Sequential Bayesian Inference for Hidden Markov Models

Pierre Jacob
Department of Statistics, University of Oxford

Journées MAS 2014
Outline

1. Hidden Markov Models
2. $\text{SMC}^2$ for sequential inference
3. Illustration on a stochastic volatility model
4. Towards online inference
Outline

1. Hidden Markov Models

2. SMC² for sequential inference

3. Illustration on a stochastic volatility model

4. Towards online inference
Figure: Graph representation of a general HMM.

\( (X_t) \): initial distribution \( \mu_\theta \), transition \( f_\theta \).
\( (Y_t) \) given \( (X_t) \): measurement \( g_\theta \).

Prior on the parameter \( \theta \in \Theta \).

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 \( \theta \),

- likelihood: \( p(y_{1:t} | \theta) = \int p(y_{1:t} | x_{0:t}, \theta) p(x_{0:t} | \theta) dx_{0:t} \).

Particle filters:

- propagate recursively \( N_x \) particles approximating \( p(x_t | y_{1:t}, \theta) \) for all \( t \),

- give likelihood estimates \( \hat{p}^{N_x}(y_{1:t} | \theta) \) of \( p(y_{1:t} | \theta) \) for all \( t \).
Sequential Monte Carlo for filtering

\[ X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_T \]

\[ y_1 \rightarrow y_2 \rightarrow \ldots \rightarrow y_T \]

\[ \theta \]

Sequential Bayesian Inference
Sequential Monte Carlo for filtering

\[ X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_T \]

\[ y_1 \rightarrow y_2 \rightarrow \cdots \rightarrow y_T \]

\[ \theta \]
Sequential Monte Carlo for filtering

\[ \theta \]

\[ X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_T \]

\[ y_1 \rightarrow y_2 \rightarrow \ldots \rightarrow y_T \]
Sequential Monte Carlo for filtering
Sequential Monte Carlo for filtering

\[ \begin{align*}
X_0 & \rightarrow X_1 \\
& \rightarrow y_1 \\
X_1 & \rightarrow X_2 \\
& \rightarrow y_2 \\
& \rightarrow \ldots \\
X_T & \rightarrow y_T
\end{align*} \]
Sequential Monte Carlo for filtering

\[ \begin{align*}
X_0 & \rightarrow X_1 & \rightarrow X_2 & \rightarrow \cdots & \rightarrow X_T \\
y_1 & \rightarrow y_2 & \rightarrow \cdots & \rightarrow y_T
\end{align*} \]

\[ \theta \]

Pierre Jacob
Sequential Monte Carlo for filtering

\[ x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow \ldots \rightarrow x_T \]

\[ y_1 \rightarrow y_2 \rightarrow \ldots \rightarrow y_T \]

\[ \theta \]
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,

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

for some constant $C$ (under some conditions!).

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 \( N_\theta \) of \( \theta \)-particles approximating
  \[ p(\theta \mid y_{1:t}) \text{ for all } t. \]
Figure: Sequence of target distributions.
First step

$\pi(\theta | y_1) \quad \pi(\theta)$

**Figure**: First distribution in black, next distribution in red.
Importance Sampling

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

Figure: Samples $\theta$ 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 $\theta$.
- Hence we equip each $\theta$-particle with its own particle filter.
One step of SMC$^2$

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

\[ \omega_{t+1}^{(m)} = \omega_t^{(m)} \times \hat{p}^N(y_{t+1} \mid y_{1:t}, \theta_t^{(m)}) \]

\[ \hat{p}^N_x(y_{t+1} \mid y_{1:t}, \theta_t^{(m)}) \] and reweight:
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 $\theta$-particles and move them by PMCMC, i.e.

- Propose $\theta^* \sim q(\cdot | \theta_t^{(m)})$ and run PF($N_x, \theta^*$) for $t + 1$ steps.
- Accept or not based using $\hat{p}^{N_x}(y_{1:t+1} \mid \theta^*)$. 
Sequential Bayesian Inference
SMC$^2$ is a standard SMC sampler on an extended space, with target distribution:

$$
\pi_t(\theta, x_{0:t}^{1:N_x}, a_{0:t-1}^{1:N_x}) = p(\theta|y_{1:t})
$$

$$
\times \frac{1}{N_x} \sum_{n=1}^{N_x} \frac{p(x_{0:t}^n|\theta, y_{1:t})}{N_x^{t-1}} \left\{ \prod_{i=1}^{N_x} q_{1,\theta}(x_1^i) \right\}
$$

$$
\times \left\{ \prod_{s=1}^{t} \prod_{i=1}^{N_x} W_{s-1, \theta}^{a_{s-1}^i} q_{s,\theta}(x_{s}^i|x_{s-1}^{a_{s-1}^i}) \right\}.
$$

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

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

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 \( t \):
\[ k \sim \text{Poi} \left( \lambda \xi^2 / \omega^2 \right) \quad c_{1:k} \sim \text{iid} \, \text{U} \left( t, t + 1 \right) \quad e_{1:k} \sim \text{iid} \, \text{Exp} \left( \xi / \omega^2 \right) \]
Figure: Synthetic data with $T = 500$. 

Numerical illustrations: Stochastic Volatility
Figure: Posterior of parameter $\xi$ at time 400.
Numerical illustrations: Stochastic Volatility

Figure: Posterior of parameter $\xi$ at time 410.
Figure: Posterior of parameter $\xi$ at time 500.
Numerical illustrations: Stochastic Volatility

Figure: Predicted $y_{t+1}^2$ given $y_{1:t}$ (90% credible region), and squared observations (line).
1. Hidden Markov Models
2. SMC\(^2\) for sequential inference
3. Illustration on a stochastic volatility model
4. Towards online inference
Scalability in $T$

### Cost if move at each time step

- A single move step at time $t$ costs $O(tN_xN_\theta)$.

- If move at every time, the total cost becomes $O(t^2N_xN_\theta)$.

- If $N_x = Ct$, the total cost becomes $O(t^3N_\theta)$.

With adaptive resampling, the cost is only $O(t^2N_\theta)$. Why?
Scalability in $T$

Figure: Typical ESS of the $\theta$-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.

\[ p(\theta | y_{1:ct}) \]

\[ p(\theta | y_{1:t}) \]

$\mathbb{E}[ESS]$ from $p(\theta | y_{1:t})$ to $p(\theta | y_{1:ct})$ 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 $\dim(X_{0:t}) = \dim(X') \times (t + 1)$ which grows . . .

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

Our specific problem
Move steps at time $t$ imply running a particle filter from time zero.
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.
• *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