2013-10-11

SMC.1.1

SMC summary 1

Goal: Focusing on state space models (SSMs) summarize the development from sequential importance sampling to auxiliary particle filters.

General setting

SSM, Markov model for state
Y_t conditionally independent given X_t

state $X_t \mid X_{t-1} = x_{t-1} \sim p(x_t \mid x_{t-1}, \theta)$

measurement $Y_t \mid X_t = x_t \sim p(y_t \mid x_t, \theta)$

Typically two distributions of interest

filtering distribution $p(x_t \mid y_{1:t}, \theta)$

joint filtering distribution $p(x_{1:t} \mid y_{1:t}, \theta)$

It holds that $p(x_{0:T} \mid \theta) = p(x_T \mid x_{0:T-1}, \theta)\, p(x_{0:T-1} \mid \theta) =$

Markov $\rightarrow$ $= p(x_T \mid x_{T-1}, \theta)\, p(x_{0:T-1}, \theta) =$

$$= \prod_{t=1}^{T} p(x_t \mid x_{t-1}, \theta)\, p(x_0 \mid \theta)\, p(\theta)$$

$$p(x_{t+1} \mid y_{1:t}, \theta) = \int p(x_{t+1}, x_t \mid y_{1:t}, \theta)\, dx_t =$$
$$= \int p(x_{t+1} \mid x_t, \theta)\, p(x_t \mid y_{1:t}, \theta)\, dx_t$$

$$p(x_t \mid y_{1:t}, \theta) = \frac{p(x_t, y_t \mid y_{1:t-1}, \theta)}{p(y_t \mid y_{1:t-1}, \theta)} =$$
$$= \frac{p(y_t \mid x_t, \theta)\, p(x_t \mid y_{1:t-1}, \theta)}{p(y_t \mid y_{1:t-1}, \theta)} =$$
$$= \frac{p(y_t \mid x_t, \theta) \int p(x_t \mid x_{t-1}, \theta)\, p(x_{t-1} \mid y_{1:t-1}, \theta)\, dx_{t-1}}{p(y_t \mid y_{1:t-1}, \theta)}$$

Additionally: (omit parameters from now on)

$$p(x_{0:T}, y_{1:T}) = p(x_T, y_T \mid x_{0:T-1}, y_{1:T-1})\, p(x_{0:T-1}, y_{1:T-1})$$
$$= p(y_T \mid x_T, x_{0:T-1}, y_{1:T-1})\, p(x_T \mid x_{0:T-1}, y_{1:T-1}) \cdot$$
$$\cdot\, p(x_{0:T-1}, y_{1:T-1}) =$$

conditional independence & Markov $= p(y_T \mid x_T) \cdot p(x_T \mid x_{T-1}) \cdot p(x_{0:T-1}, y_{1:T-1}) =$

$$= \underbrace{\prod_{t=1}^{T} p(y_t \mid x_t)}_{\text{data distribution}} \underbrace{\prod_{t=1}^{T} p(x_t \mid x_{t-1}) \cdot p(x_0)}_{\text{state distribution}}$$

Joint filtering distribution

$$p(x_{1:t} \mid y_{1:t}) = \frac{p(x_{1:t}, y_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} =$$
$$= \frac{p(x_t, y_t \mid x_{1:t-1}, y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \cdot p(x_{1:t-1} \mid y_{1:t-1})$$
$$= \frac{p(y_t \mid x_t)\, p(x_t \mid x_{t-1})}{p(y_t \mid y_{1:t-1})} \cdot p(x_{1:t-1} \mid y_{1:t-1})$$

The joint filtering distribution has a somewhat easier recursion than the filtering distribution.

SMC.1.2

Importance sampling

Let $\pi(x) = \frac{\gamma(x)}{Z}$, where π is a pdf, γ is the unnormalized density and Z is the normalisation constant.

By choosing an appropriate proposal density η we can approximate

$$\mathbb{E}_\pi[f] = \tfrac{1}{Z} \int f(x) \cdot \gamma(x) dx = \tfrac{1}{Z} \int f(x) \cdot w(x) \cdot \eta(x)\, dx$$

where $w(x) = \frac{\gamma(x)}{\eta(x)}$.

It also holds that $Z = \int w(x) \eta(x) dx$.

These integrals can be solved via Monte Carlo methods by replacing $\eta(x)$ by $\frac{1}{N}\sum_{i=1}^N \delta_{X^i}(x)$ where $X^i \sim \eta(x)$.

This leads to

$$\mathbb{E}_\pi[f] \approx \frac{1}{Z \cdot N} \sum_{i=1}^N f(x^i) w(x^i) \quad \text{and}$$

$$Z \approx \frac{1}{N} \sum_{i=1}^N w(x^i).$$

Self-normalising IS: Replace Z by approximation

$$\mathbb{E}_\pi[f] \approx \sum_{i=1}^N \tilde{w}(x^i) f(x^i)$$

where $\tilde{w}(x^i) = \frac{w(x^i)}{\sum_{j=1}^N w(x^j)}$ (normalized weights)

Applying this idea to $p(x_t | y_{1:t})$:

We want to use the recursive property of the target density:

$$p(x_t | y_{1:t}) \propto p(y_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | y_{1:t-1}) dx_{t-1}$$

1) $p(x_0 | y_{1:0}) = p(x_0) = \frac{\gamma_0(x_0)}{Z_0}$

Let η_0 be some proposal function, then

$\gamma_0(x_0) \approx \frac{1}{Z_0 \cdot N} \sum_{i=1}^N w_0^i \delta_{x_0^i}(x) = $, w_0^i: weights, unnormalized; x_0^i, $i = 1, \dots, N$ particles at $t=0$

$= \sum_{i=1}^N \tilde{w}_0^i \delta_{x_0^i}(x)$

2) $p(x_1 | y_{1:1}) \propto p(y_1 | x_1) \int \underbrace{p(x_1 | x_0) \overbrace{p(x_0)}^{= p(x_0 | y_{1:0})} dx_0}_{} = \gamma_1(x_1) \approx$

Replace $p(x_0)$: $\approx p(y_1 | x_1) \sum_{i=1}^N \tilde{w}_0^i p(x_1 | x_0^i)$.

Here: $\gamma_1(x_1) \overset{\text{approx.}}{\propto} p(y_1 | x_1) \underbrace{\sum_{i=1}^N \tilde{w}_0^i p(x_1 | x_0^i)}_{\text{mixture distribution}}$

Choose proposal η_1, then

$w_1^i = \frac{\gamma_1(x_1^i)}{\eta_1(x_1^i)}$ where $X_1^i \sim \eta_1(\cdot)$

We can choose $\eta_1(x_1) = \sum_{i=1}^N \tilde{w}_0^i p(x_1 | x_0^i)$.

This introduces resampling and propagation.

That, however, is known as the <u>bootstrap particle filter.</u>

Applying the IS idea to $\boxed{\text{SMC 1.3}}$

the joint filtering distribution: $p(x_{0:t} \mid y_{1:t})$

Try to make use of recursion again:

$$p(x_{0:t} \mid y_{1:t}) \propto p(y_t \mid x_t) \cdot p(x_t \mid x_{t-1}) \cdot p(x_{0:t-1} \mid y_{1:t-1})$$

1) $p(x_{0:0} \mid y_{1:0}) = p(x_0)$, given

Choose $\underset{\text{or similar}}{\eta_0(x_0)} = p(x_0)$ then $p(x_0) \approx \sum_{i=1}^{N} \tilde{\omega}_0^i \, \delta_{x_0^i}(x_0)$

$\tilde{\omega}_0^i$, normalized weights

2) $p(x_{0:1} \mid y_{1:1}) \propto p(y_1 \mid x_1)\, p(x_1 \mid x_0)\, p(x_0)$

Assume $\eta_1(x_{0:1}) = \eta_1(x_1 \mid x_0)\, \eta_0(x_0)$

We can reuse the samples for η_0 from step 1.

Let $\{\tilde{\omega}_0^i, x_0^i\}$ be the result from step 1.

Sampling x_1^i then means to sample from $\eta_1(x_1 \mid x_0^i)$.

We are effectively sampling a trajectory in every step, but re-using the existing particles.

Then $w(x_{0:1}^i) = w_1^i = \dfrac{p(y_1 \mid x_1^i)\, p(x_1^i \mid x_0^i)}{\eta_1(x_1^i \mid x_0^i)} \cdot \underbrace{\dfrac{p(x_0^i)}{\eta_0(x_0^i)}}_{= w_0^i}$

and $p(x_{0:1} \mid y_{1:1}) \approx \sum_{i=1}^{N} \tilde{\omega}_1^i \, \delta_{x_{0:1}^i}(x_{0:1})$

Note: We have a sequence of pdfs on a measure space of growing dimension.

In general: If for all n

$$\eta_n(x_{0:n}) = \eta_n(x_n \mid x_{n-1})\, \eta_{n-1}(x_{0:n-1})$$

then $w_n^i = \underbrace{\dfrac{p(y_n \mid x_n^i)\, p(x_n^i \mid x_{n-1}^i)}{\eta_n(x_n^i \mid x_{n-1}^i)}}_{=:\, \alpha_n(x_{n-1}^i,\, x_n^i)} \, w_{n-1}^i$ "Sequential Importance Sampling"

As we have seen, this leads to the degeneracy of the weights: Most of them almost zero and only one or very few close to 1.

<u>Problem:</u> $w_n^i = \alpha_n(x_{n-1}^i, x_n^i)\, w_{n-1}^i$

Can we get rid of w_{n-1}^i?

Resampling from $\{x_{0:n-1}^i\}$ according to $\{\tilde{\omega}_{n-1}^i\}$ ensures that $\{X_{0:n-1}^{a^i}\}$ is approximately distributed as $p(x_{0:n-1} \mid y_{1:n-1})$, instead of the weighted sample $\{\tilde{\omega}_{n-1}^i, x_{0:n-1}^i\}$!

Sample then $x_n^i \sim \eta_n(x_n \mid x_{0:n-1}^{a^i})$. We get that

$$(x_{0:n-1}^{a^i}, x_n^i) \sim \eta_n(x_n \mid x_{0:n-1})\, p(x_{0:n-1} \mid y_{1:n-1}).$$

This leads to

$$w_n^i = \frac{p(y_n \mid x_n^i)\, p(x_n^i \mid x_{n-1}^{a^i})}{\eta_n(x_n^i \mid x_{n-1}^{a^i})} = \alpha_n(x_{n-1}^{a^i}, x_n^i)$$

This gets rid of w_{n-1}^i.

This algorithm is sometimes called the "Sequential Importance Resampling".

$\boxed{\text{SMC.1.4}}$

We arrived at a natural proposal for the filtering problem, that lead to the bootstrap particle filter and already contained resampling. In case of the joint filtering problem we had to introduce another step, the resampling step, before the rather natural proposal appeared.

<u>Auxiliary particle filter: Making resampling explicit</u>

In BPF $\quad \eta_n(x_n) = \sum_{i=1}^{N} \tilde{w}_{n-1}^i \, p(x_n | x_{n-1}^i)$

To sample from this proposal we pick a mixture component by sampling $a_n^i \sim \mathrm{Cat}(\{\tilde{w}_{n-1}^i\})$ and then propagate by sampling $x_n \sim p(\cdot | x_{n-1}^{a_n^i})$.

Then $\eta_n(x_n, \underset{\text{resampling index}}{a_n}) = \eta_n(x_n | a_n)\, p(a_n) =$

$$= p(x_n | x_{n-1}^{a_n})\, \tilde{w}_{n-1}^{a_n}$$

and $\eta_n(x_n) = \sum_{i=1}^{N} \eta_n(x_n, a_n^i) =$

$$= \sum_{i=1}^{N} \tilde{w}_{n-1}^{a_n^i}\, p(x_n | x_{n-1}^{a_n^i}).$$

In case of the joint filtering problem, we introduced a resampling step $a_n^i \sim \mathrm{Cat}(\{\tilde{w}_{n-1}^i\})$ and used the proposal $\eta_n(x_{0:n}) = \eta_n(x_n | x_{n-1})\, p(x_{0:n-1} | y_{1:n-1})$

i.e. $x_n^i \sim \eta_n(\cdot | x_{n-1}^{a^i})\, p(x_{0:n-1}^{a^i} | y_{1:n-1})$

Making the resampling explicit, we get

$$\eta_n(x_{0:n}, a_n) = \eta_n(x_n | a_n)\, p(a_n) =$$
$$= \eta_n(x_n | x_{n-1}^{a_n})\, p(x_{0:n-1}^{a_n} | y_{1:n-1}) \cdot \tilde{w}_{n-1}^{a_n}$$

is a proposal for

$$p(x_{0:n}, a_n | y_{1:n}) \propto p(y_n | x_n)\, p(x_n | x_{n-1}^{a_n})\, \underset{\text{"}\tilde{w}_{n-1}^{a_n}}{p(a_n)} \cdot p(x_{0:n-1}^{a_n} | y_{1:n-1})$$

Allow some more generality in the proposal for the APF:

$$\eta_n(x_n, a_n) = \eta_n(x_n | x_{n-1}^{a_n})\, \tilde{v}_{n-1}^{a_n}$$

and in SIR:

$$\eta_n(x_{0:n}, a_n) = \eta_n(x_n | x_{n-1}^{a_n})\, \tilde{v}_{n-1}^{a_n}\, p(x_{0:n-1}^{a_n} | y_{1:n-1})$$

We then get for the new weights

Filtering problem as well as joint filtering problem: $\quad w_n^i = \dfrac{\tilde{w}_n^{a_n^i}\, p(y_n | x_n^i)\, p(x_n^i | x_{n-1}^{a_n^i})}{\tilde{v}_n^{a_n^i}\, \eta_n(x_n | x_{n-1}^{a_n})}$

This makes the particle filter algorithms used to calculate the filtering as well as the joint filtering density exactly equal.

Example: Fully Adapted Particle Filter

Choose $\quad v_n^i \propto \tilde{w}_n^i\, p(y_n | x_{n-1}^i), \quad i = 1, \dots, N$

$$\eta_n(x_n | x_{n-1}) = p(x_n | x_{n-1}, y_n)$$

Then $\tilde{w}_n^i = 1/N$ for all $i = 1, \dots, N$ and all n.

$\boxed{\text{SMC 1.5}}$

Note:
In case of the filtering pdf
we get $p(x_n, a_n | y_{1:n}) \propto p(y_n | x_n)\, p(x_n, a_n | y_{1:n-1}) =$

$$= p(y_n | x_n) \sum_{i=1}^{N} p(x_n, x_{n-1}^{i}, a_n | y_{1:n-1}) =$$

$$= p(y_n | x_n)\, p(x_n | x_{n-1}^{a_n})\, \underbrace{p(a_n)}_{\tilde{w}_n^{a_n}}$$

The idea behind the auxiliary particle filter is to incorporate the ancestor index into the problem and to embrace that we can choose past states only from the available ones, thus leading to a slightly different target distribution.
For large N both should solve the same problem.