Advanced Simulation - Lecture 13

Patrick Rebeschini

February 26th, 2018

Outline

- Sequential Importance Sampling.
- Resampling step.
- Sequential Monte Carlo / Particle Filters.

Hidden Markov Models

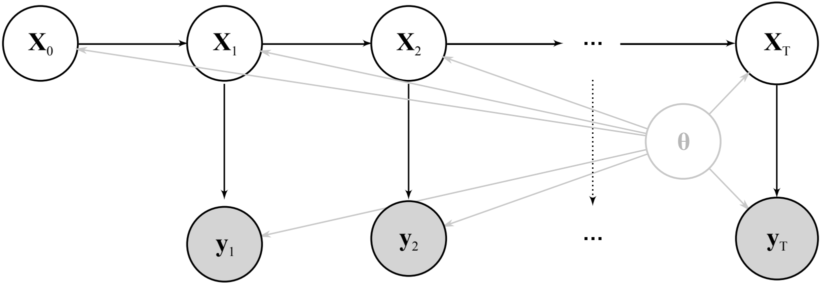


Figure: Graph representation of a general HMM.

(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.

General inference in HMM

- **Proposition**: The posterior $p\left(x_{1:t}\mid y_{1:t},\theta\right)$ satisfies

$$p\left(x_{1:t}\mid y_{1:t},\theta\right)=p\left(x_{1:t-1}\mid y_{1:t-1},\theta\right)\frac{f_{\theta}\left(x_{t}\mid x_{t-1}\right)g_{\theta}\left(y_{t}\mid x_{t}\right)}{p\left(y_{t}\mid y_{1:t-1},\theta\right)}$$

where

$$p\left(y_{t}\mid y_{1:t-1},\theta\right)=\int p\left(x_{1:t-1}\mid y_{1:t-1},\theta\right)f_{\theta}\left(x_{t}\mid x_{t-1}\right)g_{\theta}\left(y_{t}\mid x_{t}\right)dx_{1:t}.$$

- **Proposition**: The marginal posterior $p\left(x_{t}\mid y_{1:t}\right)$ satisfies the following recursion

$$p\left(x_{t}\mid y_{1:t-1}\right)=\int f\left(x_{t}\mid x_{t-1}\right)p\left(x_{t-1}\mid y_{1:t-1}\right)dx_{t-1}$$

$$p\left(x_{t}\mid y_{1:t}\right)=\frac{g\left(y_{t}\mid x_{t}\right)p\left(x_{t}\mid y_{1:t-1}\right)}{p\left(y_{t}\mid y_{1:t-1}\right)}$$

where

$$p\left(y_{t}\mid y_{1:t-1}\right)=\int g\left(y_{t}\mid x_{t}\right)p\left(x_{t}\mid y_{1:t-1}\right)dx_{t}.$$

General inference in HMM

- In general, the filtering problem is thus intractable:

$$\int \varphi(x_t)p(x_t \mid y_{1:t}, \theta)dx_t = \int \varphi(x_t)p(x_{1:t}, y_{1:t} \mid \theta)dx_{1:t}$$
$$= \int \varphi(x_t)\mu_\theta(x_1) \prod_{s=1}^{t} f_\theta(x_s \mid x_{s-1}) \prod_{s=1}^{t} g_\theta(y_s \mid x_s)dx_{1:t}.$$

- It is a $t \times \dim(\mathbb{X})$ dimensional integral.
- The likelihood is also intractable:

$$p(y_{1:t} \mid \theta) = \int p(x_{1:t}, y_{1:t} \mid \theta)dx_{1:t}$$
$$= \int \mu_\theta(x_1) \prod_{s=1}^{t} f_\theta(x_s \mid x_{s-1}) \prod_{s=1}^{t} g_\theta(y_s \mid x_s)dx_{1:t}.$$

- Thus we cannot compute it pointwise, e.g. to perform Metropolis–Hastings algorithm on the parameters.

Sequential Importance Sampling

- We now consider the parameter θ to be fixed. We want to infer $X_{1:t}$ given $y_{1:t}$.
- Two ingredients: importance sampling, and "sampling via composition", or "via condition".
- IS: if we have a weighted sample (w_1^i, X^i) approximating π_1, then (w_2^i, X^i) approximates π_2 if we define

$$w_2^i = w_1^i \times \frac{\pi_2(X^i)}{\pi_1(X^i)}.$$

In standard IS, π_1 and π_2 are defined on the same space.

Sequential Importance Sampling

- Sampling via composition: if (w^i, X^i) approximates $p_X(x)$, and if $Y^i \sim q_{Y|X}(y \mid X^i)$, then $(w^i, (X^i, Y^i))$ approximates $p_X(x)q_{Y|X}(y \mid x)$.

- The space has been extended.

- Marginally, (w^i, Y^i) approximates

$$q_Y(y) = \int p_X(x)q_{Y|X}(y \mid x)dx.$$

- Sequential Importance Sampling combines both ingredients to iteratively approximate $p(x_{1:t} \mid y_{1:t})$.

Sequential Importance Sampling: algorithm

- *At time* $t = 1$
  - Sample $X_1^i \sim q_1(\cdot)$.
  - Compute the weights

$$w_1^i = \frac{\mu(X_1^i) g(y_1 \mid X_1^i)}{q_1\left(X_1^i\right)}.$$

- *At time* $t \geq 2$
  - Sample $X_t^i \sim q_{t|t-1}(\cdot| X_{t-1}^i)$.
  - Compute the weights

$$\begin{aligned} w_t^i &= w_{t-1}^i \times \omega_t^i \\ &= w_{t-1}^i \times \frac{f\left(X_t^i \middle| X_{t-1}^i\right) g\left(y_t \middle| X_t^i\right)}{q_{t|t-1}(X_t^i \mid X_{t-1}^i)}. \end{aligned}$$

Sequential Importance Sampling: example

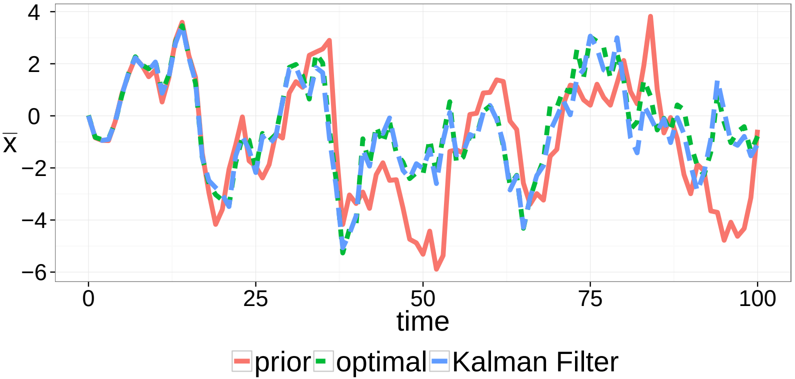


Figure: Estimation of filtering means $\mathbb{E}\left(x_t \mid y_{1:t}\right)$.

Sequential Importance Sampling: example

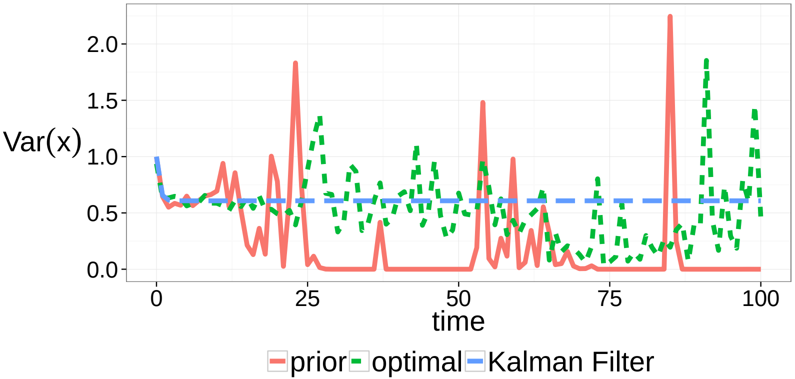


Figure: Estimation of filtering variances $\mathbb{V}(x_t \mid y_{1:t})$.

Sequential Importance Sampling: example

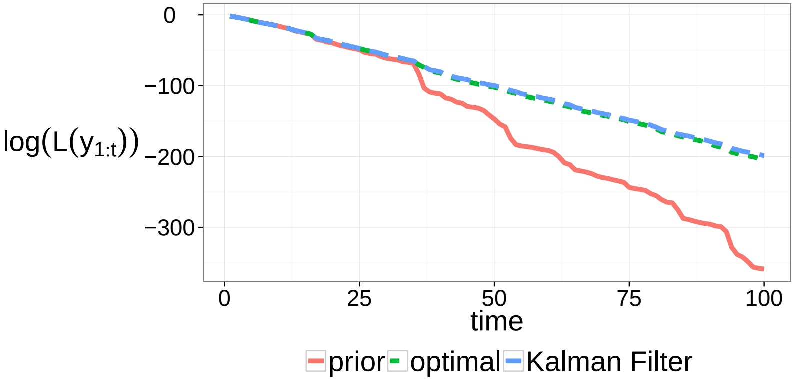


Figure: Estimation of marginal log likelihoods $\log p(y_{1:t})$.

Sequential Importance Sampling: example

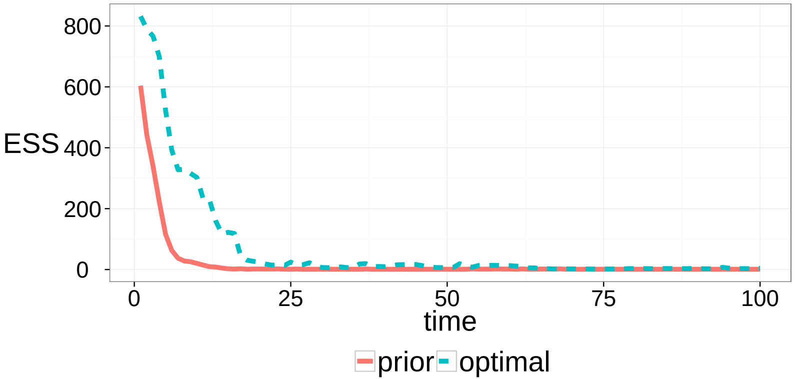


Figure: Effective sample size over time.

Sequential Importance Sampling: example

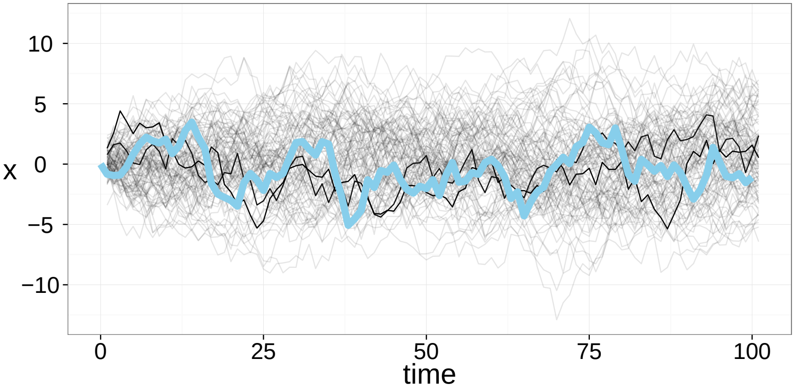


Figure: Spread of 100 paths drawn from the prior proposal, and KF means in blue. Darker lines indicate higher weights.

Resampling

- Idea: at time t, select particles with high weights, and remove particles with low weights.
- Spend the fixed computational budget "N" on the most promising paths.
- Obtain an equally weighted sample $(N^{-1}, \bar{X}^i)$ from a weighted sample (w^i, X^i).
- Resampling on empirical probability measures: input

$$\pi^N(x) = \left(\sum w^j\right)^{-1} \sum w^i \delta_{X^i}(x)$$

and output

$$\bar{\pi}^N(x) = N^{-1} \sum \delta_{\bar{X}^i}(x).$$

Multinomial resampling

- How to draw from an empirical probability distribution?

$$\pi^N(x) = \frac{1}{\sum_{j=1}^N w^j} \sum_{i=1}^N w^i \delta_{X^i}(x)$$

- Remember how to draw from a mixture model?

$$\sum_{i=1}^K \omega^i \, p^i(x)$$

- Draw k with probabilities $\omega^1, \ldots, \omega^N$, then draw from p^k.

Multinomial resampling

- Draw an "ancestry vector" $A^{1:N} = \left(A^1, \ldots, A^N\right) \in \{1, \ldots, N\}^N$ independently from a categorical distribution

$$A^{1:N} \overset{\text{i.i.d}}{\sim} \mathcal{C}at\left(w^1, \ldots, w^N\right),$$

in other words

$$\forall i \in \{1, \ldots, N\} \quad \forall k \in \{1, \ldots, N\} \quad \mathbb{P}\left[A^i = k\right] = \frac{w^k}{\sum_{j=1}^N w^j}.$$

- Define $\bar{X}^i$ to be X^{A^i} for all $i \in \{1, \ldots, N\}$. X^{A^i} is said to be the "parent" or "ancestor" of $\bar{X}^i$.
- Return $\bar{X} = \left(\bar{X}^1, \ldots, \bar{X}^N\right)$.

Multinomial resampling

- Draw an "offspring vector" $O^{1:N} = \left(O^1, \ldots, O^N\right) \in \{0, \ldots, N\}^N$ from a multinomial distribution

$$O_t^{1:N} \sim \mathcal{M}ultinomial\left(N; w^1, \ldots, w^N\right)$$

so that

$$\forall i \in \{1, \ldots, N\} \quad \mathbb{E}\left[O^i\right] = N\frac{w^i}{\sum_{j=1}^N w^j} \quad \text{and} \quad \sum_{i=1}^N O^i = N.$$

- Each particle X^i is replicated O^i times (possibly zero times) to create the sample $\bar{X}$:
  - $\bar{X} \leftarrow \{\}$
  - For $i = 1, \ldots, N$, for $k = 0, \ldots, O_t^i$, $\bar{X} \leftarrow \left\{\bar{X}, X^i\right\}$
- Return $\bar{X} = \left(\bar{X}^1, \ldots, \bar{X}^N\right)$.

Multinomial resampling

- Other strategies are possible to perform resampling.
- Some properties are desirable:

$$\mathbb{E}\left[O^i\right] = N \frac{w^i}{\sum_{j=1}^N w^j},$$

$$\text{or} \quad \mathbb{P}\left[A^i = k\right] = \frac{w^k}{\sum_{j=1}^N w^j}.$$

- This is sometimes called "unbiasedness", because then

$$\frac{1}{N}\sum_{k=1}^N \varphi\left(\bar{X}^k\right) = \frac{1}{N}\sum_{k=1}^N O^k \varphi\left(X^k\right)$$

has expectation

$$\sum_{k=1}^N \frac{w^k}{\sum_{j=1}^N w^j} \varphi\left(X^k\right).$$

Sequential MC / Sequential Importance Resampling

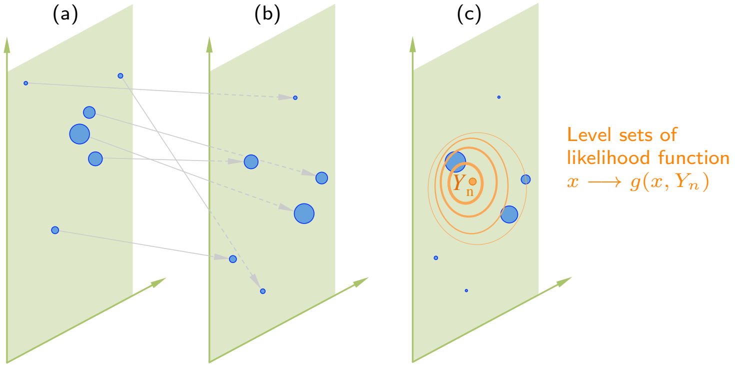
- *At time* $t = 1$
  - Sample $X_1^i \sim q_1(\cdot)$.
  - Compute the weights

$$w_1^i = \frac{\mu(X_1^i) g(y_1 \mid X_1^i)}{q_1\left(X_1^i\right)}.$$

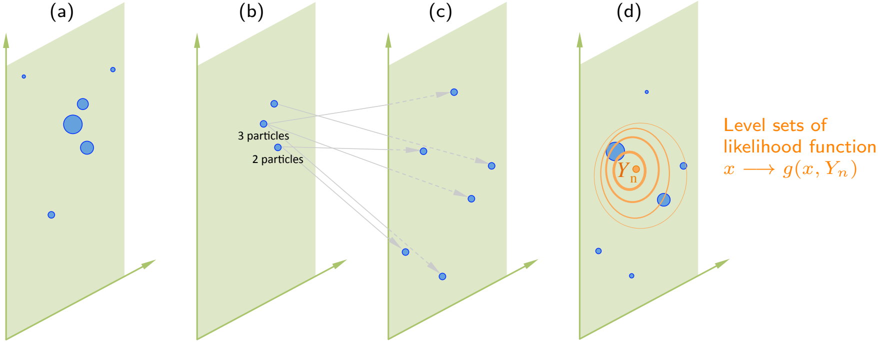
- *At time* $t \geq 2$
  - Resample $(w_{t-1}^i, X_{1:t-1}^i) \to \left(N^{-1}, \overline{X}_{1:t-1}^i\right)$.
  - Sample $X_t^i \sim q_{t|t-1}(\cdot | \bar{X}_{t-1}^i)$, $X_{1:t}^i := \left(\bar{X}_{1:t-1}^i, X_t^i\right)$
  - Compute the weights

$$w_t^i = \omega_t^i = \frac{f\left(X_t^i | X_{t-1}^i\right) g\left(y_t | X_t^i\right)}{q_{t|t-1}(X_t^i \mid X_{t-1}^i)}.$$

Sequential Importance Sampling



Sequential Importance Resampling



Sequential Monte Carlo: example

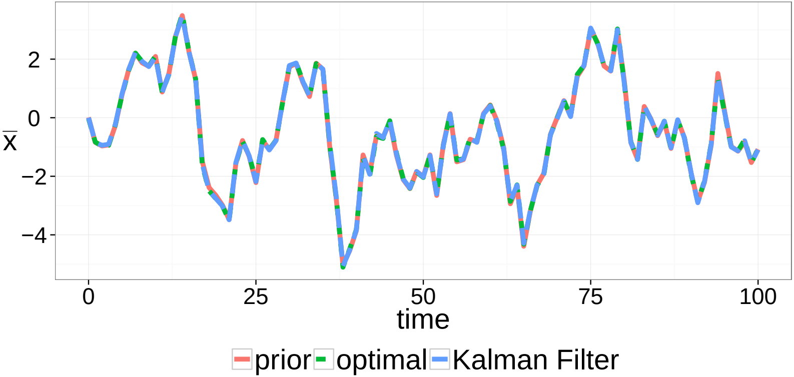


Figure: Estimation of filtering means $\mathbb{E}\left(x_t \mid y_{1:t}\right)$.

Sequential Monte Carlo: example

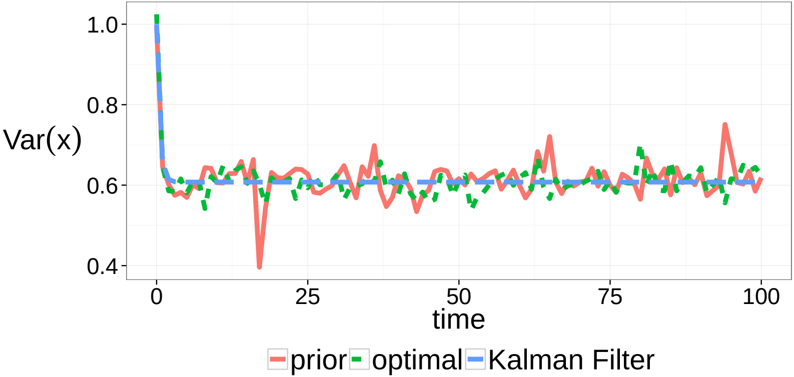


Figure: Estimation of filtering variances $\mathbb{V}(x_t \mid y_{1:t})$.

Sequential Monte Carlo: example

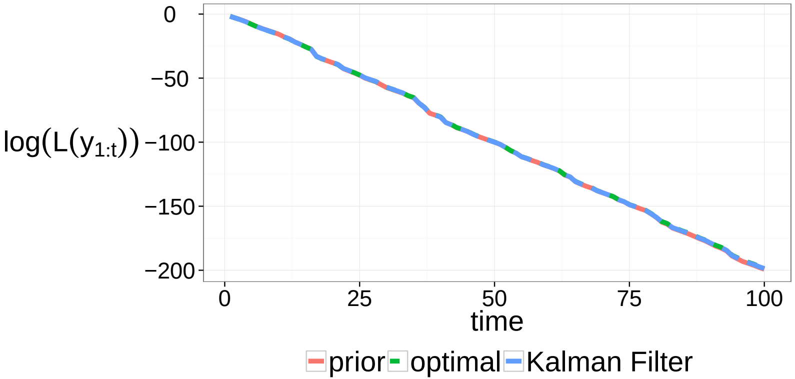


Figure: Estimation of marginal log likelihoods $\log p(y_{1:t})$.

Sequential Monte Carlo: example

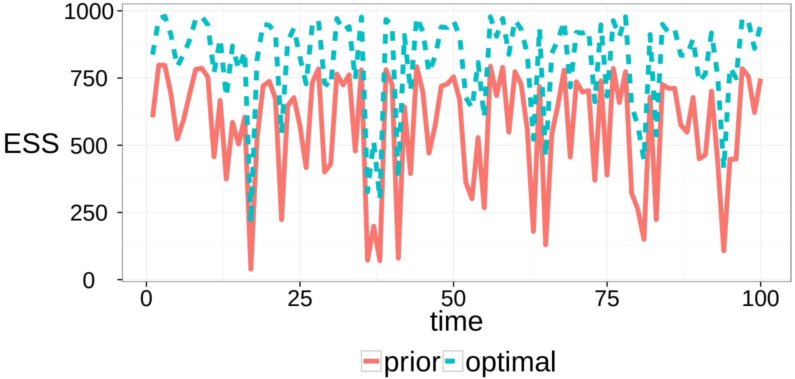


Figure: Effective sample size over time.

Sequential Monte Carlo: example

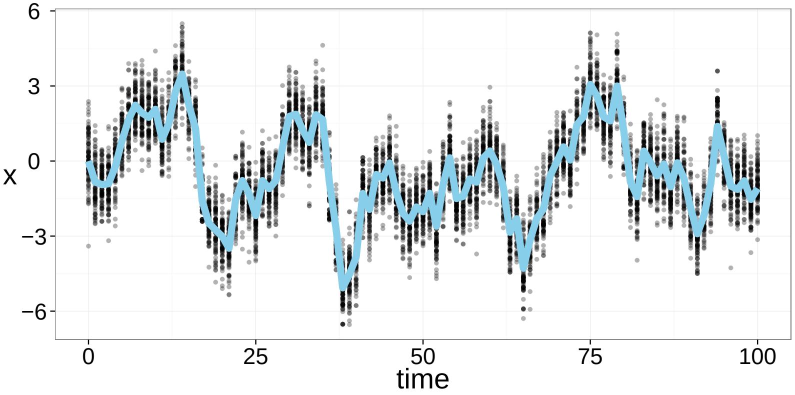


Figure: Support of the approximation of $p(x_t \mid y_{1:t})$, over time.

Sequential Monte Carlo: example

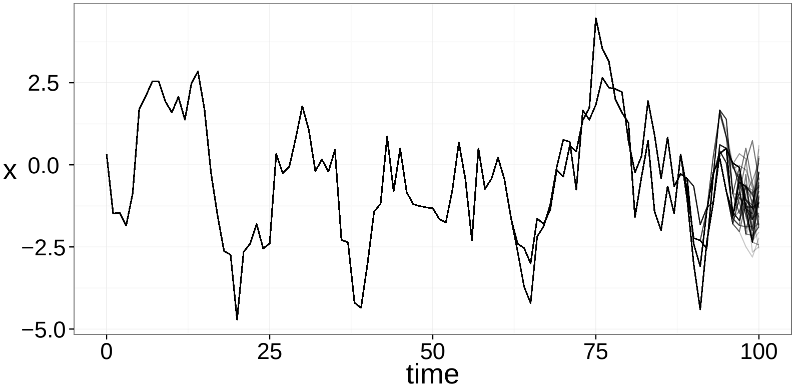


Figure: Trajectories $\bar{X}^i_{1:T}$, for $i \in \{1, \ldots, N\}$ and $N = 100$.