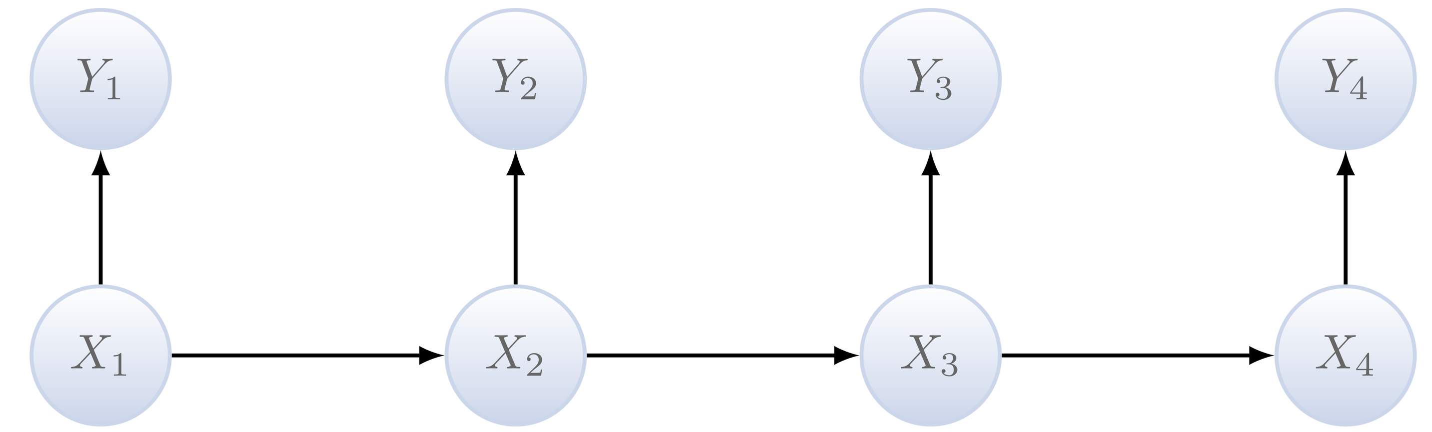
Bernoulli Race Particle Filters

Sebastian M. Schmon, George Deligiannidis and Arnaud Doucet

UNIVERSITY OF OXFORD DEPARTMENT OF STATISTICS

Introduction: Hidden Markov Models



- Latent X-valued process such that $X_1 \sim \mu(\cdot)$ and
$$X_t \mid (X_{t-1} = x) \sim f(\cdot \mid x).$$
- Y-valued observation process
$$Y_t \mid (X_t = x) \sim g(\cdot \mid x).$$
- **Aim:** Sequentially approximate the posterior $(p(x_{1:t} \mid y_{1:t}))_{t\geq 1}$ where
$$p(x_{1:t} \mid y_{1:t}) = \frac{g(y_1 \mid x_1)\mu(x_1)\prod_{k=2}^{t} g(y_k \mid x_k) f(x_k \mid x_{k-1})}{p(y_{1:t})}.$$

Particle Filter Algorithm

Given the particles $\{X^i_{1:t-1}, i = 1, \dots, N\}$ use the posterior approximation
$$\hat{p}(dx_{1:t-1}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i_{1:t-1}}(dx_{1:t-1}).$$
Propose new state $\widetilde{X}^i_t \sim q(\cdot \mid X^i_{t-1}, y_t)$ and compute weights
$$w_{t,i}(X^i_{t-1}, \widetilde{X}^i_t) = \frac{g(y_t \mid \widetilde{X}^i_t) f(\widetilde{X}^i_t \mid X^i_{t-1})}{q(\widetilde{X}^i_t \mid X^i_{t-1}, y^i_t)}. \quad (1)$$
Resample $I_{1:N} \sim \text{Multinomial}\{N; w_{t,1}, \dots, w_{t,N}\}$ and set $X^i_{1:t} = \left(X^{I_i}_{1:(t-1)}, \widetilde{X}^{I_i}_t\right)$.
What if the weight cannot be computed analytically?

Examples: Locally Optimal Proposal

The *locally optimal proposal* q^*, is
$$q^*(x_t \mid x_{t-1}, y_t) = \frac{g(y_t \mid x_t) f(x_t \mid x_{t-1})}{p(y_t \mid x_{t-1})},$$
Sampling from q^* is trivial using a rejection sampler, but the weights
$$w_t(x_{t-1}, x_t) = p(y_t \mid x_{t-1}) = \int g(y_t \mid x_t) f(x_t \mid x_{t-1}) dx_t,$$
are intractable if the integral does not have an analytic solution.

Random Weight Particle Filters

Possible solution: random weight particle filter

- Random weight particle filters replaces weight with unbiased estimator.
- This is equivalent to a particle filter on an extended space.
- Random weight particle filter results in higher asymptotic variance for particle approximations than exact weight particle filter $\Rightarrow$ **Can we perform exact resampling?**

Bernoulli Races

Problem Setting: Denote (1) as $w_i = c_i b_i$, for particles $i = 1, \dots, N$ (dropping t). Want to sample from discrete distributions
$$p(i) = \frac{w_i}{\sum_{k=1}^{N} w_k} = \frac{c_i b_i}{\sum_{k=1}^{N} c_k b_k}, \quad i = 1, \dots N \quad (2)$$
where $c_1, \dots, c_N$ denote fixed constants whereas the $b_1, \dots, b_N$ are unknown probabilities.
Assumptions:

- Assume we are able to sample coins $Z_i \sim \text{Ber}(b_i)$, $i = 1, \dots, N$.
- The $c_1, \dots, c_N$ are selected to ensure that $b_1, \dots, b_N$ take values between 0 and 1.

Bernoulli Race Algorithm: To sample from (2),

- draw
$$\mathcal{I} \sim \left(\frac{c_1}{\sum_{k=1}^{N} c_k}, \dots, \frac{c_k}{\sum_{k=1}^{N} c_k}\right), \qquad Z \sim b_{\mathcal{I}}$$
- **if** $Z = 1$ **return** $\mathcal{I}$ **else** restart.

Probability of Stopping: The probability of stopping (accepting a coin flip) is
$$\rho_N = \frac{\sum_{k=1}^{N} c_k b_k}{\sum_{k=1}^{N} c_k}.$$
For $N \in \mathbb{N}$ let $C_{1,N}, \dots, C_{N,N} \overset{iid}{\sim} \text{Geom}(\rho_N)$. An unbiased estimator for ρ_N is
$$\hat{\rho}^{\text{mvue}}_N = \frac{N-1}{\sum_{k=1}^{N} C_{k,N} - 1}. \quad (3)$$

Bernoulli Race Particle Filters

- Bernoulli race as $\mathcal{O}(N)$ resampling algorithm in particle filter.
- Choose a valid decomposition $w_i = c_i b_i$.
- Works even if only unbiased estimators $\widehat{b}^i \in [0, 1]$ of b^i are available as
$$Z^i = 1\{V \leq \widehat{b}^i\} \Rightarrow P(Z^i = 1) = b^i, \quad V \sim \text{Unif}[0, 1].$$

Likelihood Estimation: Standard estimator
$$\hat{p}(y_{1:T}) = \prod_{t=1}^{T} \hat{p}(y_t \mid y_{1:t-1}) = \prod_{t=1}^{T} \frac{1}{N}\sum_{i=1}^{N} w_{t,i}. \quad (4)$$
This suggests using (3) in (4):
$$\hat{p}_{N,T} = \prod_{t=1}^{T} \frac{1}{N}\sum_{i=1}^{N} c_{t,i} \cdot \frac{N-1}{\sum_{k=1}^{N} C_{k,N} - 1}. \quad (5)$$
Theorem The estimator (5) is unbiased for $p(y_{1:T})$, i.e. $\mathbb{E}[\hat{p}_{N,T}] = p(y_{1:T})$.

Cox Process Example

Prior: Latent Gaussian process
$$dX_t = \begin{bmatrix} 0 & 1 \\ 0 & \theta \end{bmatrix} X_t dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dB_t, \qquad t \in [0, \mathcal{T}], \quad (6)$$
where $(B_t)_{t\in[0,\mathcal{T}]}$ is a Brownian motion from 0 to $\mathcal{T}$ and θ is a negative real value.
Observation Likelihood:
$$p(s_{1:n} \mid \lambda, \mathcal{T}) = \exp\left(-\int_0^{\mathcal{T}} \lambda(s) ds\right) \prod_{i=1}^{n} \lambda(s_i).$$
with intensity
$$\lambda(t) = \frac{\lambda_{\max}}{1 + \exp(-X_{1,t})},$$
where $(X_{1,t})_{t\in[0,\mathcal{T}]}$ is the first coordinate of the SDE (6).
Inference: Inference is done in batches over intervals $[t_0, t_1]$.
Proposing from the prior (6) the weight is
$$g(s_{k\colon s_k\in[t_0,t_1]} \mid X^i) = \exp\left(-\int_{t_0}^{t_1} \lambda^i(s) ds\right) \prod_{k\colon s_k\in[t_0,t_1]} \lambda(s^i_k),$$
where
$$\lambda^i(s) = \frac{\lambda_{\max}}{1 + \exp(-X^i_{1,s})}.$$
In order to implement the Bernoulli race resampling, we can choose
$$c^i_{t_1} = \prod_{k\colon s_k\in[t_0,t_1]} \lambda^i(s_k),$$
$$b^i_{t_1} = \exp\left(-\int_{t_0}^{t_1} \lambda^i(s) ds\right).$$
Unbiased Coin Flips: For $U \sim \text{Unif}[t_0, t_1]$ we have
$$\mathbb{E}[\lambda(U) \mid X = x] = \int_{t_0}^{t_1} \frac{\lambda(s)}{t_1 - t_0} ds = \frac{1}{t_0 - t_1}\int_{t_0}^{t_1} \frac{1}{1 + \exp(-x_s)} ds.$$
Sample $\kappa \sim \text{Pois}(\lambda_{\max}(t_1 - t_0))$. Then the coin flip
$$Z = \prod_{j=1}^{\kappa} 1\left\{V_j \leq \frac{\lambda_{\max} - \lambda(U_j)}{\lambda_{\max}}\right\},$$
has success probability
$$P(Z = 1) = \exp\left(-\int_{t_0}^{t_1} \lambda(s) ds\right).$$

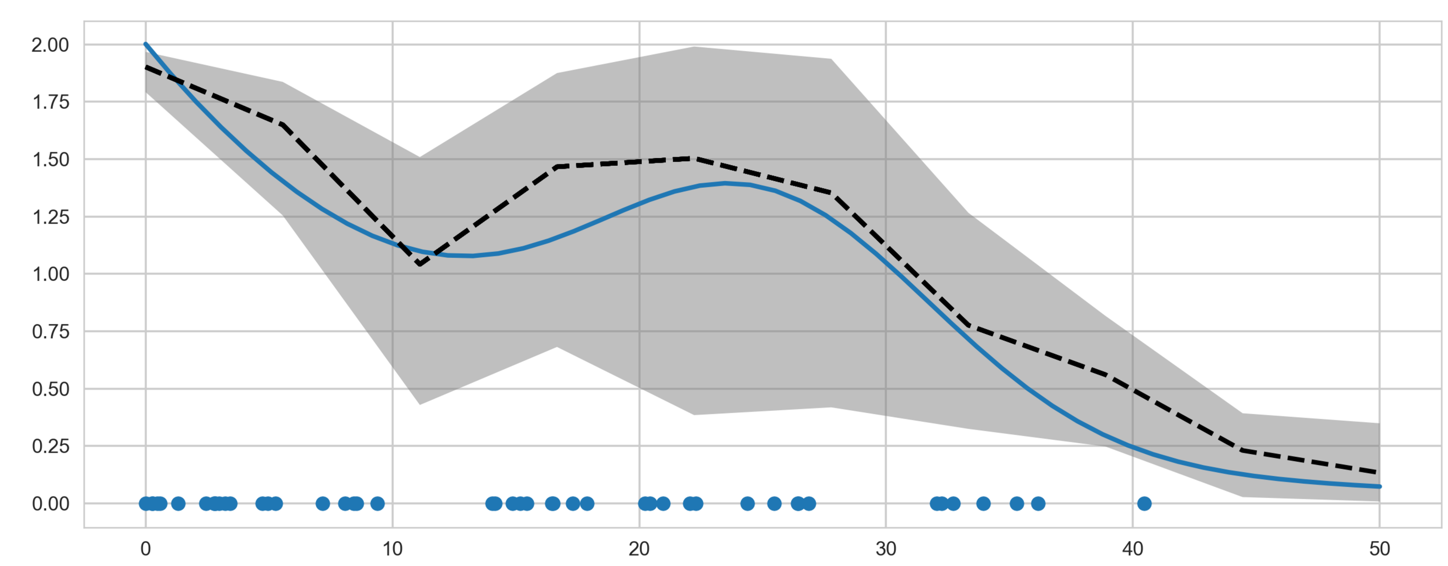


Figure: Cox process inference using Bernoulli race particle filter (30 particles). Observations (dots), true intensity (bold, blue), mean of particle approximation (dashed), 90% quantile (dashed).

Summary

- Exact resampling *even if precise weight is not known*.
- Average $\mathcal{O}(N)$ resampling complexity, where N is number of particles.
- Competitive alternative to random weight particle filter.
- Leading to guaranteed lower variance than random weight PF estimates.
- Unbiased estimation of normalizing constant possible as well.