithm.

- We cannot use these algorithms directly as $p(\theta \mid \boldsymbol{y}_{1:n})$ would need to be known up to a normalizing constant but $p(\boldsymbol{y}_{1:n} \mid \theta) \equiv p_\theta(\boldsymbol{y}_{1:n})$ is unknown.

- However, we can use a very simple Gibbs update.

$$\theta_n^{(i)} \sim p\left(\theta \mid \boldsymbol{y}_{1:n}, \boldsymbol{X}_{1:n}^{(i)}\right)$$

- Indeed note that if $\left(\boldsymbol{X}_{1:n}^{(i)}, \theta_n^{(i)}\right) \sim p(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$ then if $\theta_n^{'(i)} \sim p\left(\theta \mid \boldsymbol{y}_{1:n}, \boldsymbol{X}_{1:n}^{(i)}\right)$, we have:

$$\left(\boldsymbol{X}_{1:n}^{(i)}, \theta_n^{'(i)}\right) \sim p(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$$

- Indeed note that:

$$\int p(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})\, p(\theta' \mid \boldsymbol{y}_{1:n})\, d\theta = p(\theta', \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$$

SMC with MCMC for Parameter Estimation

- Given an approximation at time n-1:

$$\widehat{p}\left(\theta, \boldsymbol{x}_{1:n-1} \mid \boldsymbol{y}_{1:n-1}\right) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\left(\theta_{n-1}^{(i)}, \boldsymbol{X}_{1:n-1}^{(i)}\right)}\left(\theta, \boldsymbol{x}_{1:n-1}\right)$$

- Sample $\widetilde{X}_n^{(i)} \sim f_{\theta_{n-1}^{(i)}}\left(x_n \mid X_{n-1}^{(i)}\right)$ set $\widetilde{\boldsymbol{X}}_{1:n}^{(i)} \sim \left(\boldsymbol{X}_{1:n-1}^{(i)}, \widetilde{X}_n^{(i)}\right)$ and then approximate:

$$\widetilde{p}\left(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right) = \sum_{i=1}^{N} W_n^{(i)}\delta_{\left(\theta_{n-1}^{(i)}, \widetilde{\boldsymbol{X}}_{1:n}^{(i)}\right)}\left(\boldsymbol{x}_{1:n}\right), W_n^{(i)} \propto g_{\theta_{n-1}^{(i)}}\left(y_n \mid \widetilde{\boldsymbol{X}}_n^{(i)}\right)$$

- Resample $\boldsymbol{X}_{1:n}^{(i)} \sim \widetilde{p}\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right)$, then sample $\theta_n^{(i)} \sim p\left(\theta \mid \boldsymbol{y}_{1:n}, \boldsymbol{X}_{1:n}^{(i)}\right)$ to obtain

$$\widehat{p}\left(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\left(\theta_n^{(i)}, \boldsymbol{X}_{1:n}^{(i)}\right)}\left(\theta, \boldsymbol{x}_{1:n}\right)$$

SMC with MCMC for Parameter Estimation

- Consider the following model:

$$X_{n+1} = \theta X_n + \sigma_v V_{n+1}, V_n \sim \mathcal{N}(0,1)$$
$$Y_n = X_n + \sigma_w W_n, W_n \sim \mathcal{N}(0,1)$$
$$X_1 \sim \mathcal{N}\left(0, \sigma_0^2\right)$$

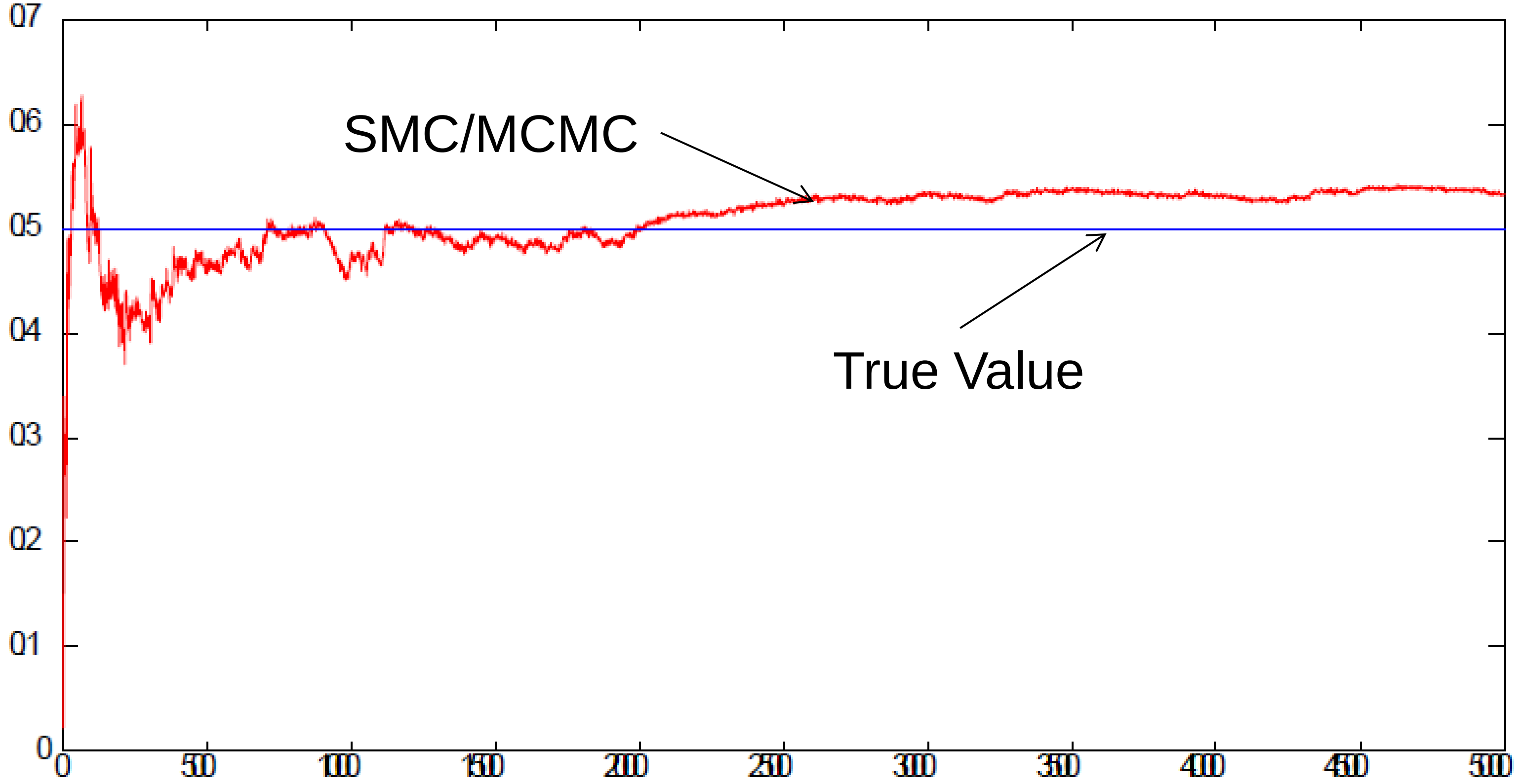
- We set the prior on θ as $\theta \sim \mathcal{U}(-1,1)$.

- In this case,

$$p(\theta \mid \boldsymbol{x}_{1:n}, \boldsymbol{y}_{1:n}) \propto \mathcal{N}\left(\theta, m_n, \sigma_n^2\right) \mathbb{I}_{(-1,1)}(\theta)$$
$$m_n = \sigma_n^2 \left(\sum_{k=2}^{n} x_k x_{k-1} \right), \sigma_n^2 = \sum_{k=2}^{n-1} x_k^2$$

SMC with MCMC for Parameter Estimation

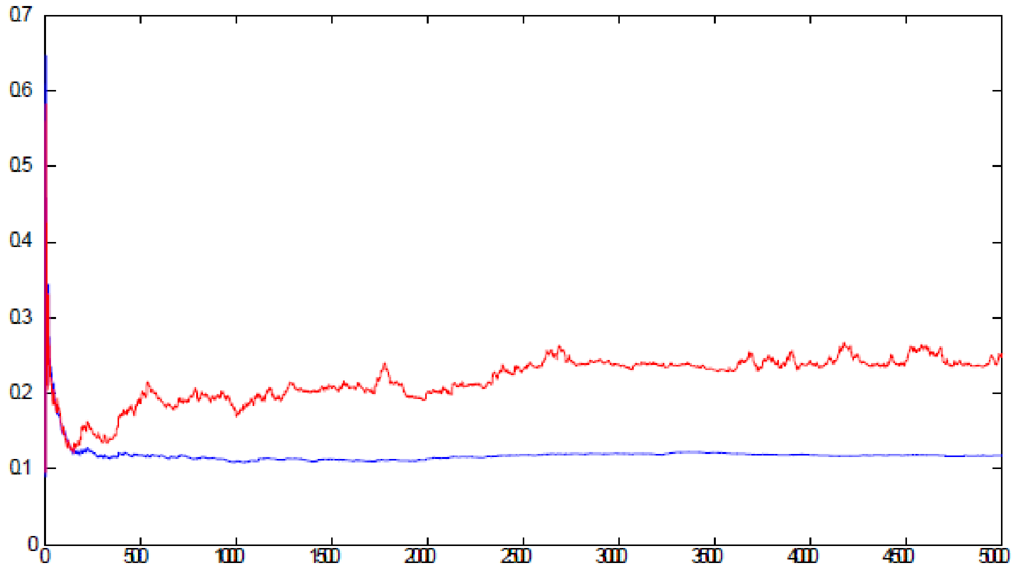
- We use SMC with Gibbs step. The degeneracy problem remains. SMC estimate is shown of $\mathbb{E}[\theta|\mathbf{y}_{1:n}]$ as n increases (From A. Doucet, lecture notes). The parameter converges to the wrong value.



SMC with MCMC for Parameter Estimation

- The problem with this approach is that although we move θ according to $p(\theta \mid \boldsymbol{x}_{1:n}, \boldsymbol{y}_{1:n})$, we are still relying implicitly on the approximation of the joint distribution $p(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$

- As n increases, the SMC approximation of $p(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$ deteriorates and the algorithm cannot converge towards the right solution.



Sufficient statistics $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[X_k^2 \mid \boldsymbol{y}_{1:n}\right]$ computed exactly through the Kalman filter (blue) vs SMC approximation (red) for a fixed value of θ.

Recursive MLE Parameter Estimation

- The log likelihood can be written as:

$$\mathrm{l}_n(\theta) = \log p_\theta(\boldsymbol{Y}_{1:n}) = \sum_{k=1}^{n} \log p_\theta\left(Y_k \mid \boldsymbol{Y}_{1:k-1}\right)$$

- Here we compute:

$$p_\theta\left(Y_k \mid \boldsymbol{Y}_{1:k-1}\right) = \int g_\theta\left(Y_k \mid x_k\right) p_\theta\left(x_k \mid \boldsymbol{Y}_{1:k-1}\right) dx_k$$

- Under regularity assumptions $\left\{X_n, Y_n, p_\theta\left(x_n \mid \boldsymbol{Y}_{1:n-1}\right)\right\}$ is an inhomogeneous Markov chain which converges towards its invariant distribution:

$$\lim_{n\to\infty} \frac{\mathrm{l}_n(\theta)}{n} = \mathrm{l}(\theta) = \int \log \int g_\theta(y \mid x)\mu(dx)\lambda_{\theta,\theta^*}(dy, d\mu)$$

Robbins-Monro Algorithm for MLE

- We can maximize l(q) by using the gradient and the Robbins-Monro algorithm:

$$\theta_n = \theta_{n-1} + \gamma_n \nabla \log p_{\theta_{1:n-1}}(Y_n \mid \boldsymbol{Y}_{1:n-1})$$

where:
$$\nabla p_\theta(Y_n \mid \boldsymbol{Y}_{1:n-1}) = \int \nabla g_\theta\left(Y_n \mid x_n\right) p_\theta\left(x_n \mid \boldsymbol{Y}_{1:n-1}\right) dx_n + \int g_\theta\left(Y_n \mid x_n\right) \nabla p_\theta\left(x_n \mid \boldsymbol{Y}_{1:n-1}\right) dx_n$$

- We thus need to approximate $\left\{\nabla p_\theta\left(x_n \mid \boldsymbol{Y}_{1:n-1}\right)\right\}$

Importance Sampling Estimation of Sensitivity

- The various proposed approximations are based on the identity:

$$\nabla p_\theta(Y_n \mid \boldsymbol{Y}_{1:n-1}) = \int \frac{\nabla p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1})}{p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1})} p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1}) d\boldsymbol{x}_{1:n-1}$$

$$= \int \nabla \log p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1}) p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1}) d\boldsymbol{x}_{1:n-1}$$

- We thus can use a SMC approximation of the form:

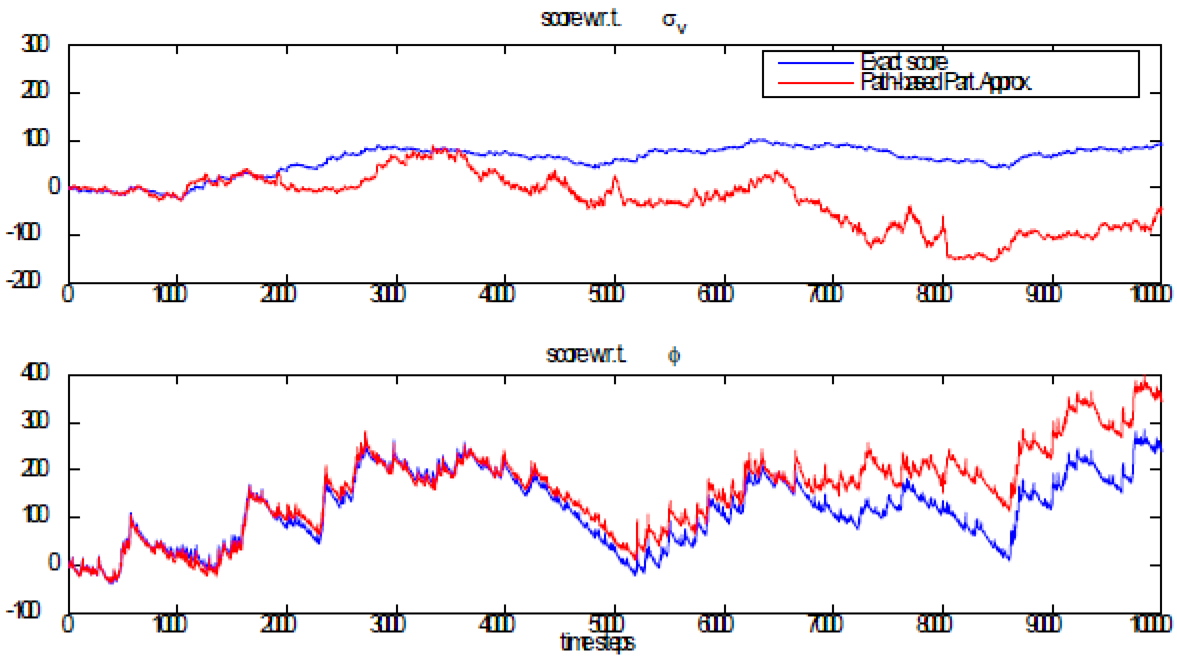
$$\widehat{\nabla p_\theta}(x_n \mid \boldsymbol{Y}_{1:n-1}) = \frac{1}{N} \sum_{i=1}^{N} \alpha_n^{(i)} \delta_{X_n^{(i)}}(\boldsymbol{x}_{1:n})$$

$$\alpha_n^{(i)} = \widehat{\nabla \log p_\theta}(\boldsymbol{X}_{1:n}^{(i)} \mid \boldsymbol{Y}_{1:n-1})$$

- This is simple but inefficient as based on IS on spaces of increasing dimension and in addition it relies on being able to obtain a good approximation of $p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{Y}_{1:n-1})$

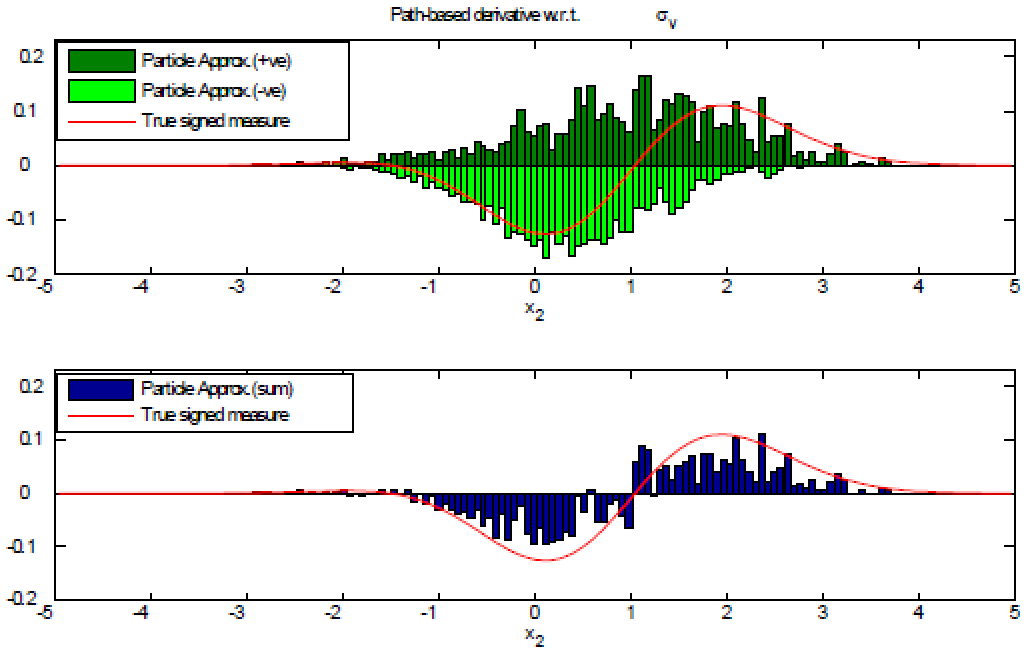
Degeneracy of the SMC Algorithm

- Score $\nabla p_\theta(\boldsymbol{Y}_{1:n})$ for linear Gaussian models: exact (blue) vs SMC approximation (red).



Degeneracy of the SMC Algorithm

- Signed measure $\nabla_\theta p(x_n | \boldsymbol{Y}_{1:n})$ for linear Gaussian models: exact (red) vs SMC (blue/green). Positive and negative particles are mixed.



Marginal Importance Sampling for Sensitivity Estimation

- An alternative identity is the following:

$$\nabla p_\theta(x_n \mid \boldsymbol{Y}_{1:n-1}) = \nabla \log p_\theta(x_n \mid \boldsymbol{Y}_{1:n-1})\, p_\theta(x_n \mid \boldsymbol{Y}_{1:n-1})$$

- It suggests using an SMC approximation of the form:

$$\widehat{\nabla p_\theta}(x_n \mid \boldsymbol{Y}_{1:n-1}) = \frac{1}{N}\sum_{i=1}^{N} \beta_n^{(i)} \delta_{X_n^{(i)}}(\boldsymbol{x}_{1:n})$$

$$\beta_n^{(i)} = \widehat{\nabla \log p_\theta}(X_n^{(i)} \mid \boldsymbol{Y}_{1:n-1}) = \frac{\widehat{\nabla p_\theta}(X_n^{(i)} \mid \boldsymbol{Y}_{1:n-1})}{\widehat{p}_\theta(X_n^{(i)} \mid \boldsymbol{Y}_{1:n-1})}$$

- Such an approximation relies on a pointwise approximation of both the filter and its derivative. The computational complexity is O(N^2) as

$$\widehat{p}_\theta(X_n^{(i)} \mid \boldsymbol{Y}_{1:n-1}) \propto \int f_\theta(X_n^{(i)} \mid x_{n-1})\, \widehat{p}_\theta(x_{n-1} \mid \boldsymbol{Y}_{1:n-1})\, dx_n = \frac{1}{N}\sum_{j=1}^{N} f_\theta(X_n^{(i)} \mid X_{n-1}^{(j)})$$

Marginal Importance Sampling for Sensitivity Estimation

- The optimal filter satisfies:

$$p_\theta(x_n \mid \boldsymbol{Y}_{1:n}) = \frac{\xi_\theta(x_n \mid \boldsymbol{Y}_{1:n})}{\int \xi_\theta(x_n \mid \boldsymbol{Y}_{1:n})dx_n}$$

$$\xi_\theta(x_n \mid \boldsymbol{Y}_{1:n}) = g_\theta\left(Y_n \mid x_n\right)\int f_\theta\left(x_n \mid x_{n-1}\right) p_\theta(x_{n-1} \mid \boldsymbol{Y}_{1:n-1})dx_{n-1}$$

- The derivatives satisfy the following:

$$\nabla p_\theta(x_n \mid \boldsymbol{Y}_{1:n}) = \frac{\nabla \xi_\theta(x_n \mid \boldsymbol{Y}_{1:n})}{\int \xi_\theta(x_n \mid \boldsymbol{Y}_{1:n})dx_n} - p_\theta(x_n \mid \boldsymbol{Y}_{1:n})\frac{\int \nabla \xi_\theta(x_n \mid \boldsymbol{Y}_{1:n})dx_n}{\int \xi_\theta(x_n, \boldsymbol{Y}_{1:n})dx_n}$$

- This way we obtain a simple recursion of $\nabla p_\theta(x_n \mid \boldsymbol{Y}_{1:n})$ as a function of $\nabla p_\theta(x_{n-1} \mid \boldsymbol{Y}_{1:n-1})$ and $p_\theta(x_{n-1} \mid \boldsymbol{Y}_{1:n-1})$.

SMC Approximation of the Sensitivity

- Sample

$$X_n^{(i)} \sim q_\theta\left(.\,|\,Y_n\right) = \sum_{j=1}^{N} \eta_n^{(j)} q_\theta\left(.\,|\,Y_n, X_{n-1}^{(j)}\right),\ \eta_n^{(j)} \propto W_{n-1}^{(j)}\,\widehat{p}\left(Y_n \,|\, X_{n-1}^{(j)}\right)$$

- Compute:

$$\alpha_n^{(i)} = \frac{\hat{\xi}_\theta(X_n^{(i)}, \boldsymbol{Y}_{1:n})}{q_\theta\left(X_n^{(i)} \,|\, Y_n\right)},\ \rho_n^{(i)} = \frac{\widehat{\nabla \xi}_\theta(X_n^{(i)}, \boldsymbol{Y}_{1:n})}{q_\theta\left(X_n^{(i)} \,|\, Y_n\right)}$$

$$W_n^{(i)} = \frac{\alpha_n^{(i)}}{\sum_{j=1}^{N} \alpha_n^{(j)}},\ W_n^{(i)} \beta_n^{(i)} = \frac{\rho_n^{(i)}}{\sum_{j=1}^{N} \alpha_n^{(j)}} - W_n^{(i)} \frac{\sum_{j=1}^{N} \rho_n^{(j)}}{\sum_{j=1}^{N} \alpha_n^{(j)}}$$

- We have:

$$\widehat{p}_\theta(x_n \,|\, \boldsymbol{Y}_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_n^{(i)}}(x_n)$$

$$\widehat{\nabla p_\theta}(x_n \,|\, \boldsymbol{Y}_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \beta_n^{(i)} \delta_{X_n^{(i)}}(x_n)$$

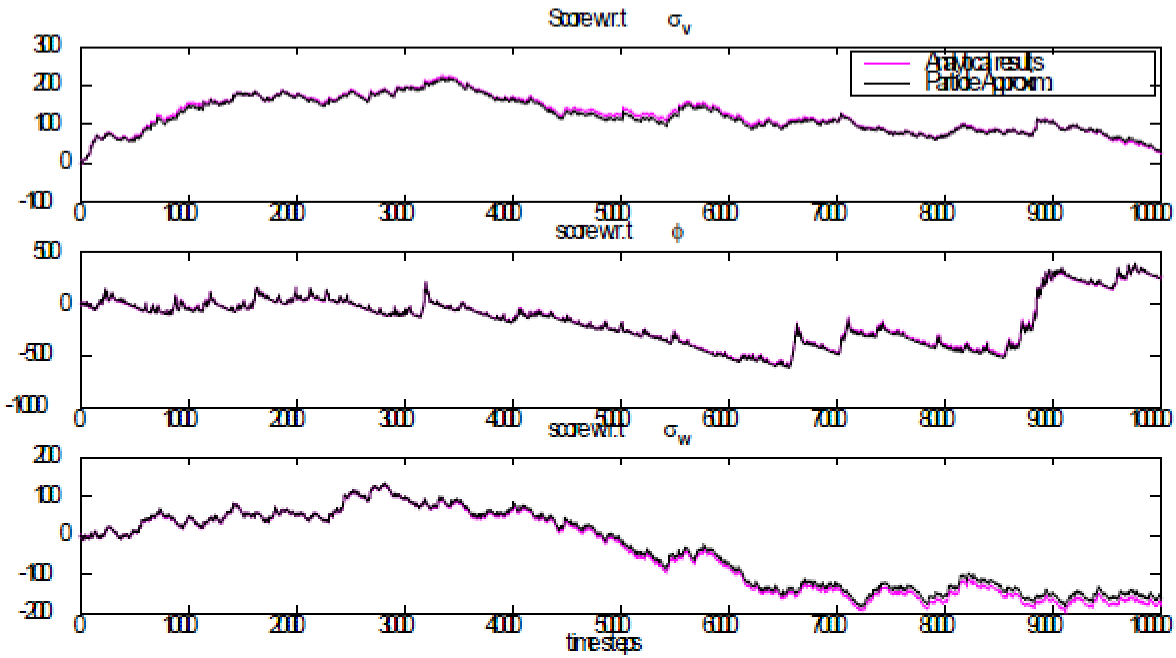
Example: Linear Gaussian Model

- Consider the following model: $$X_n = \phi X_{n-1} + \sigma_v V_n$$ $$Y_n = X_n + \sigma_w W_n$$

- We have $\theta = \{\phi, \sigma_v, \sigma_w\}$

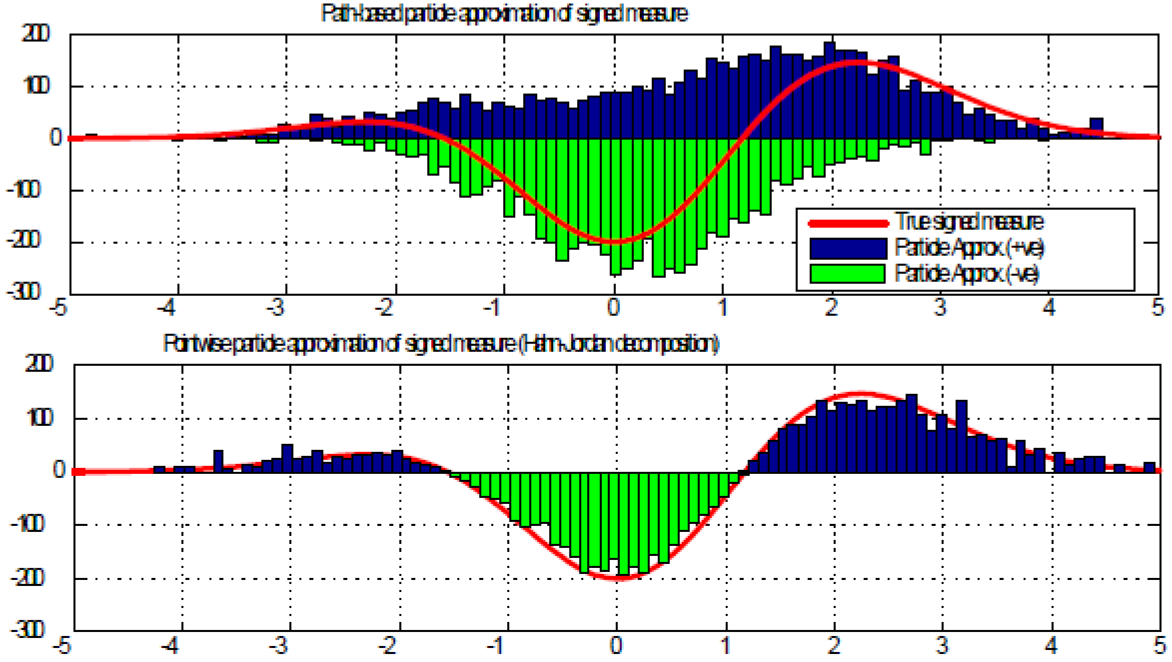
- We compare next the SMC estimate $\widehat{\nabla \log p_\theta(x_n \mid \boldsymbol{Y}_{1:n})}$ for N=1000 to its exact value given by the Kalman filter derivative (exact with cyan vs SMC with black).

$\widehat{\nabla p_\theta}(\boldsymbol{Y}_{1:n})$



Example: Linear Gaussian Model

- Marginal of the Hahn Jordan of the joint (top) and Hahn Jordan of the marginal (bottom)



Example: Stochastic Volatility Model

- Consider the following model:

$$X_n = \phi X_{n-1} + \sigma_v V_n$$

$$Y_n = \beta \exp\left(\frac{X_n}{2}\right) W_n$$

- We have $\theta = \{\phi, \sigma_v, \beta\}$. We use SMC with N=1000 for batch and on-line estimation.

- For Batch Estimation:

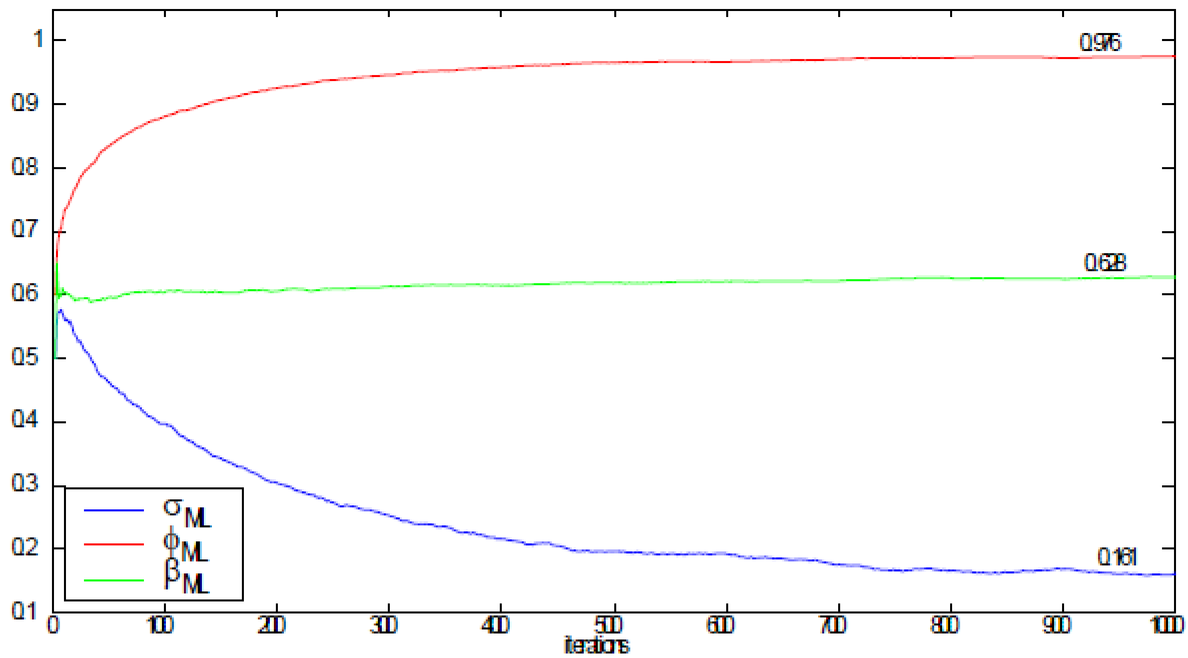
$$\theta_n = \theta_{n-1} + \gamma_n \nabla \widehat{\log p_{\theta_{n-1}}}(\boldsymbol{Y}_{1:n})$$

- For online estimation:

$$\theta_n = \theta_{n-1} + \gamma_n \nabla \widehat{\log p_{\theta_{1:n-1}}}(Y_n \mid \boldsymbol{Y}_{1:n-1})$$

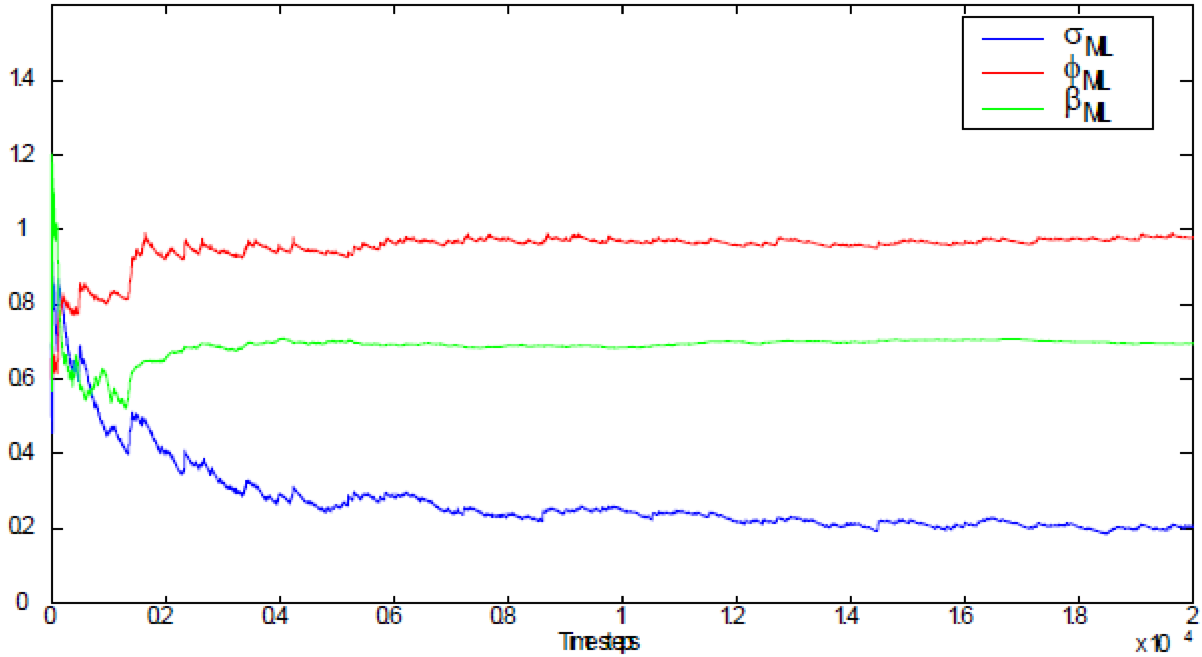
Example: Stochastic Volatility Model

- Batch Parameter Estimation for Daily Exchange Rate Pound/Dollar 81-85



Example: Stochastic Volatility Model

- Recursive parameter estimation for SV model. The estimates converge towards the true values.



SMC with MCMC for Parameter Estimation

- Given data $\mathbf{y}_{1:n}$, inference relies on

$$p(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}) = p(\theta \mid \boldsymbol{y}_{1:n})\, p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$$

where

$$p(\theta \mid \boldsymbol{y}_{1:n}) = p_\theta(\boldsymbol{y}_{1:n})\, p(\theta)$$

- We have seen that SMC are rather inefficient to sample from $p(\theta \mid \boldsymbol{y}_{1:n})$ so we look here at an MCMC approach.

- For a given parameter value θ, SMC can estimate both

$$p_\theta(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}) \text{ and } p_\theta(\boldsymbol{y}_{1:n})$$

- It will be useful if we can use SMC within MCMC to sample from

$$p(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n})$$

- C. Andrieu, R. Holenstein and G.O. Roberts, Particle Markov chain Monte Carlo methods, J. R. Statist. Soc.B (2010) 72, Part 3, pp. 269–342

Gibbs Sampling Strategy

- Using a Gibbs Sampling, we can sample iteratively from $p\left(\theta \mid \boldsymbol{x}_{1:n}, \boldsymbol{y}_{1:n}\right)$ and $p\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$

- However, it is impossible to sample from $p\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$

- We can sample from $p\left(x_k \mid \boldsymbol{y}_{1:n}, \theta, \boldsymbol{x}_{1:k-1}, \boldsymbol{x}_{k+1:n}\right)$ instead but convergence will be slow.

- Alternatively, we would like to use Metropolis Hastings step to sample from $p\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$, i.e. sample $\boldsymbol{X}_{1:n}^* \sim q\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right)$ and accept with probability

$$\min\left(1, \frac{p\left(\boldsymbol{X}_{1:n}^* \mid \boldsymbol{y}_{1:n}, \theta\right) q\left(\boldsymbol{X}_{1:n}^* \mid \boldsymbol{y}_{1:n}\right)}{p\left(\boldsymbol{X}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right) p\left(\boldsymbol{X}_{1:n}^* \mid \boldsymbol{y}_{1:n}\right)}\right)$$

Gibbs Sampling Strategy

- We will use the output of an SMC method approximating $p\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$ as a proposal distribution.

- We know that the SMC approximation $\widehat{p}\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$ degrades an n increases but we also have under mixing assumptions:

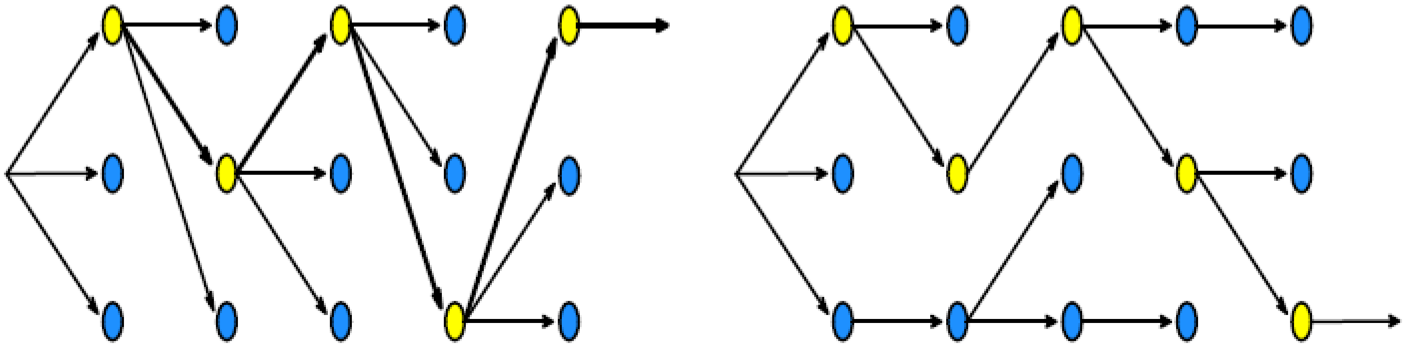
$$\left\| \mathcal{L}aw(\boldsymbol{X}_{1:n}^{(i)}) - p\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right) \right\|_{TV} \leq \frac{Cn}{N}$$

  Thus things degrade linearly rather than exponentially.

- These results are useful for small C.

Configurational Biased Monte Carlo

- The idea is related to that in the Configurational Biased MC Method (CBMC) in Molecular simulation. There are however significant differences.



CBMC for $N = 3$ particles SMC for $N = 3$ particles

- CBMC looks like an SMC method but at each step only one particle is selected that gives N offsprings. Thus if you have done the wrong selection, then you cannot recover later on.

D Frenkel, G C A M Mooij and B Smit, Novel scheme to study structural and thermal properties of continuously deformable molecules, I. Phys.: Condens. Matter 3 (1991) 3053-3076.

MCMC

- We use as proposal distribution $\boldsymbol{X}_{1:n}^{*} \sim \widehat{p}_N\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}, \theta\right)$ and store the estimate $\widehat{p}_N\left(\boldsymbol{y}_{1:n} \mid \theta\right)$

- It is impossible to compute analytically the unconditional law of a particle as it would require integrating out (N-1)n variables.

- The solution inspired by CBMC is as follows: For the current state of the Markov chain $\boldsymbol{X}_{1:n}$ run a virtual SMC method with only N-1 free particles. Ensure that at each time step k, the current state Xk is deterministically selected and compute the associated estimate $\widehat{p}_N\left(\boldsymbol{y}_{1:n} \mid \theta\right)$. Finally accept $\boldsymbol{X}_{1:n}^{*}$ with probability

$$\min\left(1, \frac{\widehat{p}_N\left(\boldsymbol{y}_{1:n} \mid \theta\right)}{\widetilde{p}_N\left(\boldsymbol{y}_{1:n} \mid \theta\right)}\right)$$

Otherwise stay where you are.

D Frenkel, G C A M Mooij and B Smit, Novel scheme to study structural and thermal properties of continuously deformable molecules, I. Phys.: Condens. Matter 3 (1991) 3053-3076.

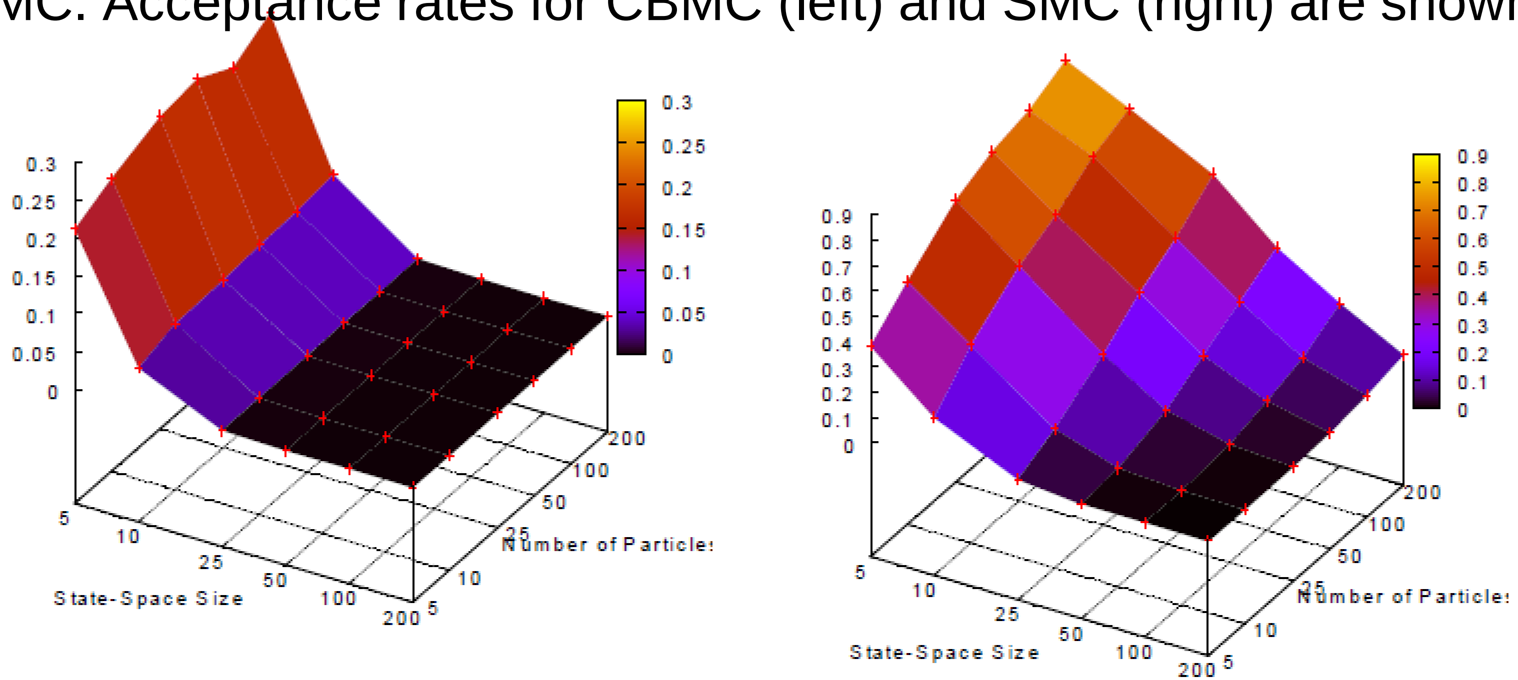
Example

- Consider the following model:

$$X_k = \frac{1}{2}X_{k-1} + 25\frac{X_{k-1}}{1+X_{k-1}^2} + 8\cos(1.2k) + V_k,\ V_k \sim \mathcal{N}(0,15)$$

$$Y_n = \frac{X_k^2}{2} + W_k,\ W_k \sim \mathcal{N}(0,0.01),\ X_1 \sim \mathcal{N}(0,5)$$

- We take the same prior proposal distribution for both CBMC and SMC. Acceptance rates for CBMC (left) and SMC (right) are shown.



Example

- We now consider the case when both variances σ_v^2, σ_w^2 are unknown and given inverse Gamma (conditionally conjugate) vague priors.

- We sample from $p\left(\boldsymbol{x}_{1:1000}, \theta \mid \boldsymbol{y}_{1:1000}\right)$ using two strategies
  - The MCMC algorithm with N=1000 and the local EKF proposal as importance distribution. Average acceptance rate 0.43.
  - An algorithm updating the components one at a time using an MH step

    of invariance distribution $p\left(x_k \mid x_{k-1}, x_{k+1}, y_k\right)$ with the same proposal.

- In both cases we update the parameters according to $p\left(\theta \mid \boldsymbol{x}_{1:500}, \boldsymbol{y}_{1:500}\right)$

- Both algorithms are run with the same computational effort.

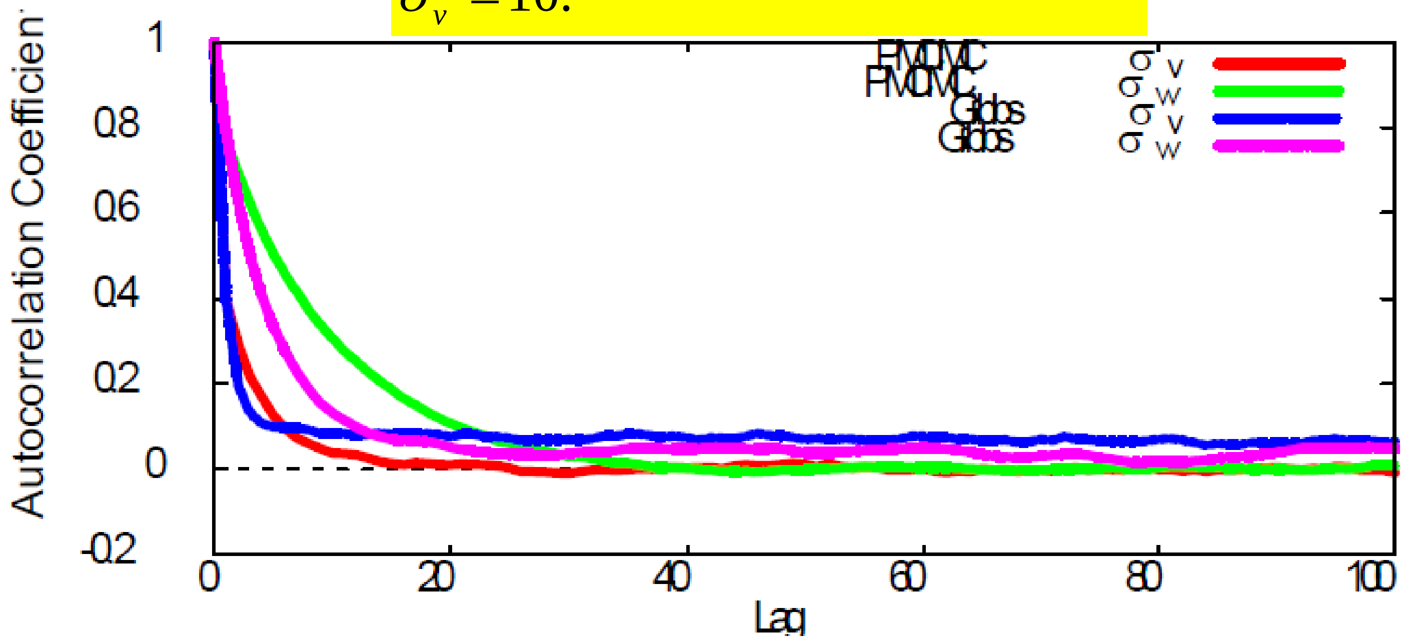
Example

- MH one at a time missed the mode and estimates

$$\mathbb{E}\left[\sigma_v^2 \mid \boldsymbol{y}_{1:500}\right] \approx 13.7\,(MH)$$

$$\mathbb{E}\left[\sigma_v^2 \mid \boldsymbol{y}_{1:500}\right] \approx 9.9\,(PF-MCMC)$$

$$\sigma_v^2 = 10.$$



- If $\mathbf{X}_{1:n}$ and θ are very correlated the MCMC algorithm is very inefficient. We will next discuss the Marginal Metropolis Hastings.

Review of Metropolis Hastings Algorithm

- As a review of MH, the proposal distribution is a Markov Chain with kernel density $q(x, y)$=q(y|x) and target distribution $\pi(x)$.

Algorithm: Generic Metropolis Hastings Sampler

- Initialization: Choose an arbitrary starting value x^0
- Iteration t $(t \geq 1)$
  1. Given x^{t-1}, generate $\tilde{x} \sim q(x^{(t-1)}, x)$
  2. Compute:

$$\rho\left(x^{(t-1)}, \tilde{x}\right) = \min\left(1, \frac{\pi(\tilde{x}) / q(x^{(t-1)}, \tilde{x})}{\pi\left(x^{(t-1)}\right) / q(\tilde{x}, x^{t-1})}\right)$$

  3. With probability $\rho(x^{t-1}, \tilde{x})$, accept $\tilde{x}$ and set $x^t = \tilde{x}$;
     Otherwise reject $\tilde{x}$ and set $x^t = x^{t-1}$.

- It can be easily shown that $\pi(x') = \int \pi(x) \underbrace{K(x, x')}_{Transition\ Kernel} dx$ and

under weak assumptions: $X^{(i)} \sim \pi(x) \; as \; i \to \infty$

Marginal Metropolis Hastings Algorithm

- Consider the target:

$$p\left(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right) = p\left(\theta \mid \boldsymbol{y}_{1:n}\right) p_{\theta}\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right)$$

- We use the following proposal distribution:

$$q\left(\left(\boldsymbol{x}_{1:n}^{*}, \theta^{*}\right) \mid \left(\boldsymbol{x}_{1:n}, \theta\right)\right) = q\left(\theta^{*} \mid \theta\right) p_{\theta^{*}}\left(\boldsymbol{x}_{1:n}^{*} \mid \boldsymbol{y}_{1:n}\right)$$

- Then the acceptance probability becomes:

$$\min\left(1, \frac{p\left(\theta^{*}, \boldsymbol{x}_{1:n}^{*} \mid \boldsymbol{y}_{1:n}\right) q\left(\left(\boldsymbol{x}_{1:n}, \theta\right) \mid \left(\boldsymbol{x}_{1:n}^{*}, \theta^{*}\right)\right)}{p\left(\theta, \boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right) q\left(\left(\boldsymbol{x}_{1:n}^{*}, \theta^{*}\right) \mid \left(\boldsymbol{x}_{1:n}, \theta\right)\right)}\right) =$$

$$\min\left(1, \frac{p_{\theta^{*}}\left(\boldsymbol{y}_{1:n}\right) p\left(\theta^{*}\right) q\left(\theta \mid \theta *\right)}{p_{\theta}\left(\boldsymbol{y}_{1:n}\right) p\left(\theta\right) q\left(\theta^{*} \mid \theta\right)}\right)$$

- We will use SMC approximations to compute

$$p_{\theta}\left(\boldsymbol{y}_{1:n}\right), \, and \;\; p_{\theta}\left(\boldsymbol{x}_{1:n} \mid \boldsymbol{y}_{1:n}\right)$$

Marginal Metropolis Hastings Algorithm

- Step 1:

$$Given: \left\{ \theta(i-1), \boldsymbol{X}_{1:n}(i-1), \widehat{p}_{\theta(i-1)}\left(\boldsymbol{y}_{1:n}\right) \right\}$$

$$Sample: \theta^* \sim q\left(\theta \,|\, \theta(i-1)\right)$$

$$Run\ an\ SMC\ to\ obtain: \widehat{p}_{\theta^*}\left(\boldsymbol{y}_{1:n}\right), p_{\theta^*}\left(\boldsymbol{x}_{1:n} \,|\, \boldsymbol{y}_{1:n}\right)$$

- Step 2:

$$Sample: \boldsymbol{X}_{1:n}^* \sim \widehat{p}_{\theta^*}\left(\boldsymbol{x}_{1:n} \,|\, \boldsymbol{y}_{1:n}\right)$$

- Step 3: With probability $\min\left(1, \dfrac{\widehat{p}_{\theta^*}\left(\boldsymbol{y}_{1:n}\right) p\left(\theta^*\right) q\left(\theta(i-1) \,|\, \theta^*\right)}{\widehat{p}_{\theta(i-1)}\left(\boldsymbol{y}_{1:n}\right) p\left(\theta(i-1)\right) q\left(\theta^* \,|\, \theta(i-1)\right)}\right)$

$$Set: \left\{ \theta(i), X_{1:n}(i), \widehat{p}_{\theta(i)}\left(\boldsymbol{y}_{1:n}\right) \right\} = \left\{ \theta^*, X_{1:n}^*, \widehat{p}_{\theta^*}\left(\boldsymbol{y}_{1:n}\right) \right\}$$

$$Otherwise: \left\{ \theta(i), X_{1:n}(i), \widehat{p}_{\theta(i)}\left(\boldsymbol{y}_{1:n}\right) \right\} = \left\{ \theta(i-1), X_{1:n}(i-1), \widehat{p}_{\theta(i-1)}\left(\boldsymbol{y}_{1:n}\right) \right\}$$

Marginal Metropolis Hastings Algorithm

- This algorithm (without sampling $\mathbf{X}_{1:n}$) was proposed as an approximate MCMC algorithm to sample from p $(\theta \,|\, \mathbf{y}_{1:n})$ in (Fernandez-Villaverde & Rubio-Ramirez, 2007).

- When N≥1, the algorithm admits exactly p $(\theta, \mathbf{x}_{1:n} \,|\, \mathbf{y}_{1:n})$ as invariant distribution (Andrieu, D. & Holenstein, 2010). A particle version of the Gibbs sampler also exists.

- The higher N, the better the performance of the algorithm: N scales roughly linearly with n.

- Useful when X_n is moderate dimensional & θ high dimensional. Admits the plug and play property (Ionides et al., 2006).

- Jesus Fernandez-Villaverde & Juan F. Rubio-Ramirez, 2007. "**On the solution of the growth model with investment-specific technological change**," Applied Economics Letters, 14(8), pages 549-553.
- C Andrieu, A Doucet, R Holenstein, Particle Markov chain Monte Carlo methods, Journal of the Royal Statistical Society: Series B (Statistical Methodology) Volume 72, Issue 3, pages 269–342, June 2010
- IONIDES, E. L., BRETO, C. AND KING, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the Nat Acad of Sciences* **103** 18438-18443. doi. Supporting online material. Pdf and supporting text.

OPEN PROBLEMS

Open Problems

- Thus SMC can be successfully used within MCMC

- The major advantage of SMC is that it builds automatically efficient very high dimensional proposal distributions based only on low-dimensional proposals.

- This is computationally expensive but it is very helpful in challenging problems.

- Several open problems remain:

  - Developing efficient SMC methods when you have E = $\mathbf{R}^{10000}$?
  - Developing a proper SMC method for recursive Bayesian parameter estimation.
  - Developing methods to avoid $O(N^2)$ for smoothing and sensitivity.
  - Developing methods for solving optimal control problems.
  - Establishing weaker assumptions ensuring uniform convergence results.

Other Topics

- In standard SMC, we only sample Xn at time n and we don't allow ourselves to modify the past.

- It is possible to modify the algorithm to take this into account.

- Arnaud DOUCET, Mark BRIERS, and Stéphane SÉNÉCAL, Efficient Block Sampling Strategies for Sequential Monte Carlo Methods, Journal of Computational and Graphical Statistics, Volume 15, Number 3, Pages 1–19