Kalman Filter and Kalman Smoother

Taisuke Nakata and Christopher Tonetti

New York University

October 25, 2010

1 Linear Gaussian State Space

Let the canonical form for a Linear Gaussian State Space (LGSS) be

$$\begin{aligned} S_t &= A_t S_{t-1} + B_t v_t && (1) \\ Y_t &= C_t z_t + D_t S_t + w_t && (2) \end{aligned}$$

where

$$\begin{aligned} \begin{pmatrix} v_t \\ w_t \end{pmatrix} &\sim iid \quad N\left(0, \begin{bmatrix} Q_t & 0 \\ 0 & R_t \end{bmatrix}\right) \\ S_0 &\sim N(S_{0|0}, P_{0|0}) \end{aligned}$$

The first equation is a state equation (or, transition equation), and the second equation is an observation equation (or, measurement equation).

S_t is a vector of latent state variables, and Y_t is what we observe. v_t is a vector of innovations to latent state variables, and w_t is a vector of measurement error. z_t is a vector of observed exogenous variables.

S_t is n_s-by-1, Y_t is n_y-by-1, and z_t is n_z-by-1. It follows that A_t is n_s-by-n_s, B is n_s-by-n_v, C_t is n_y-by-n_z, D_t is n_y-by-n_s, Q_t is n_v-by-n_v, and R_t is n_y-by-n_y.

2 Kalman Filter

Because we can write down the joint likelihood of the data as a product of conditional densities, we proceed to develop an algorithm to iteratively provide the conditional densities: $f(Y_t|Y^{t-1})$.

2.1 Useful Facts

1. Since S_0 is normal, and since $\{S_t\}$ and $\{Y_t\}$ are linear combinations of normal errors, the vector $(S_1, \cdots, S_T, Y_1, \cdots, Y_T)$ is normally distributed.

2. In general, if

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

then

$$x_1|x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

2.2 Deriving the Kalman Filter

Thus the following objects of interest are normal and can be characterized by their mean and variance. Let the following notation hold:

$$\begin{aligned} S_t|Y^{t-1} &\sim N(S_{t|t-1}, P_{t|t-1}) \\ S_t|Y^t &\sim N(S_{t|t}, P_{t|t}) \\ Y_t|Y^{t-1} &\sim N(Y_{t|t-1}, F_t) \end{aligned}$$

Then, from Equation (1):

$$S_{t|t-1} = A_t S_{t-1|t-1} \quad (3)$$
$$P_{t|t-1} = \mathbb{E}((S_t - S_{t|t-1})(S_t - S_{t|t-1})'|Y^{t-1}) = A_t P_{t-1|t-1} A_t' + B_t Q_t B_t' \quad (4)$$

From Equation (2):

$$Y_{t|t-1} = C_t Z_t + D_t S_{t|t-1} \quad (5)$$
$$F_t = \mathbb{E}((Y_t - Y_{t|t-1})(Y_t - Y_{t|t-1})'|Y^{t-1}) = D_t P_{t|t-1} D_t' + R_t \quad (6)$$

Using the general fact about normal distributions[1]:

$$\begin{pmatrix} S_t \\ Y_t \end{pmatrix} |Y^{t-1} \sim \quad N\left(\begin{bmatrix} S_{t|t-1} \\ Y_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1} D_t' \\ D_t P_{t|t-1} & F_t \end{bmatrix} \right)$$

Thus,

$$S_t|Y^t = S_t|Y_t, Y^{t-1} \sim N(S_{t|t}, P_{t|t}) \quad (7)$$
$$\sim N(s_{t|t-1} + P_{t|t-1} D_t' F_t^{-1}(Y_t - Y_{t|t-1}), P_{t|t-1} - P_{t|t-1} D_t' F_t^{-1} D_t P_{t|t-1}) \quad (8)$$

Using the initial condtions iteratively applying the updating equations derived above, we can construct the sequence of the conditional distributions of the states and observations, and thus the likelihood.

3 Kalman Smoother

The Kalman filter uses past and current observations to predict the current state, (i.e., $\{S_t|Y^t\} \quad \forall t$). While this is sufficient for computing the likelihood of the system, this is suboptimal for estimating the sequence of states. The econometrician should use all available data to estimate the sequence of states (i.e., $\{S_t|Y^T\} \quad \forall t$). The Kalman smoother produces these distributions.

Before calculating the Kalman smoother it is useful to note[2]

$$\begin{pmatrix} S_t \\ S_{t+1} \end{pmatrix} |Y^t \sim \quad N\left(\begin{bmatrix} S_{t|t} \\ S_{t+1|t} \end{bmatrix}, \begin{bmatrix} P_{t|t} & P_{t|t} A_{t+1}' \\ A_{t+1} P_{t|t} & P_{t+1|t} \end{bmatrix} \right)$$

Let $J_t := P_{t|t} A_{t+1}' P_{t+1|t}^{-1}$. Then, by the general fact about Normal distributions,

$$\mathbb{E}\left[S_t|S_{t+1}, Y^t\right] = S_{t|t} + J_t(S_{t+1} - S_{t+1|t}) \quad (9)$$
$$Var\left[S_t|S_{t+1}, Y^t\right] = P_{t|t} - P_{t|t} A_{t+1}' P_{t+1|t}^{-1} A_{t+1} P_{t|t} \quad (10)$$

We are now ready to derive the Kalman smoother:

$$\mathbb{E}[S_t|Y^T] = \mathbb{E}\left[\mathbb{E}\left[S_t|S_{t+1}, Y^T\right]|Y^T\right]; \qquad \text{(by Law of Iterated Expectations)} \quad (11)$$
$$= \mathbb{E}\left[\mathbb{E}\left[S_t|S_{t+1}, Y^t\right]|Y^T\right] \quad (12)$$
$$= \mathbb{E}\left[S_{t|t} + J_t(S_{t+1} - S_{t+1|t})|Y^T\right]; \qquad \text{(by Equation (9))} \quad (13)$$
$$= S_{t|t} + J_t(S_{t+1|T} - S_{t+1|t}) \quad (14)$$

We can go from Equation (13) to Equation (14) by realizing $S_{t|t}$, J_t, and $S_{t+1|t}$ are functions of Y^t. That is, they are all known exactly as output from the Kalman filter, which condtions only on Y^t. Thus, conditioning on any set that at least includes Y^t cannot change $S_{t|t}$, J_t, or $S_{t+1|t}$ (as they are known exactly).

[1]See A.1 for a derivation of the covariance.
[2]See A.2 for a derivation of the covariance.

We can go from Equation (11) to Equation (12) by recognizing the following fact:

$$\begin{aligned}
\text{Let} \quad Z &= Y + \epsilon \\
\mathbb{E}[X|\epsilon] &= \mathbb{E}[X] \\
\text{then,} & \\
\mathbb{E}[X|Y,Z] &= \mathbb{E}[X|Y,\epsilon] \quad \text{(because given, } Y, Z, \epsilon \text{ is known)} \\
&= \mathbb{E}[X|Y] \quad \text{(by independence of } X \text{ and } \epsilon)
\end{aligned}$$

Applying this fact to our case notice that $Y^T_{t+1} = g(S_{t+1}, \{w_s, v_s\}^T_{s=T+1})$ for some g.

$$\begin{aligned}
\mathbb{E}\left[S_t|S_{t+1}, Y^T\right] &= \mathbb{E}\left[S_t|S_{t+1}, Y^t, Y^T_{t+1}\right] \\
&= \mathbb{E}\left[S_t|S_{t+1}, Y^t, g(S_{t+1}, \{w_s, v_s\}^T_{s=t+1})\right] \\
&= \mathbb{E}\left[S_t|S_{t+1}, Y^t\right]; \qquad \text{(by independence of } S_t \text{ and } \{w_s, v_s\}^T_{s=t+1} \forall t)
\end{aligned}$$

3.1 Kalman Smoother: Time 0

The Kalman smoother formula can be used to derive $\mathbb{E}\left[S_0|S_1, Y^T\right]$. Note: $Y^0 = \emptyset$.

$$\begin{pmatrix} S_0 \\ S_1 \end{pmatrix} |Y^0 \quad \sim \quad N\left(\begin{bmatrix} S_{0|0} \\ S_{1|0} \end{bmatrix}, \begin{bmatrix} P_{0|0} & P_{0|0}A_1' \\ A_1 P_{0|0} & P_{1|0} \end{bmatrix}\right)$$

Then,

$$\begin{aligned}
\mathbb{E}\left[S_0|Y^T\right] &= \mathbb{E}\left[\mathbb{E}\left[S_0|S_1, Y^T\right]|Y^T\right] \\
&= \mathbb{E}\left[S_{0|0} + J_0(S_1 - S_{1|0})|Y^T\right] \\
&= S_{0|0} + J_0(S_{1|T} - S_{1|0}) \\
&= S_{0|0} + P_{0|0}A_1' P_{1|0}^{-1}(S_{1|T} - S_{1|0})
\end{aligned}$$

Where we use the fact that $S_{0|0}$, J_0, and $S_{1|0}$ are parameters or functions of paramters ($S_{1|0} = A_1 S_{0|0}, P_{1|0} = A_1 P_{0|0} A_1' + B_1 Q_1 B_1'$).

A Deriving Covariances

A.1 $cov(Y_t, S_t|Y^{t-1})$

$$\begin{aligned}
cov(Y_t, S_t|Y^{t-1}) &= \mathbb{E}[(Y_t - \mathbb{E}[(Y_t|Y^{t-1})])(S_t - \mathbb{E}[(S_t|Y^{t-1})])|Y^{t-1}] \\
&= \mathbb{E}[(Y_t - Y_{t|t-1})(S_t - S_{t|t-1})|Y^{t-1}] \\
&= \mathbb{E}[(D_tS_t + w_t - Y_{t|t-1})(S_t - S_{t|t-1})|Y^{t-1}] \\
&= \mathbb{E}[D_tS_tS_t' + w_tS_t - Y_{t|t-1}S_t - D_tS_tS_{t|t-1} - w_tS_{t|t-1} + Y_{t|t-1}S_{t|t-1}|Y^{t-1}] \\
&= D_t\mathbb{E}[S_tS_t'|Y^{t-1}] - D_t\mathbb{E}[S_t|Y^{t-1}]S_{t|t-1} - Y_{t|t-1}\mathbb{E}[S_t|Y^{t-1}] + Y_{t|t-1}\mathbb{E}[S_t|Y^{t-1}] \\
&= D_t\left[\mathbb{E}[S_tS_t'|Y^{t-1}] - (\mathbb{E}[S_t|Y^{t-1}])^2\right] \\
&= D_t var(S_t|Y^{t-1}) \\
&= D_tP_{t|t-1}
\end{aligned}$$

A.2 $cov(S_t, S_{t+1}|Y^t)$

$$\begin{aligned}
cov(S_t, S_{t+1}|Y^t) &= cov(S_t, A_tS_t + B_tv_t|Y^t) \\
&= cov(S_t, A_tS_t|Y^t) + cov(S_t, B_tv_t|Y^t) \\
&= A_t cov(S_t, S_t|Y^t); \qquad \text{(by independence)} \\
&= A_tP_{t|t}
\end{aligned}$$