Independent Markov chains on a Markov chain and MRS models for electricity prices.

Me

[1]Department of legends
University of legends

PG seminar, 2018

Into to HMMs

Whiteboard

Example HMM

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

Example HMM

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

- R_t evolves with probabilities p_{ij}
- $B_t \sim i.i.d.$ with pdf $f^{(1)}$
- $S_t \sim i.i.d.$ with pdf $f^{(2)}$

Example HMM

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

- R_t evolves with probabilities p_{ij}
- $B_t \sim i.i.d.$ with pdf $f^{(1)}$
- $S_t \sim i.i.d.$ with pdf $f^{(2)}$

- Suppose we observe $x_0, ..., x_T$

Classical problems for HMMs

- Filtered & Prediction inferences

$$P(R_{t-1} = i | x_{t-1}, ..., x_0), \qquad P(R_t = i | x_{t-1}, ..., x_0)$$

- Smoothed inferences

$$P(R_t = i | x_T, ..., x_0)$$

- Likelihood evaluation

$$f(\mathbf{x}; \theta) = f(x_0; \theta) \prod_{t=1}^{T} f(x_t | x_{t-1}, ..., x_0; \theta)$$

- Likelihood maximisation

$$\hat{\theta} = \arg\max_{\theta \in \Theta} f(\mathbf{x} | \theta)$$

Classical problems for HMMs

Problem

- Naive complexity is $\mathcal{O}(2^T)$
  E.g. evaluate the likelihood

$$
\begin{aligned}
f(x_0, ..., x_T; \theta) &= \sum_{\mathcal{R} \in S} f(x_0, ..., x_T, R_0, ..., R_t) \\
&= \sum_{\mathcal{R} \in S} f(x_0, ..., x_T | R_0, ..., R_t) f(R_0, ..., R_t)
\end{aligned}
$$

Classical problems for HMMs

Problem

- Naive complexity is $\mathcal{O}(2^T)$
  E.g. evaluate the likelihood

$$
\begin{aligned}
f(x_0, \dots, x_T; \theta) &= \sum_{\mathcal{R} \in S} f(x_0, \dots, x_T, R_0, \dots, R_t) \\
&= \sum_{\mathcal{R} \in S} f(x_0, \dots, x_T | R_0, \dots, R_t) f(R_0, \dots, R_t)
\end{aligned}
$$

Solution

- forward-backward algorithm
  - Filtered & Prediction inferences
  - Smoothed inferences
  - Likelihood evaluation
- Baum-Welch algorithm (EM algorithm)
  - Likelihood maximisation

Complexity $\mathcal{O}(2T)$

The forward algorithm: Calculating $P(R_t = i | \boldsymbol{x}_{0:t})$

Let $\boldsymbol{x}_{t:0} = (x_t, ..., x_0)$

- Initialise probabilities $P(R_0 = i)$; set $C = 0$;
- For $t = 1, ..., T$

The forward algorithm: Calculating $P(R_t = i|\boldsymbol{x}_{0:t})$

Let $\boldsymbol{x}_{t:0} = (x_t, \ldots, x_0)$

- Initialise probabilities $P(R_0 = i)$; set $C = 0$;
- For $t = 1, \ldots, T$
  - Calculate

$$P(R_t = i|\boldsymbol{x}_{t:0}) = \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0})P(R_t = i|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

$$= \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0}) \sum_{j=1}^{2} p_{ji} P(R_{t-1} = j|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

where

$$f(x_t|\boldsymbol{x}_{t-1:0}) = \sum_{j=1}^{2} f(x_t|R_t = j, \boldsymbol{x}_{t-1:0})P(R_t = j|\boldsymbol{x}_{t-1:0})$$

The forward algorithm: Calculating $P(R_t = i|\boldsymbol{x}_{0:t})$

Let $\boldsymbol{x}_{t:0} = (x_t, ..., x_0)$

- Initialise probabilities $P(R_0 = i)$; set $C = 0$;
- For $t = 1, ..., T$
  - Calculate

$$P(R_t = i|\boldsymbol{x}_{t:0}) = \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0})P(R_t = i|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

$$= \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0}) \sum_{j=1}^{2} p_{ji} P(R_{t-1} = j|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

    where

$$f(x_t|\boldsymbol{x}_{t-1:0}) = \sum_{j=1}^{2} f(x_t|R_t = j, \boldsymbol{x}_{t-1:0})P(R_t = j|\boldsymbol{x}_{t-1:0})$$

  - Calculate $C = C + \log f(x_t|\boldsymbol{x}_{t-1:0})$

The backward algorithm: Calculating $P(R_t = i|\boldsymbol{x}_{T:0})$

- Initialise $P(R_T = i|\boldsymbol{x}_{T:0})$
- For $t = T-1, T-2, ..., 1$
  - Calculate

$$P(R_t = i|\boldsymbol{x}_{T:0}) = \sum_{j=1}^{2} \frac{P(R_t = i|\boldsymbol{x}_{t:0}) p_{ij} P(R_{t+1} = j|\boldsymbol{x}_{T:0})}{P(R_{t+1} = j|\boldsymbol{x}_{t:0})}$$

Finding the MLE

Goal: $\hat{\theta} = \arg\max_{\theta \in \Theta} f(\mathbf{x}|\theta)$

- Could use the forward algorithm and a black-box optimiser
- EM seems to be more stable

Finding the MLE

Goal: $\hat{\theta} = \arg\max_{\theta \in \Theta} f(\mathbf{x}|\theta)$

- Could use the forward algorithm and a black-box optimiser
- EM seems to be more stable

The EM algorithm

- Is just a hill-climbing algorithm
- Idea: at each iteration construct and maximise a *minorising* function, $Q(\theta, \theta^{(n)})$
- Diagram on board

Implementing the EM algorithm

E-step: Construct $Q(\theta, \theta^{(n)})$

$$
\begin{aligned}
&Q(\theta, \theta^{(n)}) \\
&\qquad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0}, \boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\qquad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0} | \boldsymbol{R}) + \log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\qquad = \sum_{t=1}^{T} \sum_{j=1}^{2} P(R_t = j | \boldsymbol{x}_{T:0}, \theta^{(n)}) \log \left\{ f_\theta^{(j)}(x_t | \boldsymbol{x}_{t-1:0}) \right\} \\
&\qquad\quad + \mathbb{E}\left[\log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right]
\end{aligned}
$$

Implementing the EM algorithm

M-step: Maximise $Q(\theta, \theta^{(n)})$

- Explicit expressions for the maximisers of Q

Extending the HMM: The typical MRS model

Whiteboard

Example MRS model

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

- R_t evolves with probabilities p_{ij}
- $B_t \sim f^{(1)}(\cdot|X_{t-1})$
- $S_t \sim f^{(2)}(\cdot|X_{t-1})$

- Suppose we observe $x_0, ..., x_T$
- The same machinery for HMMs works for this model!

The forward algorithm: Calculating $P(R_t = i|\boldsymbol{x}_{0:t})$

Let $\boldsymbol{x}_{t:0} = (x_t, \ldots, x_0)$

- Initialise probabilities $P(R_0 = i)$; set $C = 0$;
- For $t = 1, \ldots, T$
  - Calculate

$$P(R_t = i|\boldsymbol{x}_{t:0}) = \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0})P(R_t = i|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

$$= \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0}) \sum_{j=1}^{2} p_{ji} P(R_{t-1} = j|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

where

$$f(x_t|\boldsymbol{x}_{t-1:0}) = \sum_{j=1}^{2} f(x_t|R_t = j, \boldsymbol{x}_{t-1:0})P(R_t = j|\boldsymbol{x}_{t-1:0})$$

  - Calculate $C = C + \log f(x_t|\boldsymbol{x}_{t-1:0})$

The backward algorithm: Calculating $P(R_t = i|\boldsymbol{x}_{T:0})$

- Initialise $P(R_T = i|\boldsymbol{x}_{T:0})$
- For $t = T-1, T-2, ..., 1$
  - Calculate

$$P(R_t = i|\boldsymbol{x}_{T:0}) = \sum_{j=1}^{2} \frac{P(R_t = i|\boldsymbol{x}_{t:0}) p_{ij} P(R_{t+1} = j|\boldsymbol{x}_{T:0})}{P(R_{t+1} = j|\boldsymbol{x}_{t:0})}$$

Implementing the EM algorithm

E-step: Construct $Q(\theta, \theta^{(n)})$

$$
\begin{aligned}
&Q(\theta, \theta^{(n)}) \\
&\qquad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0}, \boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\qquad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0} | \boldsymbol{R}) + \log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\qquad = \sum_{t=1}^{T} \sum_{j=1}^{2} P(R_t = j | \boldsymbol{x}_{T:0}, \theta^{(n)}) \log \left\{ f_\theta^{(j)}(x_t | \boldsymbol{x}_{t-1:0}) \right\} \\
&\qquad \quad + \mathbb{E}\left[\log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right]
\end{aligned}
$$

Implementing the EM algorithm

M-step: Maximise $Q(\theta, \theta^{(n)})$

- Explicit expressions for the maximisers of Q

Extending the HMM: The *independent* MRS model

Whiteboard

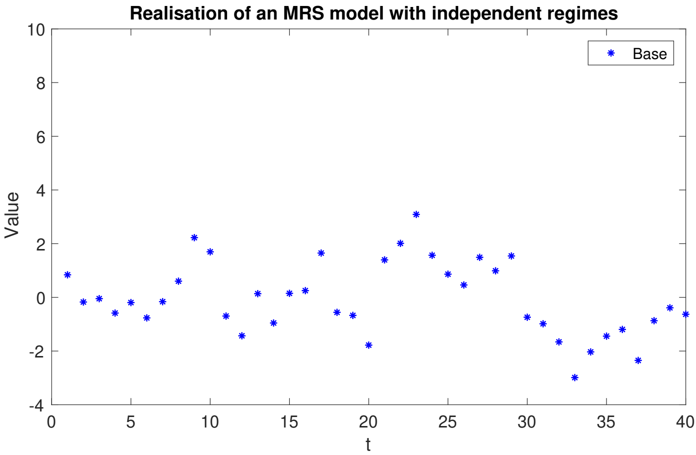
Example independent MRS model

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

- R_t evolves with probabilities p_{ij}
- $B_t \sim f^{(1)}(\cdot|B_{t-1})$
- $S_t \sim f^{(2)}(\cdot|S_{t-1})$

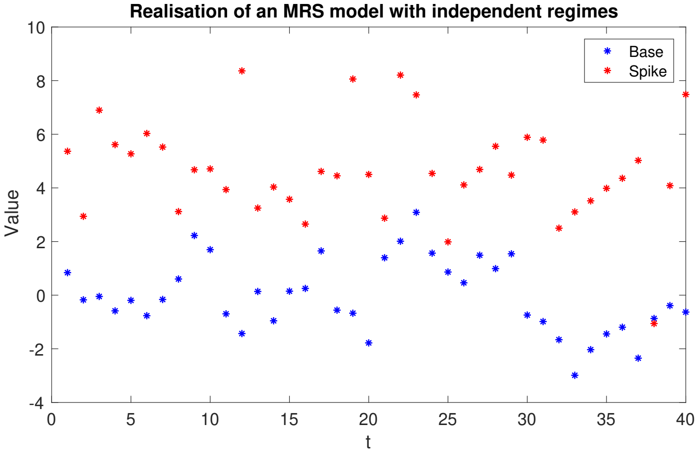
- Suppose we observe $x_0, \ldots, x_T$

Independent regimes



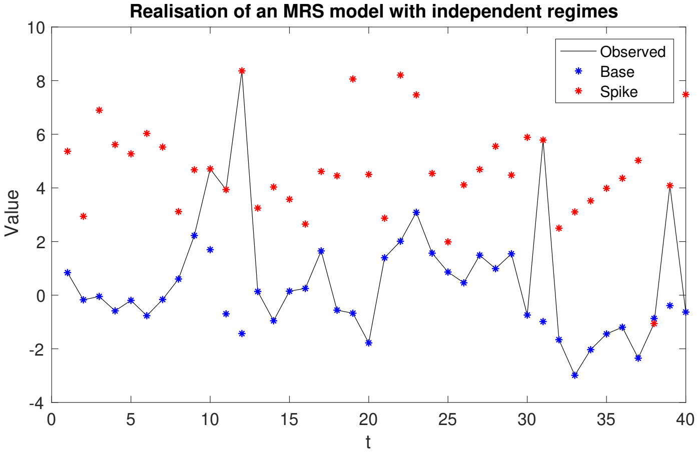
- Behold, a simulation of an AR(1) process

Independent regimes



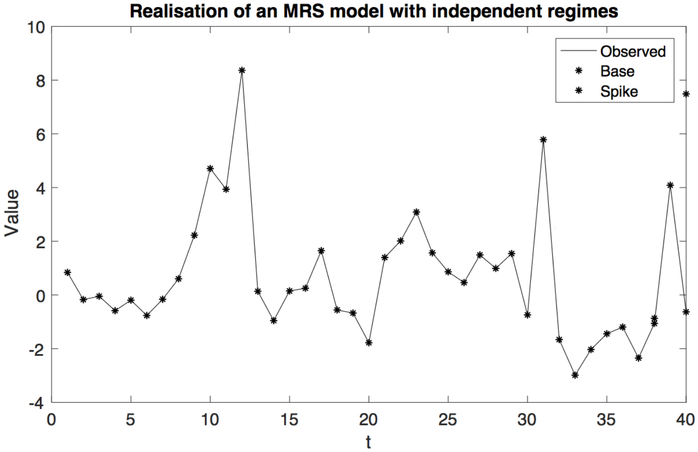
- Spike process – Independent of Base process

Independent regimes



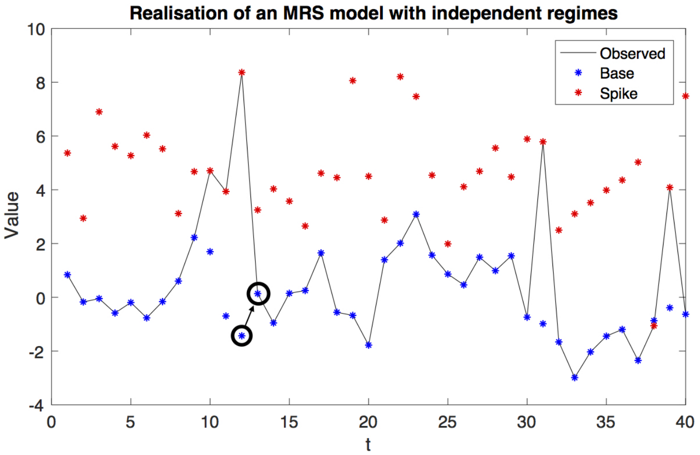
- The regime sequence determines which points we observe

Independent regimes



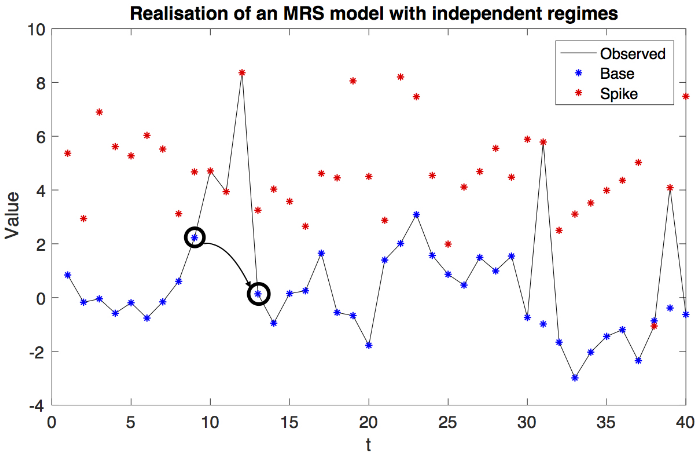
- But this is all we actually observe

Independent regimes



- B_t depends on B_{t-1}, but it might be unobserved

Independent regimes



- Can integrate unobserved prices away
- B_t depends on a random lagged observation

Extension to independent model

- Initialise probabilities $P(R_0 = (\boldsymbol{n}_t, r_t))$; set $C = 0$;
- For $t = 1, \ldots, T$
  - Calculate

$$P(R_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{t:0}) = \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0})P(R_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

$$= \frac{f(x_t|R_t = i, \boldsymbol{x}_{t-1:0}) \sum_{j=1}^{2} p_{ji} P(R_{t-1} = j|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

    where

$$f(x_t|\boldsymbol{x}_{t-1:0}) = \sum_{j=1}^{2} f(x_t|R_t = j, \boldsymbol{x}_{t-1:0})P(R_t = j|\boldsymbol{x}_{t-1:0})$$

  - Calculate $C = C + \log f(x_t|\boldsymbol{x}_{t-1:0})$

Problem

- $f(x_t|R_t = 1, \boldsymbol{x}_{t-1:0})$ is an $\mathcal{O}(2^t)$ calculation
- Could avoid this if we knew the last time in state 1

Extension to independent model

Solution

- Use the augmented chain

$$\boldsymbol{R}_t = (r_t, \boldsymbol{n}_t)$$

  - r_t is the regime at time t
  - $\boldsymbol{n}_t$ is the counter 'how long ago were we in state i'
- The augmented hidden chain remains Markovian
  - Transition probabilities

$$\begin{aligned}
P(r_t = 1, \boldsymbol{n}_t = 1 | r_{t-1} = 1, \boldsymbol{n}_{t-1} = k) &= p_{11}, \\
P(r_t = 1, \boldsymbol{n}_t = k+1 | r_{t-1} = 2, \boldsymbol{n}_{t-1} = k) &= p_{12}, \\
P(r_t = 2, \boldsymbol{n}_t = 1 | r_{t-1} = 1, \boldsymbol{n}_{t-1} = k) &= p_{21}, \\
P(r_t = 2, \boldsymbol{n}_t = k+1 | r_{t-1} = 2, \boldsymbol{n}_{t-1} = k) &= p_{22}, \quad k = 1, ..., t-1
\end{aligned}$$

  - State space $\mathcal{S} = \{1, 2\} \times \{0, 1, 2, ..., T\}$

The forward algorithm: Calculating $P(\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{0:t})$

Let $\boldsymbol{x}_{t:0} = (x_t, ..., x_0)$

- Initialise probabilities $P(\boldsymbol{R}_0 = (n_0, r_0))$; set $C = 0$;
- For $t = 1, ..., T$
  - Calculate

$$P(\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{t:0})$$

$$= \frac{f(x_t|\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t), \boldsymbol{x}_{t-1:0}) \sum_{r_{t-1}=1}^{2} \sum_{r_{t-1}=1}^{t} p_{ji} P(\boldsymbol{R}_{t-1} = (\boldsymbol{n}_{t-1}, r_{t-1})|\boldsymbol{x}_{t-1:0})}{f(x_t|\boldsymbol{x}_{t-1:0})}$$

where

$$f(x_t|\boldsymbol{x}_{t-1:0}) = \sum_{r_t=1}^{2} \sum_{n_t=1}^{t} f(x_t|\boldsymbol{R}_t = (n_t, r_t), \boldsymbol{x}_{t-1:0}) P(\boldsymbol{R}_t = (n_t, r_t)|\boldsymbol{x}_{t-1:0})$$

  - Calculate $C = C + \log f(x_t|\boldsymbol{x}_{t-1:0})$

The backward algorithm: Calculating $P(\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{T:0})$

- Initialise $P(\boldsymbol{R}_T = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{T:0})$
- For $t = T-1, T-2, ..., 1$
  - Calculate

$$P(\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{T:0}) = \sum_{r_{t-1}=1}^{2} \sum_{\boldsymbol{n}_{t-1}=1}^{t} \frac{P(\boldsymbol{R}_t = (\boldsymbol{n}_t, r_t)|\boldsymbol{x}_{t:0}) p_{ij} P(\boldsymbol{R}_{t+1} = (\boldsymbol{n}_{t-1}, r_{t-1})|\boldsymbol{x}_{T:0})}{P(\boldsymbol{R}_{t+1} = (\boldsymbol{n}_{t-1}, r_{t-1})|\boldsymbol{x}_{t:0})}$$

Implementing the EM algorithm

E-step: Construct $Q(\theta, \theta^{(n)})$

$$
\begin{aligned}
&Q(\theta, \theta^{(n)}) \\
&\quad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0}, \boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\quad = \mathbb{E}\left[\log f_\theta(\boldsymbol{x}_{T:0} | \boldsymbol{R}) + \log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right] \\
&\quad = \sum_{t=1}^{T} \sum_{j=1}^{2} \sum_{\ell=1}^{t} P(n_t = \ell, r_t = j | \boldsymbol{x}_{T:0}, \theta^{(n)}) \log \left\{ f_\theta^{(\ell,j)}(x_t | \boldsymbol{x}_{t-1:0}) \right\} \\
&\qquad + \mathbb{E}\left[\log f_\theta(\boldsymbol{R}) | \boldsymbol{x}_{T:0}, \theta^{(n)}\right]
\end{aligned}
$$

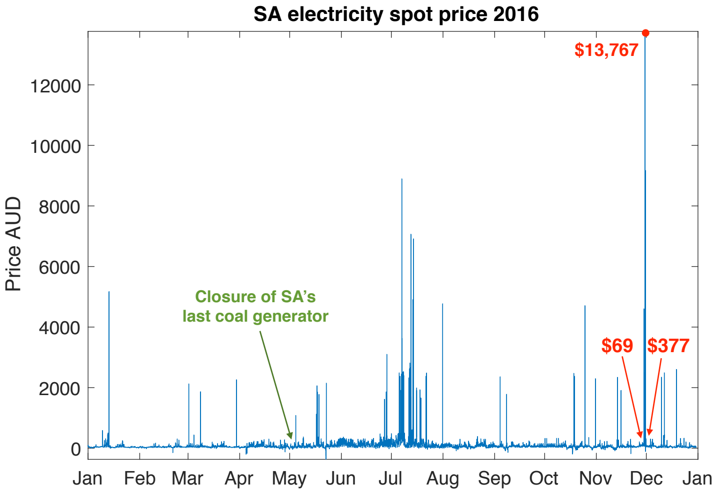
Implementing the EM algorithm

M-step: Maximise $Q(\theta, \theta^{(n)})$

- Explicit expressions for the maximisers for i.i.d. regimes
- Expressions for p_{ij} parameters

Motivation

Electricity price modelling



Motivation

Electricity price modelling

- Markov Regime Switching (MRS) models can capture this behaviour

E.g.

$$X_t = \begin{cases} B_t & \text{when } R_t = 1, \\ S_t & \text{when } R_t = 2. \end{cases}$$

- R_t evolves with probabilities $p_{ij} = \mathbb{P}(R_t = j | R_{t-1} = i)$
- $B_t = \alpha + \phi B_{t-1} + \sigma \varepsilon_t$ - AR(1) base regime
- $S_t \sim i.i.d.$ - spikes
- Note the independent regimes!

Comments

- Black-box optimisation
  - Converges to the wrong maximiser about 50% of the time
  - Often converges to the boundary

Comments

- Black-box optimisation
  - Converges to the wrong maximiser about 50% of the time
  - Often converges to the boundary
- EM seems to be more stable
  - I have never seen it hit a boundary

Comments

- Black-box optimisation
  - Converges to the wrong maximiser about 50% of the time
  - Often converges to the boundary
- EM seems to be more stable
  - I have never seen it hit a boundary
- MLEs for AR(1) parameters in this model are biased

$$B_t = \alpha + \phi B_{t-1} + \sigma \varepsilon_t$$

  - $|\phi|$ biased toward 0
  - σ is biased towards 0 if ϕ is positive
  - σ is biased towards ∞ if ϕ is negative
- The EM-like algorithm that is currently used for this model works well and we don't know why!