3. Derive the Bayesian filtering theory that is at the core of the kalman filtering + (variance learning) equations: P&W section 4.3.1, 4.3.2

DLM: $Y_t = F_t'\theta_t + r_t \qquad r_t \sim N(0, V_t)$

$\theta_t = G_t\theta_{t-1} + w_t \qquad w_t \sim N(0, W_t)$

with prior $p(\theta_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1})$

① Kalman filtering

$p(\theta_t|D_{t-1}) = p(G_t\theta_{t-1}|D_{t-1}) \sim N(a_t, R_t)$

where $a_t = E(G_t\theta_{t-1}) = G_tE(\theta_{t-1}) = G_t m_{t-1}$, $\quad R_t = Var(G_t\theta_{t-1}+w_t) = G_t V(\theta_{t-1}) G_t^T + Var(w_t)$

$p(Y_t|D_{t-1}) = p(F_t'\theta_t|D_{t-1}) \sim N(f_t, q_t) \qquad = G_t C_{t-1} G_t^T + W_t$

where $f_t = E(F_t'\theta_t) = F_t'E(\theta) = F_t'a_t$, $\quad q_t = F_t'R_tF_t + V_t$

$p(y_t|\theta_t) \sim N(F_t'\theta_t, V_t)$

$p(\theta_t|D_t) = p(\theta_t|y_t, D_{t-1})$

$\propto p(y_t|\theta_t)\,p(\theta_t|D_{t-1})$

$\propto \exp\{-\frac{1}{2V_t}(y-F_t'\theta_t)'(y-F_t'\theta_t)\}\exp\{-\frac12(\theta_t-a_t)'R_t^{-1}(\theta_t-a_t)\}$

$\propto \exp\{-\frac12[\theta_t'F_tF_t'\theta_t/V_t - 2\theta_t'F_t y + \theta_t'R_t^{-1}\theta_t - 2\theta_t'R_t^{-1}a_t]\}$

$\propto \exp\{-\frac12[\theta_t'(\frac{F_tF_t'}{V_t}+R_t^{-1})\theta_t - 2\theta_t'(\frac{F_t y}{V_t}+R_t^{-1}a_t)\}$

$\Rightarrow C_t = \left(\frac{F_tF_t'}{V_t}+R_t^{-1}\right)^{-1}, \quad m_t = C_t\left(\frac{F_t y}{V_t}+R_t^{-1}a_t\right)$

^ Statement

Suppose $A \in \mathbb{R}^{n\times n}$ is an invertible square matrix and $u, v \in \mathbb{R}^n$ are column vectors. Then $A + uv^\mathsf{T}$ is invertible iff $1 + v^\mathsf{T}A^{-1}u \neq 0$. In this case,

$$(A+uv^\mathsf{T})^{-1} = A^{-1} - \frac{A^{-1}uv^\mathsf{T}A^{-1}}{1+v^\mathsf{T}A^{-1}u}.$$

Here, uv^T is the outer product of two vectors u and v. The general form shown here is the one published by Bartlett.[5]

By Sherman - Morris formula

$C_t = R_t - \dfrac{R_tF_tF_t'R_t/V_t}{1+F_t'R_tF_t/V_t}$

$= R_t - \dfrac{A_tA_t'q_t^2/V_t}{1+q_t} = R_t - \dfrac{A_tA_t'q_t^2}{q_t}$

$= R_t - A_tA_t'q_t$

$m_t = (R_t - A_tA_t'q_t)\left(\frac{F_t y}{V_t}+R_t^{-1}a_t\right)$

$= a_t + \frac{A_tq_t}{V_t}y_t - A_tA_t'F_t\frac{q_ty_t}{V_t} - A_tA_t'R_t^{-1}q_ta_t$

$= a_t + A_t(I - A_t'F_t)\frac{q_ty_t}{V_t} - A_tA_t'R_t^{-1}q_ta_t$

$= a_t + A_t\left(\frac{F_t'R_tF_t+V_t}{q_t} - F_t'R_tF_t\right)\frac{q_ty_t}{V_t} - A_tA_t'R_t^{-1}q_ta_t$

$= a_t + A_ty_t - A_tA_t'R_t^{-1}q_ta_t$

$= a_t + A_ty_t - A_t\frac{F_t'R_t'R_t^{-1}}{q_t}q_ta_t$

$= a_t + A_ty_t - A_tF_t'a_t = a_t + A_t(y_t - f_t) = a_t + A_te_t$

② Learning Variance

DLM: $Y_t = F_t'\theta_t + v_t \qquad v_t \sim N(0, V)$

$\theta_t = G_t\theta_{t-1} + w_t \qquad w_t \sim N(0, \frac{V}{S_{t-1}} W_t)$

with prior $P(\theta_{t-1}|V, D_{t-1}) \sim N(m_{t-1}, \frac{V}{S_{t-1}} C_{t-1})$

$P(V|D_{t-1}) \sim IG(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2})$

$\phi_t = 1/V_t \quad P(\phi|D_{t-1}) \sim G(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2})$

- $P(\theta_{t-1}|D_{t-1}) = \int P(\theta_{t-1}|\phi, D_{t-1}) P(\phi|D_{t-1})\, d\phi$

$\propto \int \phi^{\frac{p+n_{t-1}}{2}} \exp\{-\frac{\phi}{2}[S_{t-1}(\theta_{t-1}-m_{t-1})^T C_{t-1}^{-1}(\theta_{t-1}-m_{t-1}) + n_{t-1}S_{t-1}]$

$\propto [n_{t-1}S_{t-1} + S_{t-1}(\theta_{t-1}-m_{t-1})^T C_{t-1}^{-1}(\theta_{t-1}-m_{t-1})]^{-(\frac{p+n_{t-1}}{2})}$

$\propto [1 + \frac{1}{n_{t-1}}(\theta_{t-1}-m_{t-1})^T C_{t-1}^{-1}(\theta_{t-1}-m_{t-1})]^{-(\frac{p+n_{t-1}}{2})}$

$\Rightarrow P(\theta_{t-1}|D_{t-1}) \sim T_{n_{t-1}}(m_{t-1}, C_{t-1})$

$P(\theta_t|V, D_{t-1}) = P(G_t\theta_{t-1} + w_t|V, D_{t-1}) \sim N(a_t, \frac{V}{S_{t-1}} R_t)$

where $a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t' + W_t$, $Var(G_t\theta_{t-1} + w_t) = G_t V(\theta_{t-1}) G_t^T + Var(w_t)$

$\Rightarrow P(\theta_t|D_{t-1}) \sim T_{n_{t-1}}(a_t, R_t)$ $\qquad = G_t \frac{V}{S_{t-1}} C_{t-1} G_t^T + \frac{V}{S_{t-1}} W_t$

$P(Y_t|V, D_{t-1}) = P(F_t'\theta_t + v_t|V, D_{t-1}) \sim N(f_t, \frac{V}{S_{t-1}} q_t)$

where $f_t = F_t' E(\theta_t) = F_t' a_t$, $\quad q_{t-1} = F_t' R_t F_t + S_{t-1}$

$Var(F_t'\theta_t + v_t|V, D_{t-1}) = F_t' \frac{V}{S_{t-1}} R_t F_t + V$

$= \frac{V}{S_{t-1}}(F_t' R_t F_t + S_{t-1}) = \frac{V}{S_{t-1}} q_t$

- $p(\theta_t, V \mid D_t) = p(\theta_t, V \mid D_{t-1}, y_t)$

$$\propto p(y_t \mid \theta_t, V, D_{t-1})\, p(\theta_t \mid V\ D_{t-1})\, p(V \mid D_{t-1})$$

$$\propto V^{-\frac{1}{2}} \exp\left\{-\frac{1}{2V}(y_t - F_t'\theta_t)^T (y_t - F_t'\theta_t)\right\} V^{-\frac{p}{2}} \exp\left\{-\frac{S_{t-1}}{2V}(\theta_t - a_t)^T R_t^{-1} (\theta_t - a_t)\right\}$$

$$\times V^{-\frac{n_{t-1}}{2}} \exp\left\{-\frac{n_{t-1}S_{t-1}}{2V}\right\}$$

$$\propto V^{-\left(\frac{n_{t-1}+1}{2}\right)} \times V^{-\frac{p}{2}} \exp\left\{-\frac{S_{t-1}}{2V}\left(\theta_t^T\left(\frac{F_tF_t'}{S_{t-1}} + R_{t-1}^{-1}\right)\theta_t - 2\theta_t^T\left(\frac{F_t y_t}{S_{t-1}} + R_t^{-1}a_{t-1}\right) + \frac{y_t^T y_t}{S_{t-1}} + a_t^T R_t^{-1} a_t + n_{t-1}\right)\right\}$$

$$\propto V^{-\frac{p}{2}} \times V^{-\frac{n_t}{2}} \exp\left\{-\frac{S_{t-1}}{2V}\left(\theta_t^T B_t^{-1}\theta_t - 2\theta_t^T B_t^{-1} m_t + m_t^T B_t^{-1} m_t\right)\right\} \exp\left\{-\frac{S_{t-1}}{2V}\left(\frac{y_t^T y_t}{S_{t-1}} + a_t^T R_t^{-1} a_t + n_{t-1} - m_t^T B_t^{-1} m_t\right)\right\}$$

$$\propto V^{-\frac{p}{2}} \exp\left\{-\frac{S_{t-1}}{2V}(\theta_t - m_t) B_t^{-1} (\theta_t - m_t)\right\} V^{-\frac{n_t}{2}} \exp\left\{-\frac{S_{t-1}}{2V}\left[n_{t-1} + a_t^T R_t^{-1} a_t + \frac{y_t^T y_t}{S_{t-1}} - m_t^T B_t^{-1} m_t\right]\right\},$$

$B_t = \left(\frac{F_tF_t'}{S_{t-1}} + R_t^{-1}\right)^{-1} = R_t - A_tA_t' q_t$ by sherman morris formula as previous deriavtion.

$$C_t = \frac{S_t}{S_{t-1}} B_t = r_t B_t = r_t(R_t - A_t A_t q_t)$$

$$m_t = B_t\left(\frac{F_t y_t}{S_{t-1}} + R_t^{-1}a_t\right) = (R_t - A_tA_t q_t)\left(\frac{F_t y_t}{S_{t-1}} + R_t^{-1}a_t\right)$$

$= a_t + A_t e_t$ as previous question

$$n_t = n_{t-1} + 1$$

$$S_t = \frac{S_{t-1}}{n_t}\left(n_{t-1} + a_t^T R_t^{-1} a_t + \frac{y_t^T y_t}{S_{t-1}} - m_t^T B_t^{-1} m_t\right)$$

$$= \frac{S_{t-1}}{n_t}\left(n_{t-1} + \frac{y_t^T y_t}{S_{t-1}} - (a_t + A_t e_t)^T\left(\frac{F_tF_t'}{S_{t-1}} + R_t^{-1}\right)(a_t + A_t e_t) + a_t R^{-1} a_t\right)$$

$$= \frac{S_{t-1}}{n_t}\left(n_{t-1} + \frac{(y - F_t'a)^T (y - F_t a)}{S_{t-1}} + (A_t e_t)^T\left(\frac{F_tF_t'}{S_{t-1}} + R_t^{-1}\right)(A_t e_t)\right)$$

$$= \frac{S_{t-1}}{n_t}\left(n_{t-1} + \frac{e_t^2}{S_{t-1}} + e_t^2\left(A_t^T\left(\frac{F_tF_t'}{S_{t-1}} + R_t^{-1}\right)A_t\right)\right)$$

$$= \frac{S_{t-1}}{n_t}\left(n_{t-1} + e_t^2\left(F_t' R_t F_t + S_{t-1}\right)^{-1}\right)$$

$$= \frac{S_{t-1}}{n_t}\left(n_{t-1} + e_t^2 / q_t\right) = S_{t-1} r_t$$

4. $y_t = F_t'\theta_t + v_t$

$\theta_t = \theta_{t-1}$

(a) $p(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})$

$p(\theta_t \mid D_{t-1}) = p(\theta_{t-1} + w_t \mid D_{t-1})$

$E(\theta_t \mid D_{t-1}) = m_{t-1} = a_t$

$V(\theta_t \mid D_{t-1}) = C_{t-1} + W_t = (1+\varepsilon)C_{t-1} = \frac{1}{\delta}C_{t-1} = R_t$

$p(y_t \mid D_{t-1}) = p(F_t'\theta_t + v_t \mid D_{t-1})$

$E(y_t \mid D_{t-1}) = F_t' E(\theta_t) = F_t' a_t = F_t' m_{t-1} = f_t$

$Var(y_t \mid D_{t-1}) = F_t' C_{t-1}/\delta F_t + V_t = q_t$

$p(\theta_t \mid D_t) \propto p(y \mid \theta_t) p(\theta_t \mid D_{t-1})$ $\qquad A_t = \frac{R_t F_t'}{q_t}$

$\sim N(m_t, C_t)$

where $m_t = a_t + A_t e_t = m_{t-1} + \frac{\delta}{F_t' C_{t-1} F_t + \delta V_t} \times \frac{C_{t-1}}{\delta} F_t (y_t - F_t' m_{t-1})$

$= m_{t-1} + \frac{C_{t-1}}{q_t \delta} e_t$

$C_t = R_t - AA^T q_t = \frac{C_{t-1}}{\delta} - AA^T q_t$

$A_t = \frac{C_{t-1}}{q_t \delta} = \frac{C_{t-1}}{F_t' C_{t-1} F_t + \delta V_t}$

(b) For one step a head distribution of θ_t given D_{t-1} ($p(\theta_t \mid D_{t-1})$)

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

For predictive dist of y_t given D_{t-1}

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

For evolved dist of θ_t given D_t:

m_t is depends on δ by A_t and as $\delta \to 0$ the effect of V_t on adaptive coefficient gets smaller $A_t \to \frac{C_{t-1}}{F_t' C_t F_t}$ which becomes larger. on the other hand $\delta \to 1$ A_t smaller.

5. (a) $C(\theta_t, \theta_{t-1} | D_{t-1})$

$= E[(\theta_t - E(\theta_t))(\theta_{t-1} - E(\theta_{t-1}))']$

$= E[(G\theta_{t-1} - Gm_{t-1} + r_t)(\theta_{t-1} - m_{t-1})']$

$= E[G(\theta_{t-1} - m_{t-1})(\theta_{t-1} - m_{t-1})'] + E(r_t(\theta_{t-1} - m_{t-1})']$

$= GV(\theta_{t-1}) + Cov(r_t, \theta_{t-1}) = GC_{t-1}$

On contrary

$C(\theta_{t-1}, \theta_t | D_{t-1}) = E((\theta_{t-1} - m_{t-1})(G\theta_{t-1} - Gm_{t-1} + r_t)']$

$= E[(\theta_{t-1} - m_{t-1})(\theta_t - m_{t-1})']G_t' + C(\theta_{t-1}, r_t)G_t'$

$= C_{t-1}G_t'$

(b) By above result,

we could find that $\theta = \begin{pmatrix} \theta_t \\ \theta_{t-1} \end{pmatrix} \sim N(\mu, \Sigma)$ where $\mu = \begin{pmatrix} a_t \\ m_{t-1} \end{pmatrix}$ $\Sigma = \begin{bmatrix} R_t & G_tC_{t-1} \\ C_{t-1}G_t' & C_{t-1} \end{bmatrix}$

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

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

$\begin{bmatrix} \theta_t \\ W \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ -X & I_p \end{bmatrix} \begin{bmatrix} \theta_t \\ \theta_{t-1} \end{bmatrix}$

Then $V\begin{pmatrix} \theta_t \\ W \end{pmatrix} = \begin{bmatrix} I_p & 0 \\ -X & I_p \end{bmatrix} \begin{bmatrix} R_t & G_tC_{t-1} \\ C_{t-1}G_t' & C_{t-1} \end{bmatrix} \begin{bmatrix} I_p & -X^T \\ 0 & I_p \end{bmatrix}$

$= \begin{bmatrix} R_t & G_tC_{t-1} \\ -XR_t + C_{t-1}G_t', & -XG_tC_{t-1} + C_{t-1} \end{bmatrix} \begin{bmatrix} I_p & -X^T \\ 0 & I_p \end{bmatrix}$

$= \begin{bmatrix} R_t, & -R_tX^T + G_tC_{t-1} \\ -XR_t + G_{t-1}G_t' & XR_tX^T - C_{t-1}G_t'X^T - XG_tC_{t-1} + C_{t-1} \end{bmatrix}$

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

$= \begin{bmatrix} R_t & 0 \\ 0 & C_{t-1} - XR_tX^T \end{bmatrix}$

Now W and θ_t is indpt

Thus $W | \theta_t \sim N(m_{t-1} - B_t a_t, C_{t-1} - B_tR_tB_t^T)$

Then $\theta_{t-1} = W + B_t\theta_t$

$\Rightarrow$ $\theta_{t-1} \sim N(m_{t-1} + B_t(\theta_t - a_t), C_{t-1} - B_tR_tB_t^T)$

(c) DLM : $Y_t = F_t' \theta_t + v_t$

$\theta_t = G_t \theta_{t-1} + w_t$

has markovian structure which means that state vector only depends on next one. Thus all future observations are irrelavant

Thus $p(\theta_{t-1} | \theta_t, D_{t-1}) = p(\theta_{t-1} | \theta_t, D_n)$

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

$p(\theta_{1:n} | D_n) = p(\theta_1 | \theta_2, D_1) \times p(\theta_2 | \theta_3, D_2) \cdots$

$= \prod_{i=1}^{n} p(\theta_i | \theta_{i+1}, D_i)$

(e) For simplified case, we have confirmed that

$a_t = m_{t-1}$, $R_t = \frac{1}{\delta} C_{t-1}$

More over simplified covariance of θ_t, θ_{t-1} is

C_{t-1}

Thus $p(\theta_t, \theta_{t-1} | D_{t-1}) \sim N\left(m_{t-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, C_{t-1} \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{\delta} \end{pmatrix} \right)$

and $p(\theta_{t-1} | \theta_t D_{t-1}) \sim N(m_{t-1} + B_t(\theta_t - a_t), C_{t-1} - B_t R_t B_t^T)$

$\Rightarrow B_t = \delta$ $\Rightarrow N(m_{t-1}(1-\delta) + \delta\theta_t , C_{t-1} \cdot \delta)$

That is, in restrospecive dist of θ_{t-1}, δ play a role of weight that averaging θ_{t-1}'s mean and new θ_t, and we can also find that $C_{t-1} \to C_{t-1} - \delta$ which becomes smaller by new θ_t

By δ computations becomes much easier.