A Kalman Filter/Smoother Overview

Engineering Methodology Working Group
SAMSI 2007

January 23, 2007

Recall from previous discussions the state-space model written as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_t \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t. \end{aligned}$$

We observe $\mathbf{y}$ and $\mathbf{x}$ is some unobserved process, and $\mathbf{G}$ and u_t are known. Assuming some noise in our system this can be written as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_t + \boldsymbol{\epsilon}_t \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \boldsymbol{\eta}_t \\ &\text{where} \\ \boldsymbol{\epsilon} &\overset{iid}{\sim} N(\mathbf{0}, \mathbf{R}_t) \\ \boldsymbol{\eta} &\overset{iid}{\sim} N(\mathbf{0}, \mathbf{Q}_t) \\ &\text{or} \\ (\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{H}, \mathbf{R}_t) &\sim N(\mathbf{H}\mathbf{x}_t, \mathbf{R}_t) \\ (\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{F}, \mathbf{G}, u, \mathbf{Q}_t) &\sim N(\mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_t, \mathbf{Q}_t) \end{aligned}$$

Given $\mathbf{H}$, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{R}_t$ and $\mathbf{Q}_t$ we can define three distributions

$$
\begin{array}{rl}
\left[\mathbf{x}_t \mid \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\right] & \text{The predictive distribution} \\
\left[\mathbf{x}_t \mid \mathbf{y}_t, \ldots, \mathbf{y}_1\right] & \text{The filter distribution} \\
\left[\mathbf{x}_t \mid \mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_t, \ldots, \mathbf{y}_1\right] & \text{The smoother distribution}
\end{array}
$$

First a few definitions

$$
\begin{aligned}
\text{Let } \mathbf{Y}_j &= (\mathbf{y}_1, \ldots, \mathbf{y}_j) \\
\mathbf{x}_{t|t} &\equiv E(\mathbf{x}_t \mid \mathbf{y}_t) \\
\mathbf{y}_{t|t-1} &\equiv E(\mathbf{x}_t \mid \mathbf{y}_{t-1}) \\
\mathbf{\Sigma}_{t|t} &\equiv E\left((\mathbf{x}_t - \mathbf{x}_{t|t})(\mathbf{x}_t - \mathbf{x}_{t|t})' \mid \mathbf{Y}_t\right) \\
\mathbf{\Sigma}_{t|t-1} &\equiv E\left((\mathbf{x}_t - \mathbf{x}_{t|t-1})(\mathbf{x}_t - \mathbf{x}_{t|t-1})' \mid \mathbf{Y}_{t-1}\right)
\end{aligned}
$$

The Filter Distribution

$$(\mathbf{x}_t \mid \mathbf{Y}_t) \sim N(\mathbf{x}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \forall\, t \ \geq 0$$

The Predictive Distribution

$$\begin{aligned}
(\mathbf{x}_t \mid \mathbf{Y}_{t-1}) &\sim N(\mathbf{x}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \\
E(\mathbf{x}_t \mid \mathbf{Y}_{t-1}) &= \mathbf{F}\mathbf{x}_{t-1|t-1} + \mathbf{G}u_t \\
Var(\mathbf{x}_t \mid \mathbf{Y}_{t-1}) &= \mathbf{Q}_t + \mathbf{F}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{F}'.
\end{aligned}$$

So

$$\mathbf{x}_{t|t-1} = \mathbf{F}\mathbf{x}_{t-1|t-1} + \mathbf{G}u_t \tag{1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q}_t + \mathbf{F}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{F}'. \tag{2}$$

Next we get

$$(\mathbf{x}_t \mid \mathbf{Y}_t) \sim N(\mathbf{x}_{t|t-1} + \mathbf{J}_t(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|t-1}), (\mathbf{I} - \mathbf{J}_t\mathbf{H})\boldsymbol{\Sigma}_{t|t-1})$$

where

$$\begin{aligned}
\mathbf{J}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{H}'(\mathbf{H}\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}' + \mathbf{R}_t)^{-1} && (3)\\
\mathbf{x}_{t|t} &= \mathbf{x}_{t|t-1} + \mathbf{J}_t(\mathbf{y}_t - \mathbf{H}\mathbf{y}_{t|t-1}) && (4)\\
\boldsymbol{\Sigma}_{t|t} &= (\mathbf{I} - \mathbf{J}_t\mathbf{H})\boldsymbol{\Sigma}_{t|t-1} && (5)
\end{aligned}$$

Given $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ and initial conditions $\mathbf{x}_{0|0}$ and $\mathbf{\Sigma}_{0|0}$, the Kalman Filter Algorithm is as follows.

1. For $t = 1$ obtain $\mathbf{x}_{t|t-1}$ and $\mathbf{\Sigma}_{t|t-1}$ from (1) and (2).
2. Obtain $\mathbf{J}_t$, $\mathbf{x}_{t|t}$, and $\mathbf{\Sigma}_{t|t}$ from (3)-(5).
3. Repeat for $t = 2, \ldots, T$.

In order to smoothe the data $\mathbf{x}_t|T,\ t \leq T$ we take initial conditions from Kalman Filter $\mathbf{x}_{T|T}$ and $\boldsymbol{\Sigma}_{T|T}$ and run a backwards algorithm for $t = T, \ldots, 1$

$$\begin{aligned}
\mathbf{x}_{t-1|T} &= \mathbf{x}_{t-1|t-1} + \mathbf{K}_{t-1}(\mathbf{x}_{t|T} - \mathbf{x}_{t|t-1}) && (6)\\
\boldsymbol{\Sigma}_{t-1|T} &= \boldsymbol{\Sigma}_{t-1|t-1} + \mathbf{K}_{t-1}(\boldsymbol{\Sigma}_{t|T} - \boldsymbol{\Sigma}_{t|t-1})\mathbf{K}'_{t-1} && (7)\\
\mathbf{K}_{t-1} &= \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{F}'\boldsymbol{\Sigma}_{t|t-1} && (8)
\end{aligned}$$

1. After running the filter for $t = T$, obtain $\mathbf{K}_{t-1}$ from (8)
2. Obtain $\mathbf{x}_{t-1|T}$ and $\boldsymbol{\Sigma}_{t-1|T}$ from (6) and (7)
3. Repeat for $t = T - 1, \ldots, 1$.

Bayesian Estimation of the Hierarchical State-Space Model

$$
\begin{aligned}
[data \mid process, parameters] = [\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{R}_t] &\sim N(\mathbf{H}\mathbf{x}_t, \mathbf{R}) \\
[process \mid parameters] = [\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{Q}] &\sim N(\mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}u_t, \mathbf{Q}_t) \\
[parameters] = [\mathbf{F}, \mathbf{Q}_t, \mathbf{R}_t] &= [\mathbf{F}]\,[\mathbf{Q}_t]\,[\mathbf{R}_t]\,.
\end{aligned}
$$

We could assume some priors and hyperparamters for $\mathbf{F}$ $\mathbf{Q}_t$ $\mathbf{R}_t$ and $\mathbf{x}_0$ such as choosing inverse-Wishart priors for $\mathbf{Q}_t$ and $\mathbf{R}_t$ and putting a Gaussian prior on $vec(\mathbf{F})$ and $\mathbf{x}_0$. A Gibbs sampler could then be implemented incorporating the Kalman filter and predictive distribution.

Convergence can be improved by jointly sampling $(\mathbf{x}_0, \ldots, \mathbf{x}_T)$ as in the forward filtering backwards sampling algorithm (Frühwirth and Schnatter 1994)