Expectation Maximization for State Space Models with Missing Data

Adam Coogan

September 29, 2017

1 State Space Models

A general state space model (SSM) has the form

$$\vec{x}_t = A_t \vec{x}_{t-1} + B_t \vec{u}_t + \vec{\varepsilon}_t \tag{1}$$
$$\vec{y}_t = C_t \vec{x}_t + D_t \vec{v}_t + \vec{\delta}_t, \tag{2}$$

where $\vec{u}_t$ and $\vec{v}_t$ are the state transition and observation controls and $\vec{\varepsilon}_t \sim N(0, Q_t)$ and $\vec{\delta}_t \sim N(0, R_t)$. The dimensions of the vectors $\vec{x}_t$, $\vec{y}_t$, $\vec{u}_t$, $\vec{v}_t$ are n_{LF}, N, L and M. T denotes the number of observations and $\mathcal{D}$ the set of obserations.

In our scenario, $N > n_{\mathrm{LF}}$ (though the equations below apply regardless of this condition), and the hidden state $\vec{x}_t$ does not have an obvious physical interpretation. This means we need to learn the parameters of the model $\theta = A, B, C, D, Q, R$. To improve numerical stability, I set $Q = I$ and take R to be diagonal. (TODO: would also be good to set largest eigenvalue of A to 1.)

2 Expectation Maximization for SSMs

To estimate the model parameters and unobserved states, we alternate between computing the expectation value of the complete-data log likelihood

$$\mathcal{Q}(\theta^{(j)}|\theta^{(j-1)}) = \mathbb{E}[\log p(\vec{x}_{1:T}, \vec{y}_{1:T})|\mathcal{D}, \theta^{(j-1)}] \tag{3}$$

and maximizing it to find a new estimate for the parameters

$$\theta^{(j)} = \arg\max_{\theta} \mathcal{Q}(\theta|\theta^{(j-1)}). \tag{4}$$

Note the notation $\vec{a}_{1:n} = \{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$. I'll use $\mathbb{E}_*[\,\cdot\,] = \mathbb{E}_*[\,\cdot\,|\mathcal{D}, \theta^{(j-1)}]$ below for convenience.

2.1 E-step

Note that $\mathbb{E}_*[\,\cdot\,] = \mathbb{E}_*[\mathbb{E}_*[\,\cdot\,|\vec{x}_{1:T}]]$. We can compute $\hat{x}_t \equiv \mathbb{E}_*[\vec{x}_t]$ and $P_t \equiv \Sigma_{t|T} - \hat{x}\hat{x}^T \equiv \mathbb{E}_*[\vec{x}_t \vec{x}_t^T]$ using the Kalman smoother (which I won't review here) with a key modification[1]. The procedure proceeds as normal

[1] see Shumway and Stoffer's 2017 book on time series

with the replacements

$$y_{ti} = \begin{cases} y_{ti} & \text{if observed} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$C_{ij}(D_{ij}) = \begin{cases} C_{ij}(D_{ij}) & y_{ti} \text{ observed} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$R_{ij} = \begin{cases} R_{ij} & y_{ti},\, y_{tj} \text{ both observed} \\ 1 & y_{ti},\, y_{tj} \text{ both unobserved} \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

With the smoothed values and covariances in hand, we can now compute the expected values involving the missing components of $\vec{y}_t$:

$$\mathbb{E}_*[y_{ti}] = \begin{cases} y_{ti} & \text{if observed} \\ [C^{(j-1)}\hat{x}_t + D^{(j-1)}\vec{v}_t]_i & \text{otherwise} \end{cases} \tag{8}$$

$$\mathbb{E}_*[y_{ti}x_{tj}] = \begin{cases} y_{ti}\hat{x}_{tj} & \text{if } y_{ti} \text{ observed} \\ [C^{(j-1)}P_t]_{ij} + [D^{(j-1)}\vec{v}_t]_i\hat{x}_{tj} & \text{otherwise} \end{cases} \tag{9}$$

$$\mathbb{E}_*[y_{ti}] = \begin{cases} (y_{ti})^2 & \text{if observed} \\ [C^{(j-1)}\hat{x} + D^{(j-1)}\vec{v}_t]_i^2 + [C^{(j-1)}\Sigma_{t|T}{C^{(j-1)}}^T]_{ii}^2 + R_{ii}^{(j-1)} & \text{otherwise} \end{cases}. \tag{10}$$

2.2 M-step

A and B must be solved for simultaneously since they do not appear in separate terms in the likelihood:

$$\begin{pmatrix} A^{(j)} & B^{(j)} \end{pmatrix} M_{AB} = N_{AB}, \tag{11}$$

$$M_{AB} = \begin{pmatrix} \mathbb{I}_{n_{\mathrm{LF}}\times n_{\mathrm{LF}}} & \left[\sum_{t=2}^T \hat{x}_{t-1}\vec{u}_t^T\right]\left[\sum_{t=2}^T \vec{u}_t\vec{u}_t^T\right]^{-1} \\ \left[\sum_{t=2}^T \vec{u}_t\hat{x}_{t-1}^T\right]\left[\sum_{t=1}^{T-1} P_t\right]^{-1} & \mathbb{I}_{L\times L} \end{pmatrix} \tag{12}$$

$$N_{AB} = \begin{pmatrix} \left[\sum_{t=2}^T P_{t,t-1}\right]\left[\sum_{t=1}^{T-1} P_t\right]^{-1} & \left[\sum_{t=2}^T \hat{x}_t\vec{u}_t^T\right]\left[\sum_{t=2}^T \vec{u}_t\vec{u}_t^T\right]^{-1} \end{pmatrix}. \tag{13}$$

Similarly, C and D are obtained by solving

$$\begin{pmatrix} C^{(j)} & D^{(j)} \end{pmatrix} M_{CD} = N_{CD}, \tag{14}$$

$$M_{CD} = \begin{pmatrix} \mathbb{I}_{n_{\mathrm{LF}}\times n_{\mathrm{LF}}} & \left[\sum_{t=1}^T \hat{x}_t\vec{v}_t^T\right]\left[\sum_{t=1}^T \vec{v}_t\vec{v}_t^T\right]^{-1} \\ \left[\sum_{t=1}^T \vec{v}_t\hat{x}_t^T\right]\left[\sum_{t=1}^{T-1} P_t\right]^{-1} & \mathbb{I}_{M\times M} \end{pmatrix} \tag{15}$$

$$N_{CD} = \begin{pmatrix} \left[\sum_{t=1}^T \mathbb{E}_*[\vec{y}_t\vec{x}_t^T]\right]\left[\sum_{t=1}^T P_t\right]^{-1} & \left[\sum_{t=1}^T \mathbb{E}_*[\vec{y}_t]\vec{v}_t^T\right]\left[\sum_{t=1}^T \vec{v}_t\vec{v}_t^T\right]^{-1} \end{pmatrix}. \tag{16}$$

Maximizing $\mathcal{Q}$ with respect to R gives

$$R_{ii}^{(j)} = \frac{1}{T}\sum_{t=1}^T \Big\{\mathbb{E}_*[(y_{ti})^2] + [C^{(j)}P_t{C^{(j)}}^T]_{ii} + [D^{(j)}\vec{v}_t]_i^2 \tag{17}$$

$$-2(\mathbb{E}_*[\vec{y}_t\vec{x}_t^T]{C^{(j)}}^T)_{ii} - 2[D^{(j)}\vec{v}_t]_i\mathbb{E}_*[y_{ti}] + 2[C^{(j)}\hat{x}_t]_i[D^{(j)}\vec{v}_t]_i\Big\}. \tag{18}$$

And finally, as usual we have

$$\pi_1^{(j)} = \hat{x}_1 \tag{19}$$

$$\Sigma_1^{(j)} = \Sigma_{1|T}. \tag{20}$$