微信扫一扫进行教学评价

第 11 讲：动态线性模型

张伟平

目录

1.1 动态线性模型

一元动态线性模型 (DLM) 的一般形式为

$$\begin{aligned} Y_t &= \mathbf{F}_t\boldsymbol{\theta}_t + \nu_t, \nu_t \sim \mathcal{N}_m\left(0, V_t\right) && \text{测量向量} \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \omega_t \sim \mathcal{N}_p\left(0, W_t\right). && \text{状态向量} \end{aligned}$$

Definition

其中 ν_t 和 $\boldsymbol{\omega}_t$ 为零均值的测量误差和状态信息 (innovations) 向量且相互独立, $\boldsymbol{\theta}_0 \sim N_p\left(m_0, C_0\right)$.

这类模型是线性**状态空间**(state space) 模型, 其中 $x_t = \mathbf{F}_t\boldsymbol{\theta}_t$ 表示信号, $\boldsymbol{\theta}_t$ 为状态向量, $\mathbf{F}_t$ 为回归向量以及 $\mathbf{G}_t$ 为状态矩阵. 时间序列常见的趋势和季节性特征都可以以这种形式进行建模.

平稳的 AR(2) 模型

平稳 AR(2) 模型可以表示为

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad \epsilon_t \sim NID\left(0, \tau^2\right).$$

定义状态向量 $\boldsymbol{\theta}_t = (y_t, y_{t-1})'$ 因此转移方程为

$$\boldsymbol{\theta}_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$$

以及测量方程为

$$y_t = (1, 0)\boldsymbol{\theta}_t + v_t.$$

按序推断例子

- 一个慢变水平模型为

$$
\begin{aligned}
y_t &= \theta_t + \nu_t \\
\theta_t &= \theta_{t-1} + \omega_t
\end{aligned}
$$

观测点绕着均值根据随机游走进行波动. 如果记 $\mathcal{D}_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$ 为在时间 $t-1$ 时刻时候的累积信息, 假设 $\theta_{t-1}|\mathcal{D}_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$, 以及误差 $\nu_t \sim \mathcal{N}(0, V_t)$, $\omega_t \sim \mathcal{N}(0, W_t)$ 和 $\boldsymbol{\theta}_0 \sim N_p(m_0, C_0)$, 则我们有

1. θ_t 的先验分布为

$$
\begin{aligned}
\theta_t|\mathcal{D}_{t-1} &\sim \mathcal{N}(m_{t-1}, R_t) \quad \text{其中} \\
R_t &= C_{t-1} + W_t
\end{aligned}
$$

2. y_t 的一步预测分布为

$$y_t|\mathcal{D}_{t-1} \sim \mathcal{N}\left(m_{t-1}, Q_t\right)$$
$$Q_t = R_t + V_t$$

3. θ_t 给定 $\mathcal{D}_t = \{\mathcal{D}_{t-1}, y_t\}$ 的后验分布为

$$\theta_t|\mathcal{D}_t \sim \mathcal{N}\left(m_t, C_t\right) \text{ 其中}$$
$$m_t = m_{t-1} + A_t e_t$$
$$A_t = R_t / Q_t$$
$$e_t = y_t - m_{t-1}$$
$$C_t = R_t - A_t^2 Q_t$$

- 动态线性回归模型

$$y_t = \mathbf{F}_t \boldsymbol{\theta}_t + \nu_t$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

如果误差项, ν_t 和 ω_t 服从正态分布, e.g. $\nu_t \sim \mathcal{N}(0, V_t), \omega_t \sim \mathcal{N}(0, \mathbf{W}_t)$ 和 $\boldsymbol{\theta}_0 \sim N_p(m_0, C_0)$, 则进行贝叶斯统计分析是直接的.

1. θ_t 的先验分布为

$$\boldsymbol{\theta}_t | \mathcal{D}_{t-1} \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{R}_t)$$

$$\mathbf{R}_t = \mathbf{C}_{t-1} + \mathbf{W}_t$$

2. y_t 的一步预测分布为

$$y_t | \mathcal{D}_{t-1} \sim \mathcal{N}(f_t, Q_t) \text{ 其中}$$

$$f_t = \mathbf{F}_t \mathbf{m}_{t-1}$$

$$Q_t = \mathbf{F}_t \mathbf{R}_t \mathbf{F}_t' + V_t$$

3. θ_t, y_t 的联合分布为

$$\begin{pmatrix} \boldsymbol{\theta}_t \\ y_t \end{pmatrix} \Big| \mathcal{D}_{t-1} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m}_{t-1} \\ \mathbf{F}_t \mathbf{m}_{t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_t & \mathbf{R}_t \mathbf{F}_t' \\ \mathbf{F}_t \mathbf{R}_t & Q_t \end{pmatrix} \right)$$

4. θ_t 给定 $D_t = \{\mathcal{D}_{t-1}, y_t\}$ 的后验分布为

$$\boldsymbol{\theta}_t | \mathcal{D}_t \sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t) \text{ 其中}$$

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{A}_t e_t$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}_t^T$$

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t' Q_t^{-1}$$

$$e_t = y_t - f_t$$

DLM 的一般定理

定理 1. *(filtering recursions)*

(1) 状态的一步向前预报密度可以基于过滤密度 $p(\theta_{t-1}|\mathcal{D}_{t-1})$ 由下式给出

$$p(\theta_t|\mathcal{D}_{t-1}) = \int p(\theta_t|\theta_{t-1})\, p(\theta_{t-1}|\mathcal{D}_{t-1})\, \mathrm{d}\nu(\theta_{t-1})$$

(2) 观测的一步向前预报密度可以基于状态的预测密度给出

$$f(y_t|\mathcal{D}_{t-1}) = \int f(y_t|\theta_t)\, p(\theta_t|\mathcal{D}_{t-1})\, \mathrm{d}\nu(\theta_t)$$

(3) 过滤密度可以基于前述密度给出为

$$p(\theta_t|\mathcal{D}_t) = \frac{f(y_t|\theta_t)\, p(\theta_t|\mathcal{D}_{t-1})}{f(y_t|\mathcal{D}_{t-1})}$$

证明. (1) 注意到 $\theta_t \perp \mathcal{D}_{t-1}|\theta_{t-1}$, 因此

$$\begin{aligned}
p\left(\theta_t|\mathcal{D}_{t-1}\right) &= \int p\left(\theta_{t-1},\theta_t|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_{t-1}\right)\\
&= \int p\left(\theta_t|\theta_{t-1},\mathcal{D}_{t-1}\right)p\left(\theta_{t-1}|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_{t-1}\right)\\
&= \int p\left(\theta_t|\theta_{t-1}\right)p\left(\theta_{t-1}|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_{t-1}\right)
\end{aligned}$$

(2) 由条件独立性 $Y_t \perp \mathcal{D}_{t-1}|\theta_t$, 我们有

$$\begin{aligned}
f\left(y_t|\mathcal{D}_{t-1}\right) &= \int f\left(y_t,\theta_t|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_t\right)\\
&= \int f\left(y_t|\theta_t,\mathcal{D}_{t-1}\right)p\left(\theta_t|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_t\right)\\
&= \int f\left(y_t|\theta_t\right)p\left(\theta_t|\mathcal{D}_{t-1}\right)\mathrm{d}\nu\left(\theta_t\right)
\end{aligned}$$

(3) 利用贝叶斯公式和条件独立性 $Y_t \perp \mathcal{D}_{t-1}|\theta_t$ 有

$$p(\theta_t|\mathcal{D}_t) = \frac{f(y_t|\theta_t, \mathcal{D}_{t-1})\, p(\theta_t|\mathcal{D}_{t-1})}{f(y_t|\mathcal{D}_{t-1})} = \frac{f(y_t|\theta_t)\, p(\theta_t|\mathcal{D}_{t-1})}{f(y_t|\mathcal{D}_{t-1})}$$

□

上述结果可以用来递归计算 k 步向前预报密度 $(k \geq 1)$:

$$p(\theta_{t+k}|\mathcal{D}_t) = \int p(\theta_{t+k}|\theta_{t+k-1})\, p(\theta_{t+k-1}|\mathcal{D}_t)\, \mathrm{d}\nu(\theta_{t+k-1})$$

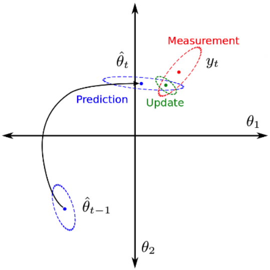
和

$$f(y_{t+k}|\mathcal{D}_t) = \int f(y_{t+k}|\theta_{t+k})\, p(\theta_{t+k}|\mathcal{D}_t)\, \mathrm{d}\nu(\theta_{t+k})$$

1.2 Kalman filter

- 卡尔曼滤波 (Kalman filter) 是一种均方误差意义下最优线性状态估计方法. 应用范围非常广泛
  - 在空中交通管制系统中使用雷达角度和范围的测量值来确定飞机接近机场的三维位置和速度
  - 使用电池单元的外部电气和热条件的测量值确定混合动力电动汽车电池组的内部“充电状态”和“健康状态”
  - 使用对地球磁场, GPS 和加速度的测量来确定我自己在三个维度上的位置, 方向和速度
  - 使用金融市场数据预测交易性证券和商品的未来价格

Kalman filter 的核心思想是利用**预测 + 测量反馈**来达到最优估计.



Kalman 滤波是一种递归过程，主要有两个更新过程：时间更新和观测更新，其中时间更新主要包括状态预测和协方差预测，主要是对系统的预测，而观测更新主要包括计算卡尔曼增益、状态更新和协方差更新，因此整个递归过程主要包括五个方面的计算：1) 状态预测；2）协方差预测；3）卡尔曼增益；4）状态更新；5）协方差更新；

定理 2. *(Kalman filter)* 对一元 *DLM*,

$$\begin{aligned} Y_t &= \mathbf{F}_t\boldsymbol{\theta}_t + \nu_t \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \mathbf{H}_{t-1}\mathbf{u}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

其中 $\nu_t \sim \mathcal{N}(0, V_t)$ 和 $\omega_t \sim \mathcal{N}(0, \mathbf{W}_t)$, 假设先验分布 $\boldsymbol{\theta}_{t-1}|\mathcal{D}_{t-1} \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$, $\mathbf{u}_t$ 为系统控制输入变量, $\boldsymbol{\theta}_0 \sim N_p(m_0, C_0)$. 我们有

1. θ_t 的先验分布为

$$\boldsymbol{\theta}_t|\mathcal{D}_{t-1} \sim \mathcal{N}(\mathbf{a}_t, \mathbf{R}_t) \text{ 其中}$$

$$\begin{aligned} \mathbf{a}_t &= \mathbf{G}_t\mathbf{m}_{t-1} + \mathbf{H}_{t-1}\mathbf{u}_{t-1} \\ \mathbf{R}_t &= \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t^T + \mathbf{W}_t \end{aligned}$$

2. 一步预测分布为

$$
\begin{aligned}
y_t|\mathcal{D}_{t-1} &\sim \mathcal{N}(f_t, Q_t) \ \text{其中} \\
f_t &= \mathbf{F}_t \mathbf{a}_t \\
Q_t &= \mathbf{F}_t \mathbf{R}_t \mathbf{F}_t + V_t
\end{aligned}
$$

3. 后验分布 $\theta_t|D_t$ 为

$$
\begin{aligned}
\boldsymbol{\theta}_t|\mathcal{D}_t &\sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t) \ \text{其中} \\
\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \\
\mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}_t^T \\
&= (\mathbf{R}_t^{-1} + \mathbf{F}_t' \mathbf{V}_t^{-1} \mathbf{F}_t)^{-1} \\
\mathbf{A}_t &= \mathbf{R}_t \mathbf{F}_t' Q_t^{-1} \\
e_t &= y_t - f_t
\end{aligned}
$$

根据定理2结果有

- 记 $\hat{\boldsymbol{\theta}}_{t|t-1} = E[\boldsymbol{\theta}|\mathcal{D}_{t-1}]$, $\boldsymbol{m}_{t-1} = \hat{\boldsymbol{\theta}}_{t-1}$, 则 t 时刻的状态预测

$$\hat{\boldsymbol{\theta}}_{t|t-1} = \mathbf{G}_t \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{H}_{t-1}\mathbf{u}_{t-1}$$

- 协方差预测

$$\mathbf{R}_{t|t-1} = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^T + \mathbf{W}_t$$

- 卡尔曼增益

$$\mathbf{A}_t = \mathbf{R}_{t|t-1} \mathbf{F}'_t Q_t^{-1}$$

- t 时刻状态更新

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t|t-1} + \mathbf{A}_t(\mathbf{y}_t - \mathbf{F}_t \hat{\boldsymbol{\theta}}_{t|t-1})$$

- t 时刻协方差更新

$$\mathbf{C}_t = \mathbf{R}_{t|t-1} - \mathbf{A}_t \mathbf{F}_t \mathbf{R}_{t|t-1}$$

1.3 光滑

在时间序列中，我们经常对 Y_t 观察一定时期, $t = 1, \ldots, T$, 记 $\mathcal{D}_t = (Y_1, \ldots, Y_T)$, 则可以从 $p(\theta_T|\mathcal{D}_T)$ 出发使用向后迭代算法计算条件密度 $\theta_t|\mathcal{D}_T, t < T$.

定理 3. *(Smoothing recursion) (1)* 给定 $\mathcal{D}_T$, 状态序列 $(\theta_0, \ldots, \theta_T)$ 具有向后转移概率

$$p\left(\theta_t|\theta_{t+1}, \mathcal{D}_T\right) = \frac{p\left(\theta_{t+1}|\theta_t\right) p\left(\theta_t|\mathcal{D}_t\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)}$$

(2) $\theta_t|\mathcal{D}_T, t = T-1, T-2, \ldots, 1$ 的光滑密度可以由下式计算

$$p\left(\theta_t|\mathcal{D}_T\right) = p\left(\theta_t|\mathcal{D}_t\right) \int \frac{p\left(\theta_{t+1}|\theta_t\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)} p\left(\theta_{t+1}|\mathcal{D}_T\right) d\mu\left(\theta_{t+1}\right)$$

证明. (1) 注意到 $\theta_{t+1} \perp \left(\theta_0, \ldots, \theta_{t-1}\right)|\theta_t, \mathcal{D}_T$ 和 $\theta_t \perp \left(Y_{t+1}, \ldots, Y_T\right)|\theta_{t+1}$,

由贝叶斯公式有

$$
\begin{aligned}
p\left(\theta_t|\theta_{t+1},\mathcal{D}_T\right) &= p\left(\theta_t|\theta_{t+1},\mathcal{D}_t\right) \\
&= \frac{p\left(\theta_t|\mathcal{D}_t\right)p\left(\theta_{t+1}|\theta_t,\mathcal{D}_t\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)} \\
&= \frac{p\left(\theta_t|\mathcal{D}_t\right)p\left(\theta_{t+1}|\theta_t\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)}
\end{aligned}
$$

(2) 对 $p\left(\theta_t,\theta_{t+1}|\mathcal{D}_T\right)$ 边际化有

$$
\begin{aligned}
p\left(\theta_t|\mathcal{D}_T\right) &= \int p\left(\theta_t,\theta_{t+1}|\mathcal{D}_T\right)d\theta_{t+1} = \int p\left(\theta_{t+1}|\mathcal{D}_T\right)p\left(\theta_t|\theta_{t+1},\mathcal{D}_T\right)d\theta_{t+1} \\
&= \int p\left(\theta_{t+1}|\mathcal{D}_T\right)p\left(\theta_t|\theta_{t+1},\mathcal{D}_t\right)d\theta_{t+1} \\
&= \int p\left(\theta_{t+1}|\mathcal{D}_T\right)\frac{p\left(\theta_{t+1}|\theta_t,\mathcal{D}_t\right)p\left(\theta_t|\mathcal{D}_t\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)}d\theta_{t+1},\ \text{using Bayes rule} \\
&= p\left(\theta_t|\mathcal{D}_t\right)\int p\left(\theta_{t+1}|\theta_t\right)\frac{p\left(\theta_{t+1}|\mathcal{D}_T\right)}{p\left(\theta_{t+1}|\mathcal{D}_t\right)}d\theta_{t+1}
\end{aligned}
$$

□

定理 4. *对定理 1 中所定义模型, 如果 $\theta_{t+1}|\mathcal{D}_T \sim \mathcal{N}(s_{t+1}, S_{t+1})$, 则 $\theta_t|\mathcal{D}_T \sim \mathcal{N}(s_t, S_t)$, 其中*

$$\begin{aligned} s_t &= m_t + C_t G'_{t+1} R^{-1}_{t+1}(s_{t+1} - a_{t+1}) \\ S_t &= C_t + C_t G'_{t+1} R^{-1}_{t+1}(S_{t+1} - R_{t+1}) R^{-1}_{t+1} G_{t+1} C_t \end{aligned}$$

证明. 由多元正态性质, 我们很容易得到 $\theta_t|\mathcal{D}_T$ 的条件密度为正态, 因此只需要计算期望和协方差矩阵. 我们有

$$s_t = \mathrm{E}(\theta_t|\mathcal{D}_T) = \mathrm{E}(\mathrm{E}(\theta_t|\theta_{t+1}, \mathcal{D}_T)|\mathcal{D}_T)$$

和

$$S_t = \mathrm{Var}(\theta_t|\mathcal{D}_T) = \mathrm{Var}(\mathrm{E}(\theta_t|\theta_{t+1}, \mathcal{D}_T)|\mathcal{D}_T) + \mathrm{E}(\mathrm{Var}(\theta_t|\theta_{t+1}, \mathcal{D}_T)|\mathcal{D}_T)$$

现在注意到, $\theta_t \perp (Y_{t+1}, \ldots, Y_T)|\theta_{t+1}$ 因此 $p(\theta_t|\theta_{t+1}, \mathcal{D}_T) = p(\theta_t|\theta_{t+1}, \mathcal{D}_t)$ 可以通过贝叶斯公式计算, 此处 $p(\theta_{t+1}|\theta_t, \mathcal{D}_t) = p(\theta_{t+1}|\theta_t)$

通过下述状态方程

$$\theta_{t+1} = G_{t+1}\theta_t + w_{t+1}, w_{t+1} \sim N\left(0, W_{t+1}\right)$$

即 $\theta_{t+1}|\theta_t \sim \mathcal{N}\left(G_{t+1}\theta_t, W_{t+1}\right)$. $p\left(\theta_t|\mathcal{D}_t\right)$ 承担先验角色, 其为 $\mathcal{N}\left(m_t, C_t\right)$.
我们有

$$\begin{aligned}
\mathrm{E}\left(\theta_t|\theta_{t+1}, \mathcal{D}_t\right) &= m_t + C_t G'_{t+1}\left(G_{t+1}C_t G'_{t+1} + W_{t+1}\right)^{-1}\left(\theta_{t+1} - G_{t+1}m_t\right) \\
&= m_t + C_t G'_{t+1} R_{t+1}^{-1}\left(\theta_{t+1} - a_{t+1}\right) \\
\mathrm{Var}\left(\theta_t|\theta_{t+1}, \mathcal{D}_t\right) &= C_t - C_t G'_{t+1} R_{t+1}^{-1} G_{t+1} C_t
\end{aligned}$$

由此

$$\begin{aligned}
s_t &= \mathrm{E}\left(\mathrm{E}\left(\theta_t|\theta_{t+1}, \mathcal{D}_t\right)|\mathcal{D}_T\right) = m_t + C_t G'_{t+1} R_{t+1}^{-1}\left(s_{t+1} - a_{t+1}\right) \\
S_t &= \mathrm{Var}\left(\mathrm{E}\left(\theta_t|\theta_{t+1}, \mathcal{D}_t\right)|\mathcal{D}_T\right) + \mathrm{E}\left(\mathrm{Var}\left(\theta_t|\theta_{t+1}, \mathcal{D}_t\right)|\mathcal{D}_T\right) \\
&= C_t - C_t G'_{t+1} R_{t+1}^{-1} G_{t+1} C_t + C_t G'_{t+1} R_{t+1}^{-1} S_{t+1} R_{t+1}^{-1} G_{t+1} C_t \\
&= C_t + C_t G'_{t+1} R_{t+1}^{-1}\left(S_{t+1} - R_{t+1}\right) R_{t+1}^{-1} G_{t+1} C_t
\end{aligned}$$

根据假设即为 $\mathrm{E}\left(\theta_{t+1}|\mathcal{D}_T\right) = s_{t+1}$ 和 $\mathrm{Var}\left(\theta_{t+1}|\mathcal{D}_T\right) = S_{t+1}$.

$\square$

1.4 预报

很多情形下我们感兴趣 k 步向前预报:

$$\begin{array}{ccccccc} \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots & \longrightarrow & \theta_{t+k} \\ | & & & & & & | \\ (Y_1,\ldots,Y_t) & & & & & & Y_{t+k} \end{array}$$

信息从 $(Y_1,\ldots,Y_t)$ 流到 Y_{t+k}.

对 $k \geq 1$, 定义

$$a_t(k) = \mathrm{E}\left(\theta_{t+k}|\mathcal{D}_t\right)$$

$$R_t(k) = \mathrm{Var}\left(\theta_{t+k}|\mathcal{D}_t\right)$$

$$f_t(k) = \mathrm{E}\left(Y_{t+k}|\mathcal{D}_t\right)$$

$$Q_t(k) = \mathrm{Var}\left(Y_{t+k}|\mathcal{D}_t\right)$$

定理 5. 令 $a_t(0)=m_t$ 和 $R_t(0)=C_t$, 则对 $k\geq 1$ 下面结论成立:

(1) θ_{t+k} 给定 $\mathcal{D}_t$ 的分布是正态, 均值和协方差分别为

$$a_t(k)=G_{t+k}a_t(k-1)$$
$$R_t(k)=G_{t+k}R_t(k-1)G'_{t+k}+W_{t+k}$$

(2) Y_{t+k} 给定 $\mathcal{D}_t$ 的条件分布为正态分布, 均值和协方差分别为

$$f_t(k)=F_{t+k}a_t(k)$$
$$Q_t(k)=F_{t+k}R_t(k)F'_{t+k}+V_t$$

证明. 条件分布为正态分布容易得出, 因此我们只需要计算均值和协

方差. 下面我们对 k 进行归纳. $k=1$ 显然, 对 $k>1$ 有

$$
\begin{aligned}
a_t(k) =& \mathrm{E}\left(\theta_{t+k}|\mathcal{D}_t\right) = \mathrm{E}\left(\mathrm{E}\left(\theta_{t+k}|\mathcal{D}_t, \theta_{t+k-1}\right)|\mathcal{D}_t\right) \\
=& \mathrm{E}\left(G_{t+k}\theta_{t+k-1}|\mathcal{D}_t\right) = G_{t+k}a_t(k-1) \\
R_t(k) =& \mathrm{Var}\left(\theta_{t+k}|\mathcal{D}_t\right) = \mathrm{Var}\left(\mathrm{E}\left(\theta_{t+k}|\mathcal{D}_t, \theta_{t+k-1}\right)|\mathcal{D}_t\right) \\
& + \mathrm{E}\left(\mathrm{Var}\left(\theta_{t+k}|\mathcal{D}_t, \theta_{t+k-1}\right)|\mathcal{D}_t\right) \\
=& G_{t+k}R_{t,k-1}G'_{t+k} + W_{t+k} \\
f_t(k) =& \mathrm{E}\left(Y_{t+k}|\mathcal{D}_t\right) = \mathrm{E}\left(\mathrm{E}\left(Y_{t+k}|\mathcal{D}_t, \theta_{t+k}\right)|\mathcal{D}_t\right) \\
=& \mathrm{E}\left(F_{t+k}\theta_{t+k}|\mathcal{D}_t\right) = F_{t+k}a_t(k) \\
Q_t(k) =& \mathrm{Var}\left(Y_{t+k}|\mathcal{D}_t\right) = \mathrm{Var}\left(\mathrm{E}\left(Y_{t+k}|\mathcal{D}_t, \theta_{t+k}\right)|\mathcal{D}_t\right) \\
& + \mathrm{E}\left(\mathrm{Var}\left(Y_{t+k}|\mathcal{D}_t, \theta_{t+k}\right)|\mathcal{D}_t\right) \\
=& F_{t+k}R_t(k)F'_{t+k} + V_{t+k}
\end{aligned}
$$

$\square$

1.5 应用例子

1. **卡车位置和速度** 考虑一辆在无摩擦的直线导轨上的卡车. 最初卡车固定在位置 0, 但其受到随机不受控制的外力抖动. 我们每 $\triangle t$ 秒测量一次卡车的位置, 但是这些测量结果并不精确. 我们构建了刻画卡车位置和速度的模型.

- 位置和速度通过线性状态空间

$$\mathbf{x}_k = \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$$

表示, 其中 $\dot{x}$ 为速度, 即位置对时间的导数.

- 我们假设在 $k-1$ 和 k 时刻之间, 不可控制外力导致一个常数加速度 a_k, 其服从均值 0, 方差 σ_a^2 的正态分布, 因此

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{G}a_k$$

其中

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \frac{1}{2}\Delta t^2 \\ \Delta t \end{bmatrix}$$

因此等价地

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{w}_k$$

其中

$$\mathbf{w}_k \sim N(0, \mathbf{Q})$$

$$\mathbf{Q} = \mathbf{G}\mathbf{G}^\top \sigma_a^2 = \begin{bmatrix} \frac{1}{4}\Delta t^4 & \frac{1}{2}\Delta t^3 \\ \frac{1}{2}\Delta t^3 & \Delta t^2 \end{bmatrix} \sigma_a^2$$

矩阵 Q 不是满秩的 (秩为 1, 如果 $\triangle t \neq 0$). 因此 $N(0, Q)$ 不是绝对连续的, 即不存在概率密度. 另外一种避免这种情况的表示为 $\mathbf{w}_k \sim \mathbf{G} \cdot N\left(0, \sigma_a^2\right)$

- 在每个时间步, 卡车真实位置的一个带噪观测为

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k$$

  其中 $\mathbf{H} = [1, 0]$, $\mathbf{v}_k \sim N(0, \sigma_v^2)$ 为随机噪音.

- 我们知道初始时候

$$\hat{\mathbf{x}}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  如果我们知道初始时候真实的位置和速度, 我们可以指定一个值为 0 的初始协方差矩阵 $\mathbf{P}_{0|0}$, 否则可以指定一个合适的协方差矩阵.

- 则卡尔曼滤波步骤为

$$
\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k \left[\mathbf{z}_k - \mathbf{H}_k \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}\right] \\
\mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^\top + \mathbf{Q}_k \\
\mathbf{S}_k &= \mathbf{R}_k + \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k^\top \mathbf{S}_k^{-1} \\
\mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}
\end{aligned}
$$

2. Xu & Zhang (2015) 预测股票价格

记 x_t 表示价格, 状态变量到当前时间是可观测的 ($z_t = x_t$)

$$\begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} + w_t \begin{bmatrix} T^2/2 \\ T \end{bmatrix}$$

$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} + v_t$$

或者表示为矩阵形式

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + w_t$$

$$z_t = \mathbf{H}\mathbf{x}_t + v_t$$

其中 $T = 1$ 为简单起见.

- 初始化:

$$\hat{\mathbf{x}}_{0|0} = \mathbb{E}[\mathbf{x}_0] = \mu_0$$
$$\mathbf{P}_{0|0} = \mathbb{E}[(\mathbf{x}_0 - \mu_0)^T(\mathbf{x}_0 - \mu_0)] = \mathbf{P}_0$$

- 预测

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t}$$
$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^T + \ \mathbf{Q}\ ^T$$

- 卡尔曼增益

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t+1|t}\mathbf{H}^T + R)^{-1}$$

- 更新

$$
\begin{aligned}
y_{t+1} &= z_{t+1} - \mathbf{H}\hat{\mathbf{x}}_{t+1|t} \\
\hat{\mathbf{x}}_{t+1|t+1} &= \hat{\mathbf{x}}_{t+1|t} + y_{t+1}\mathbf{K}_{t+1} \\
\mathbf{P}_{t+1|t+1} &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H})\mathbf{P}_{t+1|t}
\end{aligned}
$$