Lecture Notes: Recursive Bayesian Estimation

The Kalman filter is only intended for linear systems. The extended Kalman filter works on nonlinear systems. However, both filters assume that the state distribution, dynamic noise and observation noise are all Gaussian. Figure 1 illustrates an example.
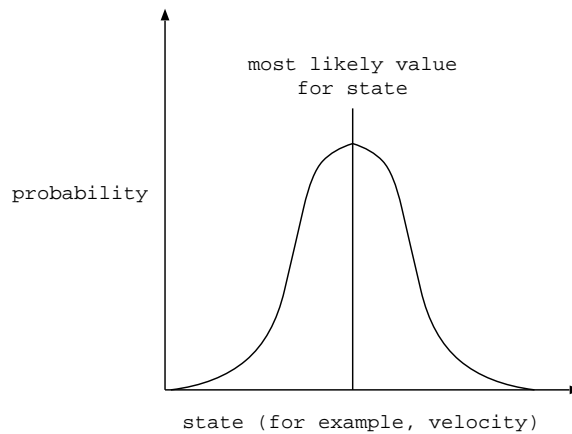

Figure 1: An example Gaussian distribution for a state variable.

Mathematically, this is modeled by the covariance matrices used in the filters. For example, in the 1D constant velocity model from previous lectures, the dynamic noise modeling acceleration can be written as:

$$\dot{x}_{t+1} = \dot{x}_t + a_t \tag{1}$$

The term a_t is a zero-mean Gaussian noise with variance σ_a^2. Another way to write this equation is

$$p(\dot{x}_{t+1}|\dot{x}_t) = \exp\left(\frac{-(\dot{x}_{t+1} - \dot{x}_t)^2}{\sigma_a^2}\right) \tag{2}$$

Similarly, we can take the observation equation

$$y_t = x_t + n_t \tag{3}$$

where n_t is a zero-mean Gaussian noise with variance σ_n^2, and rewrite it as

$$p(y_t|x_t) = \exp\left(\frac{-(y_t - x_t)^2}{\sigma_n^2}\right) \tag{4}$$
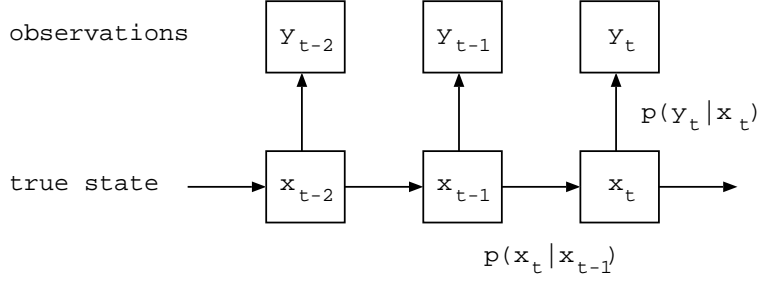
Figure 2: Illustration of a discrete process x and observations (y) of it.

However, equations 2 and 4 are only appropriate if the distributions are Gaussian. When the distributions are not Gaussian, the Kalman and extended Kalman filters do not apply. Instead, we need a more general theory that applies to generic distributions.

Consider the problem as illustrated in figure 2. At discrete time intervals t, the true state x_t is observed by measurements y_t. It then transitions to a next state at $t + 1$. In general form, we wish to determine

$$p(x_{0:t}|y_{0:t}) \tag{5}$$

which is the probability of all the transitions over all time given all the observations. However, realistically, we do not want to base our calculations on all the previous estimates of state and all the previous measurements. We want to set up a recursive relationship where we base our next estimate on the previous estimate and the latest measurement:

$$p(x_{0:t}|y_{0:t}) = f\left[p(x_{0:t-1}|y_{0:t-1}), y_t\right] \tag{6}$$

So, we set out to identify such a relationship.

Before doing so, we define the notation:

$$
\begin{align}
x_{0:t} &= x_0 \text{ AND } x_1 \text{ AND } ... x_t \tag{7}\\
&= x_0, x_1, ... x_t \tag{8}\\
&= x_0 \cap x_1 \cap ... x_t \tag{9}
\end{align}
$$

Also, we need to establish two identities used in proving the recursion. **Identity #1** is defined as

$$p(x, y) = p(x|y)p(y) \tag{10}$$

This is a basic relationship in probability. We use it to derive a second, somewhat less obvious, identity:

$$
\begin{align}
p(a, b, c) &= p(\{a, b\}, c) = p(\{a, b\}|c)p(c) \tag{11}\\
p(a, b, c) &= p(a, \{b, c\}) = p(a|\{b, c\})p(b, c) = p(a|b, c)p(b|c)p(c) \tag{12}
\end{align}
$$

Setting the results equal to each other, and eliminating the common term $p(c)$, we obtain

$$p(a, b|c) = p(a|b, c)p(b|c) \tag{13}$$

2

We will call this **identity #2**.

Finally, we make two assumptions. First, we assume that we are dealing with a **Markov process**, so that

$$p(x_t|x_{0:t-1}) = p(x_t|x_{t-1}) \tag{14}$$

This means that the next state depends only upon the current state, and not upon all the previous history of state. This can be imagined in considering a plane flying about. The next position and velocity of the plane depends upon its current position and velocity, but does not depend upon the whole history of the plane before it flew to its current position and velocity. The second assumption is one of **independent observations**:

$$p(y_t|x_t, \text{ANY}) = p(y_t|x_t) \tag{15}$$

This means that the current observation only depends upon the current state. It does not depend upon any previous states, nor does it depend upon any previous observations.

Now, we set out to establish the function desired in equation 6. From Bayes' rule, we know that $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$. Therefore we can write

$$p(x_{0:t}|y_{0:t}) = \frac{p(y_{0:t}|x_{0:t})p(x_{0:t})}{p(y_{0:t})} \tag{16}$$

The observation terms can be sub-grouped as

$$= \frac{p(y_t, y_{0:t-1}|x_{0:t})p(x_{0:t})}{p(y_t, y_{0:t-1})} \tag{17}$$

Applying identity #2 to the first term in the numerator yields

$$= \frac{p(y_t|y_{0:t-1}, x_{0:t})p(y_{0:t-1}|x_{0:t})p(x_{0:t})}{p(y_t, y_{0:t-1})} \tag{18}$$

Applying identity #1 to the denominator term yields

$$= \frac{p(y_t|y_{0:t-1}, x_{0:t})p(y_{0:t-1}|x_{0:t})p(x_{0:t})}{p(y_t|y_{0:t-1})p(y_{0:t-1})} \tag{19}$$

We now apply Bayes' rule to the middle numerator term, producing

$$= \frac{p(y_t|y_{0:t-1}, x_{0:t})p(x_{0:t}|y_{0:t-1})p(y_{0:t-1})p(x_{0:t})}{p(y_t|y_{0:t-1})p(y_{0:t-1})p(x_{0:t})} \tag{20}$$

Simplifying common terms in the numerator and denominator yields

$$= \frac{p(y_t|y_{0:t-1}, x_{0:t})p(x_{0:t}|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{21}$$

Applying the assumption of independent observations to the first numerator term produces

$$= \frac{p(y_t|x_t)p(x_{0:t}|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{22}$$

The states in the second numerator term can be sub-grouped as

$$= \frac{p(y_t|x_t)p(x_t, x_{0:t-1}|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{23}$$

Applying identity #2 to the second numerator term yields

$$= \frac{p(y_t|x_t)p(x_t|x_{0:t-1}, y_{0:t-1})p(x_{0:t-1}|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{24}$$

Applying the assumption of a Markov process to the middle numerator term produces

$$= \frac{p(y_t|x_t)p(x_t|x_{t-1})p(x_{0:t-1}|y_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{25}$$

Rearranging this last result, we obtain

$$p(x_{0:t}|y_{0:t}) = \frac{p(x_t|x_{t-1})p(y_t|x_t)}{p(y_t|y_{0:t-1})} p(x_{0:t-1}|y_{0:t-1}) \tag{26}$$

This is the solution for equation 6. It is called recursive Bayesian estimation. Note that is is applicable for ANY distribution, not just Gaussians.