This week we continue our discussion of state space models, focusing on the particle method approach for nonlinear models.\textsuperscript{1} Besides its practical application, this will give us some insight into the structure of state space models.

The nonlinear state space model has an evolution equation

\begin{equation}
    x_t = F_t(x_{t-1}, w_t).
\end{equation}

In general, $x_t$ could live in any kind of space, but for simplicity we take $x_t \in \mathbb{R}^n$. Here $w_t$ is a noise vector that could have any distribution, but we will take successive values $w_t, w_{t+1}, \ldots$ to be independent.

We don’t observe $x_t$ but only $y_t$, which is given in terms of $x_t$ by

\begin{equation}
    y_t = G_t(x_t, u_t)
\end{equation}

with a noise term $u_t$. $u_t$ is another noise term, with arbitrary distribution but serially independent; we take $u_t$ independent of $w_t$.

The filtering problem is to deduce information about $x_t$, given the series of observations to time $t$: $y_t, y_{t-1}, \ldots$, which we denote collectively by $y_{\leq t}$. The prediction problem would extrapolate information about $x_{t+1}, x_{t+2}, \ldots$ given information only through time $t$, and the smoothing problem would try to retroactively improve the estimates of $x_{t-1}, x_{t-2}, \ldots$ using information through $t$.

Since $x_t$ is a random variable, information about it means constructing some model for its entire distribution

\begin{equation}
    \pi_t(x_t) = \mathbb{P}(x_t | y_{\leq t})
\end{equation}

\textsuperscript{1}This material and the presentation is taken from René A. Carmona, \textit{Statistical Analysis of Financial Data in S-Plus}, Springer 2004, Section 7.6.
conditional on the observations through time $t$. We want to determine an update formula
\[ \pi_{t-1} \rightarrow \pi_t \]
incorporating the evolution equation (1) and the information (2).

In the linear case, it is reasonable to take all distributions to be Gaussian, or at least fully described by their mean and covariance structure. The linear Kalman filter that we discussed earlier is just an update formula for the best mean-square estimator $\hat{x}_t$ and the covariance of its estimator $\Sigma_{x,t} = \text{Cov}(x_t - \hat{x}_t)$. For general nonlinear models, mean and variance are not enough to describe the distribution $\pi_t$.

**Particle method**

The idea of a particle method is that we represent the density of interest by a collection of “particles” $x^i_t$ for $i = 1, \ldots, m$, which will be interpreted as independent samples from the distribution $\pi_t$. From such a sample we can compute most quantities of interest. We can reconstruct the density using a kernel method or simply point masses
\[ \pi_t(dx) \approx \frac{1}{m} \sum_{j=1}^{m} \delta_{x^j_t}(dx) \]
and we can directly estimate averages
\[ \mathbb{E}(f(x_t)) \approx \frac{1}{m} \sum_{j=1}^{m} f(x^j_t). \]

Particle methods have been widely used in fluid dynamics (“vortex methods”) as well as many other fields. If we assume that we have a way to determine the initial values $x^i_0$, then the problem is how to do the update step: determine the $\{x^i_t\}_{i=1}^{m}$ from the $\{x^i_{t-1}\}_{i=1}^{m}$. The novel feature of state space models, which does not appear in the physical simulation applications, is the incorporation of the observation variable $y_t$ along with the system dynamics.

We do this one step at a time. Let us suppose that $x^1_{t-1}, \ldots, x^m_{t-1}$ are independent samples from the distribution $\pi_{t-1}$, incorporating the information $y_{<t} = \{y_{t-1}, y_{t-2}, \ldots\}$ through time $t - 1$. We set
\[ z^i_t = F_t(x^i_{t-1}, w^i_t), \quad i = 1, \ldots, m, \]
with the \( w_i \) independent samples from the distribution of \( w_t \). The values \( z_i \) are our best estimates for the \( x_i \) using only the information \( y_{<t} \) through time \( t - 1 \). If we had no observations, then we would set \( x_i = z_i \), which would be a Monte Carlo simulation of the dynamics (1). With some sloppiness of notation, we write

\[
P(x_{t-1} = x_{t-1}^i \mid y_{<t}) = \frac{1}{m}
\]

and

\[
P(x_t = z_t^i \mid y_{<t}) = \frac{1}{m}.
\]

This means, not that the true value of \( x_t \) takes one of the \( m \) specific numbers \( x_i \), but rather that all the \( x_i \) are equally probable. Our goal is to determine the \( x_i \) so that

\[
P(x_t = x_t^i \mid y_{\leq t}) = \frac{1}{m}.
\]

(Updating \( z_t^i \) to \( x_t^i \) simply means evolving \( y_{<t} \) to \( y_{\leq t} \).)

We incorporate \( y_t \) using a maximum-likelihood approach. We denoting by \( Y_t \) the random variable whose observed value is \( y_t \) (and the corresponding \( X_t = x_t \) for symmetry), we define

\[
\alpha_i^t = P(Y_t = y_t \mid X_t = z_t^i), \quad i = 1, \ldots, m,
\]

which is the probability, per unit \( dy_t \), that we would have observed \( y_t \), conditional on a particular value of \( X_t \). To calculate the \( \alpha_i^t \), we assume that we can invert the observation equation (2) to write

\[
u_t = H_t(x_t, y_t).
\]

This is not too difficult in many cases of interest, for example, if the noise is additive \( y_t = G_t(x_t) + u_t \) or multiplicative \( y_t = G_t(x_t)u_t \). Then from \( P(u) \, du = P(y) \, dy \) we have \( P(y) = P(u) \, \partial u / \partial y \) or

\[
\alpha_i^t = \psi(u) \mid \text{det} \nabla_y H_t \mid
\]

where \( \psi_t(u) \) is the distribution of \( u_t \). In the above expression, \( \psi(u) \) is evaluated at \( u = H_t(z_t^i, y_t) \), and \( \nabla_y H_t \) is evaluated at \( (z_t^i, y_t) \). Of course if \( u_t \) and \( y_t \) are scalars (regardless of the dimension of \( x_t \)) then this is simply

\[
\alpha_i^t = \psi(H_t(z_t^i, y_t)) \left| \frac{\partial H_t}{\partial y}(z_t^i, y_t) \right|.
\]
Next we calculate the probability that \( X_t = z_t^i \) conditional on \( y_{<t} \), after which we can adjust the values \( z_t^i \) to get \( x_t^i \). Recalling that (loosely speaking) \( y_{\leq t} = y_{<t} \cup y_t \), it is straightforward to calculate

\[
\mathbb{P}(X_t = z_t^i \mid y_{\leq t}) = \frac{\mathbb{P}(X_t = z_t^i \text{ and } Y_t = y_t \mid y_{<t})}{\mathbb{P}(Y_t = y_t \mid y_{<t})} = \frac{\frac{\mathbb{P}(Y_t = y_t \mid X_t = z_t^i) \cdot \mathbb{P}(X_t = z_t^i \mid y_{<t})}{\sum_{j=1}^m \mathbb{P}(Y_t = y_t \mid X_t = z_t^j) \cdot \mathbb{P}(X_t = z_t^j \mid y_{<t})}}{\sum_j \alpha_t^j \cdot (1/m)} = \frac{\alpha_t^i \cdot (1/m)}{\sum \alpha_t^j}.
\]

In the first two lines, we have used standard conditional probability. In the third line, we have used the definition of \( \alpha_t^i \). We have also used the assumption that the \( x_t^{i-1} \) are equally probable samples from the distribution \( \pi_{t-1} \) (incorporating the information \( y_{<t} \)), and hence that the \( z_t^i \) are equally probable samples from the distribution of \( x_t \) without using the new information \( y_t \).

Now, if all the \( \alpha_t^i \) were equal, then the above probabilities would be equal and we would be done with this step: by setting \( x_t^i = z_t^i \) we would have an equally probable sample from the distribution of \( x_t \) conditional on \( y_t \). In general, the values of the \( \alpha_t^i \) tell us how to adjust the \( z_t^i \) to get the new particle positions \( x_t^i \): sample points \( z_t^i \) with large values of \( \alpha_t^i \) correspond to large likelihood of \( y_t \), and should have more weight in the final distribution.

The key idea is that we do not attempt to compute new particle positions. Rather, we resample the existing collection to get the desired probabilities. Specifically, for \( i = 1, \ldots, m \) we set

\[
x_t^i = \begin{cases} 
  z_t^1 & \text{with probability } \frac{\alpha_t^1}{\sum \alpha_t^j}, \\
  \vdots \\
  z_t^m & \text{with probability } \frac{\alpha_t^m}{\sum \alpha_t^j}.
\end{cases}
\]

Of course, the selection is to be made independently for each \( i \). In this way, the resulting points \( x_t^i \) will be equally probable samples from the distribution \( \pi_t \) using the observation \( y_t \).
It may seem that we are narrowing our collection of samples at each step. Even if the $\alpha_i^t$ were all equal, in the random resampling procedure we will likely set several $x_i^t$ equal to the same value $z_i^t$, and some $z_i^t$ will not survive into the collection $x_i^t$. (If you were clever, you might be able to figure out some quasi-MC procedure to preserve the above distribution while minimizing the uncertainty of the selection.) But note that on the next iteration, we will use independent samples from the noise $u_{t+1}$, and so values $x_i^t$ that are equal at time $t$ will no longer be equal at time $t+1$.

**Example: stochastic volatility**

We suppose that we have an asset price $S_t$ evolving according to the continuous-time process

$$ \frac{dS_t}{S_t} = \mu dt + \sigma_t dW_t $$

where $\sigma_t$ is some volatility process and $W_t$ is a Wiener process. In a discrete-time approximation we can write

$$ r_t = \frac{S_{t+\tau}}{S_t} - 1 = \mu \tau + \sigma_t \sqrt{\tau} \epsilon_t $$

where $r_t$ is the return across time step $\tau = \Delta t$ and $\epsilon_t \sim \mathcal{N}(0, 1)$.\(^2\)

This is the observation equation \(^2\) and our empirically observed data consists of the return series $r_1, r_2, \ldots$.

The volatility process could be whatever we like, even with hidden multi-dimensional components. But for simplicity, let us take a simple mean-reverting process

$$ d\sigma_t = -\lambda (\sigma_t - \bar{\sigma}) dt + \gamma d\tilde{W}_t $$

where $\bar{\sigma}$ is the long-term mean level, $\lambda$ is the rate of mean reversion, $\gamma$ is the volatility of volatility, and $\tilde{W}_t$ is another Wiener process, which we take to be independent of $W_t$. In a discrete-time approximation we may write

$$ \sigma_t = e^{-\lambda \tau} \sigma_{t-1} + \left(1 - e^{-\lambda \tau}\right) \bar{\sigma} + \gamma \sqrt{\frac{1}{2\lambda} \left(1 - e^{-2\lambda \tau}\right)} \tilde{\epsilon}_t $$

$$ \approx \sigma_{t-1} + \lambda (\bar{\sigma} - \sigma_{t-1}) + \gamma \sqrt{\tau} \tilde{\epsilon}_t \quad \text{for small } \tau, $$

\(^2\)I apologize that we are using $t$ both for continuous time and for the discrete index—hopefully in each case the meaning is clear from context.
where $\tilde{\epsilon}_t \sim \mathcal{N}(0, 1)$. We assume that we know the values of all the parameters $\overline{\sigma}$, $\lambda$, and $\gamma$. This is our evolution equation (1).

Our task is to take an observed historical series of returns, and from that generate a sequence of estimates of the entire distribution of $\sigma_t$ at each time, not just its estimated mean $\hat{\sigma}_t$ and variance. In doing this, we assume that we know not only the structure of the above model, but also all the coefficients $\mu$, $\overline{\sigma}$, $\lambda$, and $\gamma$ (the time step $\tau$ is ours to choose). The only problem is to determine the instantaneous value of the state variable $\sigma_t$.

The particle method above can be directly applied, with the hidden variable $x_t \rightarrow \sigma_t$, the observed variable $y_t \rightarrow r_t$, the evolution noise $w_t \rightarrow \tilde{\epsilon}_t$, and the observation noise $u_t \rightarrow \epsilon_t$. At each time step we maintain a collection $\{\tilde{\sigma}_t^i\}_{i=1}^m$, which represent an equi-probable sample from the distribution of $\sigma_t$ conditional on the observations through time $t$. To determine the new values $\{\tilde{\sigma}_t^i\}$ from the previous values $\{\tilde{\sigma}_{t-1}^i\}$, we first determine unconditional values $\tilde{\sigma}_t^i$ using the evolution equation above, with $m$ independent samples $\{\tilde{\epsilon}_t^i\}$.

Then, to incorporate the new observed return $r_t$, we observe that the inverse of the observation equation $r_t = G_t(\sigma_t, \epsilon_t)$ is

$$
\epsilon_t = H_t(\sigma_t, r_t) = \frac{r_t - \mu \tau}{\sigma_t \sqrt{\tau}}
$$

whose derivative with respect to the observation variable is $\partial H_t / \partial r_t = 1 / \sigma_t \sqrt{\tau}$, and so the likelihood coefficients evaluated at the unconditional values $\tilde{\sigma}_t^i$ are

$$
\alpha_t^i = \Phi \left( \frac{r_t - \mu \tau}{\tilde{\sigma}_t^i \sqrt{\tau}} \right) \frac{1}{\tilde{\sigma}_t^i \sqrt{\tau}}
$$

where $\Phi(z) = (1 / \sqrt{2\pi}) \exp(-z^2/2)$ is the standard normal density. We then determine the conditional particle values $\{\tilde{\sigma}_t^i\}$ as random samples from the unconditional values $\{\tilde{\sigma}_t^i\}$ by sampling with probabilities proportional to the $\alpha_t^i$.

Of course, in practice part of the problem will be to estimate values for all the parameters, especially $\lambda$ and $\gamma$. One technique that is commonly used is to include the parameters in the state equation, and determine their updates along with the state variable. If we assume that $\mu$ and $\overline{\sigma}$ can be estimated in advance (say by taking the mean and standard deviation of the entire return sequence, then we
would use the system of update equations

\[
\begin{align*}
\sigma_t &= e^{-\lambda \tau} \sigma_{t-1} + \left( 1 - e^{-\lambda \tau} \right) \sigma + \gamma \sqrt{\frac{1}{2\lambda} \left( 1 - e^{-2\lambda \tau} \right)} \tilde{\epsilon}_t \\
\lambda_t &= \lambda_{t-1} \\
\gamma_t &= \gamma_{t-1}
\end{align*}
\]

(It might be simpler and more numerically stable to use the small-\(\tau\) approximation above, so that \(\lambda\) appears linearly.) That is, the “true” values of the coefficients do not change from one step to another, only our estimates of their values. We can simply carry along these values as components of the state variable. The likelihood function does not get any more complicated since it includes only the observation variable. In doing this, we will get not only estimated values for the parameters but also approximations to their distributions.