

第10章 状态空间模型

2025年11月

本讲介绍

机器学习领域可根据模型的学习目标和建模方式分为生成式和判别式。

- 判别式模型专注于学习条件概率分布 $P(Y|X)$ ，给定输入数据 X 时，如何准确地预测输出标签 Y ，如神经网络、决策树等。
- 生成式模型以学习数据的联合概率分布 $P(X, Y)$ 为目标，理解数据的生成过程，并预测生成全新的、与训练数据具有相似特征的数据点 (X, Y) 。

本讲介绍

状态空间模型是生成式机器学习中一种重要的概率模型。

本章将介绍状态空间模型的定义、联合分布、滤波、预测、平滑及边际似然计算等，线性高斯状态空间模型及卡尔曼滤波与平滑算法，非线性和非高斯状态空间模型的近似推断方法，以及状态空间模型中的参数估计方法。

难点内容：后验分布计算过程、高维积分处理、EM算法在参数估计中的应用原理。

本讲关键问题

- (1) 状态空间模型的基本定义与结构是什么？
- (2) 什么是滤波、预测、平滑？
- (3) 推导状态空间模型中的滤波、平滑算法步骤。
- (4) 线性高斯状态空间模型的假设是什么？
- (5) 推导卡尔曼滤波与平滑算法的步骤。
- (6) 对于非线性和非高斯状态空间模型，有哪些近似推断方法？什么原理？
- (7) 状态空间模型中如何进行参数估计？

关键词：状态空间模型、滤波、预测、平滑

一、状态空间模型 - 模型定义

状态空间模型 (SSM) 是一种部分可观测的马尔可夫模型，可以表示为如下形式的随机离散时间非线性动态系统：

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{u}_t, \mathbf{q}_t) \quad (1)$$

$$\mathbf{y}_t = h(\mathbf{z}_t, \mathbf{u}_t, \mathbf{y}_{1:t-1}, \mathbf{r}_t) \quad (2)$$

其中：

- $\mathbf{z}_t \in \mathbb{R}^{N_z}$ 是隐藏状态，
- $\mathbf{u}_t \in \mathbb{R}^{N_u}$ 是可能存在的观测输入，
- $\mathbf{y}_t \in \mathbb{R}^{N_y}$ 是观测输出，
- f 是转移函数，
- \mathbf{q}_t 是过程噪声，
- h 是观测函数，
- \mathbf{r}_t 是观测噪声。

隐藏状态 \mathbf{z}_t 按照马尔可夫过程随时间演化，每个隐藏状态在每个时间生成一些观测值 \mathbf{y}_t

一、状态空间模型 - 联合分布

状态空间模型 $\mathbf{z}_{1:T}$ 与 $\mathbf{y}_{1:T}$ 的联合分布：

$$p(\mathbf{y}_{1:T}, \mathbf{z}_{1:T} | \mathbf{u}_{1:T}) = \left[p(\mathbf{z}_1 | \mathbf{u}_1) \prod_{t=2}^T p(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{u}_t) \right] \\ \times \left[\prod_{t=1}^T p(\mathbf{y}_t | \mathbf{z}_t, \mathbf{u}_t, \mathbf{y}_{1:t-1}) \right]$$

一、状态空间模型 - 联合分布

- 如果当前观测值 \mathbf{y}_t 仅依赖于当前的 $\mathbf{u}_t, \mathbf{z}_t$ 以及前一个观测值 \mathbf{y}_{t-1} ，那么我们得到一个自回归状态空间模型，如图 11.1(a) 所示。
- 如果没有外部输入 \mathbf{u}_t ， \mathbf{y}_t 在给定隐藏状态 \mathbf{z}_t 的条件下彼此独立，则该模型进一步简化为无条件生成模型，如图 11.1(b) 所示。联合分布简化为：

$$p(\mathbf{y}_{1:T}, \mathbf{z}_{1:T}) = \left[p(\mathbf{z}_1) \prod_{t=2}^T p(\mathbf{z}_t | \mathbf{z}_{t-1}) \right] \left[\prod_{t=1}^T p(\mathbf{y}_t | \mathbf{z}_t) \right]$$

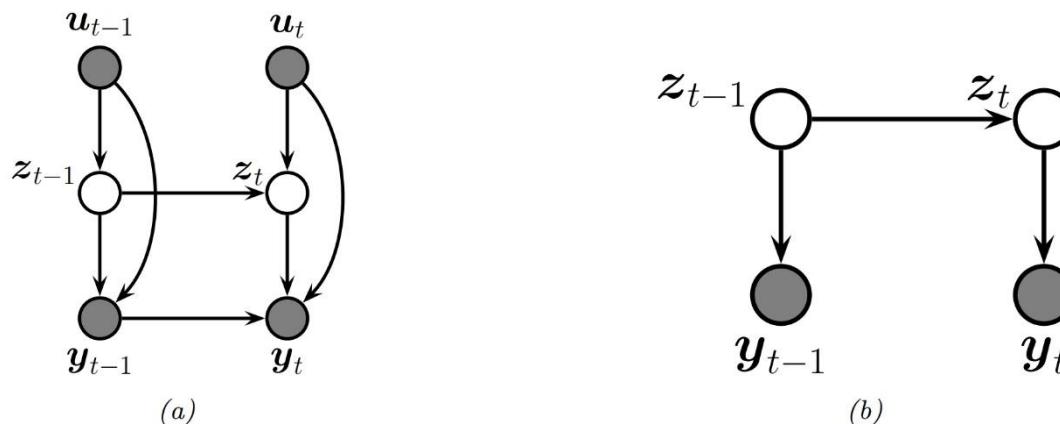


图11.1 状态空间模型示例

一、状态空间模型

在状态空间模型中，通常需要解决以下几个推断：

- 滤波 (Filtering)
- 预测 (Prediction)
- 平滑 (Smoothing)
- 边际似然

一、状态空间模型 - 滤波

滤波的目标是在给定过去和当前观测值 $\mathbf{y}_{1:t}$ 的情况下，推断隐藏状态 \mathbf{z}_t 的后验分布：

$$\alpha_t(\mathbf{z}_t) = p(\mathbf{z}_t | \mathbf{y}_{1:t}).$$

滤波需要使用贝叶斯定理和马尔可夫性质进行递归计算，包括以下三个步骤：

1. 初始话：设置先验分布 $p(\mathbf{z}_0)$ ；
2. 预测：

$$p(\mathbf{z}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{z}_t | \mathbf{z}_{t-1}) \alpha_{t-1}(\mathbf{z}_{t-1}) d\mathbf{z}_{t-1},$$

其中 $\alpha_{t-1}(\mathbf{z}_{t-1})$ 是时间 $t-1$ 时刻的后验分布。

3. 更新：

$$\begin{aligned} \alpha_t(\mathbf{z}_t) &= \frac{p(\mathbf{y}_t | \mathbf{z}_t) p(\mathbf{z}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})} \\ &\propto p(\mathbf{y}_t | \mathbf{z}_t) \int p(\mathbf{z}_t | \mathbf{z}_{t-1}) \alpha_{t-1}(\mathbf{z}_{t-1}) d\mathbf{z}_{t-1}. \end{aligned}$$

一、状态空间模型 - 预测

- 预测的目标是在给定当前和过去观测数据 $\mathbf{y}_{1:t}$ 的情况下，估计未来观测值 $\mathbf{y}_{t'} (t' > t)$ 的边际分布：

$$p(\mathbf{y}_{t'} | \mathbf{y}_{1:t}).$$

- 对隐藏状态进行边际化处理：

$$p(\mathbf{y}_{t'} | \mathbf{y}_{1:t}) = \int p(\mathbf{y}_{t'} | \mathbf{z}_{t'}) p(\mathbf{z}_{t'} | \mathbf{y}_{1:t}) d\mathbf{z}_{t'}.$$

其中 $p(\mathbf{z}_{t'} | \mathbf{y}_{1:t})$ 可通过状态转移模型递归计算：

$$p(\mathbf{z}_{t'} | \mathbf{y}_{1:t}) = \int p(\mathbf{z}_{t'} | \mathbf{z}_{t'-1}) p(\mathbf{z}_{t'-1} | \mathbf{y}_{1:t}) d\mathbf{z}_{t'-1}.$$

一、状态空间模型 - 平滑

平滑的目标是在给定整个观测序列 $\mathbf{y}_{1:T}$ 的情况下，估计过去隐藏状态 $\mathbf{z}_t (t < T)$ 的后验分布：

$$\gamma_t(\mathbf{z}_t) = p(\mathbf{z}_t | \mathbf{y}_{1:T}).$$

平滑利用未来的观测信息，因此比滤波更加准确。

一、状态空间模型 - 平滑

平滑分布可以使用前向-后向算法计算：

1. 前向计算：使用滤波算法得到 $\alpha_t(\mathbf{z}_t)$ 。
2. 后向计算：后向分布 $\beta_t(\mathbf{z}_t)$ 是给定 \mathbf{z}_t 时，观测值 $\mathbf{y}_{t+1:T}$ 的条件概率。根据贝叶斯法则，后向分布 $\beta_t(\mathbf{z}_t)$ 满足：

$$\begin{aligned}\beta_t(\mathbf{z}_t) &= p(\mathbf{y}_{t+1:T} | \mathbf{z}_t), \\ &= \int p(\mathbf{y}_{t+2:T}, \mathbf{y}_{t+1}, \mathbf{z}_{t+1} | \mathbf{z}_t) d\mathbf{z}_{t+1} \\ &= \int p(\mathbf{y}_{t+2:T} | \mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{z}_t) p(\mathbf{y}_{t+1} | \mathbf{z}_{t+1}, \mathbf{z}_t) p(\mathbf{z}_{t+1} | \mathbf{z}_t) d\mathbf{z}_{t+1}, \\ &= \int p(\mathbf{y}_{t+2:T} | \mathbf{z}_{t+1}) p(\mathbf{y}_{t+1} | \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} | \mathbf{z}_t) d\mathbf{z}_{t+1}, \\ &= \int \beta_{t+1}(\mathbf{z}_{t+1}) p(\mathbf{y}_{t+1} | \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1} | \mathbf{z}_t) d\mathbf{z}_{t+1}.\end{aligned}$$

初始条件为 $\beta_T(\mathbf{z}_T) = 1$ 。

3. 平滑计算：平滑分布可计算为：

$$\gamma_t(\mathbf{z}_t) = \frac{\alpha_t(\mathbf{z}_t) \beta_t(\mathbf{z}_t)}{\int \alpha_t(\mathbf{z}_t) \beta_t(\mathbf{z}_t) d\mathbf{z}_t} \propto \alpha_t(\mathbf{z}_t) \beta_t(\mathbf{z}_t).$$

一、状态空间模型 - 平滑

2*. 如果已知 $\gamma_{t+1}(\mathbf{z}_{t+1})$, 则有:

$$p(\mathbf{z}_t, \mathbf{z}_{t+1} | \mathbf{y}_{1:T}) = p(\mathbf{z}_t | \mathbf{z}_{t+1}, \mathbf{y}_{1:T}) \gamma_{t+1}(\mathbf{z}_{t+1}).$$

利用马尔可夫性质, 可以得到:

$$\begin{aligned} p(\mathbf{z}_t | \mathbf{z}_{t+1}, \mathbf{y}_{1:T}) &= p(\mathbf{z}_t | \mathbf{z}_{t+1}, \mathbf{y}_{1:t}), \\ &= \frac{p(\mathbf{z}_t, \mathbf{z}_{t+1} | \mathbf{y}_{1:t})}{p(\mathbf{z}_{t+1} | \mathbf{y}_{1:t})}, \\ &= \frac{p(\mathbf{z}_{t+1} | \mathbf{z}_t) \alpha_t(\mathbf{z}_t)}{\int p(\mathbf{z}_{t+1} | \mathbf{z}_t) \alpha_t(\mathbf{z}_t) d\mathbf{z}_t}. \end{aligned}$$

因此, $\gamma_t(\mathbf{z}_t)$ 也满足以下递归关系:

$$\begin{aligned} \gamma_t(\mathbf{z}_t) &= \int p(\mathbf{z}_t, \mathbf{z}_{t+1} | \mathbf{y}_{1:T}) d\mathbf{z}_{t+1}, \\ &= \alpha_t(\mathbf{z}_t) \int \frac{p(\mathbf{z}_{t+1} | \mathbf{z}_t) \gamma_{t+1}(\mathbf{z}_{t+1})}{\int p(\mathbf{z}_{t+1} | \mathbf{z}_t) \alpha_t(\mathbf{z}_t) d\mathbf{z}_t} d\mathbf{z}_{t+1}. \end{aligned}$$

平滑分布的初始条件是 $\gamma_T(\mathbf{z}_T) = \alpha_T(\mathbf{z}_T)$ 。

一、状态空间模型 - 边际似然

观测序列 $\mathbf{y}_{1:T}$ 的边际分布为：

$$p(\mathbf{y}_{1:T}|\theta) = \prod_{t=1}^T p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta).$$

每个边际分布可以使用预测算法求得：

$$\begin{aligned} & p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta) \\ &= \iint p(\mathbf{y}_t|\mathbf{z}_t, \theta) p(\mathbf{z}_t|\mathbf{z}_{t-1}, \theta) p(\mathbf{z}_{t-1}|\mathbf{y}_{1:t-1}, \theta) d\mathbf{z}_{t-1} d\mathbf{z}_t. \end{aligned}$$

或者对 $p(\mathbf{y}_{1:T}, \mathbf{z}_{1:T}|\theta)$ 在 $\mathbf{z}_{1:T}$ 上积分得到 $\mathbf{y}_{1:T}$ 的边际分布

$$p(\mathbf{y}_{1:T}|\theta) = \int \cdots \int p(\mathbf{y}_{1:T}, \mathbf{z}_{1:T}|\theta) d\mathbf{z}_1 \cdots d\mathbf{z}_t$$

二、线性高斯状态空间模型

本节将讨论一种特定的状态空间模型：线性高斯状态空间模型（Linear Gaussian State-space Model, LG-SSM）或线性动态系统（linear dynamical system, LDS）。

该模型的假设是，转移模型和观测模型均为线性关系，同时隐藏状态和观测值服从高斯分布。LS-SSM模型因其推断具有解析形式从而特别重要，并为更复杂的模型提供了基础。

二、线性高斯状态空间模型 - 模型定义

线性高斯状态空间模型定义如下：

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + B\mathbf{u}_t + \mathbf{q}_t, \quad \mathbf{q}_t \sim \mathcal{N}(0, Q) \quad (3)$$

$$\mathbf{y}_t = C\mathbf{z}_t + D\mathbf{u}_t + \mathbf{r}_t, \quad \mathbf{r}_t \sim \mathcal{N}(0, R) \quad (4)$$

其中：

- A, B, C, D 是系统的线性变换矩阵，
- \mathbf{q}_t 是过程噪声，其协方差矩阵为 Q ，
- \mathbf{r}_t 是观测噪声，其协方差矩阵为 R ，
- 初始隐藏状态服从正态分布： $\mathbf{z}_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ 。

由于初始状态 \mathbf{z}_1 和所有噪声变量都是正态分布，整个系统都服从多元正态分布。

二、线性高斯状态空间模型 - 多元正态分布性质

- 定理1

假设 x 在给定 z 的条件下服从正态分布：

$$x|z \sim \mathcal{N}(\mu, \Sigma_1),$$

且 y 在给定 x, z 的条件下服从正态分布：

$$y|(x, z) \sim \mathcal{N}(Ax + B, \Sigma_2),$$

那么 x, y 在给定 z 的条件下服从联合正态分布：

$$\begin{pmatrix} x \\ y \end{pmatrix} \Big| z \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ A\mu + B \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_1 A^T \\ A\Sigma_1 & A\Sigma_1 A^T + \Sigma_2 \end{pmatrix} \right).$$

二、线性高斯状态空间模型 - 多元正态分布性质

- 定理2

假设 x, y 是联合正态分布的随机变量,

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right),$$

那么:

$$1. E(x|y) = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y);$$

$$2. Var(x|y) = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}.$$

二、线性高斯状态空间模型 - 滤波

对于 LG-SSM, 滤波称为卡尔曼滤波, 滤波分布仍然是高斯分布:

$$\mathbf{z}_t | \mathbf{y}_{1:t} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t}).$$

递归滤波包含以下两个步骤:

1. 预测: 给定

$$\mathbf{z}_{t-1} | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\mu_{t-1|t-1}, \Sigma_{t-1|t-1}),$$

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + B\mathbf{u}_t + \mathbf{q}_t, \quad \mathbf{q}_t \sim \mathcal{N}(0, Q),$$

由定理 1 得:

$$\begin{pmatrix} \mathbf{z}_{t-1} \\ \mathbf{z}_t \end{pmatrix} \Big| \mathbf{y}_{1:t-1} \sim \mathcal{N} \left(\begin{pmatrix} \mu_{t-1|t-1} \\ \mu_{t|t-1} \end{pmatrix}, \begin{pmatrix} \Sigma_{t-1|t-1} & \Sigma_{t-1|t-1} A^T \\ A \Sigma_{t-1|t-1} & \Sigma_{t|t-1} \end{pmatrix} \right), \quad (5)$$

$$\mu_{t|t-1} = A\mu_{t-1|t-1} + B\mathbf{u}_t, \quad (6)$$

$$\Sigma_{t|t-1} = A\Sigma_{t-1|t-1}A^T + Q. \quad (7)$$

二、线性高斯状态空间模型 - 滤波

2. 更新: 给定

$$\mathbf{z}_t | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\mu_{t|t-1}, \Sigma_{t|t-1}),$$

$$\mathbf{y}_t = C\mathbf{z}_t + D\mathbf{u}_t + \mathbf{r}_t, \quad \mathbf{r}_t \sim \mathcal{N}(0, R),$$

由定理 1 得:

$$\begin{pmatrix} \mathbf{z}_t \\ \mathbf{y}_t \end{pmatrix} \Bigg| \mathbf{y}_{1:t-1} \sim \mathcal{N} \left(\begin{pmatrix} \mu_{t|t-1} \\ C\mu_{t|t-1} + D\mathbf{u}_t \end{pmatrix}, \begin{pmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1}C^T \\ C\Sigma_{t|t-1} & C\Sigma_{t|t-1}C^T + R \end{pmatrix} \right)$$

然后由定理 2 得:

$$\mathbf{z}_t | (\mathbf{y}_t, \mathbf{y}_{1:t-1}) \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t}),$$

$$\mu_{t|t} = \mu_{t|t-1} + K_t(\mathbf{y}_t - C\mu_{t|t-1} - D\mathbf{u}_t),$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t C \Sigma_{t|t-1},$$

$$K_t = \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + R)^{-1}.$$

其中 K_t 称为卡尔曼增益矩阵。

二、线性高斯状态空间模型 - 平滑

对于 LG-SSM, 平滑过程称为 Rauch-Tung-Striebel 平滑器或卡尔曼平滑器。平滑分布仍然是高斯分布:

$$\mathbf{z}_t | \mathbf{y}_{1:T} \sim \mathcal{N}(\mu_{t|T}, \Sigma_{t|T}).$$

平滑分布的递归计算如下:

1. 先执行卡尔曼滤波, 从前向后计算 $\mu_{t+1|t}$ 、 $\mu_{t|t}$ 、 $\Sigma_{t+1|t}$ 以及 $\Sigma_{t|t}$ 。
2. 根据马尔可夫性质和公式(5): $p(\mathbf{z}_t | \mathbf{z}_{t+1}, \mathbf{y}_{1:T}) = p(\mathbf{z}_t | \mathbf{z}_{t+1}, \mathbf{y}_{1:t}) \sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t)$,

$$\bar{\mu}_t = \mu_{t|t} + J_t(\mathbf{z}_{t+1} - \mu_{t+1|t}),$$

$$\bar{\Sigma}_t = \Sigma_{t|t} - J_t \Sigma_{t+1|t} J_t^T,$$

$$J_t = \Sigma_{t|t} A^T \Sigma_{t+1|t}^{-1}.$$

其中 J_t 称为平滑增益 (smooth gain)。

给定 $\mathbf{z}_{t+1} | \mathbf{y}_{1:T} \sim \mathcal{N}(\mu_{t+1|T}, \Sigma_{t+1|T})$, 由定理1:

$$\mathbf{z}_t | \mathbf{y}_{1:T} \sim \mathcal{N}(\mu_{t|T}, \Sigma_{t|T}),$$

$$\mu_{t|T} = \mu_{t|t} + J_t(\mu_{t+1|T} - \mu_{t+1|t}),$$

$$\Sigma_{t|T} = J_t \Sigma_{t+1|T} J_t^T + \bar{\Sigma}_t = \Sigma_{t|t} + J_t(\Sigma_{t+1|T} - \Sigma_{t+1|t}) J_t^T.$$

二、线性高斯状态空间模型 - 联合分布与边际似然

基于线性高斯假设，我们可以写出 $\mathbf{z}_{1:T}$ 和 $\mathbf{y}_{1:T}$ 的联合分布：

$$p(\mathbf{y}_{1:T}, \mathbf{z}_{1:T} | \mathbf{u}_{1:T}) = p(\mathbf{z}_1) \prod_{t=2}^T p(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{u}_t) \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{z}_t, \mathbf{u}_t).$$

每个条件分布都是多元正态分布：

$$p(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{u}_t) = \mathcal{N}(A\mathbf{z}_{t-1} + B\mathbf{u}_t, Q),$$

$$p(\mathbf{y}_t | \mathbf{z}_t, \mathbf{u}_t) = \mathcal{N}(C\mathbf{z}_t + D\mathbf{u}_t, R).$$

观测序列 $\mathbf{y}_{1:T}$ 的边际似然为：

$$p(\mathbf{y}_{1:T} | \mathbf{u}_{1:T}) = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:t}),$$

该边际似然由预测分布计算：

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:t}) = \mathcal{N}(C\mu_{t|t-1} + D\mathbf{u}_t, C\Sigma_{t|t-1}C^T + R).$$

三、非线性与非高斯状态空间模型 - 模型定义

非线性和非高斯状态空间模型可以表示为以下一般形式：

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{u}_t, \mathbf{q}_t),$$

$$\mathbf{y}_t = h(\mathbf{z}_t, \mathbf{u}_t, \mathbf{r}_t),$$

其中 f 为状态转移函数, h 为观测函数, 且 \mathbf{q}_t 与 \mathbf{r}_t 分别为过程噪声与观测噪声。

尽管线性高斯状态空间模型具有解析解且计算高效, 但其假设的线性动态和高斯分布在许多实际应用中不成立, 卡尔曼滤波器和平滑器不能直接应用于这些系统。

为处理非线性或非高斯模型, 我们将探讨几种近似方法, 包括扩展卡尔曼滤波/平滑 (Extended Kalman Filter/Smooth) 与无迹卡尔曼滤波/平滑 (Unscented Kalman Filter/Smoothes)。

三、非线性与非高斯状态空间模型 - 扩展卡尔曼滤波/平滑

扩展卡尔曼滤波 (EKF) 的思想是通过一阶泰勒展开对非线性系统进行线性化，从而获得一个近似的线性模型。

假设

$$y = f(z, q),$$

$$z \sim \mathcal{N}(\mu, \Sigma),$$

$$q \sim \mathcal{N}(0, Q).$$

对 $f(z, q)$ 进行一阶泰勒展开，得到近似表达式：

$$y \approx f(\mu, 0) + F_z(\mu, 0)(z - \mu) + F_q(\mu, 0)q.$$

y 和 z 的联合分布可近似为：

$$\begin{pmatrix} z \\ y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ f(\mu, 0) \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma F_z(\mu, 0)^T \\ F_z(\mu, 0) \Sigma & \Sigma_y \end{pmatrix} \right),$$

$$\Sigma_y = F_z(\mu, 0) \Sigma F_z(\mu, 0)^T + F_q(\mu, 0) Q F_q(\mu, 0)^T,$$

其中 $F_z(\mu, 0)$ 和 $F_q(\mu, 0)$ 分别是 f 关于 z 与 q 的雅可比矩阵（在 $z = \mu, q = 0$ 取值）。

三、非线性与非高斯状态空间模型 - 扩展卡尔曼滤波

扩展卡尔曼滤波 (EKF) 在每一时刻都使用局部的线性近似来更新状态估计和协方差，其步骤与标准卡尔曼滤波器类似。以下给出噪声可加的非线性状态空间模型的示例：

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{u}_t) + \mathbf{q}_t, \quad \mathbf{q}_t \sim \mathcal{N}(0, Q),$$

$$\mathbf{y}_t = h(\mathbf{z}_t, \mathbf{u}_t) + \mathbf{r}_t, \quad \mathbf{r}_t \sim \mathcal{N}(0, R).$$

1. 预测：给定 $\mathbf{z}_{t-1} | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\mu_{t-1|t-1}, \Sigma_{t-1|t-1})$ ，预测 \mathbf{z}_t 的分布：

$$\mu_{t|t-1} = f(\mu_{t-1|t-1}, \mathbf{u}_t),$$

$$\Sigma_{t|t-1} = F_{t-1} \Sigma_{t-1|t-1} F_{t-1}^T + Q,$$

其中 $F_{t-1} = \frac{\partial f}{\partial z} \Big|_{\mu_{t-1|t-1}, \mathbf{u}_t}$ 是 f 关于 z 的雅可比矩阵。

2. 更新：给定 $\mathbf{z}_t | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\mu_{t|t-1}, \Sigma_{t|t-1})$ ，对观测函数 h 的线性化更新 \mathbf{z}_t 的分布，

$$K_t = \Sigma_{t|t-1} H_t^T (H_t \Sigma_{t|t-1} H_t^T + R)^{-1},$$

$$\mu_{t|t} = \mu_{t|t-1} + K_t (y_t - h(\mu_{t|t-1}, \mathbf{u}_t)),$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t H_t \Sigma_{t|t-1},$$

其中 $H_t = \frac{\partial h}{\partial z} \Big|_{\mu_{t|t-1}, \mathbf{u}_t}$ 为 h 关于 z 的雅可比矩阵。

三、非线性与非高斯状态空间模型 - 扩展卡尔曼平滑

类似地，扩展卡尔曼平滑器为：

1. 前向滤波：执行扩展卡尔曼滤波，估计 $\mu_{t+1|t}$ 、 $\mu_{t|t}$ 、 $\Sigma_{t+1|t}$ 以及 $\Sigma_{t|t}$ 。
2. 后向平滑：计算平滑增益矩阵 J_t ：

$$J_t = \Sigma_{t|t} F_t^T \Sigma_{t+1|t}^{-1}.$$

修正当前时刻的平滑均值与协方差：

$$\mu_{t|T} = \mu_{t|t} + J_t (\mu_{t+1|T} - \mu_{t+1|t}),$$

$$\Sigma_{t|T} = \Sigma_{t|t} + J_t (\Sigma_{t+1|T} - \Sigma_{t+1|t}) J_t^T.$$

三、非线性与非高斯状态空间模型 - EKF的局限性

扩展卡尔曼滤波器（EKF）因其简单且相对高效而被广泛使用，但在以下两种情况下表现较差：

- 当先验协方差较大时，此时先验分布较为宽泛，线性化会将大量的概率信息传递到距离 $\mu_{t-1|t-1}$ 较远的函数部分
- 当函数在当前均值附近高度非线性时

更精确的方法是使用二阶泰勒展开，称为二阶扩展卡尔曼滤波器（第二阶 EKF）。

我们还可以通过反复围绕 $\mu_{t|t}$ （而不是 $\mu_{t|t-1}$ ）重新线性化观测方程来进一步提高性能，这种方法称为迭代扩展卡尔曼滤波器（iterated EKF）。

三、非线性与非高斯状态空间模型 - 无迹卡尔曼滤波 (UKF)

无迹卡尔曼滤波 (Unscented Kalman Filter, UKF) 则用一种不同的近似方法替换了模型的局部线性化，其关键思想是匹配一阶与二阶矩，利用数值积分计算矩条件，从而获得近似的高斯分布。

UKF 不需要计算观测与转移模型的雅可比矩阵，因此可应用于不可微分的模型或具有约束的模型。

与 EKF 相比，UKF 的主要优势在于更高的精度与稳定性；然而，由于需要对模型进行多次迭代，UKF 的计算速度可能较慢。

四、状态空间模型中的参数估计 - 贝叶斯参数估计

贝叶斯估计将参数视为随机变量，并试图根据观测数据估计参数的后验分布 $p(\theta | \mathbf{y}_{1:T})$ ，反映了我们对参数分布的信念。

贝叶斯估计的目标是计算参数的后验分布：

$$p(\theta | \mathbf{y}_{1:T}) = \frac{p(\mathbf{y}_{1:T} | \theta) p(\theta)}{p(\mathbf{y}_{1:T})},$$

其中： $p(\mathbf{y}_{1:T} | \theta)$ 是似然函数， $p(\theta)$ 是参数的先验分布，而 $p(\mathbf{y}_{1:T})$ 是 $\mathbf{y}_{1:T}$ 的边际分布，用于归一化。

由于计算 $p(\mathbf{y}_{1:T} | \theta)$ 涉及高维积分，直接求解后验分布通常不可行，因此需要利用数值方法对后验分布进行近似：数值积分、随机采样或MCMC。

四、状态空间模型中的参数估计 - 点估计

在某些应用中，我们可能对参数的点估计感兴趣，例如：

- **最大后验 (MAP) 估计**: 寻找使后验分布 $p(\theta | \mathbf{y}_{1:T})$ 最化的参数值:

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\theta | \mathbf{y}_{1:T}).$$

- **最大似然 (ML) 估计**: 寻找使似然函数 $p(\mathbf{y}_{1:T} | \theta)$ 最化的参数值:

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p(\mathbf{y}_{1:T} | \theta).$$

MAP 和 ML 估计的主要区别在于，前者包含了参数的先验分布，而后者仅依赖于观测数据。实际上，当先验改为为 $p(\theta) \propto 1$ 时，ML 估计等价于 MAP 估计。

四、状态空间模型中的参数估计 - 状态空间模型的EM算法

假设我们试图利用极大似然来估计 θ , $p(\mathbf{y}_{1:T} | \theta)$ 涉及高维积分, 直接最大化 $\log p(\mathbf{y}_{1:T} | \theta)$ 通常不可行, EM 算法通过引入 $\mathbf{z}_{1:T}$ 来解决这一问题。

状态空间模型的 EM 算法可简化为: 从初始猜测 $\theta^{(0)}$ 开始, 对于 $n = 0, 1, 2, \dots$, 执行下列步骤:

- E 步计算

$$Q(\theta, \theta^{(n)}) = \int p(\mathbf{z}_{1:T} | \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{z}_{1:T}, \mathbf{y}_{1:T} | \theta) d\mathbf{z}_{1:T};$$

- M 步计算

$$\theta^{(n+1)} = \arg \max_{\theta} Q(\theta, \theta^{(n)})$$

四、状态空间模型中的参数估计 - 状态空间模型的EM算法

$$\log p(\mathbf{z}_{1:T}, \mathbf{y}_{1:T} | \theta) = \log p(\mathbf{z}_1 | \theta) + \sum_{k=2}^T \log p(\mathbf{z}_k | \mathbf{z}_{k-1}, \theta)$$

$$+ \sum_{k=1}^T \log p(\mathbf{y}_k | \mathbf{z}_k, \theta)$$

将 $\log p(\mathbf{z}_{1:T}, \mathbf{y}_{1:T} | \theta)$ 代入 $Q(\theta, \theta^{(n)})$ 后可化简为：

$$Q(\theta, \theta^{(n)}) = I_1(\theta, \theta^{(n)}) + I_2(\theta, \theta^{(n)}) + I_3(\theta, \theta^{(n)}),$$

$$I_1(\theta, \theta^{(n)}) = \int p(\mathbf{z}_1 | \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{z}_1 | \theta) d\mathbf{z}_1,$$

$$I_2(\theta, \theta^{(n)}) = \sum_{k=2}^T \int p(\mathbf{z}_k, \mathbf{z}_{k-1} | \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{z}_k | \mathbf{z}_{k-1}, \theta) d\mathbf{z}_k d\mathbf{z}_{k-1},$$

$$I_3(\theta, \theta^{(n)}) = \sum_{k=1}^T \int p(\mathbf{z}_k | \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{y}_k | \mathbf{z}_k, \theta) d\mathbf{z}_k.$$

四、状态空间模型中的参数估计 - 状态空间模型的EM算法

因此，我们无需对 $\mathbf{z}_{1:T}$ 联合后验分布求期望，而只需对 $p(\mathbf{z}_k | \mathbf{y}_{1:T}, \theta^{(n)})$ 和 $p(\mathbf{z}_k, \mathbf{z}_{k-1} | \mathbf{y}_{1:T}, \theta^{(n)})$ 求期望。

在线性状态空间模型中，可以用卡尔曼平滑结果以闭合形式计算这些积分。在非线性情况下，可以使用扩展卡尔曼平滑近似积分。在更一般的概率状态空间模型中，可以使用粒子平滑进行近似。

五、案例 - 季节性与趋势

某公司股票1960-1980年的季度每股盈利 (EPS) 和对数盈利 ($\log EPS$) 如图11.2所示, EPS呈指数增长, $\log EPS$ 呈线性增长, 同时存在季节性。

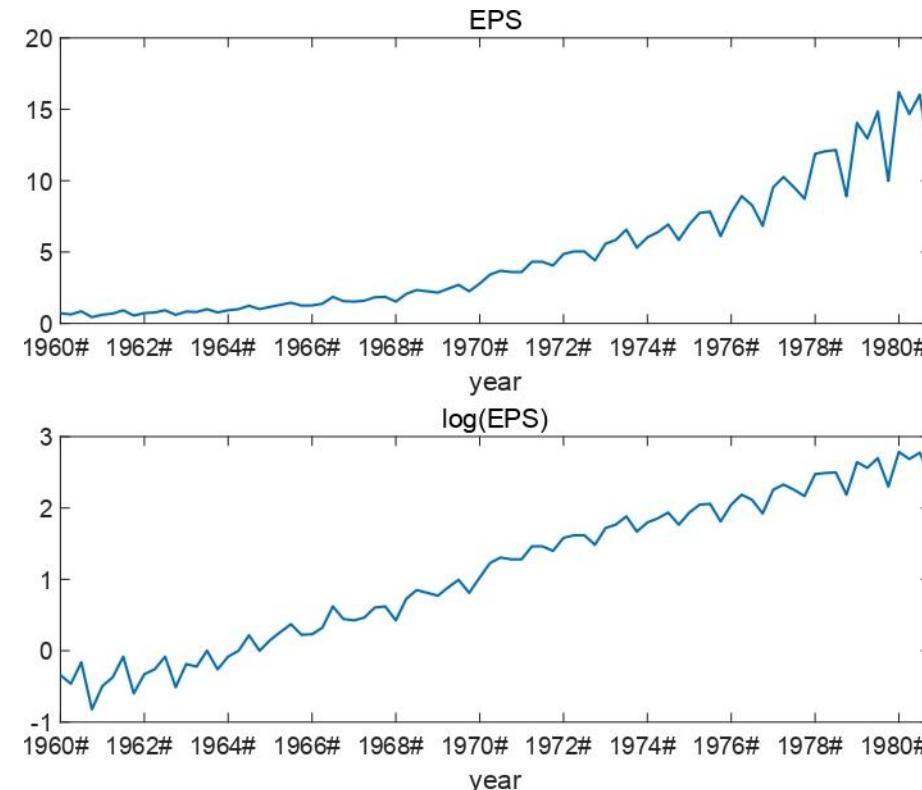


图11.2 季度每股盈利/每股对数盈利

五、案例 - 季节性与趋势

定义logEPS序列为 y_t ,

$$y_t = \mu_t + \gamma_t + e_t, \quad e_t \sim N(0, \sigma_e^2)$$

其中 μ_t 是局部趋势,

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

γ_t 是季节项

$$\gamma_{t+1} + \gamma_t + \gamma_{t-1} + \gamma_{t-2} = \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2)$$

y_t 可以由状态空间模型表示

$$\begin{bmatrix} \mu_{t+1} \\ \gamma_{t+1} \\ \gamma_t \\ \gamma_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t \\ \omega_t \end{bmatrix}$$

$$y_t = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} + e_t,$$

五、案例 - 季节性与趋势

这是一个线性高斯状态空间模型，边际似然可通过Kalman滤波迭代计算得到。

$$p(y_{1:T}|\theta) = \prod_{t=1}^T p(y_t|y_{1:t-1}, \theta),$$

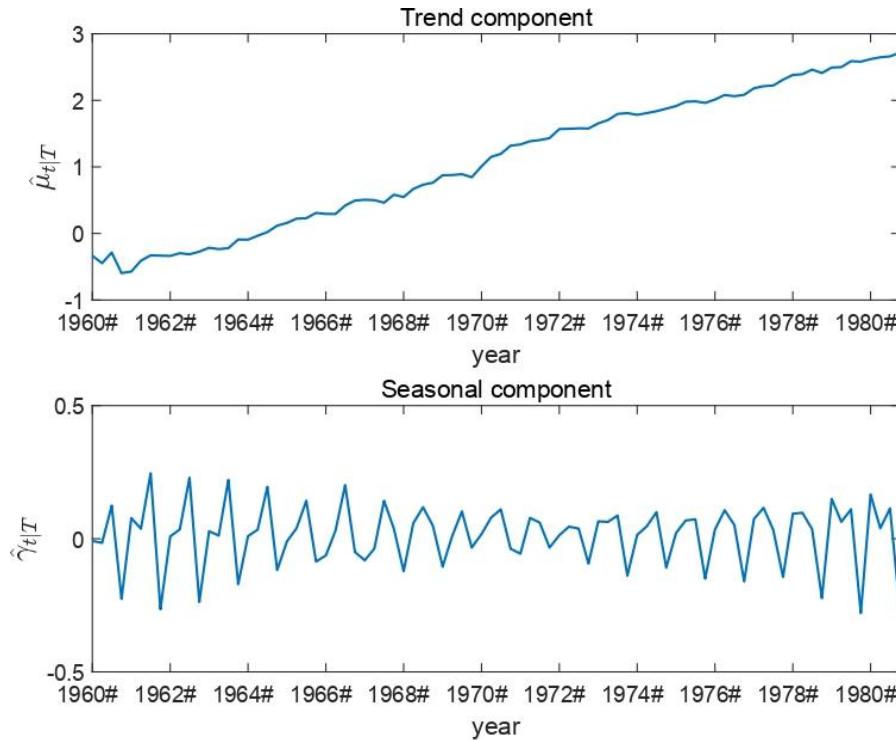
$$p(y_t|y_{1:t-1}, \theta) = \mathcal{N}(C\mu_{t|t-1}, C\Sigma_{t|t-1}C^T + R).$$

该模型的未知参数是 $\sigma_e^2, \sigma_\eta^2, \sigma_\omega^2$ 。在Matlab里利用极大似然估计得到

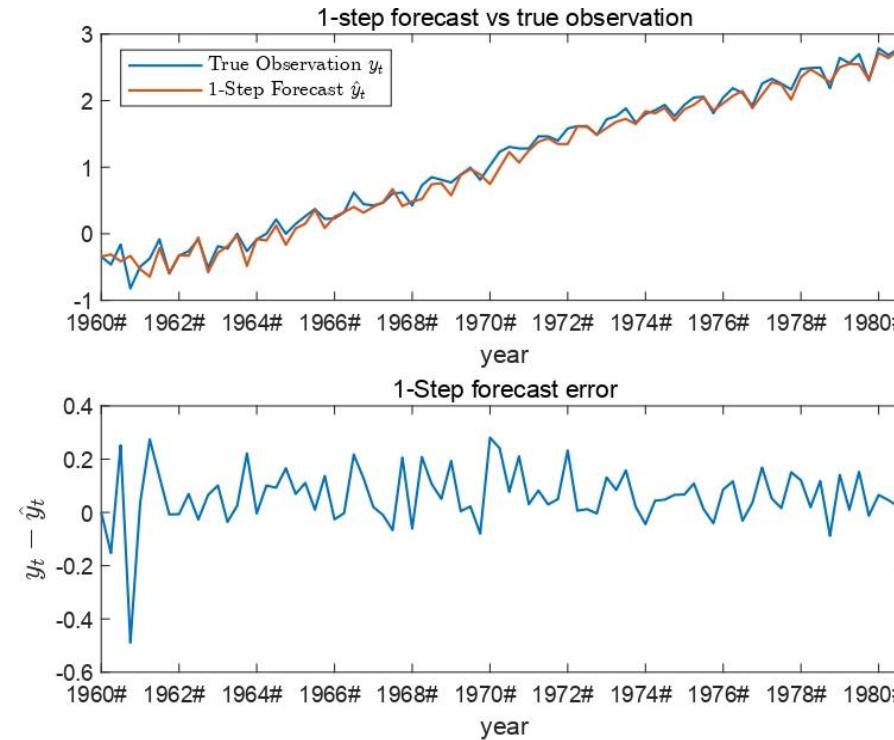
$$(\hat{\sigma}_e^2, \hat{\sigma}_\eta^2, \hat{\sigma}_\omega^2) = (7.83 \times 10^{-14}, 5.74 \times 10^{-3}, 2.05 \times 10^{-3})$$

五、案例 - 季节性与趋势

下图 (a) 画出了趋势项和季节项的平滑估计, (b) 画出了 y_t 的一步预测和预测误差。可以看出, Kalman 滤波和平滑能够较好地拟合未知状态变量并对观测变量做出预测。



(a) 状态变量的平滑估计



(b) 一步预测

图11.3 估计结果

谢 谢!

