# Module 14
# The Particle Filter

To completely understand and appreciate the particle filter, it is first critical that we review the two key components of the Kalman filter. Recall that the state-space equations for the Kalman filter are

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{V}_t \qquad \text{where } \mathbf{V}_t \sim N(\mathbf{0}, \mathbf{Q}) \tag{1}$$
$$\mathbf{Y}_t = \mathbf{G}\mathbf{X}_t + \mathbf{W}_t \qquad \text{where } \mathbf{W}_t \sim N(\mathbf{0}, \mathbf{R}). \tag{2}$$

Recall that Equation (1) is the "state" equation and Equation (2) is the "measurement" equation. They are both linear equations and assuming the relationship between $\mathbf{X}_t$ and $\mathbf{X}_{t-1}$ and $\mathbf{Y}_t$ and $\mathbf{X}_t$ is linear is a rather strict and and strong assumption. If the relationships are not linear, the state-space equations can be written as

$$\mathbf{X}_t = f(\mathbf{X}_{t-1}, \mathbf{V}_t) \tag{3}$$
$$\mathbf{Y}_t = g(\mathbf{X}_t, \mathbf{W}_t), \tag{4}$$

where $f$ and $g$ are "non-linear" functions.

When $f$ and $g$ are non-linear functions, how should one filter? How can one approximate/predict the value of $\mathbf{X}_t$ given the measurements $\mathbf{Y}_t$? A common solution to this problem is to linearlize the functions $f$ and $g$ (using Taylor linearization) and proceed with the standard Kalman filter. This works *only* in the case where this linear approximation to $f$ and $g$ are good. When they are not good approximations, other approaches need to be considered. This is where the particle filter comes in.

Before we discuss the details of the particle filter, it is critical that we review some of the fundamentals of probability (specifically conditional probability). Keep in mind that what we eventually want to calculate is

$$p(\mathbf{X}_t|\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_0).$$

From the laws of conditional propbability, we know that

$$p(\mathbf{X}_t|\mathbf{Y}_t) = \underbrace{p\left(\mathbf{Y}_t|\mathbf{X}_t\right)}_{\text{Easy to get with (4)}} \times \underbrace{p(\mathbf{X}_t|\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_0)}_{\text{Prediction of } \mathbf{X}_t \text{ based on previous obsrvtns.}} \Big/ \underbrace{p(\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots, \mathbf{Y}_0)}_{\text{normalizing constant}}.$$

Think about this: the normalizing constant in the equation above does **not** depend on $\mathbf{X}_t$, but the left side of the equation does. So it would be fair to write

$$p(\mathbf{X}_t|\mathbf{Y}_t) = c \times p\left(\mathbf{Y}_t|\mathbf{X}_t\right) \times p(\mathbf{X}_t|\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_0),$$

where $c$ is a normalizing constant. Yet another way to write this would be

$$p(\mathbf{X}_t|\mathbf{Y}_t) \propto p\left(\mathbf{Y}_t|\mathbf{X}_t\right) \times p(\mathbf{X}_t|\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_0),$$

where $\propto$ means "proportional to".

The equation above is what the particle filter depends on. Let us assume that we have $N$ samples of $\mathbf{X}_{t+1}$ from the distribution $p(\mathbf{X}_{t-1}|\mathbf{Y}_{t-1})$. The article I recommend denotes this sample as

$$\mathbf{x}_{t-1}^{1,*}, \mathbf{x}_{t-1}^{2,*}, \ldots, \mathbf{x}_{t-1}^{N,*} \tag{5}$$

(shorthand for this is $\left\{\mathbf{x}_{t-1}^{i,*}\right\}_{i=1}^{N}$). One thing to absolutely keep in mind is that the sample in (5) is the only information we have of the distribution $p(\mathbf{X}_{t-1}|\mathbf{Y}_{t-1})$. The nonlinear functions $g$ and $h$ make calculating the actual distribution of $p(\mathbf{X}_t|\mathbf{Y}_t)$ too mathematically challenging.

The question now is this: How is it possible to go from the sample in (5) (a sample from $p\left(\mathbf{X}_{t-1}|\mathbf{Y}_{t-1}\right)$ ) to a sample from $p\left(\mathbf{X}_t|\mathbf{Y}_t\right)$? The sample from $p\left(\mathbf{X}_t|\mathbf{Y}_t\right)$ of size $N$ would be denoted as

$$\mathbf{x}_t^{1,*}, \mathbf{x}_t^{2,*}, \ldots, \mathbf{x}_t^{N,*}. \tag{6}$$

Here's the answer to the question: for each value of $\mathbf{x}_{t-1}$ in (5), generate a predicted value of $\mathbf{x}_t$. All this requires is simulating a value of $\mathbf{V}_t$ and then applying Equation (3). The predicted value of $\mathbf{x}_t$ based on $\mathbf{x}_{t-1}^{j,*}$ is

$$\mathbf{x}_t^j = f(\mathbf{x}_{t-1}^{j,*}, \mathbf{V}_{t-1}).$$

This generates a sample of predicted values

$$\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^N. \tag{7}$$

Note that this value of $\mathbf{x}_t$ in (7) does not have an asterisk (since it is simply the prediction). With $\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^N$, and with the newly observed value of $\mathbf{Y}$ at time $t$ ($\mathbf{Y}_t$), it is possible to generate the sample in (6). For each value of $\mathbf{x}$ in (6), assign a weight $\tilde{w}_t^i$. This weight is calculated as

$$\tilde{w}_t^i = p(\mathbf{Y}_t | \mathbf{X}_t = \mathbf{x}_t^i).$$

Those values of $\mathbf{x}_t^i$ that lead to a large value of $\tilde{w}_t^i$ are more likely values of $\mathbf{x}_t$ given the observed value of $\mathbf{Y}$ at time $t$. Now normalize these weights so that

$$w_t^i = \tilde{w}_t^i / \sum_{i=1}^{N} \tilde{w}_t^i.$$

The sample in (6) is then just the sample in (7) resampled (with replacement) such that the probability of sampling $\mathbf{x}_t^i$ and putting it in (**??**) is $w_t^i$.