Lecture 33

E-M algorithm for state-space models

Arnab Hazra

DLM with covariates (recap)

- In this case, we suppose we have an $r \times 1$ vector of inputs $\boldsymbol{u}_t$, and write the model as
$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \gamma \boldsymbol{u}_t + \boldsymbol{W}_t$$
$$\boldsymbol{Y}_t = \boldsymbol{A}_t \boldsymbol{X}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{V}_t$$
- Here γ is $p \times r$ and Γ is $q \times r$; either of these matrices may be the zero matrix.
- In the DLM, we assume the process starts with a normal vector $\boldsymbol{X}_0$, such that $\boldsymbol{X}_0 \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0)$.
- Here $\boldsymbol{W}_t \overset{IID}{\sim} \mathcal{N}_p(\mathbf{0}, \boldsymbol{Q})$ and the additive observation noise is $\boldsymbol{V}_t \overset{IID}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{R})$.

EM algorithm

- A conceptually simpler estimation procedure based on the EM algorithm is also possible.
- The basic idea is that if we could observe the states, $\mathcal{X}_{0:T} = \{\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_T\}$, in addition to the observations $\mathcal{Y}_{1:T} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T\}$, then we would consider $\{\mathcal{X}_{0:T}, \mathcal{Y}_{1:T}\}$ as the complete data.
- For the sake of brevity, we ignore the covariates, i.e., $\gamma \boldsymbol{u}_t$ and $\Gamma \boldsymbol{u}_t$ are ignored.
- The complete data have joint density

$$p_{\boldsymbol{\Theta}}(\mathcal{X}_{0:T}, \mathcal{Y}_{1:T}) = p_{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0}(\boldsymbol{X}_0) \prod_{t=1}^{T} p_{\boldsymbol{\Phi}, \boldsymbol{Q}}(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}) \prod_{t=1}^{T} p_{\boldsymbol{R}}(\boldsymbol{Y}_t | \boldsymbol{X}_t).$$

- The EM algorithm gives us an iterative method for finding the MLEs of $\boldsymbol{\Theta}$ based on the incomplete data $\mathcal{Y}_{1:T}$ by successively maximizing the conditional expectation of the complete data likelihood.

E-step

- Here we have

$$
\begin{aligned}
-2\log L_{\mathcal{X},\mathcal{Y}}(\boldsymbol{\Theta}) &= \log|\boldsymbol{\Sigma}_0| + (\boldsymbol{X}_0 - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{X}_0 - \boldsymbol{\mu}_0) \\
&+ T\log|\boldsymbol{Q}| + \sum_{t=1}^{T}(\boldsymbol{X}_t - \boldsymbol{\Phi}\boldsymbol{X}_{t-1})'\boldsymbol{Q}^{-1}(\boldsymbol{X}_t - \boldsymbol{\Phi}\boldsymbol{X}_{t-1}) \\
&+ T\log|\boldsymbol{R}| + \sum_{t=1}^{T}(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t)'\boldsymbol{R}^{-1}(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t).
\end{aligned}
$$

- To implement the EM algorithm, at iteration j, we write

$$
Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)}) = E[-2\log L_{\mathcal{X},\mathcal{Y}}(\boldsymbol{\Theta}|\mathcal{Y}_{1:T}, \boldsymbol{\Theta}^{(j-1)})].
$$

E-step (contd.)

- $$E[(\boldsymbol{X}_0-\mu_0)'\Sigma_0^{-1}(\boldsymbol{X}_0-\mu_0)|\mathcal{Y}_{1:T},\boldsymbol{\Theta}^{(j-1)}] = \mathrm{tr}\left[\Sigma_0^{-1}\left\{\boldsymbol{P}_0^T+(\boldsymbol{X}_0^T-\mu_0)(\boldsymbol{X}_0^T-\mu_0)'\right\}\right]$$
- $$E[(\boldsymbol{X}_t-\Phi\boldsymbol{X}_{t-1})'\boldsymbol{Q}^{-1}(\boldsymbol{X}_t-\Phi\boldsymbol{X}_{t-1})|\mathcal{Y}_{1:T},\boldsymbol{\Theta}^{(j-1)}] = \mathrm{tr}\left[\boldsymbol{Q}^{-1}\left\{\boldsymbol{S}_{11}-\boldsymbol{S}_{10}\Phi'-\Phi\boldsymbol{S}_{10}+\Phi\boldsymbol{S}_{00}\Phi'\right\}\right]$$
- $$E[(\boldsymbol{Y}_t-\boldsymbol{A}_t\boldsymbol{X}_t)'\boldsymbol{R}^{-1}(\boldsymbol{Y}_t-\boldsymbol{A}_t\boldsymbol{X}_t)|\mathcal{Y}_{1:T},\boldsymbol{\Theta}^{(j-1)}] = \mathrm{tr}\left[\boldsymbol{R}^{-1}\left\{(\boldsymbol{Y}_t-\boldsymbol{A}_t\boldsymbol{X}_t^T)(\boldsymbol{Y}_t-\boldsymbol{A}_t\boldsymbol{X}_t^T)'+\boldsymbol{A}_t\boldsymbol{P}_t^T\boldsymbol{A}_t'\right\}\right]$$

where $\boldsymbol{S}_{11}=\sum_{t=1}^T(\boldsymbol{X}_t^T[\boldsymbol{X}_t^T]'+\boldsymbol{P}_t^T)$, $\boldsymbol{S}_{00}=\sum_{t=1}^T(\boldsymbol{X}_{t-1}^T[\boldsymbol{X}_{t-1}^T]'+\boldsymbol{P}_{t-1}^T)$, $\boldsymbol{S}_{10}=\sum_{t=1}^T(\boldsymbol{X}_t^T[\boldsymbol{X}_{t-1}^T]'+\boldsymbol{P}_{t,t-1}^T)$.

M-step

- $\mu_0^{(j)} = \boldsymbol{X}_0^T$
- $\Sigma_0^{(j)} = \boldsymbol{P}_0^T$
- $\Phi^{(j)} = \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}$
- $\boldsymbol{Q}^{(j)} = T^{-1}\left\{\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{10}'\right\}$
- $\boldsymbol{R}^{(j)} = T^{-1}\sum_{t=1}^{T}\{(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t^T)(\boldsymbol{Y}_t - \boldsymbol{A}_t\boldsymbol{X}_t^T)' + \boldsymbol{A}_t\boldsymbol{P}_t^T\boldsymbol{A}_t'\}$

Overall algorithm

- Initialize by choosing starting values for the parameters, say $\Theta^{(0)}$, and compute the incomplete-data likelihood $\log L_{\mathcal{Y}}(\Theta)$.
- On iteration j, perform the E-Step: Using the parameters $\Theta^{(j-1)}$ to obtain the smoothed values $\boldsymbol{X}_t^T$, $\boldsymbol{P}_t^T$, and $\boldsymbol{P}_{t,t-1}^T$ for $t = 1, \ldots, T$, and calculate $\boldsymbol{S}_{11}, \boldsymbol{S}_{10}, \boldsymbol{S}_{00}$.
- On iteration j, perform the M-Step: Update the estimates to obtain $\{\boldsymbol{\mu}_0^{(j)}, \Sigma_0^{(j)}, \Phi^{(j)}, \boldsymbol{Q}^{(j)}, \boldsymbol{R}^{(j)}\}$.
- Compute the negative incomplete-data likelihood, $-\log L_{\mathcal{Y}}(\Theta)$.
- Repeat Steps until convergence.

Thank you!