Sequential Bayesian Inference for Hidden Markov Models

Pierre Jacob
Department of Statistics, University of Oxford

Journées MAS 2014

Outline

1. Hidden Markov Models
2. SMC^2 for sequential inference
3. Illustration on a stochastic volatility model
4. Towards online inference

Outline

1. Hidden Markov Models

2. SMC^2 for sequential inference

3. Illustration on a stochastic volatility model

4. Towards online inference

Hidden Markov Models

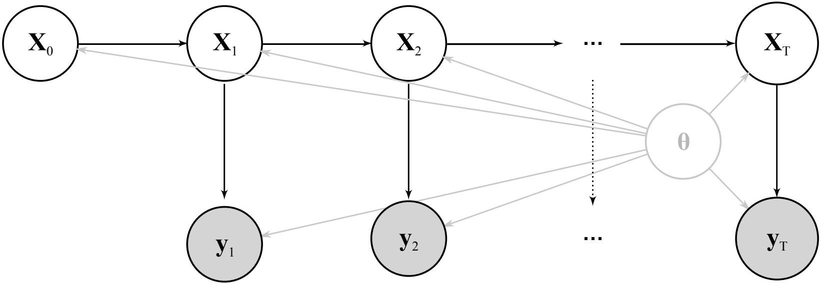


Figure : Graph representation of a general HMM.

(X_t): initial distribution μ_θ, transition f_θ.
(Y_t) given (X_t): measurement g_θ.
Prior on the parameter $\theta \in \Theta$.

Inference in HMMs, Cappé, Moulines, Ryden, 2005.

Example: battery voltage

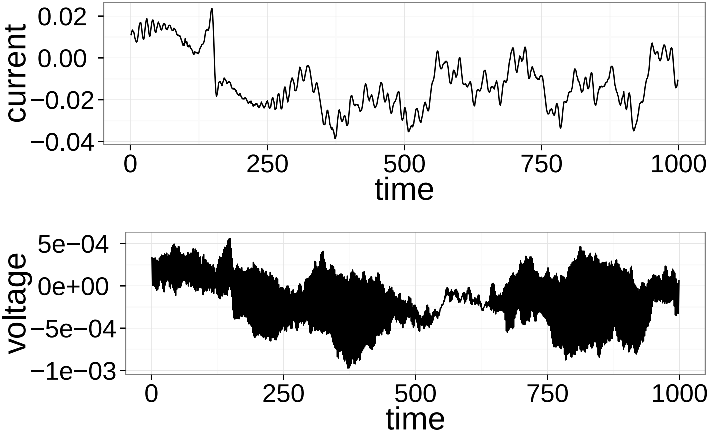
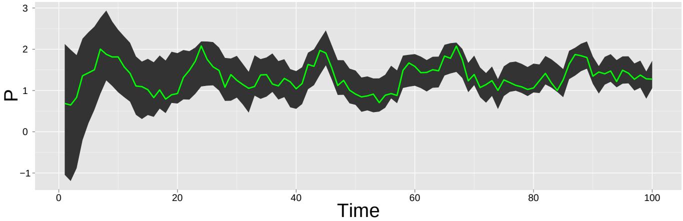
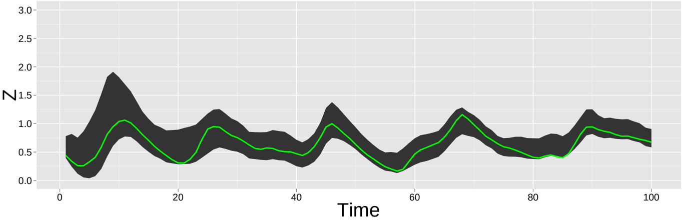


Figure : Current (input) and measured voltage (output) of a battery.

Example: phytoplankton – zooplankton



(a) Phytoplankton $+90\%$ credible interval of filtering distributions.



(b) Zooplankton $+90\%$ credible interval of filtering distributions.

Example: athletic records

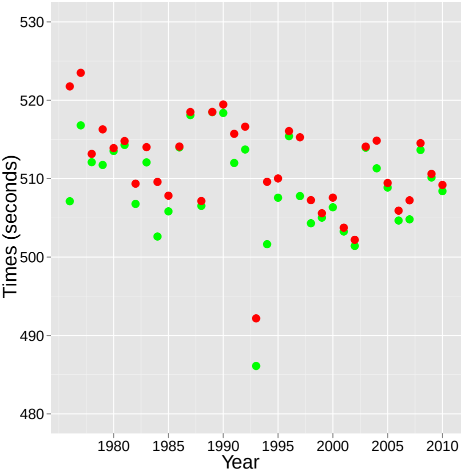


Figure : Best two times of each year in women's 3000m events.

Example: stochastic volatility

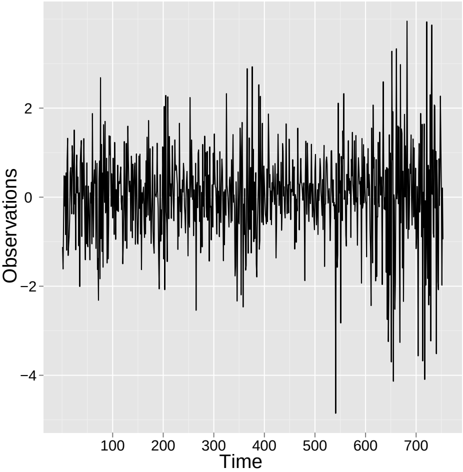


Figure : Daily log returns of S&P 500 between 2005 and 2007.

Sequential Monte Carlo for filtering

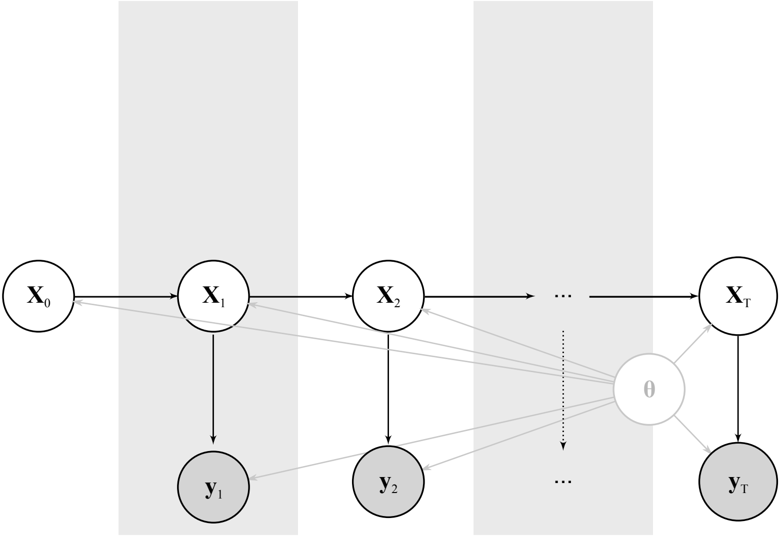
Objects of interest:

- filtering distributions: $p(x_t|y_{1:t},\theta)$, for all t, for a given θ,
- likelihood: $p(y_{1:t} \mid \theta) = \int p(y_{1:t} \mid x_{0:t},\theta)p(x_{0:t} \mid \theta)dx_{0:t}$.

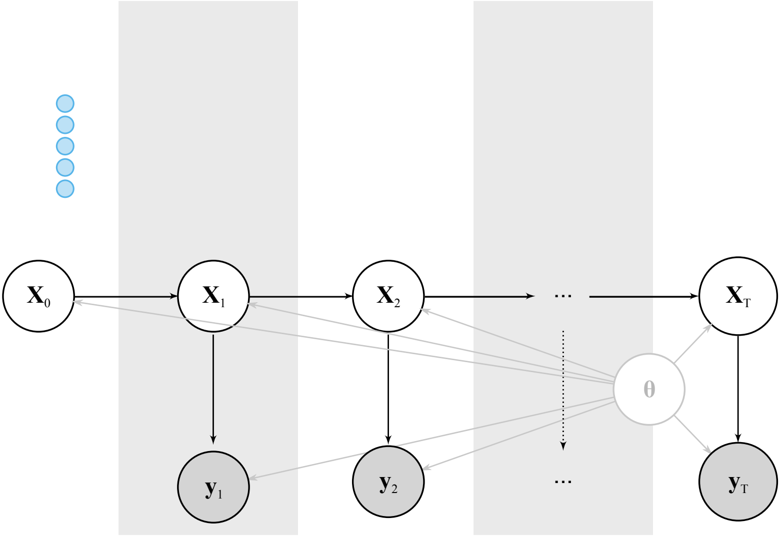
Particle filters:

- propagate recursively N_x particles approximating $p(x_t \mid y_{1:t},\theta)$ for all t,
- give likelihood estimates $\hat{p}^{N_x}(y_{1:t} \mid \theta)$ of $p(y_{1:t} \mid \theta)$ for all t.

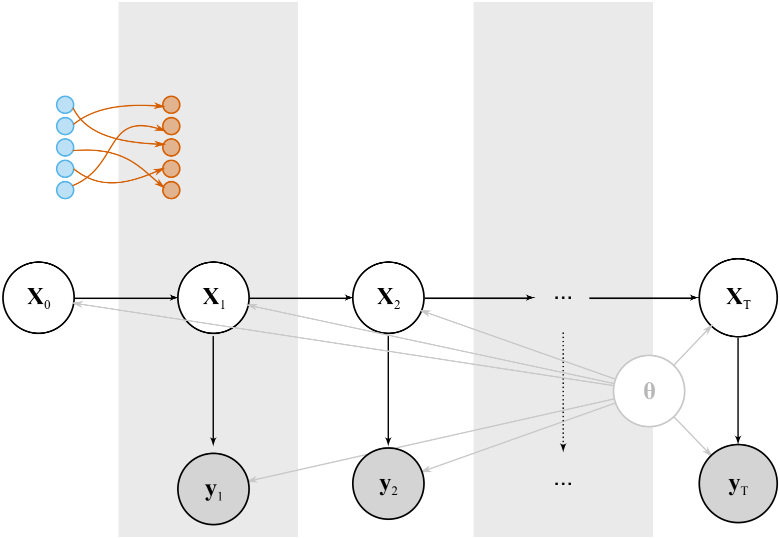
Sequential Monte Carlo for filtering



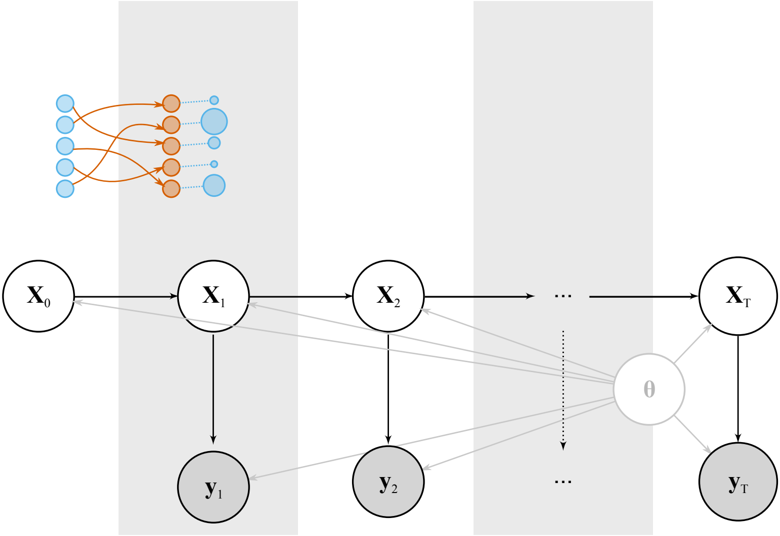
Sequential Monte Carlo for filtering



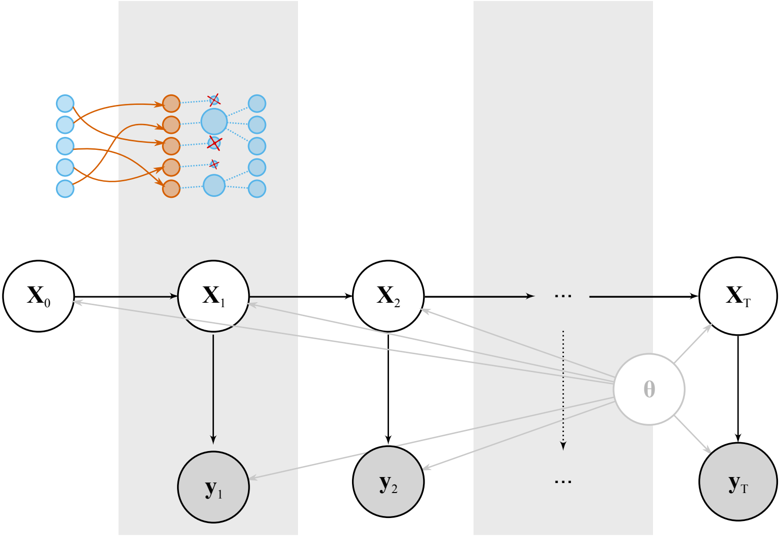
Sequential Monte Carlo for filtering



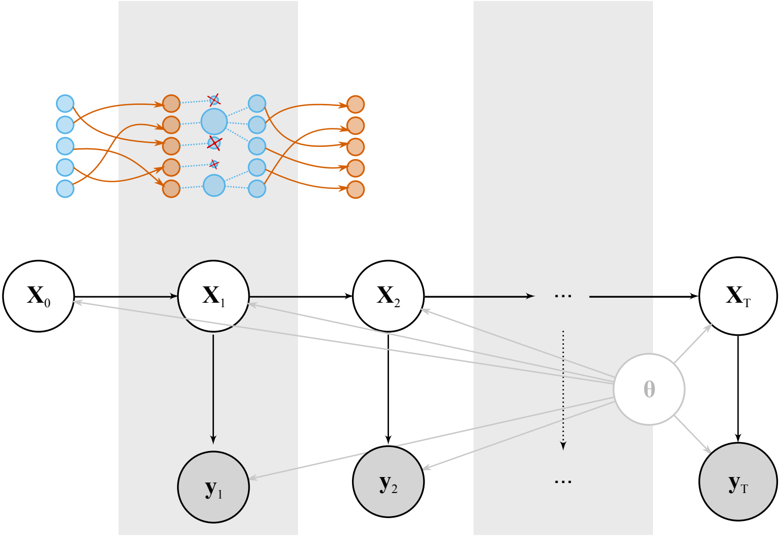
Sequential Monte Carlo for filtering



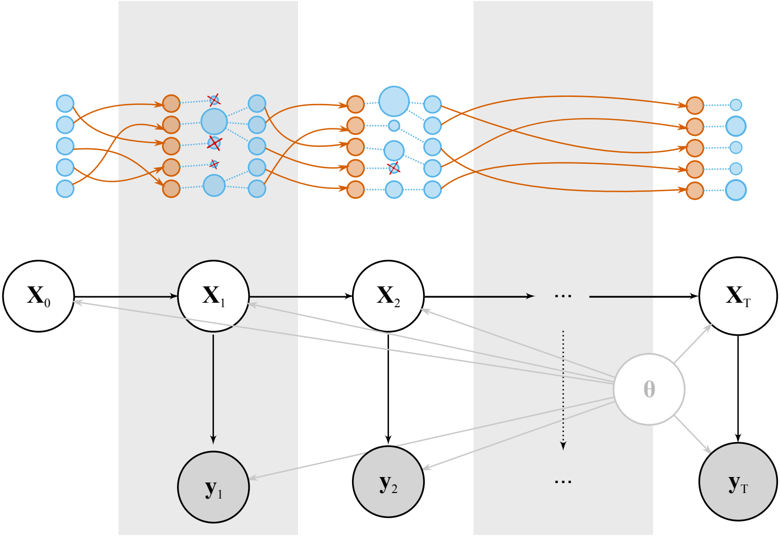
Sequential Monte Carlo for filtering



Sequential Monte Carlo for filtering



Sequential Monte Carlo for filtering



Sequential Monte Carlo for filtering

Properties of the likelihood estimator

The likelihood estimator is unbiased,

$$\mathbb{E}\left[\hat{p}^{N_x}(y_{1:T} \mid \theta)\right] = \mathbb{E}\left[\prod_{t=1}^{T} \frac{1}{N_x} \sum_{k=1}^{N_x} w_t^k\right] = p(y_{1:T} \mid \theta)$$

and the relative variance is bounded linearly in time,

$$\mathbb{V}\left[\frac{\hat{p}^{N_x}(y_{1:T} \mid \theta)}{p(y_{1:T} \mid \theta)}\right] \leq C \frac{T}{N_x}$$

for some constant C (under some conditions!).

Del Moral 2004, 2013 for books on the topic.

Outline

1 Hidden Markov Models

2 SMC^2 for sequential inference

3 Illustration on a stochastic volatility model

4 Towards online inference

Ideal approach

- The goal is now to approximate sequentially

$$p(\theta), p(\theta|y_1), \ldots, p(\theta|y_{1:T}).$$

- Sequential Monte Carlo samplers.

  Jarzynski 1997, Neal 2001, Chopin 2004, Del Moral, Doucet & Jasra 2006...

- Propagates a number N_θ of θ-particles approximating $p(\theta \mid y_{1:t})$ for all t.

Targets

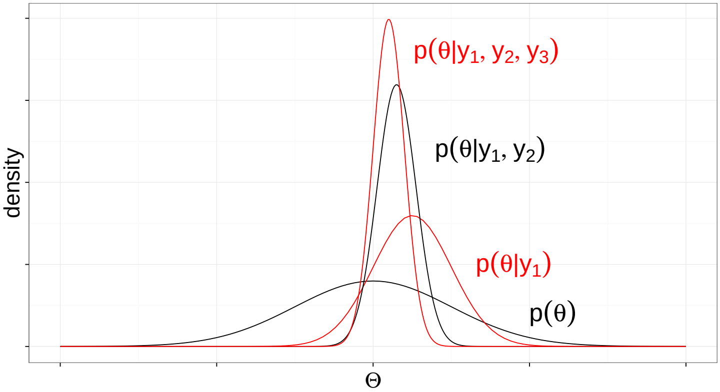


Figure : Sequence of target distributions.

First step

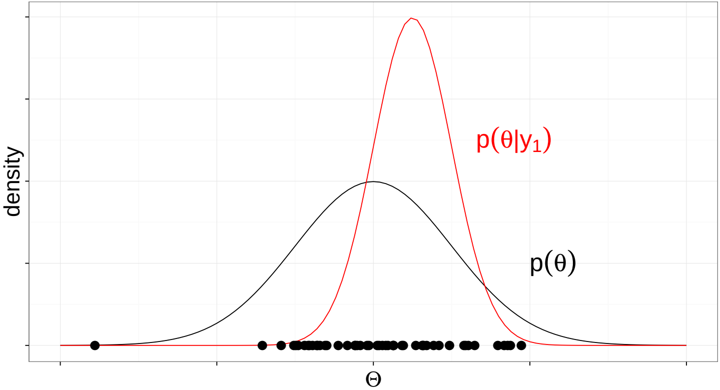


Figure : First distribution in black, next distribution in red.

Importance Sampling

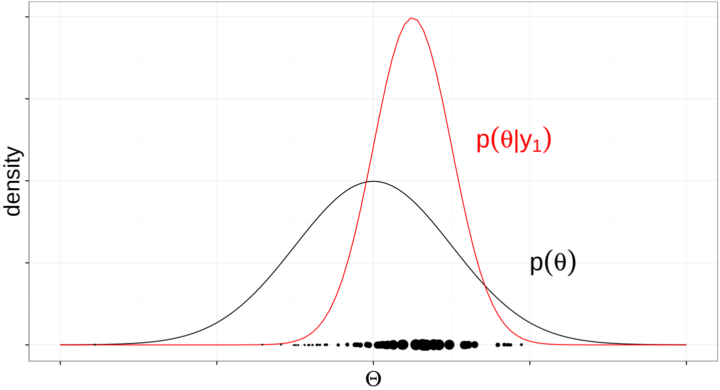


Figure : Samples θ weighted by $p(\theta \mid y_1)/p(\theta) \propto p(y_1 \mid \theta)$.

Resampling and move

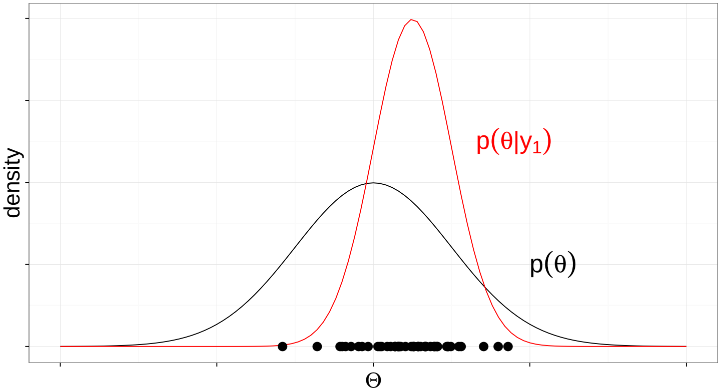


Figure : Samples θ after resampling and MCMC move.

Proposed method

SMC samplers require

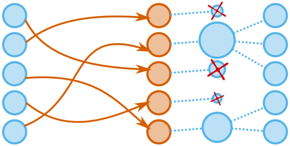
- pointwise evaluations of $p(y_t \mid y_{1:t-1}, \theta)$,
- MCMC moves leaving each intermediate distribution invariant.

For Hidden Markov models, the likelihood is intractable.

- Particle filters provide likelihood approximations for a given θ.
- Hence we equip each θ-particle with its own particle filter.

One step of SMC^2

For each θ-particle $\theta_t^{(m)}$, perform one step of its particle filter:



to obtain $\widehat{p}^{N_x}(y_{t+1} \mid y_{1:t}, \theta_t^{(m)})$ and reweight:

$$\omega_{t+1}^{(m)} = \omega_t^{(m)} \times \widehat{p}^{N_x}(y_{t+1}|y_{1:t}, \theta_t^{(m)}).$$

One step of SMC^2

Whenever

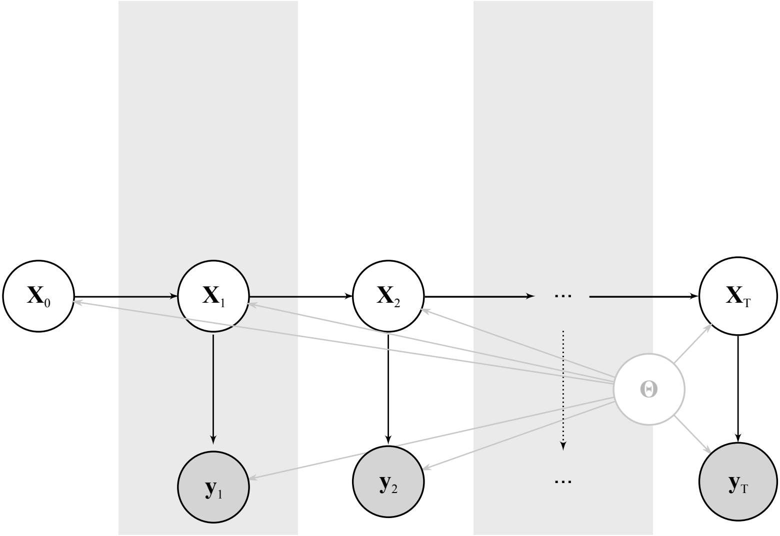
$$\text{Effective sample size} = \frac{\left(\sum_{m=1}^{N_\theta} \omega_{t+1}^{(m)}\right)^2}{\sum_{m=1}^{N_\theta} \left(\omega_{t+1}^{(m)}\right)^2} < \text{threshold} \times N_\theta$$

(Kong, Liu & Wong, 1994)

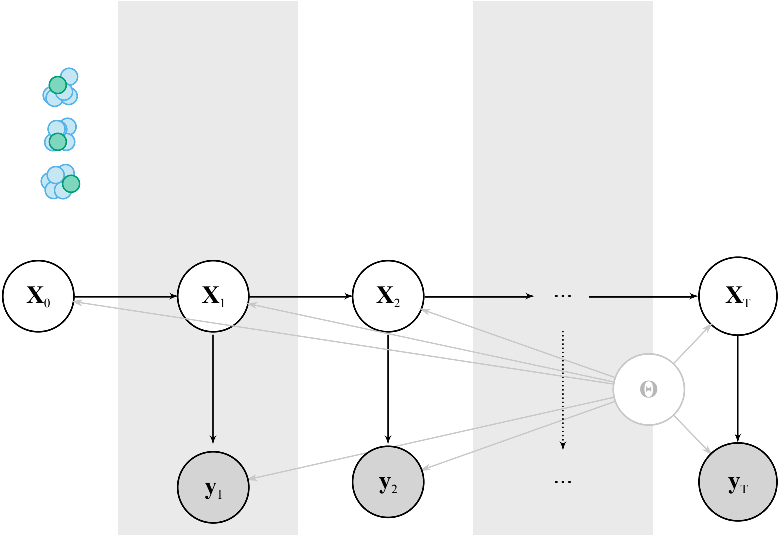
resample the θ-particles and move them by PMCMC, i.e.

- Propose $\theta^\star \sim q(\cdot|\theta_t^{(m)})$ and run PF$(N_x, \theta^\star)$ for $t+1$ steps.
- Accept or not based using $\hat{p}^{N_x}(y_{1:t+1} \mid \theta^\star)$.

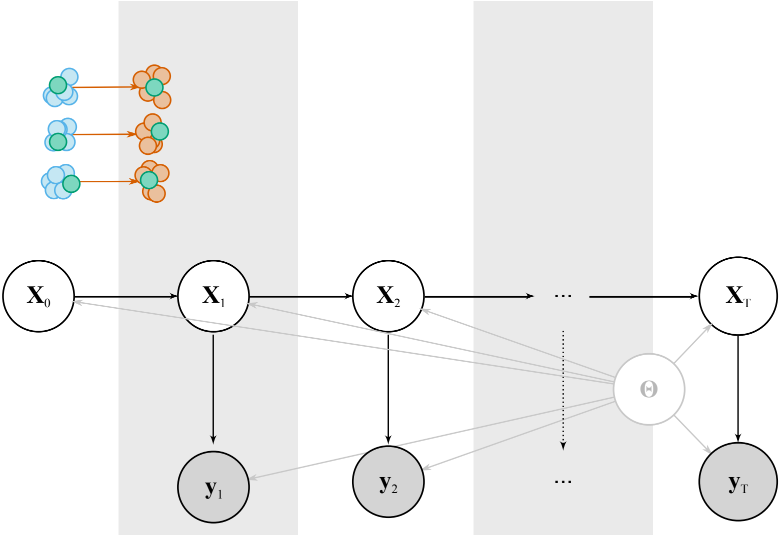
SMC^2



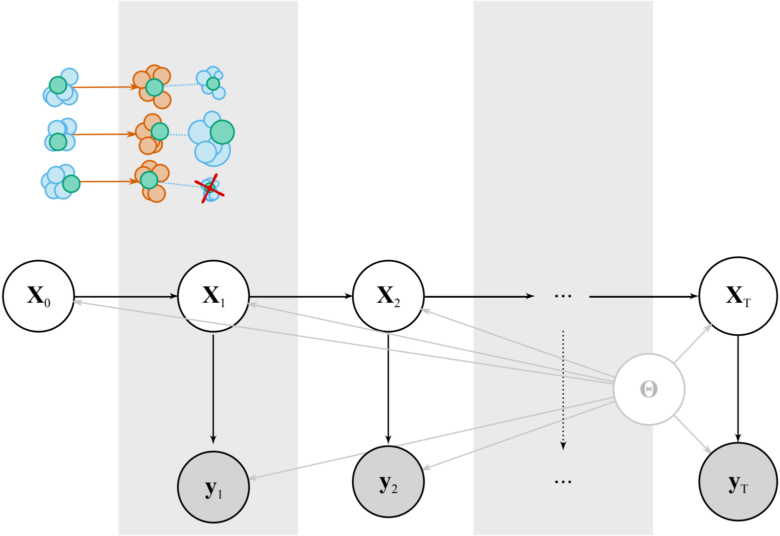
SMC^2



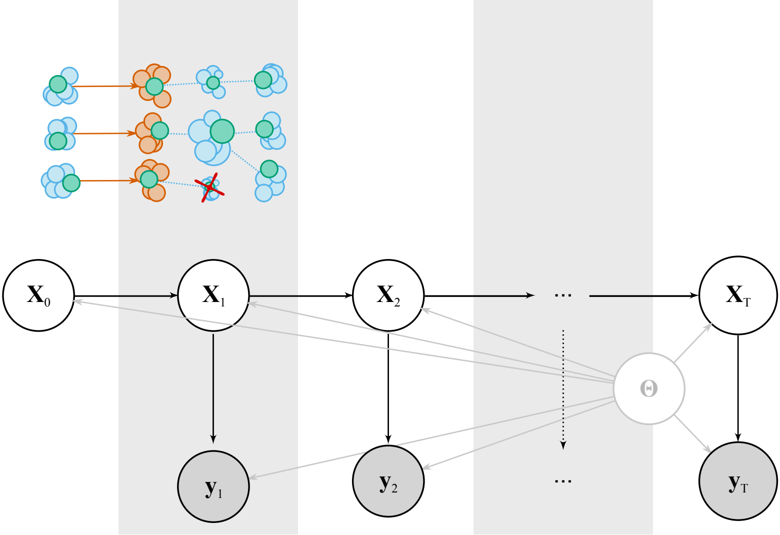
SMC^2



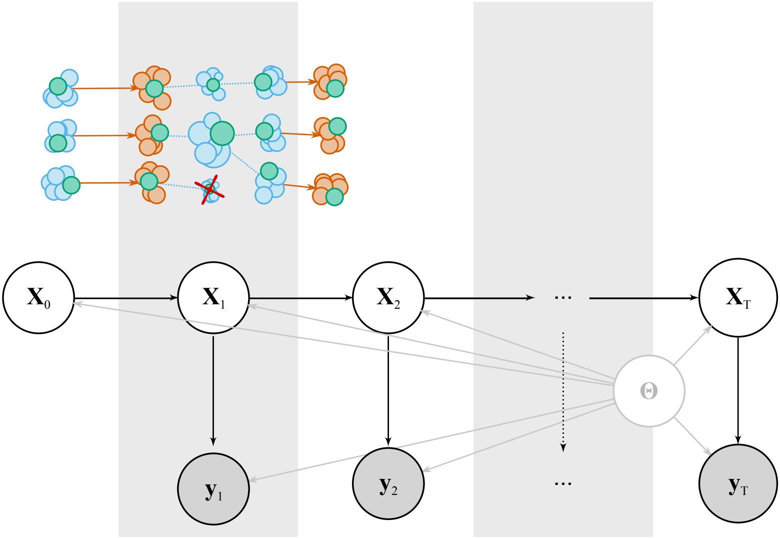
SMC^2



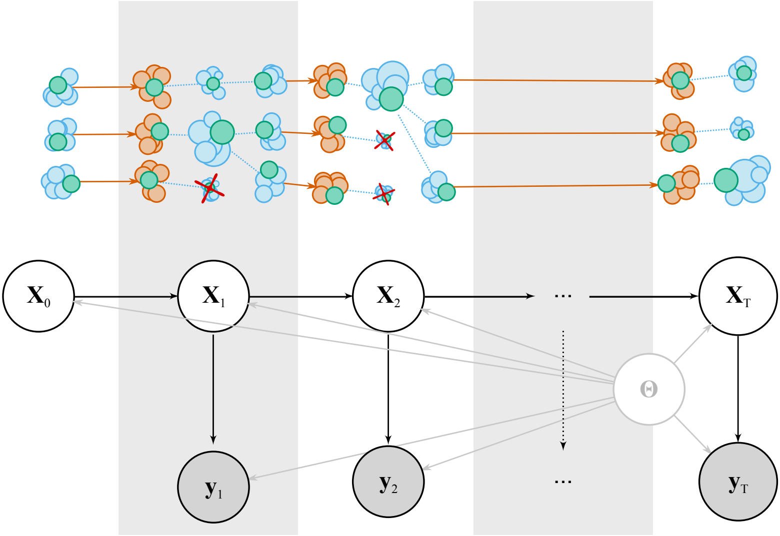
SMC^2



SMC^2



SMC^2



Exact approximation

SMC^2 is a standard SMC sampler on an extended space, with target distribution:

$$\pi_t(\theta, x_{0:t}^{1:N_x}, a_{0:t-1}^{1:N_x}) = p(\theta|y_{1:t})$$
$$\times \quad \frac{1}{N_x}\sum_{n=1}^{N_x}\frac{p(\mathbf{x}_{0:t}^{n}|\theta, y_{1:t})}{N_x^{t-1}}\left\{\prod_{\substack{i=1\\ i\neq \mathbf{h}_t^n(1)}}^{N_x} q_{1,\theta}(x_1^i)\right\}$$
$$\times \quad \left\{\prod_{s=1}^{t}\prod_{\substack{i=1\\ i\neq \mathbf{h}_t^n(s)}}^{N_x} W_{s-1,\theta}^{a_{s-1}^i} q_{s,\theta}(x_s^i|x_{s-1}^{a_{s-1}^i})\right\}.$$

For any N_x, the target admits the correct marginal on θ
$\Rightarrow$ consistency when $N_\theta \to \infty$.

Outline

Numerical illustrations: Stochastic Volatility

- Goal: model log returns $\log(p_{t+1}/p_t)$ of a series of prices (p_t).
- Daily log returns assumed to follow:

$$y_t = \mu + \beta v_t + v_t^{1/2}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1).$$

- Hidden states: actual volatility (v_t).
- Actual volatility (v_t) is the integral of the spot volatility over daily intervals.
- Spot volatility (z_t) is modeled as a Lévy process.

Barndorff-Nielsen and Shephard (2001, 2002)

Numerical illustrations: Stochastic Volatility

Transition kernel of the Markov chain

spot volatility $$z_{t+1} = e^{-\lambda} z_t + \sum_{j=1}^{k} e^{-\lambda(t+1-c_j)} e_j$$

actual volatility $$v_{t+1} = \frac{1}{\lambda}\left(z_t - z_{t+1} + \sum_{j=1}^{k} e_j\right)$$

where, at each time t:

$$k \sim \mathrm{Poi}\left(\lambda\xi^2/\omega^2\right) \quad c_{1:k} \overset{iid}{\sim} \mathrm{U}(t, t+1) \quad e_{1:k} \overset{iid}{\sim} \mathrm{Exp}\left(\xi/\omega^2\right)$$

Numerical illustrations: Stochastic Volatility

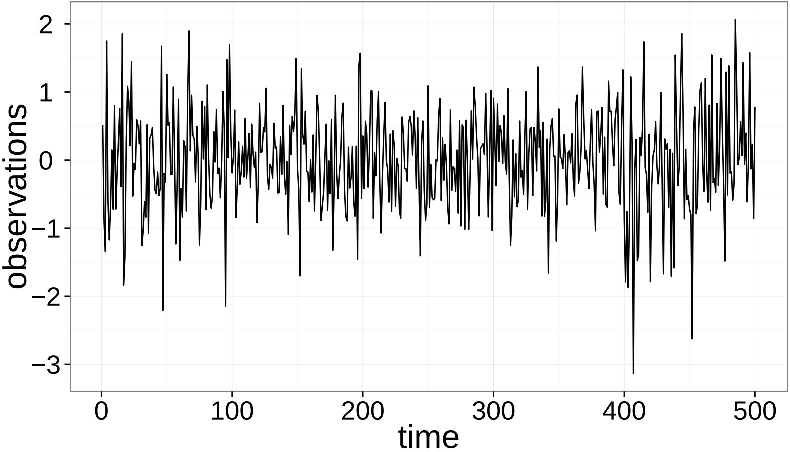


Figure : Synthetic data with $T = 500$.

Numerical illustrations: Stochastic Volatility

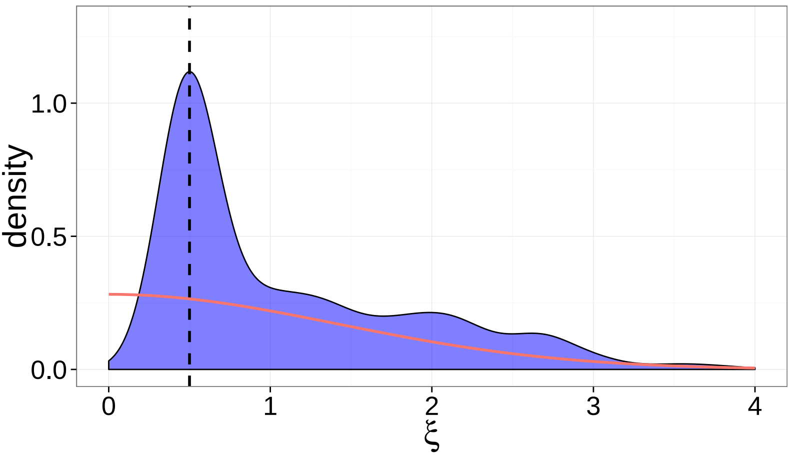


Figure : Posterior of parameter ξ at time 400.

Numerical illustrations: Stochastic Volatility

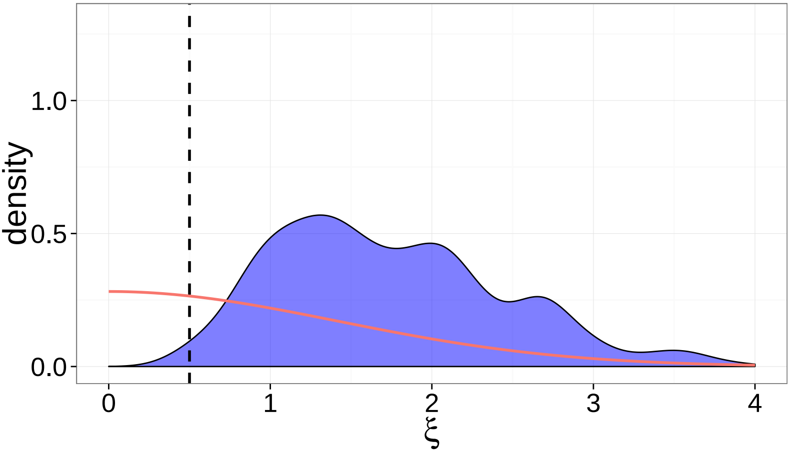


Figure : Posterior of parameter ξ at time 410.

Numerical illustrations: Stochastic Volatility

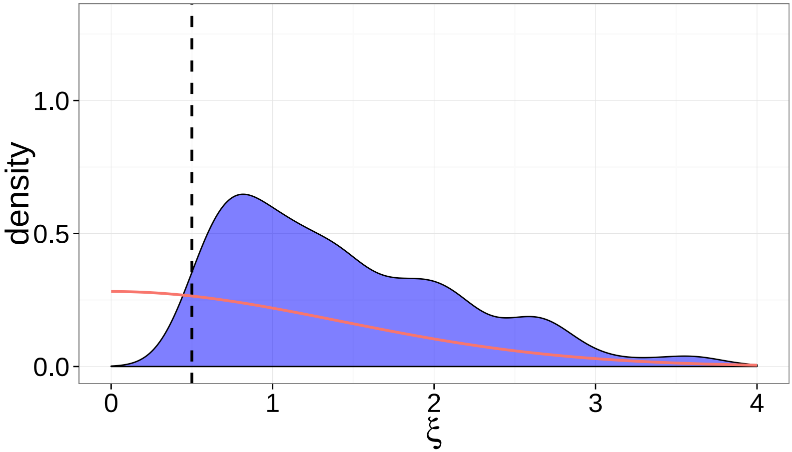


Figure : Posterior of parameter ξ at time 500.

Numerical illustrations: Stochastic Volatility

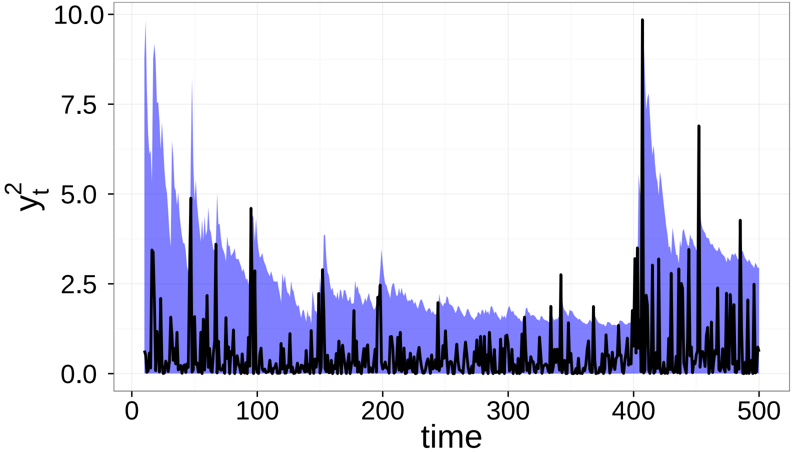


Figure : Predicted y_{t+1}^2 given $y_{1:t}$ (90% credible region), and squared observations (line).

Outline

Scalability in T

Cost if move at each time step

- A single move step at time t costs $\mathcal{O}(tN_xN_\theta)$.
- If move at every time, the total cost becomes $\mathcal{O}\left(t^2N_xN_\theta\right)$.
- If $N_x = Ct$, the total cost becomes $\mathcal{O}\left(t^3N_\theta\right)$.

With adaptive resampling, the cost is only $\mathcal{O}\left(t^2N_\theta\right)$. Why?

Scalability in T

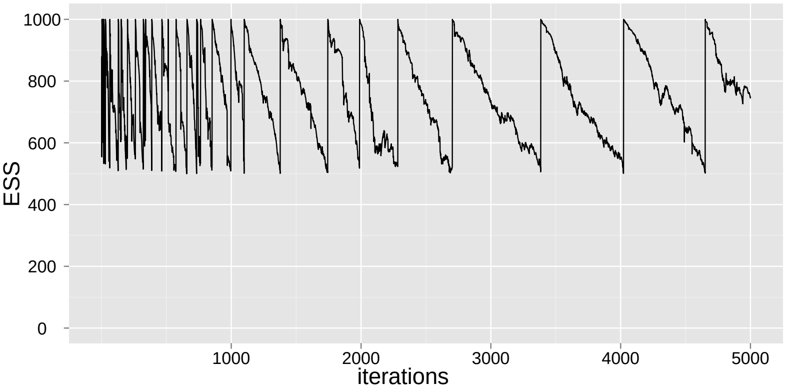


Figure : Typical ESS of the θ-particles on a long run.

Scalability in T

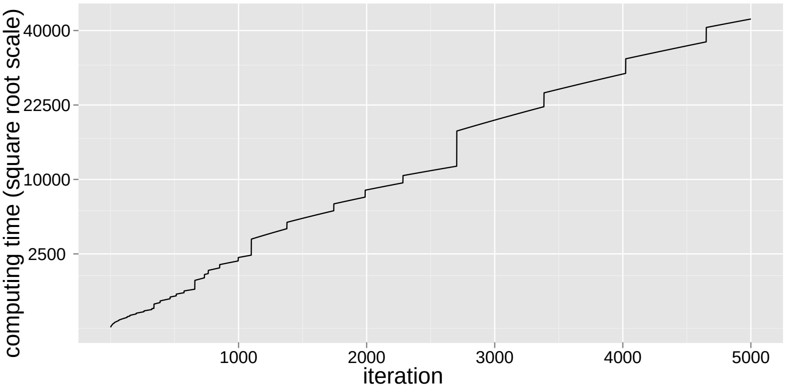
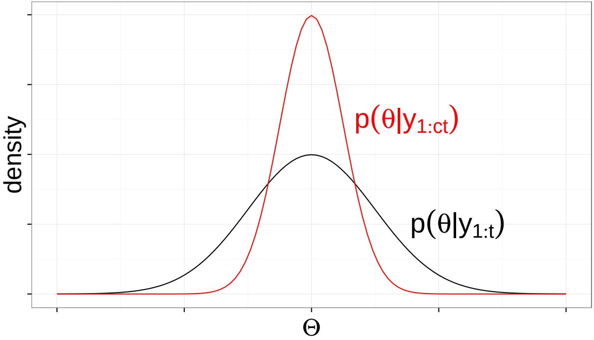


Figure : $\sqrt{\text{computing time}}$ vs iteration

Scalability in T

Under Bernstein-Von Mises, the posterior becomes Gaussian.



$\mathbb{E}[ESS]$ from $p(\theta \mid y_{1:t})$ to $p(\theta \mid y_{1:ct})$ becomes independent of t.
Hence resampling times occur geometrically: $\tau_k \approx c^k$ with $c > 1$.

Towards online inference

Open problem

Sequential Bayesian inference in linear time?

On one hand $\dim(X_{0:t}) = \dim(\mathcal{X}) \times (t+1)$ which grows ...

... but θ itself is of fixed dimension and $p(\theta \mid y_{1:t}) \approx \mathcal{N}(\theta^\star, v^\star/t)$!

Our specific problem

Move steps at time t imply running a particle filter from time zero.

Discussion

- SMC^2 allows sequential exact approximation in HMMs.
- Properties of posterior distributions could help achieving online inference, or prove that it is impossible?
- One step towards plug and play inference for time series.
- Implementation in `LibBi`, with GPU support.

Links

- *Particle Markov chain Monte Carlo*,
  Andrieu, Doucet, Holenstein, 2010 (JRSS B)

- *SMC^2: an algorithm for sequential analysis of HMM*,
  Chopin, Jacob, O. Papaspiliopoulos, 2013 (JRSS B)

- *Rethinking resampling in the particle filter on GPUs*,
  Murray, Lee, Jacob, 2013 (arXiv)

- `www.libbi.org`