State-Space Models

April 9, 2019

State space models give us yet another way of writing out a model for our time series. In this section, the notation we use will differ slightly. The observed time series that we intend to analyze has values $y_1, \ldots, y_n$. Let's work through a small example! For a univariate time series of length n we assume

$$\begin{aligned} y_t &= a x_t + v_t && \text{Observation Equation} \\ x_t &= \phi x_{t-1} + w_t && \text{State Equation,} \end{aligned}$$

where $v_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_v^2)$, $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$ and $x_0 = \mu$. The state equation is just an $\mathbf{AR}(1)$ model, but values of the $\mathbf{AR}(1)$ process x_t are not observed. Rather, a linear transformation y_t of the $\mathbf{AR}(1)$ states x_t are observed with some additional normal noise. To emphasize, we observe values of y_t, but we do *not* observe values of x_t. For now, let's assume that we know σ_v^2, σ_w^2, and μ. Having assumed a state-space model, we might be interested in:

- **Predicting**, i.e. estimating future values of x_t given past values of $y_{t-1}, \ldots, y_1$
  - For example, computing $\mathbb{E}[x_2|y_1]$ and $\mathbb{V}[x_2|y_1]$
- **Filtering**, i.e. estimating future values of x_t given current and past values $y_t, \ldots, y_1$
  - For example, computing $\mathbb{E}[x_2|y_1, y_2]$ and $\mathbb{V}[x_2|y_1, y_2]$

- **Smoothing**, i.e. estimating a past value of x_{t-1} given all observed past, current, and future values $y_n, \dots, y_1$
  - For example, computing $\mathbb{E}[x_1|y_1, y_2]$ and $\mathbb{V}[x_1|y_1, y_2]$

As usual, we might also be interested in forecasting future values y_{n+k} given $y_1, \dots, y_n$, which will require **prediction** of the states because we will need to compute $\mathbb{E}[x_{n+k}|y_1, \dots, y_n]$. Even if we do know the values of these parameters, how do we go about predicting, filtering, and smoothing?

The first step is to realize that we can write out the joint probability distribution of $\boldsymbol{y}$ and $\boldsymbol{x}$. Because we have assumed that the errors $\boldsymbol{v}$ and $\boldsymbol{w}$ and the initial value x_0 is fixed, we know that the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{x}$ is given by:

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbb{E}[\boldsymbol{y}] \\ \mathbb{E}[\boldsymbol{x}] \end{pmatrix}, \begin{pmatrix} \mathbb{V}[\boldsymbol{y}] & \mathrm{Cov}[\boldsymbol{y}, \boldsymbol{x}] \\ \mathrm{Cov}[\boldsymbol{y}, \boldsymbol{x}] & \mathbb{V}[\boldsymbol{x}] \end{pmatrix}\right). \tag{1}$$

Why is this useful? Well, predicting, filtering, smoothing, and forecasting all correspond to different conditional mean and variance estimation problems, and when we have a normal joint distribution and know its parameters, it's easy to figure out any conditional distribution we want! We can see this most easily in the case of smoothing, which corresponds to the problem of estimating the conditional means $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]$ and variances $\mathbb{V}[\boldsymbol{x}|\boldsymbol{y}]$. Before we go further, a very useful property of the multivariate normal distribution is that:

If $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}, \begin{pmatrix} \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{D}' & \boldsymbol{E} \end{pmatrix}\right)$, then

- The marginal distributions of $\boldsymbol{u}$ and $\boldsymbol{v}$ are given by

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{a}, \boldsymbol{C}) \text{ and } \boldsymbol{v} \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{E})$$

– The conditional distributions of $\boldsymbol{u}|\boldsymbol{v}$ and $\boldsymbol{v}|\boldsymbol{u}$

$$\boldsymbol{u}|\boldsymbol{v} \sim \mathcal{N}\left(\boldsymbol{a} + \boldsymbol{D}'\boldsymbol{E}^{-1}\left(\boldsymbol{v} - \boldsymbol{b}\right), \boldsymbol{C} - \boldsymbol{D}'\boldsymbol{E}^{-1}\boldsymbol{D}\right)$$
$$\boldsymbol{v}|\boldsymbol{u} \sim \mathcal{N}\left(\boldsymbol{b} + \boldsymbol{D}\boldsymbol{C}^{-1}\left(\boldsymbol{u} - \boldsymbol{a}\right), \boldsymbol{A} - \boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{D}'\right)$$

We can use these facts to find the conditional distributions of the states $\boldsymbol{x}$ given the data $\boldsymbol{y}$ and accordingly, the smoothed estimates of the states, but first we'll simplify the joint distribution a bit. We can simplify $\mathbb{E}\left[\boldsymbol{y}\right]$, $\mathbb{V}\left[\boldsymbol{y}\right]$, and $\operatorname{Cov}\left[\boldsymbol{x}, \boldsymbol{y}\right]$.

- It's pretty straightforward to see that $\mathbb{E}\left[\boldsymbol{y}\right] = a\mathbb{E}\left[\boldsymbol{x}\right]$
- A bit of algebra lets us rewrite $\mathbb{V}\left[\boldsymbol{y}\right]$

$$\begin{aligned}
\mathbb{V}\left[\boldsymbol{y}\right] &= \mathbb{E}\left[\left(\boldsymbol{y} - \mathbb{E}\left[\boldsymbol{y}\right]\right)\left(\boldsymbol{y} - \mathbb{E}\left[\boldsymbol{y}\right]\right)'\right] \\
&= \mathbb{E}\left[\left(a\left(\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x}\right]\right) + \boldsymbol{v}\right)\left(\left(a\left(\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x}\right]\right) + \boldsymbol{v}\right)'\right] \\
&= a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n.
\end{aligned}$$

- A bit more algebra lets us rewrite $\operatorname{Cov}\left[\boldsymbol{x}, \boldsymbol{y}\right]$

$$\begin{aligned}
\operatorname{Cov}\left[\boldsymbol{x}, \boldsymbol{y}\right] &= \mathbb{E}\left[\left(\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x}\right]\right)\left(\boldsymbol{y} - \mathbb{E}\left[\boldsymbol{y}\right]\right)'\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{x} - \mathbb{E}\left[\boldsymbol{x}\right]\right)\left(a\boldsymbol{x} - a\mathbb{E}\left[\boldsymbol{x}\right] + \boldsymbol{w}\right)'\right] \\
&= a\mathbb{V}\left[\boldsymbol{x}\right]
\end{aligned}$$

Plugging these expressions into 2 yields a nicely structured normal distribution,

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} a\mathbb{E}\left[\boldsymbol{x}\right] \\ \mathbb{E}\left[\boldsymbol{x}\right] \end{pmatrix}, \begin{pmatrix} a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n & a\mathbb{V}\left[\boldsymbol{x}\right] \\ a\mathbb{V}\left[\boldsymbol{x}\right] & \mathbb{V}\left[\boldsymbol{x}\right] \end{pmatrix}\right). \tag{2}$$

Using our normal distribution facts and (2), we can return to the smoothing problem

$$\begin{aligned}
\boldsymbol{x}|\boldsymbol{y} \sim \mathcal{N}\Big(&\mathbb{E}\left[\boldsymbol{x}\right] + a\mathbb{V}\left[\boldsymbol{x}\right]\left(a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n\right)^{-1}\left(\boldsymbol{y} - a\mathbb{E}\left[\boldsymbol{x}\right]\right), \\
&\mathbb{V}\left[\boldsymbol{x}\right] - a^2\mathbb{V}\left[\boldsymbol{x}\right]\left(a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n\right)^{-1}\mathbb{V}\left[\boldsymbol{x}\right]\Big).
\end{aligned}$$

The smoothed values of the states are given by the conditional mean

$$\mathbb{E}\left[\boldsymbol{x}\right] + a\mathbb{V}\left[\boldsymbol{x}\right]\left(a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n\right)^{-1}\left(\boldsymbol{y} - a\mathbb{E}\left[\boldsymbol{x}\right]\right),$$

and our uncertainty about them is quantified by the conditional variance,

$$\mathbb{V}\left[\boldsymbol{x}\right] - a^2\mathbb{V}\left[\boldsymbol{x}\right]\left(a^2\mathbb{V}\left[\boldsymbol{x}\right] + \sigma_v^2\boldsymbol{I}_n\right)^{-1}\mathbb{V}\left[\boldsymbol{x}\right].$$

Predicting and filtering can be done similarly by identifying the marginal joint distribution of a state and the corresponding observed values, and computing the corresponding conditional distribution.

- We can obtain predicted means and variances of the states x_t given the past observed values of the time series $y_1, \dots, y_{t-1}$ by letting $u = x_t$ and $\boldsymbol{v} = (y_1, \dots, y_{t-1})$, using our marginal and conditional normal distribution facts;
- We can obtain filtered means and variances of the states x_t given the past and current observed values of the time series $y_1, \dots, y_t$ using our marginal and conditional normal distribution facts, letting $u = x_t$ and $\boldsymbol{v} = (y_1, \dots, y_t)$.

Forecasts follow straightforwardly. We have assumed $y_t = ax_t + v_t$, so it is natural to define a forecasted value of y_{n+k} as

$$\mathbb{E}\left[y_{n+k}|y_1, \dots, y_n\right] = a\mathbb{E}\left[x_{n+k}|y_1, \dots, y_n\right],$$

with variance

$$\mathbb{V}\left[y_{n+k}|y_1, \dots, y_n\right] = a^2\mathbb{V}\left[x_{n+k}|y_1, \dots, y_n\right] + \sigma_w^2.$$

In general, even though we *can* figure out any conditional distribution we want, it would be very computationally burdensome to figure out each conditional expectation one by one. Fortunately, the **Kalman filter** allows for fast, recursive estimation of:

- Predicted means and variances

- Filtered means and variances
- Smoothed means and variances
- Forecast means and variances.

We won't get into the details here, but the existence of a fast algorithm to compute all of these quantities is essential in practice.