Chapter 3

Particle Filtering

3.1 Filtering and smoothing

Let us use the notation $\mathcal{X}_n = (X_1, \ldots, X_n)$ and $\mathcal{Y}_n = (Y_1, \ldots, Y_n)$. Consider the Markov model given by

$$X_1 \sim f(X_1), \quad X_n \,|\, X_{n-1} \sim f(X_n \,|\, X_{n-1}), \quad Y_n \,|\, X_n \sim g(Y_n \,|\, X_n).$$

The filtering problem looks for the conditional density $p(\mathcal{X}_n \,|\, \mathcal{Y}_n)$, which can be written as

$$p(\mathcal{X}_n \,|\, \mathcal{Y}_n) = \frac{p(\mathcal{X}_n, \mathcal{Y}_n)}{p(\mathcal{Y}_n)} = \frac{p(\mathcal{X}_{n-1}, \mathcal{Y}_{n-1})}{p(\mathcal{Y}_{n-1})} \frac{f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)}{p(Y_n \,|\, \mathcal{Y}_{n-1})} \tag{3.1}$$

$$= p(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) \frac{f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)}{p(Y_n \,|\, \mathcal{Y}_{n-1})}, \tag{3.2}$$

where

$$p(Y_n \,|\, \mathcal{Y}_{n-1}) = \int p(X_{n-1} \,|\, \mathcal{Y}_{n-1}) f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n) \, d(X_{n-1}, X_n). \tag{3.3}$$

Often we are only interested in $p(X_n \,|\, \mathcal{Y}_n)$, this can be easily obtained from (3.2) via marginalisation

$$p(X_n \,|\, \mathcal{Y}_n) = \frac{g(Y_n \,|\, X_n) p(X_n \,|\, \mathcal{Y}_{n-1})}{p(Y_n \,|\, \mathcal{Y}_{n-1})} \tag{3.4}$$

$$= \frac{g(Y_n \,|\, X_n)}{p(Y_n \,|\, \mathcal{Y}_{n-1})} \int f(X_n \,|\, X_{n-1}) p(X_{n-1} \,|\, \mathcal{Y}_{n-1}) \, dX_{n-1}, \tag{3.5}$$

The smoothing problem looks for conditional density $p(X_t \,|\, \mathcal{Y}_n)$ where $t < n$. In general, this can also be obtained from (3.2) via marginalisation. However, as seen later in particle filtering techniques, for $t << n$ marginalisation is problematic due to sample degeneracy. Instead, we may use the forward-backward recursion which only

depends on filtering distribution of the type $p(X_t \mid \mathcal{Y}_t)$. The key is to note that, by the Markov property of the model,

$$p(X_t \mid X_{t+1}, \mathcal{Y}_n) = p(X_t \mid X_{t+1}, \mathcal{Y}_t) = \frac{f(X_{t+1} \mid X_t) p(X_t \mid \mathcal{Y}_t)}{p(X_{t+1} \mid \mathcal{Y}_t)}.$$

Thus

$$p(X_t \mid \mathcal{Y}_n) = \int p(X_t \mid X_{t+1}, \mathcal{Y}_n) p(X_{t+1} \mid \mathcal{Y}_n)\, dX_{t+1} \tag{3.6}$$

$$= p(X_t \mid \mathcal{Y}_t) \int \frac{f(X_{t+1} \mid X_t)}{p(X_{t+1} \mid \mathcal{Y}_t)} p(X_{t+1} \mid \mathcal{Y}_n)\, dX_{t+1}. \tag{3.7}$$

3.2 Importance sampling

Suppose that we would like to compute some form of expectation $I(\gamma) = \mathbb{E}(\gamma(X))$ where X follows the density function $p(X)$. If analytical computation is too difficult, often *Monte Carlo simulation methods* are used to approximated the required expectation. In particular, if we can generate an i.i.d. sample $X^1, \ldots, X^N \sim p(X)$, then $p(X)$ and $I(\gamma)$ can be estimated by

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i}(x), \quad \hat{I}(\gamma) = \frac{1}{N} \sum_{i=1}^{N} \gamma(X^i), \tag{3.8}$$

where $\delta_{X^i}(x)$ is the Dirac delta function with the unit mass at X^i. However, in practice it may not be feasible to directly sample from $p(X)$ (e.g., $p(X)$ is overly complicated, or the dimension is too high, or maybe $p(X)$ is only known up to some constant which is difficult to compute). In this case, the approach is to introduce an *importance density* $q(X)$ such that

$$p(X) > 0 \implies q(X) > 0,$$

and it relates to $p(X)$ via

$$p(X) = \frac{w(X) q(X)}{\int w(X) q(X)\, dX} \propto w(X) q(X). \tag{3.9}$$

The *unnormalised weight* function $w(X)$ is known and chosen so that $w(X)q(X)$ is equal to $p(X)$ up to a (possibly unknown) constant. The key is to sample from the important density $q(X)$ and then weight the resulting sample by $w(X)$. So $q(X)$ must be chosen so that it is easy to sample from (e.g., a multivariate Gaussian). After drawing an i.i.d. sample $X^1, \ldots, X^N \sim q(X)$, we obtain

$$\hat{p}(x) \propto w(x) \hat{q}(x) \propto \sum_{i=1}^{N} w(X^i) \delta_{X^i}(x). \tag{3.10}$$

After normalising,

$$\hat{p}(x) = \frac{\sum_{i=1}^{N} w(X^i)\delta_{X^i}(x)}{\sum_{j=1}^{N} w(X^j)} = \sum_{i=1}^{N} W^i \delta_{X^i}(x). \tag{3.11}$$

where

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

are the *normalised weights*. The estimate for $I(\gamma) = \mathbb{E}(\gamma(X))$ is thus

$$\hat{I}(\gamma) = \sum_{i=1}^{N} W^i \gamma(X^i). \tag{3.12}$$

Note that due the normalisation, we required no knowledge of the constant $Z = w(X)q(X)/p(X)$ (which holds for any X). However this constant is in fact the expected value of the weight $w(X^i)$,

$$\mathbb{E}w(X^i) = \int w(X)q(X)\,dX = \frac{w(X)q(X)}{p(X)} = Z. \tag{3.13}$$

The variance of $w(X^i)$ is given by

$$\operatorname{Var} w(X^i) = Z^2\bigg(\int \frac{w(X)^2 q(X)}{Z^2}\,dX - 1\bigg) = Z^2\bigg(\int \frac{p(X)^2}{q(X)}\,dX - 1\bigg). \tag{3.14}$$

3.3 Sequential importance resampling

Let us again focus on the original problem of

$$X_1 \sim f(X_1), \quad X_n \,|\, X_{n-1} \sim f(X_n \,|\, X_{n-1}), \quad Y_n \,|\, X_n \sim g(Y_n \,|\, X_n).$$

We would like to perform sequential importance sampling to generate $X_1, X_2, \ldots, X_n$. Recall that the density $p(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ has the recursive structure of

$$p(\mathcal{X}_n \,|\, \mathcal{Y}_n) \propto p(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n). \tag{3.15}$$

The importance density should have a similar recursive structure

$$q(\mathcal{X}_n) = q(\mathcal{X}_{n-1}) q(X_n \,|\mathcal{X}_{n-1}).$$

At step n, we would like to generate X_n. Due to the nature of the problem, a sensible choice would be of the form

$$q(X_n \,|\mathcal{X}_{n-1}) = q(X_n \,|Y_n, X_{n-1}),$$

i.e., to have X_n only depend on the previous state variable X_{n-1} and the current observation Y_n. So we focus on importance densities of the form

$$q(\mathcal{X}_n) = q(\mathcal{X}_{n-1})q(X_n \,|Y_n, X_{n-1}). \tag{3.16}$$

The existing samples $\mathcal{X}_{n-1}^i, i = 1, \ldots, N$, known as the *particles*, will be used to generate X_n^i, extending the particles to time n. Dividing (3.15) by (3.16), the required weights should also be computed sequentially

$$w(\mathcal{X}_n) \propto w(\mathcal{X}_{n-1}) \frac{f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)}{q(X_n \,|Y_n, X_{n-1})}.$$

Again we are not concerned about missing constants as these weights will be normalised to

$$W_n^i = \frac{w(\mathcal{X}_n^i)}{\sum_{j=1}^N w(\mathcal{X}_n^j)}.$$

Now we have all the required ingredient to estimate the particle filter $p(\mathcal{X}_n \,|\, \mathcal{Y}_n)$,

$$\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n) = \sum_{i=1}^N W_n^i \delta_{\mathcal{X}_n^i}(\mathcal{X}_n).$$

Resampling

As shown in (3.14), the relative variance of the weights are given by

$$\frac{\operatorname{Var} w(\mathcal{X}_n^i)}{(\mathbb{E} w(\mathcal{X}_n^i))^2} = \int \frac{p(\mathcal{X}_n \,|\, \mathcal{Y}_n)^2}{q(\mathcal{X}_n)} \, d\mathcal{X}_n - 1. \tag{3.17}$$

Due to the sequential nature of $p(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ and $q(\mathcal{X}_n)$, the relative variance typically grows exponentially as n increases. This poses a problem since the weights of many particles will be close to zero, causing the simulation to expend most of its computation power on regions where the $p(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ is close to 0. Ideally, we would like the simulation to concentrate on "more interesting" regions of the density function in order to provide the most information.

This issue can be overcome by the process of *resampling*. The idea is simple, after constructing $\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ using $(W_n^i, \mathcal{X}_n^i), i = 1, \ldots, N$, we simply generate a sample directly from $\bar{\mathcal{X}}_n^i \sim \hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n)$. This can always be achieved since $\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ is simply a discrete multinomial distribution. Since we are directly sampling from the (estimated) desired distribution, the corresponding weights of the new sample are simply equal to $\bar{W}_n^i = 1/N$. Together, $(\bar{W}_n^i, \bar{\mathcal{X}}_n^i), i = 1, \ldots, N$ replaces $(W_n^i, \mathcal{X}_n^i), i = 1, \ldots, N$ for the next step onwards.

Even though all weights are equal to $1/N$, we still have a good representation of $\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n)$ since the sample most likely contains multiplicity. In particular, any previous particle $\mathcal{X}_n^i$ with high weights will appear multiple times in $\bar{\mathcal{X}}_n^j$. On the

other hand, many particles with weights close to zero won't appear at all and are effectively discarded. Particles which coincide in value are still meaningful as they will evolve separately in future steps.

Despite eliminating weights which are effectively zero, resampling has disadvantages. In most cases we have lost information since particles with low but non-zero weights could be discarded. If performed many times through the duration of the particle filter, the earlier values such as $X_1, X_2, \ldots$ could suffer from sample degeneracy as they may collapse to very few values. Finally, resampling is computationally expensive for large values of N. So even though resampling is useful in improving the robustness of the particle filter, it may not be the best idea to use it every step. The approach of *adaptive resampling* computes the effective sample size (ESS) at each step, where

$$ESS = \left(\sum_{i=1}^{N} (W_n^i)^2 \right)^{-1}.$$

The ESS takes values between 1 and N, with larger values indicating a healthier set of weights. Then resampling is only carried out if the ESS falls below a certain threshold.

Sequential importance resampling algorithm

At time $n = 1$

- Sample $X_1^i \sim q(X_1 \,|\, Y_1)$.
- Compute $w(X_1^i) = \frac{f(X_1^i) g(Y_1 \,|\, X_1^i)}{q(X_1^i \,|\, Y_1)}$ and normalise $W_1^i \propto w(X_1^i)$.
- The estimated density at time 1 is $\hat{p}(\mathcal{X}_1 \,|\, \mathcal{Y}_1) = \sum_{i=1}^{N} W_1^i \delta_{\mathcal{X}_1^i}(\mathcal{X}_1)$.

At time $n \geq 2$

- If resampling, replace $\{W_{n-1}^i, \mathcal{X}_{n-1}^i\}$ by $\{1/N, \bar{\mathcal{X}}_{n-1}^i\}$ where $\bar{\mathcal{X}}_{n-1}^i \sim \hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) = \sum_{i=1}^{N} W_{n-1}^i \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1})$.
- Sample $X_n^i \sim q(X_n \,|\, Y_n, X_{n-1}^i)$ and set $\mathcal{X}_n^i$ to be $(\mathcal{X}_{n-1}^i, X_n^i)$.
- Compute $w(\mathcal{X}_n^i) = w(\mathcal{X}_{n-1}^i) \frac{f(X_n^i \,|\, X_{n-1}^i) g(Y_n \,|\, X_n^i)}{q(X_n^i \,|\, Y_n, X_{n-1}^i)}$ and normalise $W_n^i \propto w(\mathcal{X}_n^i)$.
- The estimated density at time n is $\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n) = \sum_{i=1}^{N} W_n^i \delta_{\mathcal{X}_n^i}(\mathcal{X}_n)$.

One simple example, known as the *bootstrap filter* or the *prior kernel*, uses the state transitional density as the importance density

$$q(X_n \,|Y_n, X_{n-1}) = f(X_n \,|\, X_{n-1}).$$

In this case the weights simplify to

$$w(\mathcal{X}_n) \propto w(\mathcal{X}_{n-1}) g(Y_n \,|\, X_n),$$

which only depends on the likelihood of the observation. This choice is popular since sampling and weight computation are usually straightforward.

A more elaborate example, which is refer to as the *optimal kernel*, conditions on both the previous state variable as well as the current observation,

$$\begin{aligned} q(X_n \,|Y_n, X_{n-1}) = p(X_n \,|Y_n, X_{n-1}) &= \frac{f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)}{p(Y_n \,|\, X_{n-1})} \\ &= \frac{f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)}{\int f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)\, dX_n}. \end{aligned}$$

The weights are then given by

$$w(\mathcal{X}_n) \propto w(\mathcal{X}_{n-1}) p(Y_n \,|\, X_{n-1}) = w(\mathcal{X}_{n-1}) \int f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)\, dX_n.$$

This is optimal in the sense that $w(\mathcal{X}_n)$ have zero conditional variance given the past samples. It also has the advantage of taking the observations into account when sampling, as opposed to the bootstrap filter which only depends on the state variables. However, it is less tractable analytically.

3.4 Auxiliary particle filter

The key difference between the auxiliary particle filter and the sequential importance resampling is in how they handle the resampling step. In the sequential importance resampling algorithm, the particles are resampled at time n using the estimated filtering density $\hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1})$ in order to eliminate the presence of low weight particles. The auxiliary particle filter takes it one step further and incorporates the value of Y_n during resampling. The intuition is that we should favour particles with higher weights from previous step, as well as particles which are more likely to survive in the current step based on the likelihood value of Y_n.

Ideally we would like to resample from the posterior density $p(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_n)$, which, by Bayes' theorem, has the representation

$$\begin{aligned} p(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_n) = \frac{p(\mathcal{X}_{n-1}, Y_n \,|\, \mathcal{Y}_{n-1})}{p(Y_n \,|\, \mathcal{Y}_{n-1})} &\propto p(\mathcal{X}_{n-1}, Y_n \,|\, \mathcal{Y}_{n-1}) \\ &= \int p(\mathcal{X}_{n-1}, X_n, Y_n \,|\, \mathcal{Y}_{n-1})\, dX_n \\ &= p(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) \int f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)\, dX_n \end{aligned}$$

At time n, we already have the estimated density

$$\hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) = \sum_{i=1}^{N} W_{n-1}^i \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1}).$$

However, the integral in general cannot always be easily evaluated. Instead, we will simply approximate $f(X_n \,|\, X_{n-1})$ by the Dirac delta function $\delta_{\mu(X_{n-1})}(X_n)$ where $\mu(X_{n-1})$ is a well-chosen representative of the density $f(X_n \,|\, X_{n-1})$, such as its mean or mode. Hence the estimated posterior density is

$$\begin{aligned}
\hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_n) &\propto \hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_{n-1}) \int f(X_n \,|\, X_{n-1}) g(Y_n \,|\, X_n)\, dX_n \\
&\approx \sum_{i=1}^{N} W_{n-1}^i \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1}) \int \delta_{\mu(X_{n-1})}(X_n) g(Y_n \,|\, X_n)\, dX_n \\
&= \sum_{i=1}^{N} W_{n-1}^i \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1}) g(Y_n \,|\, \mu(X_{n-1})) \\
&= \sum_{i=1}^{N} W_{n-1}^i g(Y_n \,|\, \mu(X_{n-1}^i)) \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1}).
\end{aligned}$$

So the resampling should be performed on the density

$$\hat{p}(\mathcal{X}_{n-1} \,|\, \mathcal{Y}_n) = \sum_{i=1}^{N} V_{n-1}^i \delta_{\mathcal{X}_{n-1}^i}(\mathcal{X}_{n-1}), \quad \text{where } V_{n-1}^i \propto W_{n-1}^i g(Y_n \,|\, \mu(X_{n-1}^i)).$$

In other words, the weights V_{n-1}^i (instead of W_{n-1}^i) should be used for resampling the particles. They can be computed by normalising $W_{n-1}^i g(Y_n \,|\, \mu(X_{n-1}^i))$ where $\mu(X_{n-1}^i)$ is the mean or mode of $f(X_n \,|\, X_{n-1}^i)$.

Since we are resampling using a differently method to the previous section, it shouldn't be a surprise that the particle filtering weights should no longer be updated to $1/N$. In order to work out the correct weights, recall that the "correct" distribution for the newly resampled particles $\bar{\mathcal{X}}_{n-1}$ should be $\hat{p}(\bar{\mathcal{X}}_{n-1} \,|\, \mathcal{Y}_n)$, even though we sampled them from $\hat{p}(\bar{\mathcal{X}}_{n-1} \,|\, \mathcal{Y}_n)$. This situation is similar to the standard importance sampling algorithm, where the weights should be the ratio of the two densities (see (3.8) and (3.9)).

$$\begin{aligned}
\sum_{i=1}^{N} w(\bar{\mathcal{X}}_{n-1}^i) \delta_{\bar{\mathcal{X}}_{n-1}^i}(\mathcal{X}_{n-1}) &\propto \sum_{i=1}^{N} \frac{\hat{p}(\bar{\mathcal{X}}_{n-1}^i \,|\, \mathcal{Y}_{n-1})}{\hat{p}(\bar{\mathcal{X}}_{n-1}^i \,|\, \mathcal{Y}_n)} \delta_{\bar{\mathcal{X}}_{n-1}^i}(\mathcal{X}_{n-1}) \\
&= \sum_{i=1}^{N} \frac{\sum_{j=1}^{N} W_{n-1}^j \delta_{\mathcal{X}_{n-1}^j}(\bar{\mathcal{X}}_{n-1}^i)}{\sum_{j=1}^{N} V_{n-1}^j \delta_{\mathcal{X}_{n-1}^j}(\bar{\mathcal{X}}_{n-1}^i)} \delta_{\bar{\mathcal{X}}_{n-1}^i}(\mathcal{X}_{n-1})
\end{aligned}$$

This looks complicated, but in practice, since $\bar{\mathcal{X}}_{n-1}^i = \mathcal{X}_{n-1}^{j_i}$ for some resampled index j_i, the most of the terms will disappear and we will simply have

$$\sum_{i=1}^N w(\bar{\mathcal{X}}_{n-1}^i)\delta_{\bar{\mathcal{X}}_{n-1}^i}(\mathcal{X}_{n-1}) \propto \sum_{i=1}^N \frac{W_{n-1}^{j_i}}{V_{n-1}^{j_i}}\delta_{\bar{\mathcal{X}}_{n-1}^i}(\mathcal{X}_{n-1}), \quad \text{or } w(\bar{\mathcal{X}}_{n-1}^i) \propto \frac{W_{n-1}^{j_i}}{V_{n-1}^{j_i}}.$$

Thus the filtering weights used will be simply the ratio of $W_{n-1}^{j_i}$ and $V_{n-1}^{j_i}$, where the index j_i is defined by $\bar{\mathcal{X}}_{n-1}^i = \mathcal{X}_{n-1}^{j_i}$.

Aside from the resampling step, the remainder of the algorithm is identical to the sequential importance resampling algorithm. In fact, The sequential importance resampling algorithm can be seen as a special case of the auxiliary particle filter where $V_{n-1}^i = W_{n-1}^i$, in which case the new filtering weights after resampling are indeed $W_{n-1}^{j_i}/V_{n-1}^{j_i} \propto 1/N$. The two methods have similar performance in general, but the auxiliary particle tends to "eliminate" more particles during resampling, since the surviving particles must have both high original weights W_{n-1}^i and high posterier likelihood $g(Y_n \,|\, \mu(X_{n-1}^i))$.

Auxiliary particle filter algorithm

At time $n = 1$

- Sample $X_1^i \sim q(X_1 \,|\, Y_1)$.
- Compute $w(X_1^i) = \frac{f(X_1^i)g(Y_1 \,|\, X_1^i)}{q(X_1^i \,|\, Y_1)}$ and normalise $W_1^i \propto w(X_1^i)$.
- The estimated density at time 1 is $\hat{p}(\mathcal{X}_1 \,|\, \mathcal{Y}_1) = \sum_{i=1}^N W_1^i \delta_{\mathcal{X}_1^i}(\mathcal{X}_1)$.

At time $n \geq 2$

- If resampling:
  - Compute the resampling weights $V_{n-1}^i \propto W_{n-1}^i g(Y_n \,|\, \mu(X_{n-1}^i))$ where $\mu(X_{n-1}^i)$ is the mean or mode of $f(X_n \,|\, X_{n-1}^i)$.
  - Sample a new set of indices $j_1, \ldots, j_N$ from the set $\{1, \ldots, N\}$ using the weights $\{V_{n-1}^1, \ldots, V_{n-1}^N\}$.
  - Set $\bar{\mathcal{X}}_{n-1}^i = \mathcal{X}_{n-1}^{j_i}$ and normalise $\bar{W}_{n-1}^i \propto W_{n-1}^{j_i}/V_{n-1}^{j_i}$.
  - Replace $(\mathcal{X}_{n-1}^i, W_{n-1}^i)$ by $(\bar{\mathcal{X}}_{n-1}^i, \bar{W}_{n-1}^i)$.
- Sample $X_n^i \sim q(X_n \,|\, Y_n, X_{n-1}^i)$ and set $\mathcal{X}_n^i$ to be $(\mathcal{X}_{n-1}^i, X_n^i)$.
- Compute $w(\mathcal{X}_n^i) = w(\mathcal{X}_{n-1}^i)\frac{f(X_n^i \,|\, X_{n-1}^i)g(Y_n \,|\, X_n^i)}{q(X_n^i \,|\, Y_n, X_{n-1}^i)}$ and normalise $W_n^i \propto w(\mathcal{X}_n^i)$.
- The estimated density at time n is $\hat{p}(\mathcal{X}_n \,|\, \mathcal{Y}_n) = \sum_{i=1}^N W_n^i \delta_{\mathcal{X}_n^i}(\mathcal{X}_n)$.