Machine learning under physical constraints
Unsupervised learning in DA

Sixin Zhang
(sixin.zhang@toulouse-inp.fr)

Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

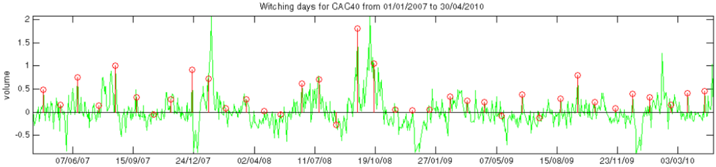
Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

A special ODS for financial time series

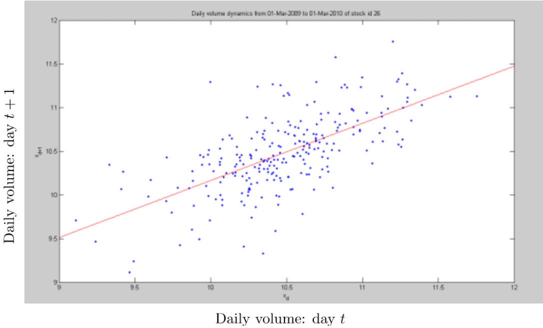
- Build a statistical model of financial time series to analyze stock volume changes under news impact



- Model dynamics using **Linear dynamical systems** (Kalman Filter).

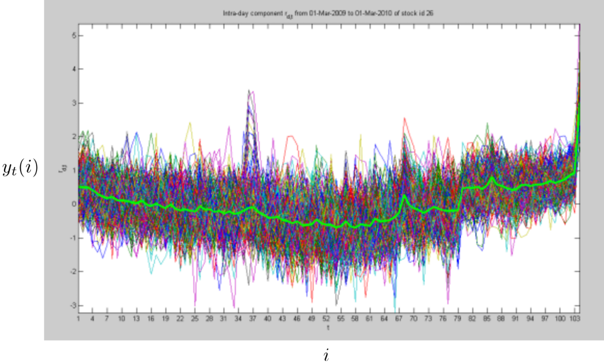
A special ODS for financial time series

- **Goal**: model daily volume dynamics x_t over t using a probabilistic model p_θ with $\theta = (\mu_0, \Sigma_0, A, B, C, E, F)$.
- Assume $x_0 \overset{p_\theta}{\sim} \mathcal{N}(\mu_0, \Sigma_0), A \in \mathbb{R}, B \in \mathbb{R}, C > 0$
- Dynamics: $x_{t+1} | x_t \overset{p_\theta}{\sim} \mathcal{N}(Ax_t + B, C)$



A special ODS for financial time series

- Model within-day volume $y_t \in \mathbb{R}^d$ for day t,



- Assume $E \in \mathbb{R}^d$, F p.d matrix, $\mathbb{1}$ vector of ones
- Observation: $y_t | x_t \overset{p_\theta}{\sim} \mathcal{N}(\mathbb{1} x_t + E, F)$

Problem formulation

- Can we estimate the densities $p_\theta(x_t|x_{t-1})$ and $p_\theta(y_t|x_t)$ for $t = 1, \cdots, T$ from $y = (y_1, \cdots, y_T)$?
- **Problem**: estimate $\theta = (\mu_0, \Sigma_0, A, B, C, E, F)$ from y.
- Challenge: x_t is not observed (y_t may be impacted by news)
- Idea: Maximum likelihood estimation with EM algorithm
- Reference: Pattern Recognition and Machine Learning by Christopher M. Bishop

Outline

Unsupervised learning of Linear Dynamical systems

MLE approach and EM algorithm

MLE approach and EM algorithm

- Identifiability condition (to restrict E):

$$E \perp \mathbb{1}$$

- Estimate θ from $y = (y_1, \cdots, y_T)$, based on MLE

$$\max_\theta \log p_\theta(y)$$

- Estimating of the latent variables $x = (x_0, \cdots, x_T)$ makes the maximization of θ easier,

$$\begin{aligned} \text{E step}: &\quad \max_\phi \int \log \frac{p_\theta(x|y)}{q_\phi(x|y)} q_\phi(x|y) dx \\ \text{M step}: &\quad \max_\theta \int \log p_\theta(x, y) q_\phi(x|y) dx \end{aligned}$$

EM algorithm

- Compute an optimal θ iteratively: At iteration k,

$$\begin{aligned} \text{E step}: &\quad \text{optimal} \quad q_\phi(x|y) = p_{\theta^{(k)}}(x|y), \\ \text{M step}: &\quad \theta^{(k+1)} \in \arg\max_\theta \int \log p_\theta(x,y) p_{\theta^{(k)}}(x|y) dx. \end{aligned}$$

- **Factorization** of $\log p$ makes EM simpler:

$$\log p_\theta(x,y) = \log p_\theta(x_0) + \sum_{t=0}^{T-1} \log p_\theta(x_{t+1}|x_t) + \sum_{t=1}^{T} \log p_\theta(y_t|x_t).$$

- Therefore in the E step, it suffices to compute

$$p_{\theta^{(k)}}(x_t|y_1,\cdots,y_T)$$

$$p_{\theta^{(k)}}(x_t, x_{t+1}|y_1,\cdots,y_T)$$

Road-map: 3 sub-problems in M step

- Decompose the M-step using factorization,

$$\theta^{(k+1)} \in \arg\max_{\theta} \int \log p_{\theta}(x, y) p_{\theta^{(k)}}(x|y) dx.$$

- **Problem 1**: estimate (μ_0, Σ_0)

$$\max_{\theta} \int \log p_{\theta}(x_0) p_{\theta^{(k)}}(x_0|y) dx_0$$

- Problem 2: estimate (A, B, C)
- Problem 3: estimate (E, F)

Problem 1: analytical formula of M step

$$\text{Problem 1:} \quad \ell_1(\theta) = \int \log p_\theta(x_0) p_{\theta^{(k)}}(x_0|y) dx_0.$$

- First solve the **E step**: compute $p_{\theta^{(k)}}(x_0|y)$.
- Assume under $p_{\theta^{(k)}}(x_0|y)$, x_0 follows $\mathcal{N}(\mu_{0|T}, \Sigma_{0|T})$
- Develop the log-likelihood,

$$-\ell_1 \propto \mathbb{E}_{x_0}((x_0 - \mu_0)^\intercal \Sigma_0^{-1}(x_0 - \mu_0)) + \log |\Sigma_0|$$

- To maximize ℓ_1, compute the critical points (exercise)

$$\frac{\partial \ell_1}{\partial \mu_0} = 0, \frac{\partial \ell_1}{\partial \Sigma_0} = 0 \quad \Rightarrow \quad \mu_0 = \mu_{0|T}, \Sigma_0 = \Sigma_{0|T}$$

E step: How to compute $\mathcal{N}(\mu_{0|T}, \Sigma_{0|T})$?

- To compute $p_{\theta^{(k)}}(x_0|y)$, we use the KF smoother.
- Backward computation: to compute recursively

$$x_{t+1}|y_1, \cdots, y_T \Rightarrow x_t|y_1, \cdots, y_T$$

- Start from $p(x_T|y)$ by KF propagation and analysis recursively.
  - Analysis step: $x_t|y_1, \cdots, y_t \overset{p_{\theta^{(k)}}}{\sim} \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$
  - Propagation step: $x_{t+1}|y_1, \cdots, y_t \overset{p_{\theta^{(k)}}}{\sim} \mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$

Backward computation in KF smoother

- Backward computation: given

$$x_{t+1}|y_1, \cdots, y_T \overset{p_{\theta(k)}}{\sim} \mathcal{N}(\mu_{t+1|T}, \Sigma_{t+1|T})$$

compute

$$x_t|y_1, \cdots, y_T \overset{p_{\theta(k)}}{\sim} \mathcal{N}(\mu_{t|T}, \Sigma_{t|T})$$

- We verify that (exercise)

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

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

where $J_t = \Sigma_{t|t}^\intercal A^\intercal(C + A\Sigma_{t|t}A^\intercal)^{-1}$

E step: KF smoother (exercise)

- Verify that $p_{\theta^{(k)}}(x_t | x_{t+1}, y_1, \cdots, y_T)$ follows

$$\mathcal{N}(\mu_{t|t+1|T}, \Sigma_{t|t+1|T})$$

where

$$\mu_{t|t+1|T} = \mu_{t|t} + J_t[x_{t+1} - (A\mu_{t|t} + B)]$$
$$\Sigma_{t|t+1|T} = \Sigma_{t|t} - J_t A \Sigma_{t|t}$$

- This implies the joint distribution $p_{\theta^{(k)}}(x_{t+1}, x_t | y)$ is

$$\mathcal{N}\left(\begin{pmatrix} \mu_{t+1|T} \\ \mu_{t|T} \end{pmatrix}, \begin{pmatrix} \Sigma_{t+1|T}, & (J_t \Sigma_{t+1|T})^\mathsf{T} \\ J_t \Sigma_{t+1|T}, & \Sigma_{t|T} \end{pmatrix}\right)$$

M step: analytical formula of Problem 2

- Consider $\ell_2 = \sum_t \ell_{2,t}$, where

$$\begin{aligned}\ell_{2,t}(\theta) &= \int \log p_\theta(x_{t+1}|x_t) p_{\theta^{(k)}}(x_{t+1}, x_t|y) dx_{t+1} dx_t \\ &\propto -\mathbb{E}_{(x_{t+1},x_t)}[(x_{t+1} - Ax_t - B)^\intercal C^{-1}(x_{t+1} - Ax_t - B)] \\ &\quad - \log|C|\end{aligned}$$

- Compute critical points of ℓ_2 using matrix calculus,

$$\frac{\partial \ell_2}{\partial A} = 0, \quad \frac{\partial \ell_2}{\partial B} = 0, \quad \frac{\partial \ell_2}{\partial C} = 0$$

- Exercise: find out the optimal solution.

M step: analytical formula of Problem 3

- Consider $\ell_3 = \sum_t \ell_{3,t}$, where

$$\ell_{3,t}(\theta) = \int \log p_\theta(y_t|x_t) p_{\theta^{(k)}}(x_t|y) dx_t$$
$$\propto -\mathbb{E}_{x_t}[(y_t - \mathbb{1}x_t - E)^\intercal F^{-1}(y_t - \mathbb{1}x_t - E)] - \log|F|$$

- Due to the constraint on E, we compute critical points of ℓ_3 using the Lagrangian method.
- Exercise: find out the optimal solution.