2.160 IDENTIFICATION, ESTIMATION, AND LEARNING

LECTURE NOTES NO. 9

9. Particle Filters

9.1 Non-Parametric Representation of a Probability Distribution

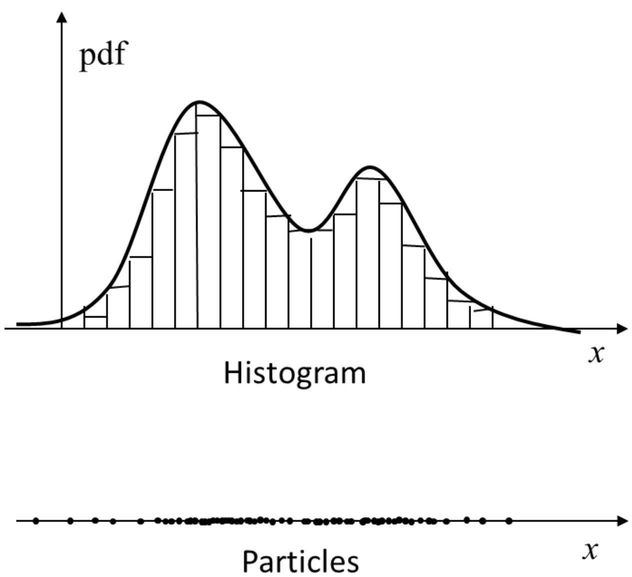

Figure 9-1 Non-parametric representations of a probability distribution: Histogram (top) and Particles (bottom)

There are two types of representation of probability distribution: Parametric and Non-parametric. The former examples include Gaussian, Poisson, binomial, Laplace, and chi-squared distributions, each of which is represented with particular parameters. A Gaussian distribution, for example, is completely characterized with two parameters: mean and covariance. The latter, non-parametric representation can represent an arbitrary distribution. A histogram, for example, represents an arbitrary distribution by a collection of adjacent rectangles, each indicating the frequency of occurrence within the interval, called a bin. It approximates the original (continuous) distribution, as shown in Figure 9-1.

"Particles" are another non-parametric representation of probability distribution. Simply, particles are a collection of samples, as shown in Figure 9-1. Suppose that we want to approximate a distribution given by pdf $f(x)$. Particles are generated by drawing M samples from $f(x)$:

$$\tilde{X} = \left\{ x^{(1)}, x^{(2)}, \cdots, x^{(M)} \right\} \tag{1}$$

Note that $x^{(i)}$ may be populated densely where $f(x)$ is large, reflecting the probability density of $f(x)$.

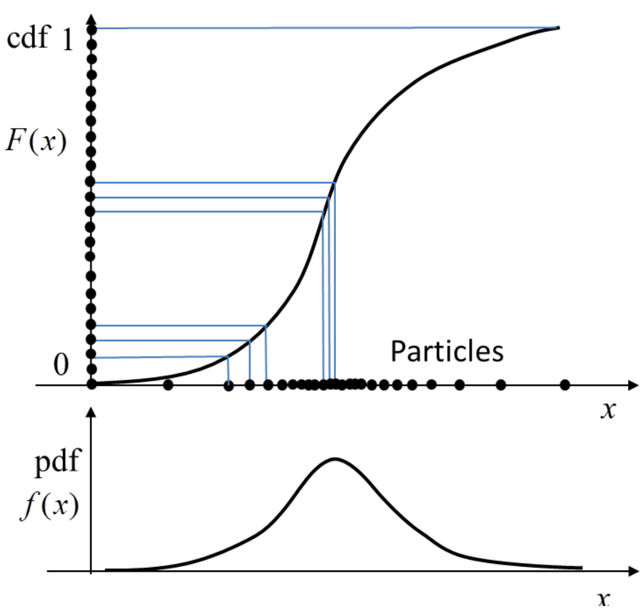

Figure 9-2 Generation of particles from a uniform distribution

Now how can we draw samples with a specific pdf? Consider the algorithm depicted in Figure 9-2.

a) Construct the cumulative distribution function (cdf) of pdf:

$$F(x) = \int_{-\infty}^{x} f(\xi) d\xi \tag{2}$$

b) Draw M samples from a uniform distribution between 0 and 1; $0 \le y^{(i)} \le 1$. This can be easily done in MATLAB.

c) Convert $y^{(i)}$ to $x^{(i)}$ by solving:

$$x^{(i)} = F(y^{(i)}) . \tag{3}$$

Note that cdf $F(x)$ is a monotonically non-decreasing function between 0 and 1. Therefore, (3) has a unique solution where $f(x) \neq 0$.

9.2 Implementation of Bayesian Filter Using Particles

The Recursive Bayesian Filter we have discussed in the previous chapter can be implemented by using "Particles". The Bayesian Filter algorithm consists of two steps of computation:

Given initial conditions: $g_0(x_0)$: Initial belief of state, i.e. probability density of x_0,
Step 1 (Propagation: the Chapman – Kolmogorov equation) Compute:

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} p(x_t \mid x_{t-1}, u_{t-1}) \cdot g_{t-1}(x_{t-1}) dx_{t-1} \tag{4}$$

Step 2 (Update: Bayes' Rule) Assimilate new data y_t and compute:

$$g_t(x_t) = \eta p(y_t \mid x_t) \cdot g_{t|t-1}(x_t) \tag{5}$$

Set $t = t+1$ and repeat the process.

The feature of Bayesian Filtering is that it is applicable to nonlinear dynamical systems with non-Gaussian process and measurement noise.

Stochastic discrete-time, nonlinear dynamical system:

$$x_t = \mathbf{f}\left(x_{t-1}, u_{t-1}\right) + w_{t-1} \tag{6}$$

where w_{t-1} is uncorrelated process noise with pdf $f_W(w_{t-1})$, and an observation equation

$$y_t = \mathbf{h}\left(x_t\right) + v_t \tag{7}$$

where v_t is uncorrelated measurement noise with pdf $f_V(v_t)$. Using these pdfs we can write

$$p(x_t \mid x_{t-1}, u_{t-1}) = f_W(x_t - \mathbf{f}(x_{t-1}, u_{t-1})) \tag{8}$$

and

$$p(y_t \mid x_t) = f_V(y_t - \mathbf{h}(x_t)) \tag{9}$$

in (4) and (5).

In applying Particles to the above Bayesian Filtering, let us first represent the posteriori belief $g_{t-1}(x_{t-1})$ by a set of M particles:

$$\tilde{X}_{t-1} = \left\{ x_{t-1}^{(1)}, x_{t-1}^{(2)}, \cdots, x_{t-1}^{(M)} \right\} \tag{10}$$

drawn from $g_{t-1}(x_{t-1})$. The number of samples is typically: M = 1,000.

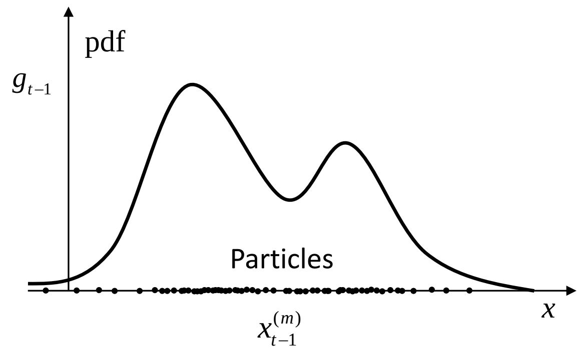

Figure 9-3 A set of particles approximating $g_{t-1}(x_{t-1})$

Instead of computing (4) directly, we stochastically propagate each sample in $\tilde{X}_{t-1}$ based on (6) to generate a set of particles approximating the a priori belief $g_{t|t-1}(x_t)$. See Figure 9-4.

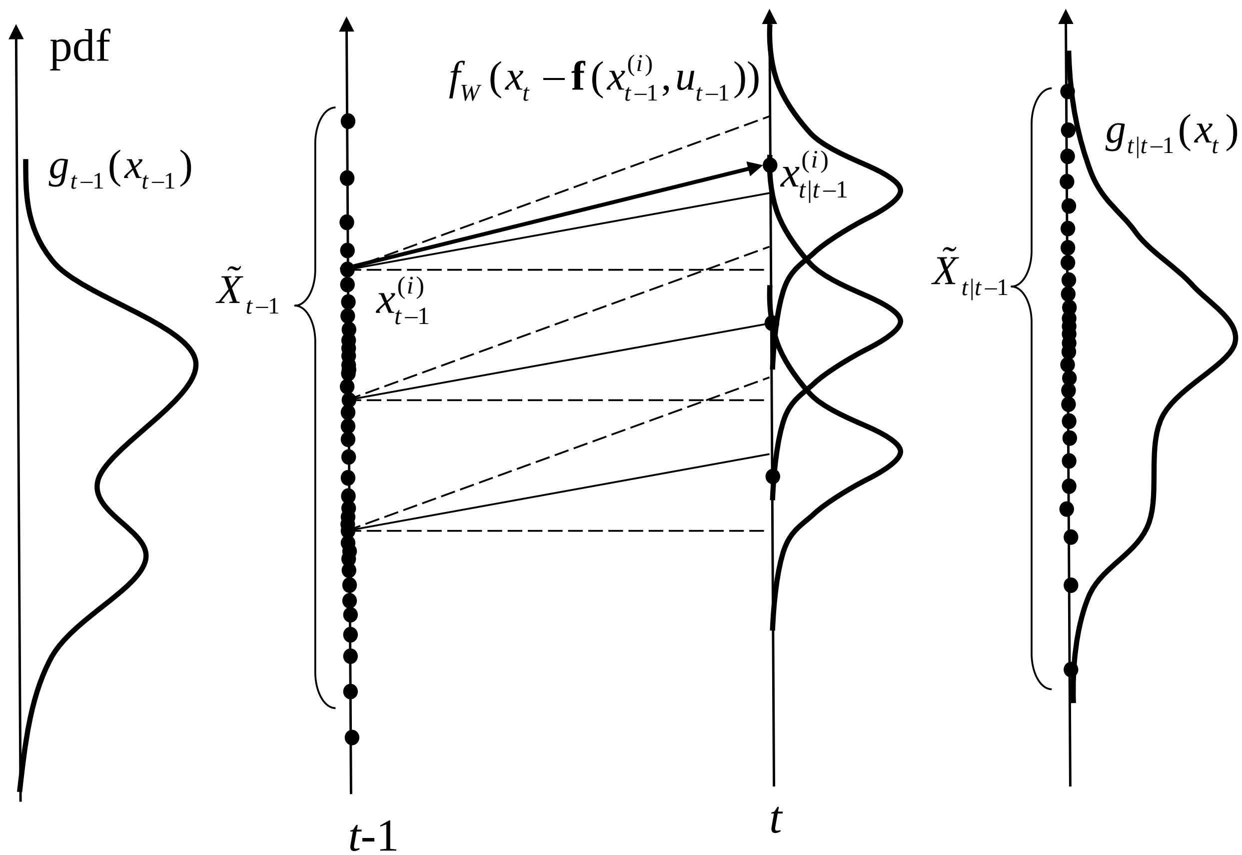

Figure 9-4 Propagation of particles through the nonlinear stochastic dynamic eq.

For i = 1 to M, draw $x_{t|t-1}^{(i)}$ from $f_W(x_t - \mathbf{f}(x_{t-1}, u_{t-1}))$. This includes a shift in x due to the deterministic part of the state transition $\mathbf{f}(x_{t-1}, u_{t-1})$ with input u_{t-1} and the random part due to the process noise. Collecting the propagated particles we can form :

$$\tilde{X}_{t|t-1} = \left\{ x_{t|t-1}^{(1)}, x_{t|t-1}^{(2)}, \cdots, x_{t|t-1}^{(M)} \right\} \tag{11}$$

which approximates the a priori belief $g_{t|t-1}(x_{t|t-1})$.

The second step of Bayesian Filter computation is belief update. The key technique for this implementation is "Importance Sampling".

<u>Importance Sampling</u>

Consider two pdfs $f(x)$ and $g(x)$, where

$$g(x) > 0 \quad \forall x \in X \quad s.t. \quad f(x) > 0. \tag{12}$$

See Figure 9-5 (a). Suppose that sampling from $f(x)$ is difficult to perform, while it is easy from $g(x)$. The following algorithm allows us to sample from $f(x)$ by sampling from $g(x)$ with a proper weight. The cumulative distribution function of $f(x)$ can be written as

$$F(x) = \int_{-\infty}^{x} f(\xi) d\xi = \int_{-\infty}^{x} \frac{f(\xi)}{g(\xi)} g(\xi) d\xi \tag{13}$$

We define:

$$W(x) = \frac{f(x)}{g(x)} \tag{14}$$

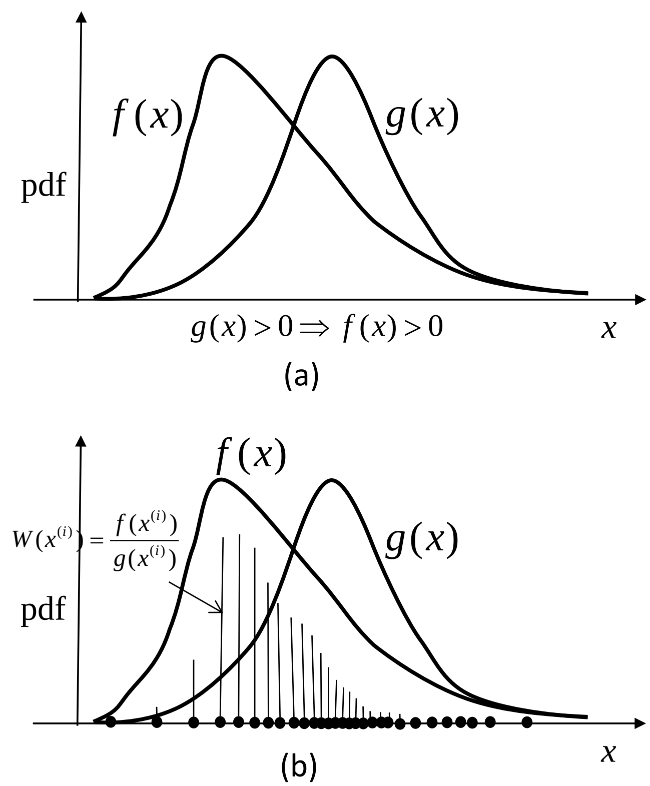

Figure 9-5 Importance sampling

called importance weight or importance factor. This suggest that, although we cannot sample particles directly from $f(x)$, we can obtain the cdf $F(x)$ from

$$F(x) = \int_{-\infty}^{x} W(\xi) g(\xi) d\xi \tag{15}$$

Let $x^{(1)} \cdots x^{(M)}$ be sample particles drawn from $g(x)$, and $I(\xi, x)$ be a membership function given by

$$I(\xi, x) = \begin{cases} 1; & \xi \le x \\ 0; & \xi > x \end{cases} . \tag{16}$$

Then the integral (15) can be written as

$$F(x) \cong \frac{1}{M} \sum_{i=1}^{M} \frac{f(x^{(i)})}{g(x^{(i)})} I(x^{(i)}, x) \tag{17}$$

As shown in Figure 9-5 (b), the locations of the particles are approximating $g(x)$, but with the importance weights they represent the distribution of $f(x)$.

This importance sampling technique can be applied to the second step of Bayesian Filter computation. Comparing (5),(9) and (15) or (17), we can relate $g(x)$, $f(x)$, and $W(x)$ to

$$\begin{aligned} & g(x) \Leftrightarrow g_{t|t-1}(x_t) \sim \tilde{X}_{t|t-1} \\ & f(x) \Leftrightarrow g_t(x_t) \\ & W(x) \Leftrightarrow W^{(i)} = f_V(y_t - \mathbf{h}(x_{t|t-1}^{(i)})) \end{aligned} \tag{17}$$

Therefore, the cumulative distribution function of $f(x) = g_t(x_t)$ can be computed as

$$G_t(x_t) = \int_{-\infty}^{x} W(\xi) g_{t|t-1}(\xi) d\xi \cong \frac{1}{M} \sum_{i=1}^{M} f_V(y_t - \mathbf{h}(x_{t|t-1}^{(i)})) I(x_{t|t-1}^{(i)}, x_t) \tag{18}$$

After constructing the cdf $G_t(x_t)$, draw M samples from $G_t(x_t)$ and place them in

$$\tilde{X}_t = \left\{ x_t^{(1)}, x_t^{(2)}, \cdots, x_t^{(M)} \right\} \tag{19}$$

This algorithm is called “Particle Filter”.

9.3 Re-sampling

There are many variations to the above Particle Filter. One well known technique is “Re-sampling”, which is an alternative to the second step of the above algorithm.

Sample an integer m ($1 \le m \le M$) with a probability proportional to importance factor $W^{(m)}$, and include the corresponding particle $x_{t|t-1}^{(m)}$ into $\tilde{X}_t$. Repeat this process M times to form:

$$\tilde{X}_t = \{ x_t^{(1)}, \cdots, x_t^{(M)} \} . \tag{20}$$

Each of these particles is taken from $\tilde{X}_{t|t-1} = \{x_{t|t-1}^{(1)}, \cdots, x_{t|t-1}^{(M)}\}$. Particle $x_{t|t-1}^{(i)}$ having a high importance factor $W^{(i)}$ is likely to be taken repeatedly, while the ones with lower importance factors may be eliminated in $\tilde{X}_t$.

The sampling of integer *m* with $W^{(m)}$ can be implemented by computing the cumulative of $W^{(m)}$:

$$\Sigma^m = \frac{1}{\overline{W}} \sum_{i=1}^{m} W^{(i)}, \quad \overline{W} = \sum_{i=1}^{M} W^{(i)} \tag{21}$$

As particles are re-sampled, the posteriori particles look, for example,

$$\tilde{X}_t = \{x_{t|t-1}^{(1)}, x_{t|t-1}^{(2)}, x_{t|t-1}^{(3)}, x_{t|t-1}^{(3)}, x_{t|t-1}^{(3)}, x_{t|t-1}^{(4)}, x_{t|t-1}^{(4)}, x_{t|t-1}^{(5)}, x_{t|t-1}^{(7)}, x_{t|t-1}^{(7)}, x_{t|t-1}^{(9)}, \cdots\}$$

Note that $x_{t|t-1}^{(3)}$ was sampled three times, since the corresponding importance factor was large. Thus, the probability of $\hat{x}_t$ conditioned by observation y_t and input u_t is reflected in the posterior belief.