

# EM with Filter

Renhe W.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>EM with Filter</b>	<b>3</b>

## List of Figures

## List of Tables

# 1 Introduction

状态空间模型的求解有很多方式，这里介绍一种比较特殊的方法，这个方法由Shumway and Stoffer (1982) 提出，这个方法结合了通常的 Kalman Filter (Kalman, 1960) 以及 EM 算法 (Dempster et al., 1977)，详细的内容如下：

考虑一个状态空间模型，在这个模型中，感兴趣的  $p \times 1$  理想化向量序列  $\mathbf{x}_t$  并没有被直接观测到，而是作为随机回归模型中的一个组成部分：

$$\mathbf{y}_t = \mathbf{M}_t \mathbf{x}_t + \mathbf{v}_t, \quad t = 1, 2, \dots, n \quad (1.1)$$

其中  $M_t$  是一个已知的  $q \times p$  设计矩阵，它表达了将未观测的随机向量  $\mathbf{x}_t$  转换为  $q \times 1$  观测序列  $\mathbf{y}_t$  的模式。误差或噪声项  $\mathbf{v}_t, t = 1, \dots, n$  被假定为具有公共  $q \times q$  协方差矩阵  $R$  的零均值非相关正态分布噪声向量。随机序列  $\mathbf{x}_t$  被认为是主要兴趣的对象；它被建模为一阶多变量过程，形式为

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t, \quad t = 1, \dots, n, \quad (1.2)$$

其中  $\Phi$  是一个  $p \times p$  过渡矩阵，描述了底层序列在连续时间段内的移动方式。由于我们没有对  $\Phi$  的特征方程的根作出特别假设，因此过程  $\mathbf{x}_t$  可能是非平稳的。初始值  $\mathbf{x}_0$  被假定为具有均值向量  $\boldsymbol{\mu}$  和  $p \times p$  协方差矩阵  $\Sigma$  的正态随机向量。 $p \times 1$  噪声项  $\mathbf{w}_t, t = 1, \dots, n$  是具有公共协方差矩阵  $Q$  的零均值非相关正态向量。

模型(1.1)和(1.2)的动机源于希望分别考虑模型误差  $\mathbf{w}_t$  定义的模型中的不确定性和通过测量噪声过程  $\mathbf{v}_t$  表达的对模型进行测量的不确定性。将(1.1)视为时间序列的一种随机效应模型可能会有所帮助，其中效应向量  $\mathbf{x}_t$  具有由多变量自回归模型(1.2)强加的时间相关结构。在这个背景下，它是普通自回归 AR 模型的推广，考虑了观测噪声以及模型引入的噪声。人们可以将  $M_t$  视为固定的设计矩阵，它定义了我们观察向量  $\mathbf{x}_t$  的组成部分的方式。例如，在这篇论文中，它提供了一种方便的方法来处理因缺失观测而引入的不完整数据问题。

平滑或预测程序的主要目标是使用观测序列  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  来估计未观测序列  $\mathbf{x}_t$ ，对于  $t = 1, 2, \dots, n$  (平滑) 和  $t = n + 1, n + 2, \dots$  (预测)。如果知道参数  $\boldsymbol{\mu}, \Sigma, \Phi, Q$  和  $R$  的值，可以将传统的卡尔曼平滑估计器计算为条件期望，并且将具

有**最小均方误差**. 这相当于将过程  $\mathbf{x}_t$  视为在贝叶斯意义上依赖于假设的先验值的随机参数向量.

### Note

通常, 与 Newton-Raphson 或评分校正的高度非线性外观相比, EM 方程呈现出一种简单、直观上吸引人的形式. 当然, 由于在 EM 程序中从未计算过二阶偏导数矩阵, 因此它无法提供估计的标准误差; 这些偏导数仍可以通过在最大值附近扰动似然函数来近似. 另一个缺点是, 在迭代过程的后期, EM 算法可能收敛缓慢; 在这个阶段, 人们可能希望切换到另一种算法.

本文将在缺失数据背景下的平滑或预测任务解释为基本上是估计状态空间模型(1.1)和(1.2)中的随机过程  $\mathbf{x}_t$  的问题. 如果参数  $\mu, \Sigma, \Phi, Q$  和  $R$  是已知的, 那么基于观测数据的条件均值提供了最小均方误差解. 如果这些参数事先未指定, 将使用 EM 算法通过最大似然进行估计. 这需要传统的条件均值和协方差的递归形式, 以及在附录 A 中给出的新递归. 我们还展示了可以容忍非常普遍的缺失观测模式, 并指出了一种用于调整估计器的校正程序. 提供了一个例子, 涉及对经济序列的平滑和预测, 该序列在感兴趣的时间段内只被两个不同的来源部分观测到.

## 2 EM with Filter

根据定义的状态空间模型(1.1)和(1.2), 则对数完整数据  $x_0, x_1, \dots, x_n, y_1, \dots, y_n$  似然函数可以写成以下形式:

$$\begin{aligned} \log L \triangleq & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{x}_0 - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}) \\ & - \frac{n}{2} \log |Q| - \frac{1}{2} \sum_{t=1}^n (\mathbf{x}_t - \Phi \mathbf{x}_{t-1})' Q^{-1} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1}) \\ & - \frac{n}{2} \log |R| - \frac{1}{2} \sum_{t=1}^n (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t)' R^{-1} (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t) \end{aligned} \quad (2.1)$$

其中  $\log L$  需要针对参数  $\mu, \Sigma, \Phi, Q$  和  $R$  进行最大化. 由于上述对数似然依赖于未观测数据序列  $\mathbf{x}_t, t = 0, 1, \dots, n$ , 我们考虑对观测序列  $y_1, y_2, \dots, y_n$  应用条件 EM

算法. 也就是说, 定义在第  $(r+1)$  次迭代中的估计参数为最大化

$$G(\mu, \Sigma, \Phi, Q, R) = E_r(\log L \mid \mathbf{y}_1, \dots, \mathbf{y}_n) \quad (2.2)$$

的值  $\mu, \Sigma, \Phi, Q, R$ , 其中  $E_r$  表示相对于包含第  $r$  次迭代值  $\mu(r), \Sigma(r), \Phi(r), Q(r)$  和  $R(r)$  的密度的条件期望. Dempster et al. (1977) 已经展示, 定义为这样一系列步骤的迭代程序能产生非递减的似然值, 且固定点被定义为似然函数的一个稳定点.

为了计算(2.2)中定义的条件期望, 定义条件均值

$$\mathbf{x}_t^s = E(\mathbf{x}_t \mid \mathbf{y}_1, \dots, \mathbf{y}_s) \quad (2.3)$$

和协方差函数

$$P_t^s = \text{cov}(\mathbf{x}_t \mid \mathbf{y}_1, \dots, \mathbf{y}_s) \quad (2.4)$$

以及

$$P_{t,t-1}^s = \text{cov}(\mathbf{x}_t, \mathbf{x}_{t-1} \mid \mathbf{y}_1, \dots, \mathbf{y}_s). \quad (2.5)$$

例如, 随机向量  $\mathbf{x}_t^n$  是常用的卡尔曼滤波估计器, 而  $\mathbf{x}_t^n, t = 0, 1, \dots, n$  是基于所有观测数据的  $\mathbf{x}_t$  的最小均方误差平滑估计器. 对于  $t > n$  的随机向量  $\mathbf{x}_t^n$  是底层序列的预测值. 使用标准卡尔曼滤波结果 (参见Jazwinski (2007)) 计算  $\mathbf{x}_t^n$  和  $P_t^n$  的一组递归公式在附录 A 中给出; 我们还提供了一种新的递归方法来计算协方差  $P_{t,t-1}^n$ .

现在, 在(2.1)中取条件期望得到

$$\begin{aligned} G(\mu, \Sigma, \Phi, Q, R) = & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} (P_0^n + (x_0^n - \mu)(x_0^n - \mu)') \right\} \\ & - \frac{n}{2} \log |Q| - \frac{1}{2} \text{tr} \left\{ Q^{-1} (C - B\Phi' - \Phi B' + \Phi A\Phi') \right\} \\ & - \frac{n}{2} \log |R| \\ & - \frac{1}{2} \text{tr} \left\{ R^{-1} \sum_{t=1}^n [(y_t - M_t x_t^n)(y_t - M_t x_t^n)' + M_t P_t^n M_t'] \right\} \end{aligned} \quad (2.6)$$

其中  $\text{tr}$  表示迹, 公式如下:

$$A = \sum_{t=1}^n (P_{t-1}^n + x_{t-1}^n x_{t-1}^{n'}) \quad (2.7)$$

$$B = \sum_{t=1}^n (P_{t,t-1}^n + x_t^n x_{t-1}^{n'}), \quad (2.8)$$

和

$$C = \sum_{t=1}^n (P_t^n + x_t^n x_t^{n'}). \quad (2.9)$$

卡尔曼滤波项  $x_t^n, P_t^n$  和  $P_{t,t-1}^n$  是在参数值  $\mu(r), \Phi(r), Q(r), R(r)$  下使用附录 A 中的递归计算的. 此外, 不难看出选择

$$\Phi(r+1) = BA^{-1}, \quad (2.10)$$

$$Q(r+1) = n^{-1} (C - BA^{-1}B'), \quad (2.11)$$

和

$$R(r+1) = n^{-1} \sum_{t=1}^n [(y_t - M_t x_t^n)(y_t - M_t x_t^n)' + M_t P_t^n M_t']$$

可以最大化似然函数(2.6)的最后两行. 第一项类似于多元正态似然的单次复制, 因此可以取  $\mu_0(r+1) = x_0^n$  并将  $\Sigma$  的值固定在某个合理的基线水平. 关于在复制的卡尔曼滤波模型中估计  $\mu$  和  $\Sigma$  的研究可见于 Shumway 等人 (1981).

在某些情况下, 人们可能想要限制  $\Phi$  的元素.

例如, 在限制条件

$$\Phi F = G,$$

下, 其中  $F$  和  $G$  是指定的  $p \times s (s \leq p)$  矩阵, 我们得到约束估计器

$$\Phi_c = \Phi - (\Phi F - G) (F' A^{-1} F)^{-1} F' A^{-1}$$

和

$$Q_c = Q + (\Phi F - G) (F' A^{-1} F)^{-1} (\Phi F - G)'$$

对应于(2.10)和(2.11), 其中参数  $(r+1)$  出于记号方便已被省略. 可以使用“创新”

形式（参见 Gupta 和 Mehra（1974））在每个阶段计算对数似然函数的值

$$\begin{aligned} \log L \doteq & -\frac{1}{2} \sum_{t=1}^n \log |M_t P_t^{t-1} M_t' + R_t| \\ & -\frac{1}{2} \sum_{t=1}^n (y_t - M_t x_t^{t-1})' (M_t P_t^{t-1} M_t' + R_t)^{-1} (y_t - M_t x_t^{t-1}). \end{aligned}$$

## References

- Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.
- Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.
- Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960.
- Robert H Shumway and David S Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis*, 3(4):253–264, 1982.