状态空间模型

陈堰平

chen@yanping.me

中国人民大学 统计学院

Nov.28,2007

模型形式

设y_n为q维时间序列。下面的模型被称为**状态空间模型**（state space model）

$$\underset{(p\times 1)}{x_t} = \underset{(p\times p)}{F_t}\ \underset{(p\times 1)}{x_{t-1}} + \underset{(p\times 1)}{v_t} \qquad (\text{状态方程}) \tag{1}$$

$$\underset{(q\times 1)}{y_t} = \underset{(q\times p)}{H_t}\ \underset{(p\times 1)}{x_t} + \underset{(q\times 1)}{w_t} \qquad (\text{观测方程}) \tag{2}$$

- x_t为不能直接观测的p维向量，称为**状态**(state)。v_t称为系统噪声或状态噪声，w_t称为观测噪声。设$v_t \overset{i.i.d}{\sim} N(0, Q_t)_p$，$w_t \overset{i.i.d}{\sim} N(0, R_t)_q$
- 用于时序分析的各种线性模型几乎都能用该状态空间模型表示，从而用统一的算法进行处理。
- 假设初始状态$x_0 \sim N(\mu_0, \Sigma_0)_p$

模型解释

1. 如果将(2)式看成时间序列y_t的回归模型的话，x_t就可以看成回归系数。这时(1)式就可以看成是表达回归系数x_t 随时间变化规律的模型。
2. 把x_t看成要估计的信号，则状态方程(1)式就成为表达信号x_t发生机制的模型，而观测模型(2)式就表示在进行信号的观测时，将信号进行某种变换时加进噪声的观测机制。

模型举例

局部水平模型（local level model）

观测方程为：

$$X_t = \mu_t + n_t \tag{3}$$

不可观测的局部水平μ_t假设服从随机游走模型：

$$\mu_t = \mu_{t-1} + w_t \tag{4}$$

方程(4)是状态方程。假设n_t和w_t方差分别是σ_n^2和σ_w^2，σ_w^2/σ_n^2被称为信噪比（signal-to-noise ratio）。主要用来描述没有长期趋势或者季节性但是有短期相关性的数据。

$$
\begin{aligned}
X_t &= \mu_t + n_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1,t} \\
\beta_t &= \beta_{t-1} + w_{2,t}
\end{aligned}
\tag{5}
$$

改写成矩阵形式：

$$
\begin{aligned}
X_t &= (1, 0)\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + n_t \\
\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix}
\end{aligned}
\tag{6}
$$

- β_t是趋势项，如果β_t为常数，则局部水平μ_t随时间变化线性增长
- X_t的二阶差分平稳，自相关函数和MA(2)的相同

$$
\begin{aligned}
X_t &= \mu_t + i_t + n_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1,t} \\
\beta_t &= \beta_{t-1} + w_{2,t} \\
i_t &= -\sum_{j=1}^{s-1} i_{t-j} + W_{3,t}
\end{aligned} \tag{7}
$$

μ_t表示局部水平，β_t表示局部趋势，i_t表示局部季节指数，s表示一年（或一季度）的期数

- 关于趋势与周期，见Harvey(1985)，Andrews(1994)

以$s=4$为例，即一年分四个季度的情况。设状态向量$\theta_t=(\mu_t,\beta_t,i_t,i_{t-1},i_{t-2})'$ 将基本结构模型方程改写成：

$$
\begin{aligned}
X_t &= (1,0,1,0,0)\theta_t + n_t \\
\theta_t &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix} \theta_{t-1} + \begin{pmatrix} w_{1,t} \\ w_{2,t} \\ w_{3,t} \\ 0 \\ 0 \end{pmatrix}
\end{aligned} \tag{8}
$$

$$X_t = a_t + b_t u_t +_t \tag{9}$$

设$\theta_t = (a_t, b_t)'$，$h_t = (1, u_t)$，则上式可以写成：

$$\begin{aligned} X_t &= h_t\theta_t + n_t \\ \theta_t &= \theta_{t-1} + w_t \end{aligned} \tag{10}$$

$$
\begin{aligned}
r_t &= \alpha_t + \beta_t r_{M,t} + e_t, \qquad e_t \sim N(0, \sigma_e^2) \\
\alpha_{t+1} &= \alpha_t + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2) \\
\beta_{t+1} &= \beta_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)
\end{aligned} \tag{11}
$$

这里r_t是某种资产的超出收益率（excess return），$r_{M,t}$是市场组合的超出收益率。$\{e_t, \eta_t, \epsilon_t\}$相互独立。

把模型改写成矩阵形式：

$$
\begin{aligned}
r_t &= (1, r_{M,t}) \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + e_t \\
\begin{pmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}
\end{aligned}
\tag{12}
$$

预测、滤波与平滑

状态空间模型的重要问题之一是用时间序列y_t的观测值来估计状态x_t。考虑用观测值$Y_j = \{y_1, y_2, \cdots, y_j\}$来求时刻$t$的状态$x_t$的估计值的问题

- 预测： $j < t$，利用到目前为止的观测数据估计未来状态
- 滤波： $j = t$，估计观测终点的状态，即估计现在状态
- 平滑： $j > t$，用到现在为止的观测值去估计过去状态

进行状态估计的Kalman滤波算法

要解决这些状态估计问题，只要求出在给定观测值Y_j时状态x_t的条件分布$p(x_t|Y_j)$即可。(1)式和(2)式都是线性的，v_t、w_t及x_0全都服从正态分布，所以$p(x_t|Y_j)$也服从正态分布，因此状态估计问题实质上就是求条件分布的均值和协方差阵的问题。
设状态x_t的条件均值向量和协方差阵为：

$$
\begin{aligned}
x_{t|j} &\equiv \mathrm{E}(x_t|Y_j) \\
V_{t|j} &\equiv \mathrm{E}[(x_t - x_{t|j})(x_t - x_{t|j})']
\end{aligned}
\tag{13}
$$

用Kalman滤波处理的是$j = t - 1$ (1期预测)和$j = t$ (滤波)这两种情况。反复交替计算1期预测和滤波，就能依次求出所需的各量。

进行状态估计的Kalman滤波算法

一步预测

$$
\begin{aligned}
x_{t|t-1} &= F_t x_{t-1|t-1} \\
V_{t|t-1} &= F_t V_{t-1|t-1} F_t' + Q_t
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
K_t &= V_{t|t-1} H_t^{'} (H_t V_{t|t-1} H_t^{'} + R_t)^{-1} \\
x_{t|t} &= x_{t|t-1} + K_t (y_t - H_t x_{t|t-1}) \\
&= K_t y_t + (I - K_t H_t) x_{t|t-1} \\
V_{t|t} &= (I - K_t H_t) V_{t|t-1} \\
&= V_{t|t-1} - K_t H_t V_{t|t-1}
\end{aligned} \tag{15}
$$

滤波只用到时刻t为止的观测值来估计x_t，而平滑则是用所有观测值来进行状态的估计。因此，通过平滑一般来说可以得到比滤波精度更高的状态估计。

固定区间平滑：

$$\begin{aligned} A_t &= V_{t|t} F'_{t+1} V^{-1}_{t+1|t} \\ x_{t|N} &= x_{t|t} + A_t(x_{t+1|N} - x_{t+1|t}) \\ V_{t|N} &= V_{t|t} + A_t(V_{t+1|N} - V_{t+1|t})A'_t \end{aligned} \tag{16}$$

其中$t < N$.为了进行平滑，首先要用滤波计算出$\{x_{t|t-1}, x_{t|t}, V_{t|t-1}, V_{t|t}\}$ $(t = 1, 2, \ldots, N)$，然后用(16)式所示的算法，从$x_{N-1|N}$，$V_{N-1|N}$出发用后向前依次计算，直至求得$x_{1|N}$，$V_{1|N}$

考虑用时序数据$Y_t = \{y_1, y_2, \cdots, y_t\}$来估计$h$期以后的状态$x_{t+h}$，$(h > 1)$这一长期预测的问题。

- 还未观测到y_{t+1}，则在形式上有$Y_{t+1} = Y_t$。显然有$x_{t+1|t+1} = x_{t+1|t}$，$V_{t+1|t}$成立，这意味着为了第二步预测省略了对y_{t+1}的滤波，只执行预测这一步就可以了。
- 一般地，因$Y_t = Y_{t+1} = \cdots = Y_{t+h}$成立，所以只要将1期预测的算法反复多次即可。

$$
\begin{aligned}
x_{t+i|t} &= F_{t+i} x_{t+i-1|t} \\
V_{t+i|t} &= F_{t+i} V_{t+i-1|t} F'_{t+i} + Q_{t+i} \\
&(i = 1, 2, \cdots, h)
\end{aligned}
\tag{17}
$$

用Kalman滤波和平滑算法进行递推计算的过程

$$
\begin{array}{cccccccccc}
x_{1|0} & \longrightarrow & x_{2|0} & \longrightarrow & x_{3|0} & \longrightarrow & x_{4|0} & \longrightarrow & x_{5|0} & \longrightarrow \\
\Downarrow & & & & & & & & & \\
x_{1|1} & \Rightarrow & x_{2|1} & \longrightarrow & x_{3|1} & \longrightarrow & x_{4|1} & \longrightarrow & x_{5|1} & \longrightarrow \\
 & & \Downarrow & & & & & & & \\
x_{1|2} & \longleftarrow & x_{2|2} & \Rightarrow & x_{3|2} & \longrightarrow & x_{4|2} & \longrightarrow & x_{5|2} & \longrightarrow \\
 & & & & \Downarrow & & & & & \\
x_{1|3} & \longleftarrow & x_{2|3} & \longleftarrow & x_{3|3} & \Rightarrow & x_{4|3} & \longrightarrow & x_{5|3} & \longrightarrow \\
 & & & & & & \Downarrow & & & \\
x_{1|4} & \longleftarrow & x_{2|4} & \longleftarrow & x_{3|4} & \longleftarrow & x_{4|4} & \Rightarrow & x_{5|4} & \longrightarrow \\
 & & & & & & & & \Downarrow &
\end{array}
$$

$\Rightarrow$表示预测，$\Downarrow$表示滤波，$\longleftarrow$表示平滑，$\longrightarrow$表示长期预测

时间序列的预测

设在给定Y_t的条件下y_{t+h}的均值向量和协方差阵分别为$y_{t+h|t} = \mathrm{E}(y_{t+h}|Y_t)$ 和$d_{t+h|t} = \mathrm{Cov}(y_{t+h}|Y_t)$，则

$$\begin{aligned} y_{t+h|t} &= \mathrm{E}(H_{t+h}x_{t+h} + w_{t+h}|Y_t) \\ &= H_{t+h}x_{t+h|t} \end{aligned} \tag{18}$$

$$\begin{aligned} d_{t+h|t} &= \mathrm{Cov}(H_{t+h}x_{t+h} + w_{t+h}|Y_t) \\ &= H_{t+h}\mathrm{Cov}(x_{t+h|Y_t})H'_{t+h} + H_{t+h}\mathrm{Cov}(x_{t+h}, w_{t+h}|Y_t) \\ &\quad + \mathrm{Cov}(w_{t+h}, x_{t+h}|Y_t)H'_{t+h} + \mathrm{Cov}(w_{t+h}|Y_t) \\ &= H_{t+h}V_{t+h|t}H'_{t+h} + R_{t+h} \end{aligned} \tag{19}$$

时间序列欠测值的处理

- 在观测过程中，由于观测仪器的故障等偶然因素，或者由于观测对象和观测机制的物理性质的限制等原因，导致部分数据观测不到，这种实际上没有观测到的数据叫做欠测值（missing observation）。
- 如果只用连续观测的那部分数据，就会使数据的实际利用率变低。
- 用0或数据的均值来替代、或用直线插值法估计，会使分析结果产生较大偏差。
- 如果用时间序列的状态空间模型，对含有欠测值的数据也能精密地计算其似然函数。因此，在这种情况下也能求得参数的最大似然估计。

设$I(t)$为到时刻t为止有观测值处的观测时刻的集合。显然，在没有欠测值时，$I(t)=\{1,2,\cdots,t\}$。设$Y_t\equiv\{y_i|i\in I(t)\}$，则似然函数可以表达如下：

$$\begin{aligned} L(\theta) &= \log p(Y_N) \\ &= \sum_{n\in I(N)} \log p(y_n|Y_{n-1},\theta) \end{aligned} \tag{20}$$

其中$p(y_n|Y_{n-1},\theta)$是给定观测值Y_{n-1}时y_n的分布。

提示

$$\text{由条件概率公式} \qquad \mathrm{P}(A\cap B)=\mathrm{P}(A|B)\mathrm{P}(B)$$

$$\mathrm{P}(y_1,y_2,\cdots,y_n|\theta)=\mathrm{P}(y_n|y_1,y_2,\cdots,y_{n-1},\theta)\mathrm{P}(y_1,y_2,\cdots,y_{n-1}|\theta)$$

- 与长期预测类似，在y_n为欠测值时，因为有$Y_n \equiv Y_{n-1}$成立，所以在Kalman滤波算法中，只要省略滤波值的计算即可
- 对所有的时刻n都实施预测计算，而只对满足$n \in I(N)$的时刻n实施滤波计算
- 这样，在有欠测值的情况下也能依次求出状态的预测分布$p(x_n|Y_{n-1})$

如果用(18)式和(19)式求得$y_{n|n-1}$和$d_{n|n-1}$，则时间序列的对数似然函数可以由下式给出

$$
\begin{aligned}
L(\theta) = -\frac{1}{2} \sum_{n \in I(N)} \Big\{ & q \log 2\pi + \log |d_{n|n-1}| \\
& + (y_n - y_{n|n-1})' d_{n|n-1}^{-1} (y_n - y_{n|n-1}) \Big\}
\end{aligned}
\tag{21}
$$

欠测值的插值步骤：

- 对有欠测值之处省略滤波计算，同时用Kalman滤波求出预测值的分布$\{x_{n|n-1}, V_{n|n-1}\}$ 和滤波值的分布$\{x_{n|n}, V_{n|n}\}$
- 用平滑算法(16)即可求出对应欠测值的状态的平滑值$x_{n|N}$
- 欠测值y_n的估计值由$y_{n|N} = H_n x_{n|N}$给出，估计误差的方差为$d_{n|N} = H_n V_{n|N} H_n' + R_n$

参数估计

- Newton-Raphson算法
- EM算法
  - 见shumway and Stoffer(1982)

外生变量（或固定的输入）会进入状态或观测中，假设输入向量为u_t

$$\underset{(p\times 1)}{x_t} = \underset{(p\times p)}{F_t}\underset{(p\times 1)}{x_{t-1}} + \underset{(p\times r)}{\Upsilon_t}\underset{(r\times 1)}{u_t} + \underset{(p\times 1)}{v_t} \qquad \text{(状态方程)}$$

$$\underset{(q\times 1)}{y_t} = \underset{(q\times p)}{H_t}\underset{(p\times 1)}{x_t} + \underset{(q\times r)}{\Gamma_t}\underset{(r\times 1)}{u_t} + \underset{(q\times 1)}{w_t} \qquad \text{(观测方程)}$$

模型扩展
非高斯型状态空间模型

$$\underset{(p\times 1)}{x_t} = \underset{(p\times p)}{F_t}\underset{(p\times 1)}{x_{t-1}} + \underset{(p\times 1)}{v_t} \qquad (\text{状态方程})$$

$$\underset{(q\times 1)}{y_t} = \underset{(q\times p)}{H_t}\underset{(p\times 1)}{x_t} + \underset{(q\times 1)}{w_t} \qquad (\text{观测方程})$$

- 假设系统噪声v_t和观测噪声w_t服从密度函数为$\varphi(v)$ 和$\psi(w)$的白噪声。与前面的状态空间模型不同的是，系统噪声或观测噪声不一定服从正态分布。在这种情况下，一般来说状态向量x_t也是非高斯的，在这样的假设下，模型被称为非高斯型状态空间模型（non-Gaussian state space model）。
- 对于非高斯型状态空间模型，必需用Kalman滤波计算状态向量的分布。

$$\boxed{p(x_t|x_{t-1}, Y_{t-1}) = p(x_t|x_{t-1}), \quad p(y_t|x_t, Y_{t-1}) = p(y_t|x_t)}$$

一步预测：

$$
\begin{aligned}
p(x_t|Y_{t-1}) &= \int_{-\infty}^{\infty} p(x_t, x_{t-1}|Y_{t-1})\, dx_{t-1} \\
&= \int_{-\infty}^{\infty} p(x_t|x_{t-1}, Y_{t-1})p(x_{t-1}|Y_{t-1})\, dx_{t-1} \\
&= \int_{-\infty}^{\infty} p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})\, dx_{t-1}
\end{aligned} \tag{22}
$$

滤波：

$$p(x_t|Y_t) = p(x_t|y_t, Y_{t-1}) = \frac{p(y_t|x_t, Y_{t-1})p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} = \frac{p(y_t|x_t)p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} \tag{23}$$

其中$p(y_t|Y_{t-1})$可由$\int_{-\infty}^{\infty} p(y_t|x_t)p(x_t|Y_{t-1})\, dx_t$ 求得

设$N > t$，利用$p(x_t|x_{t+1}, Y_N) = p(x_t|x_{t+1}, Y_n)$可得下式：

$$\begin{aligned} p(x_t, x_{t+1}|Y_N) &= p(x_{t+1}|Y_N)p(x_t|x_{t+1}, Y_N) \\ &= p(x_{t+1}|Y_N)p(x_t|x_{t+1}, Y_n) \\ &= p(x_{t+1}|Y_N)\frac{p(x_t|Y_t)p(x_{t+1}|x_t, Y_t)}{p(x_{t+1}|Y_t)} \\ &= p(x_{t+1}|Y_N)\frac{p(x_t|Y_t)p(x_{t+1}|x_t)}{p(x_{t+1}|Y_t)} \end{aligned} \tag{24}$$

由此，得出平滑公式：

$$\begin{aligned} p(x_t|Y_N) &= \int_{-\infty}^{\infty} p(x_t, x_{t+1}|Y_N)\, dx_{t+1} \\ &= p(x_t|Y_t)\int_{-\infty}^{\infty} p(x_{t+1}|Y_N)\frac{p(x_{t+1}|Y_N)p(x_{t+1}|x_t)}{p(x_{t+1}|Y_t)}\, dx_{t+1} \end{aligned} \tag{25}$$

其他扩展

- 状态空间模型的Bootstrap方法
  - 见Stoffer and Wall(1991)
- 重要性抽样
  - Jungbacker and Koopman(2005)
- DLM with Switching
- Monte Carlo 方法与状态空间模型
  - 非线性模型
  - 非高斯型模型
  - 见Carlin,Polson and Stoffer(1992)，Durbin and Koopman(1997)
- Stochastic Volatility 模型
- ……

Thank You!