STA 642 Homework4

Jae Hyun Lee, jl914

16 February, 2020

HW3 for STA-642

Exercise 3

Derive the Bayesian filtering theory that is at the core of the kalma filtering + (variance learning) equations:

$$\begin{aligned} DLM :& y_t = F_t'\theta_t + \nu_t \quad \nu_t \sim N(0, v_t) \\ & \theta_t = G_t\theta_{t-1} + w_t \quad w_t \sim N(0, W_t) \\ with\ prior\ s& P(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1}) \end{aligned}$$

1. kalamn filtering

For one step ahead distribution of θ_t given D_{t-1} is as follow:

$$\begin{aligned} & P(\theta_t \mid D_{t-1}) = P(G_T\theta_{t-1} \mid D_{t-1}) \sim N(a_t, R_t) \\ where\ & a_t = E(G_t\theta_{t-1} \mid D_{t-1}) = G_t m_{t-1}, \\ & R_t = V(G_t\theta_{t-1} + w_t) = G_t V(\theta_{t-1}) G_T' + V(w_t) = G_t C_{t-1} G_t' + W_t \end{aligned}$$

For predictive distribution of y_t given D_{t-1} is as follow:

$$\begin{aligned} & P(y_t \mid D_{t-1}) = P(F_t'\theta_t \mid D_{t-1}) \sim N(f_t, q_t) \\ where\ & f_t = E(F_t'\theta_t \mid D_{t-1}) = F_t' a_t \\ & q_t = F_t' R_t F_t + v_t \end{aligned}$$

For evolved distribution of θ_t given D_t is as follow: and $P(y_t \mid \theta_t) \sim N(F_t'\theta_t, v_t)$

$$\begin{aligned} P(\theta_t \mid D_t) &= P(\theta_t \mid y_t, D_{t-1}) \\ &\propto P(y_t \mid \theta_t) P(\theta_t \mid D_{t-1}) \\ &\propto exp\{-\frac{1}{2v_t}(y_t - F_t'\theta_t)'(y_t - F_t'\theta_t)\} \times exp\{-\frac{1}{2}(\theta_t - a_t)' R_t^{-1}(\theta_t - a_t)\} \\ &\propto exp\{-\frac{1}{2}[\theta_t' F_t F_t' \theta_t / v_t - 2\theta_t' F_t y_t + \theta_t' R_t^{-1}\theta_t - 2\theta_t' R_t^{-1} a_t]\} \\ &\propto exp\{-\frac{1}{2}[\theta_t'(F_t F_t'/v_t + R_t^{-1})\theta_t - 2\theta_t'(F_t y_t / v_t + R_t^{-1} a_t)]\} \\ \rightarrow & C_t = (F_t F_t'/v_t + R_t^{-1})^{-1}, m_t = C_t(F_t y_t / v_t + R_t^{-1} a_t) \end{aligned}$$

By Sherman - Morris formula,

$$\begin{aligned} C_t &= R_t - \frac{R_t F_t F_t' R_t / v_t}{1 + F_t' R_t F_t / v_t} \\ &= R_t - \frac{A_t A_t' q_t^2 / v_t}{q_t / v_t} \\ &= R_t - A_t A_t' q_t \end{aligned}$$

$$
\begin{aligned}
m_t &= (R_t - A_t A_t' q_t)(F_t y_t / v_t + R_t^{-1} a_t) \\
&= a_t + \frac{A_t q_t}{v_t} y_t - A_t A_t' F_t \frac{q_t y_t}{v_t} - A_t A_t' R_t^{-1} q_t a_t \\
&= a_t + A_t (I - A_t' F_t) \frac{q_t y_t}{v_t} - A_t A_t' R_t^{-1} q_t a_t \\
&= a_t + A_t (\frac{F_t' R_t F_t + v_t - F_t' R_t F_t}{q_t}) - A_t A_t' R_t^{-1} q_t a_t \\
&= a_t + A_t y_t - A_t A_t' R_t^{-1} q_t a_t \\
&= a_t + A_t y_t - A_t \frac{F_t' R_t' R_t^{-1}}{q_t} q_t a_t \\
&= a_t + A_t y_t - A_t F_t' a_t = a_t + A_t (y_t - f_t) = a_t + A_t e_t
\end{aligned}
$$

2. Learning Variance

$$
\begin{aligned}
DLM : & y_t = F_t' \theta_t + \nu_t \quad \nu_t \sim N(0, v) \\
& \theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim N(0, \frac{v}{s_{t-1}} W_t) \\
with\ prior\ & P(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, \frac{v}{s_{t-1}} C_{t-1}) \\
& P(v \mid D_{t-1}) \sim IG(n_{t-1}/2, n_{t-1} s_{t-1}/2)
\end{aligned}
$$

We can figure out that

$$
\begin{aligned}
P(\theta_{t-1} \mid D_{t-1}) &= \int P(\theta_{t-1} \mid \phi, D_{t-1}) P(\phi \mid D_{t-1}) d\phi \\
&\propto \int \phi^{\frac{p+n_{t-1}}{2}} exp\{-\frac{\phi}{2}[s_{t-1}(\theta_{t-1} - m_{t-1})' C_{t-1}^{-1} (\theta_{t-1} - m_{t-1}) + n_{t-1} s_{t-1}]\} \\
&\propto [n_{t-1} s_{t-1} + s_{t-1}(\theta_{t-1} - m_{t-1})' C_{t-1}^{-1} (\theta_{t-1} - m_{t-1}))]^{-\frac{p+n_{t-1}}{2}} \\
&\propto [1 + \frac{1}{n_{t-1}} (\theta_{t-1} - m_{t-1})' C_{t-1}^{-1} (\theta_{t-1} - m_{t-1}))]^{-\frac{p+n_{t-1}}{2}} \\
&\rightarrow P(\theta_{t-1} \mid D_{t-1}) \sim T_{n_{t-1}}(m_{t-1}, C_{t-1})
\end{aligned}
$$

For one step ahead distribution of θ_t given v, D_{t-1} is as follow:

$$
\begin{aligned}
P(\theta_t \mid v, D_{t-1}) = P(G_t \theta_{t-1} + w_t \mid v, D_{t-1}) \sim N(a_t, \frac{v}{s_{t-1}} R_t) \\
where\ a_t = G_t m_{t-1},\ R_t = G_t C_{t-1} G_t' + W_t \\
because\ V(G_t \theta_{t-1} + w_t) = G_t V(\theta_{t-1}) G_t' + V(w_t) = G_t \frac{v}{s_{t-1}} C_{t-1} G_t' + \frac{v}{s_{t-1}} W_t
\end{aligned}
$$

by above procedure $P(\theta_t \mid D_{t-1}) \sim T_{n_{t-1}}(a_t, R_t)$.

For predictive distribution of y_t given v, D_{t-1} is as follow:

$$
\begin{aligned}
P(y_t \mid v, D_{t-1}) = P(F_t' \theta_t + \nu_t \mid v, D_{t-1}) \sim N(f_t, \frac{v}{s_{t-1}} q_t) \\
where\ f_t = F_t' E(\theta_t) = F_t' a_t,\ q_t = F_t' R_t F_t + s_{t-1} \\
because\ V(F_t' \theta_t + \nu_t \mid v, D_{t-1}) = F_t' \frac{v}{s_{t-1}} R_t F_t + v = \frac{v}{s_{t-1}} (F_t' R_t F_t + s_{t-1}) = \frac{v}{s_{t-1}} q_t
\end{aligned}
$$

For evolved joint distribution of θ_t, v given D_t is as follow: and $P(y_t \mid v, \theta_t) \sim N(F_t'\theta_t, v)$

$$
\begin{aligned}
P(\theta_t, v \mid D_t) =& P(\theta_t, v \mid D_{t-1}, y_t)\\
&\propto P(y_t \mid \theta_t, v, D_{t-1})P(\theta_t \mid v, D_{t-1})P(v \mid D_{t-1})\\
&\propto v^{-1/2} exp\{-\frac{1}{2v}(y_t - F_t'\theta_t)'(y_t - F_t'\theta_t)\}\\
&\quad \times v^{-p/2} exp\{-\frac{s_{t-1}}{2v}(\theta_t - a_t)^t R_t^{-1}(\theta_t - a_t)\} \times v^{-n_{t-1}/2} exp\{-\frac{n_{t-1}s_{t-1}}{2v}\}\\
&\propto v^{-\frac{n_{t-1}+1}{2}} \times v^{-p/2}\\
&\quad \times exp\{-\frac{s_{t-1}}{2v}[\theta_t'(F_tF_t'/s_{t-1} + R_t^{-1})\theta_t - 2\theta_t'(F_t y_t/s_{t-1} + R_t^{-1}a_t) + y_t'y_t/s_{t-1} + a_t'R_t^{-1}a_t + n_{t-1}]\}\\
&\propto v^{-n_t/2} \times v^{-p/2} exp\{-\frac{s_{t-1}}{2v}(\theta_t'B_t^{-1}\theta_t - 2\theta_t'B_t^{-1}m_t + m_t'B_t^{-1}m_t)\}\\
&\quad \times exp\{-\frac{s_{t-1}}{2v}(y_t'y_t/s_{t-1} + a_t'R_t^{-1}a_t + n_{t-1} - m_t'B_t^{-1}m_t)\}\\
&\propto v^{-p/2} exp\{-\frac{s_{t-1}}{2v}(\theta_t - m_t)'B_t^{-1}(\theta_t - m_t)\}\\
&\quad \times v^{-n_t/2} exp\{-\frac{s_{t-1}}{2v}(y_t'y_t/s_{t-1} + a_t'R_t^{-1}a_t + n_{t-1} - m_t'B_t^{-1}m_t)\}
\end{aligned}
$$

$$
\begin{aligned}
where \quad & B_t = (F_tF_t'/s_{t-1} + R_t^{-1})^{-1} = R_t - A_tA_t'q_t \quad by\ Sherman\ Morris\ formula\ as\ previous\ derivation.\\
& C_t = \frac{s_t}{s_{t-1}}B_t = r_tB_t = r_t(R_t - A_tA_t'q_t)\\
& m_t = B_t(F_ty_t/s_{t-1} + R_t^{-1}a_t) = (R_t - A_tA_t'q_t)(\frac{F_ty_t}{s_{t-1}} + R_t^{-1}a_t) = a_t + A_te_t \quad as\ previous\ question.\\
& n_t = n_{t-1} + 1\\
& s_t = \frac{s_{t-1}}{n_t}(n_{t-1} + a_t'R_t^{-1}a_t + y_t'y_t/s_{t-1} - m_t'B_t^{-1}m_t)\\
& = \frac{s_{t-1}}{n_t}(n_{t-1} + y_t'y_t/s_{t-1} - (a_t + A_te_t)'(F_tF_t'/s_{t-1} + R_t^{-1})(a_t + A_te_t) + a_tR_t^{-1}a_t)\\
& = \frac{s_{t-1}}{n_t}(n_{t-1} + \frac{(y - F_t'a_t)'(y - F_t'a_t)}{s_{t-1}} - (A_te_t)'(F_tF_t'/s_{t-1} + R_t^{-1})(A_te_t))\\
& = \frac{s_{t-1}}{n_t}(n_{t-1} + e_t^2/s_{t-1} - e_t^2A_t'((F_tF_t'/s_{t-1} + R_t^{-1}))A_t\\
& = \frac{s_{t-1}}{n_t}(n_{t-1} + e_t^2(F_t'R_tF_t + s_{t-1})^{-1})\\
& = \frac{s_{t-1}}{n_t}(n_{t-1} + e_t^2/q_t) = s_{t-1}r_t
\end{aligned}
$$

Exercise 4.

$$
\begin{aligned}
DLM :& y_t = F_t'\theta_t + \nu_t \quad \nu_t \sim N(0, v_t)\\
& \theta_t = \theta_{t-1} + w_t \quad w_t \sim N(0, W_t)\\
with\ prior\ & P(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})
\end{aligned}
$$

(a)

For one step ahead distribution of θ_t given D_{t-1} is as follow:

$$
\begin{aligned}
&P(\theta_t \mid D_{t-1}) = P(\theta_{t-1} + w_t \mid D_{t-1}) \\
&E(\theta_t \mid D_{t-1}) = m_{t-1} = a_t \\
&V(\theta_t \mid D_{t-1}) = C_{t-1} + W_t = (1+\epsilon)C_{t-1} = C_{t-1}/\delta = R_t \\
&\rightarrow P(\theta_t \mid D_{t-1}) \sim N(a_t, R_t)
\end{aligned}
$$

For predictive distribution of y_t given D_{t-1} is as follow:

$$
\begin{aligned}
&P(y_t \mid D_{t-1}) = P(F_t\theta_t + \nu_t \mid D_{t-1}) \\
&E(y_t \mid D_{t-1}) = F_t' E(\theta_t) = F_t' a_t = F_t' m_{t-1} = f_t \\
&V(y_t \mid D_{t-1}) = F_t' C_{t-1} F_t/\delta + \nu_t = q_t \\
&\rightarrow P(y_t \mid D_{t-1}) \sim N(f_t, q_t)
\end{aligned}
$$

For evolved distribution of θ_t given D_t is as follow: and $P(y_t \mid \theta_t) \sim N(F_t'\theta_t, v_t)$

$$
\begin{aligned}
&P(\theta_t \mid D_t) \propto P(y_t \mid \theta_t) P(\theta_t \mid D_{t-1}) \sim N(m_t, C_t) \\
where \quad &m_t = a_t + A_t e_t = m_{t-1} + \frac{\delta}{F_t' C_{t-1} F_t + \delta v_t} \times C_{t-1}/\delta \times e_t = m_{t-1} + \frac{C_{t-1}}{q_t \delta} e_t \\
&C_t = R_t - A_t A_t' q_t = \frac{C_{t-1}}{\delta} - A_t A_t' q_t \\
&A_t = \frac{C_{t-1}}{q_t \delta}
\end{aligned}
$$

(b)

For prior distribution of θ_{t-1} given D_{t-1} is as follow:

The initial prior distribution is not affected by simplified structure of DLM. However, given m_{t-1}, C_{t-1} might be affected by discount factor δ

For one step ahead distribution of θ_t given D_{t-1} is as follow:

The mean is same as before a_{t-1} but covariance matrix becomes larger by $1/\delta$ where $\delta \in (0, 1)$. This distribution depends on δ with negative relationship because if $\delta \rightarrow 0$, $R_t = C_{t-1}/\delta \rightarrow \infty$. On contrary if $\delta \rightarrow 1, R_t \rightarrow C_{t-1}$.

For predictive distribution of y_t given D_{t-1} is as follow:

δ does not affect on its mean. But it has negative relationship with its variance as previous case.

For evolved distribution of θ_t given D_t is as follow:

m_t depends on δ by A_t and as $\delta \rightarrow 0$, the effect of v_t on adaptive coefficient gets smaller with $A_t \rightarrow \frac{C_{t-1}}{F_t' C_t F_t}$ becomes larger. On the other hand, $\delta \rightarrow 1, A_t$ becomes smaller. That is, if $\delta \rightarrow 0$, the correction effect from error on mean becomes larger.

(c)

In this simplified structure, statistics which need to be computed are reduced. As a result, computation is also reduced and we can analize the data more efficiently.

Exercise 5

(a)

$$\begin{aligned}
C(\theta_t, \theta_{t-1} \mid D_{t-1}) &= E[(\theta_t - E(\theta_t))(\theta_{t-1} - E(\theta_{t-1}))'] \\
&= E[(G_t\theta_{t-1} - G_t m_{t-1} + \nu_t)(\theta_{t-1} - m_{t-1})'] \\
&= G_t E[(\theta_{t-1} - m_{t-1})(\theta_{t-1} - m_{t-1})'] + E(\nu_t(\theta_{t-1} - m_{t-1})') \\
&= G_t V(\theta_{t-1}) + Cor(\nu_t, \theta_{t-1}) = G_t C_{t-1}
\end{aligned}$$

On contrary,

$$\begin{aligned}
C(\theta_{t-1}, \theta_t \mid D_{t-1}) &= E[(\theta_{t-1} - m_{t-1})(G_t\theta_{t-1} - G_t m_{t-1} + \nu_t)'] \\
&= E[(\theta_{t-1} - m_{t-1})(\theta_{t-1} - m_{t-1})']G_t' + Cor(\theta_{t-1}, \nu_t)G_t' \\
&= C_{t-1}G_t'
\end{aligned}$$

(b)

By above result, we could find that

$$(\theta_t, \theta_{t-1})' = \boldsymbol{\theta} \sim N(\mu, \Sigma) \quad where \quad \mu = \begin{bmatrix} a_t \\ m_{t-1} \end{bmatrix}, \Sigma = \begin{bmatrix} R_t & G_t C_{t-1} \\ C_{t-1}G_t' & C_{t-1} \end{bmatrix}$$

Then, we can deduce that $P(\theta_{t-1} \mid \theta_t, D_{t-1})$ is normal as follow:

Let $W = \theta_{t-1} - X\theta_t$ and X is chosen so that W and θ_t is independent.

$$\begin{aligned}
\begin{bmatrix} \theta_t \\ W \end{bmatrix} &= \begin{bmatrix} I_p & 0 \\ -X & I_p \end{bmatrix} \begin{bmatrix} \theta_t \\ \theta_{t-1} \end{bmatrix} \\
Then\ V(\begin{bmatrix} \theta_t \\ W \end{bmatrix}) &= \begin{bmatrix} I_p & 0 \\ -X & I_p \end{bmatrix} \begin{bmatrix} R_t & G_t C_{t-1} \\ C_{t-1}G_t' & C_{t-1} \end{bmatrix} \begin{bmatrix} I_p & -X' \\ 0 & I_p \end{bmatrix} \\
&= \begin{bmatrix} R_t & G_t C_{t-1} \\ -XR_t + C_{t-1}G_t' & -XG_tC_{t-1} + C_{t-1} \end{bmatrix} \begin{bmatrix} I_p & -X' \\ 0 & I_p \end{bmatrix} \\
&= \begin{bmatrix} R_t & -R_tX' + G_tC_{t-1} \\ -XR_t + C_{t-1}G_t' & XR_tX' - C_{t-1}G_t'X' - XG_tC_{t-1} + C_{t-1} \end{bmatrix}
\end{aligned}$$

We can choose X as $C_{t-1}G_t'R_t^{-1}$, then $X = B_t$

$$\begin{bmatrix} R_t & -R_tX' + G_tC_{t-1} \\ -XR_t + C_{t-1}G_t' & XR_tX' - C_{t-1}G_t'X' - XG_tC_{t-1} + C_{t-1} \end{bmatrix} = \begin{bmatrix} R_t & 0 \\ 0 & C_{t-1} - B_tR_tB_t' \end{bmatrix}$$

Now W and θ_t is independent. Thus

$$\begin{aligned}
& W \mid \theta_t \sim N(m_{t-1} - B_t a_t, C_{t-1} - B_tR_tB_t') \\
and \quad & \theta_{t-1} = W + B_t\theta_t \\
& \rightarrow \theta_{t-1} \sim N(m_{t-1} + B_t(\theta_t - a_t), C_{t-1} - B_tR_tB_t')
\end{aligned}$$

(c)

$$\begin{aligned} DLM : & y_t = F_t'\theta_t + \nu_t \\ & \theta_t = G_t\theta_{t-1} + w_t \end{aligned}$$

has markovian structure with means that state vector only depends on next one. Thus all feature observations are irrelavant. Thus $P(\theta_{t-1} \mid \theta_t, D_{t-1}) = P(\theta_{t-1} \mid \theta_t, D_n)$.

(d)

By this theory, we can easily quantify the a full trajectory states $P(\theta_{1:n} \mid D_n)$ because since θ_t only depends on data of time point and next step state, we can infer that

$$\begin{aligned} P(\theta_{1:n} \mid D_n) &= P(\theta_1 \mid \theta_2, D_1) \times P(\theta_2 \mid \theta_3, D_2) \cdots P(\theta_{n-1} \mid \theta_n, D_{n-1}) \\ &= \prod P(\theta_i \mid \theta_{i+1}, D_i) \end{aligned}$$

(e)

For simplified case, we have confirmed that $a_t = m_{t-1}, R_t = C_{t-1}/\delta$ from previous question. Moreover, simplified covariance of θ_t, θ_{t-1} is C_{t-1}. Thus

$$\begin{aligned} & P(\theta_t, \theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, C_{t-1} \begin{bmatrix} 1 & 1 \\ 1 & 1/\delta \end{bmatrix}) \quad and \\ & P(\theta_{t-1} \mid \theta_t, D_{t-1}) \sim N(m_{t-1} + B_t(\theta_t - a_t), C_{t-1} - B_t R_t B_t'), \quad and\ B_t = \delta \\ \rightarrow & P(\theta_{t-1} \mid \theta_t, D_{t-1}) \sim N(m_{t-1}(1-\delta) + \delta\theta_t, C_{t-1} - \delta) \end{aligned}$$

That is, in retrospective distribution of θ_{t-1}, δ plays a role of weight that averaging θ_{t-1}'s mean and new θ_t and we can also find that $C_{t-1} \rightarrow C_{t-1} - \delta$ which becomes smaller by new θ_t. By δ, computations becomes much easier.