Data Mining Techniques

CS 6220 - Section 3 - Fall 2016

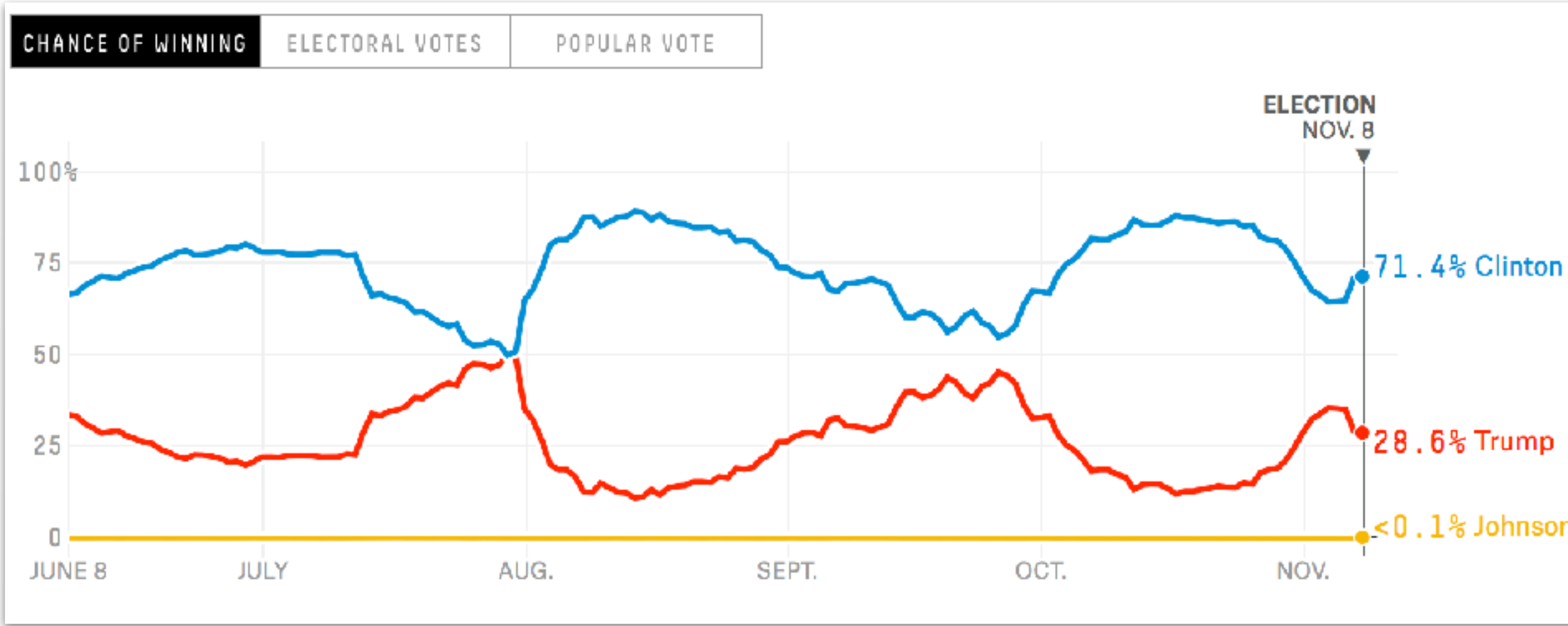
Lecture 18: Time Series

Jan-Willem van de Meent
(*credit:* Aggarwal Chapter 14.3)

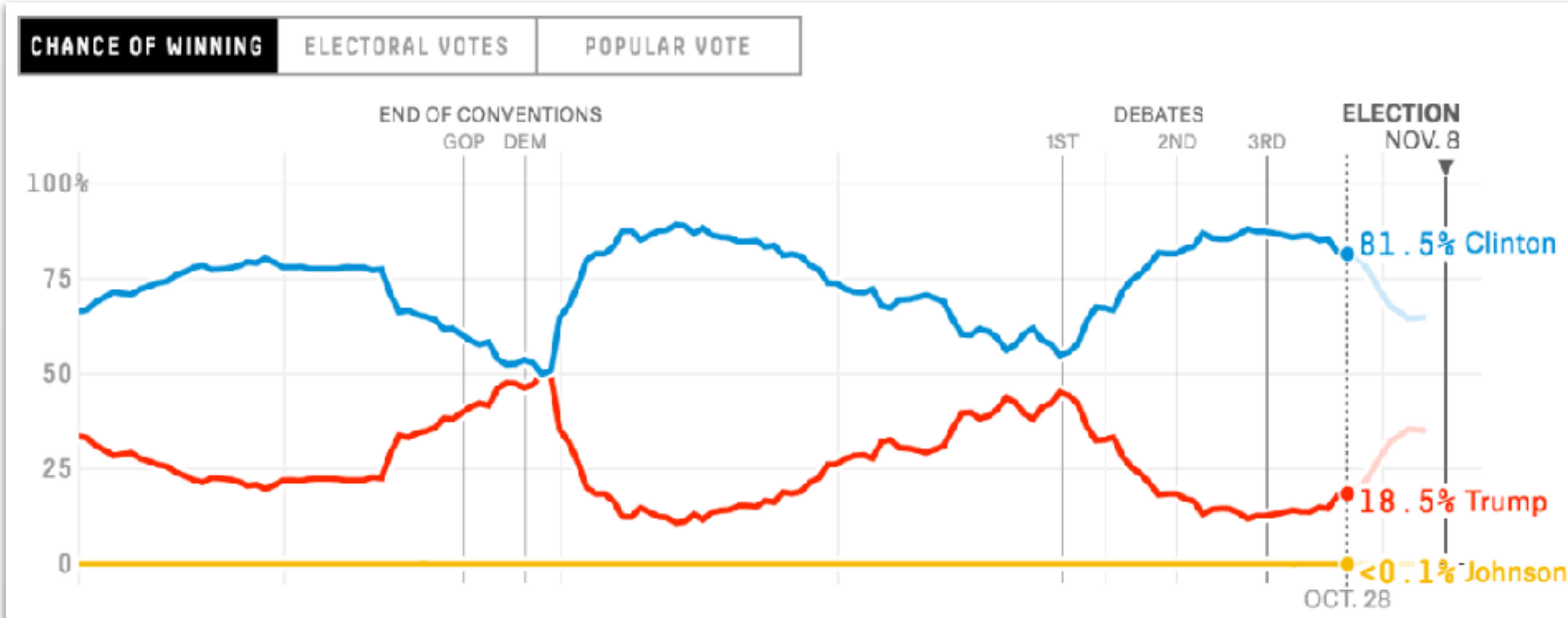
Time Series Data

http://www.capitalhubs.com/2012/08/the-correlation-between-apple-product.html

Time Series Data

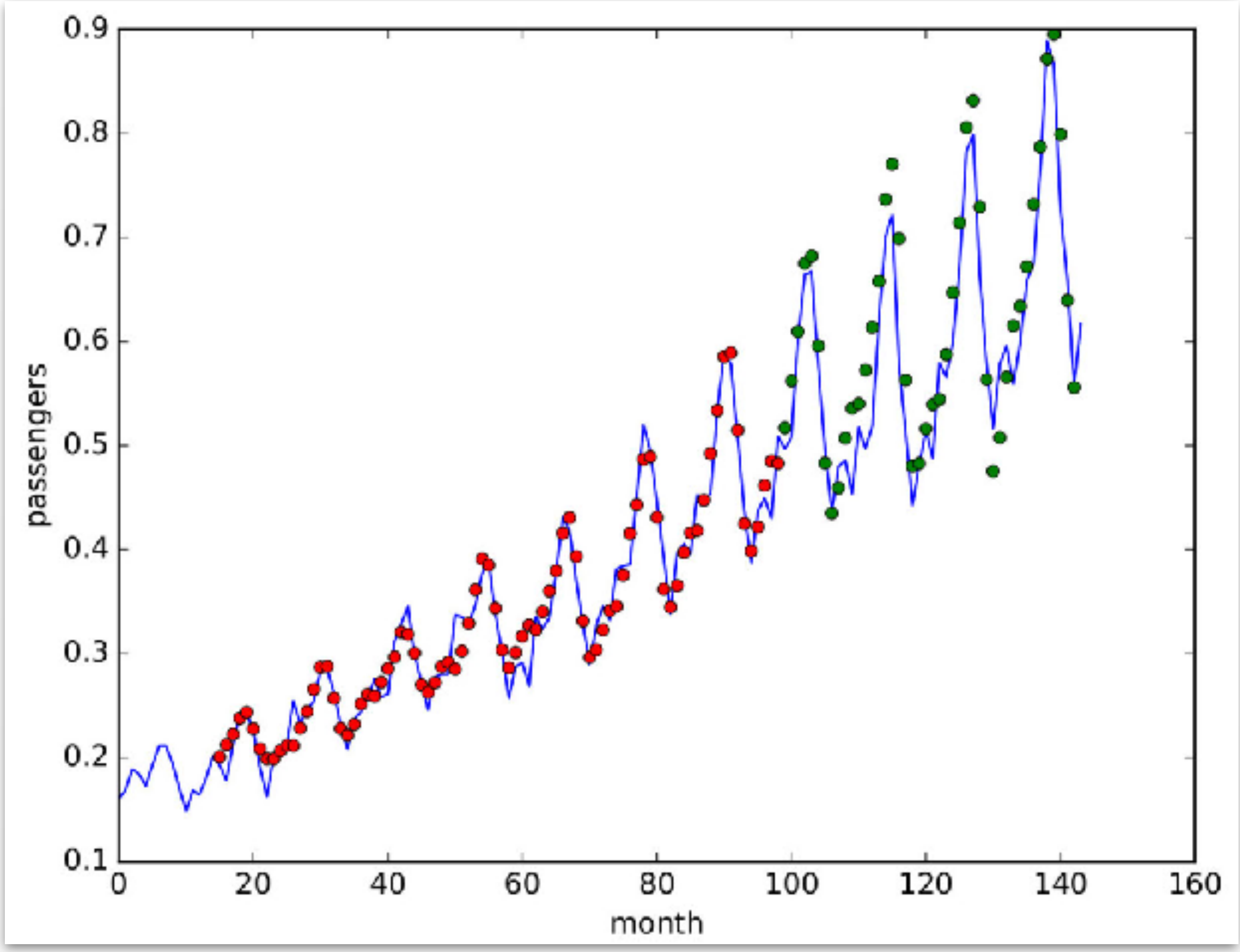


Time Series Data



- Time series forecasting is fundamentally hard
- Rare events often play a big role in changing trends
- Impossible to know how events will affects trends (and often when such events will occur)

Time Series Data



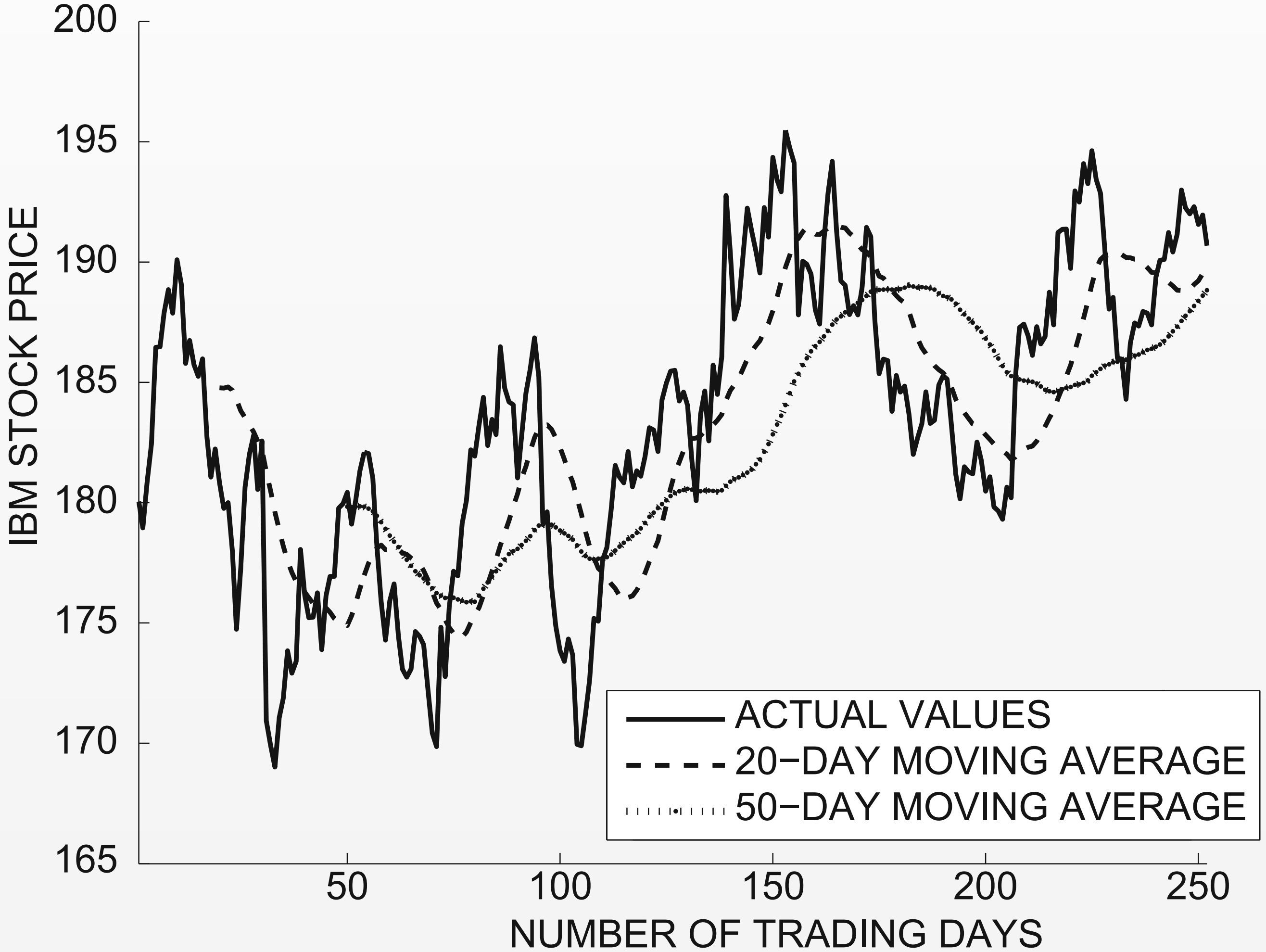
source: https://am241.wordpress.com/tag/time-series/

- In some cases there are clear trends (here: seasonal effects + growth)

Autoregressive Models

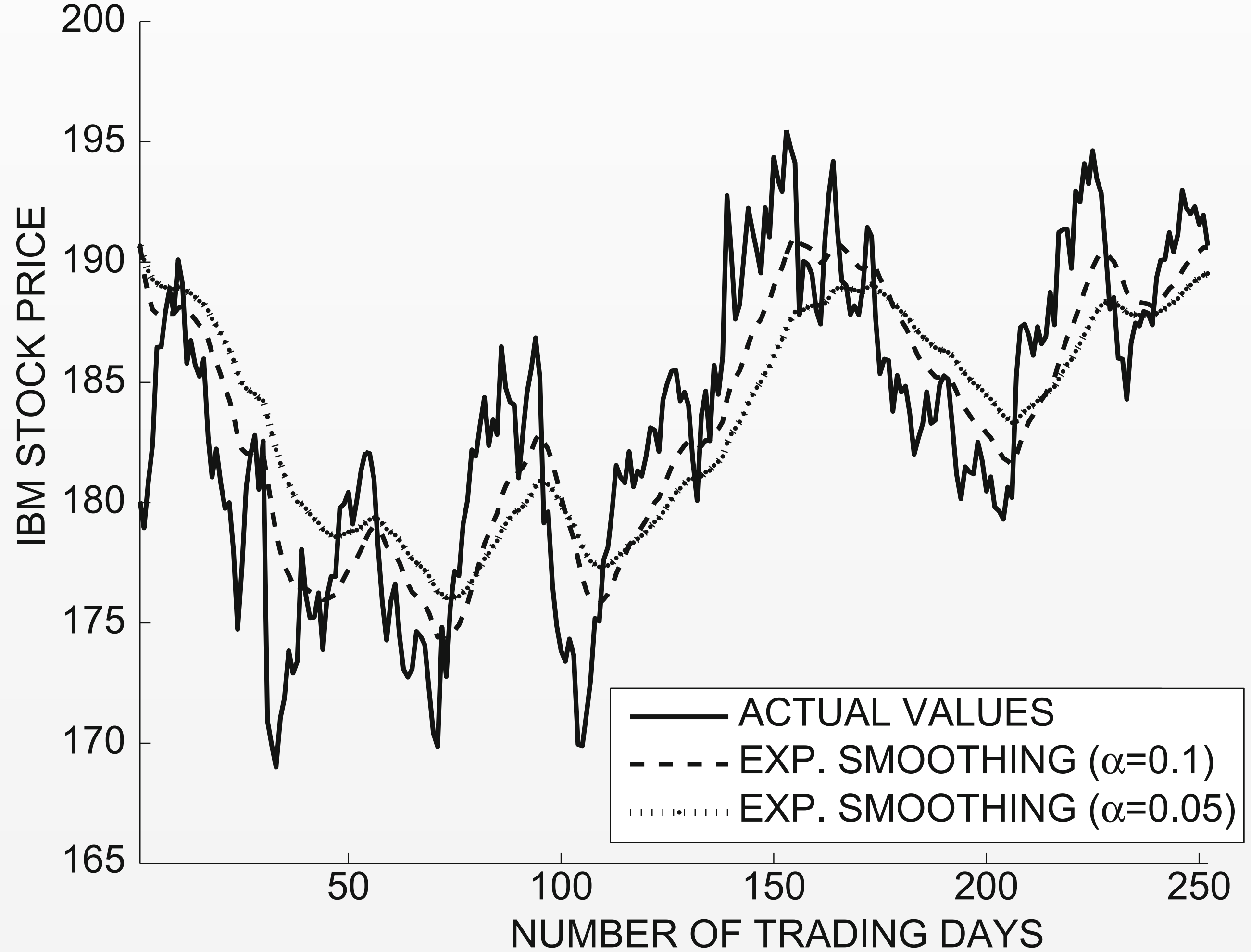
Time Series Smoothing

Moving Average



$$y'_i = \frac{1}{k} \sum_{n=0}^{k-1} y_{i-n}$$

Exponential



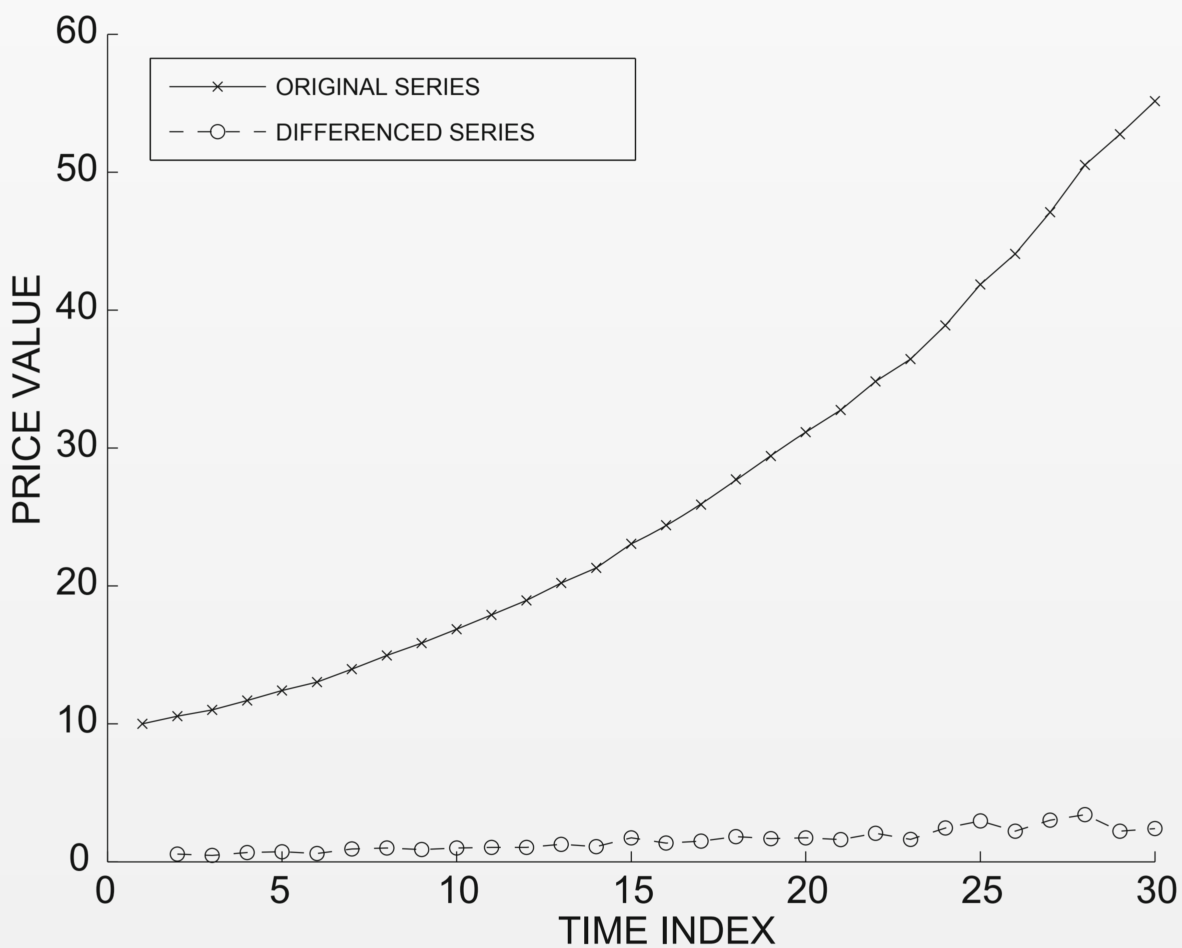
$$y'_i = \alpha \cdot y_i + (1 - \alpha) \cdot y'_{i-1}$$

Stationary Time Series

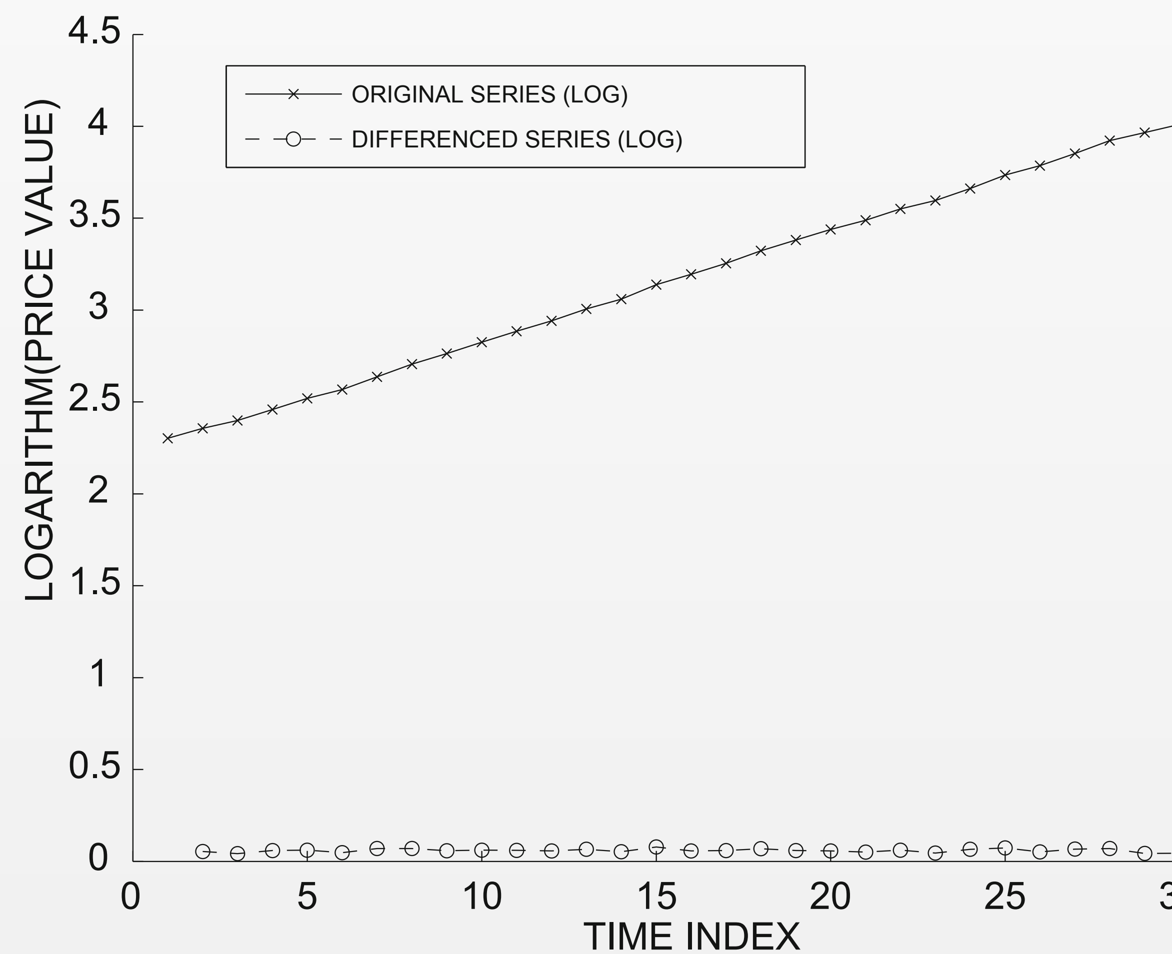
$$y_t = c + \epsilon_t \qquad\qquad E[\epsilon_t] = 0$$

Definition 14.3.1 (Strictly Stationary Time Series) *A strictly stationary time series is one in which the probabilistic distribution of the values in any time interval $[a, b]$ is identical to that in the shifted interval $[a + h, b + h]$ for any value of the time shift h.*

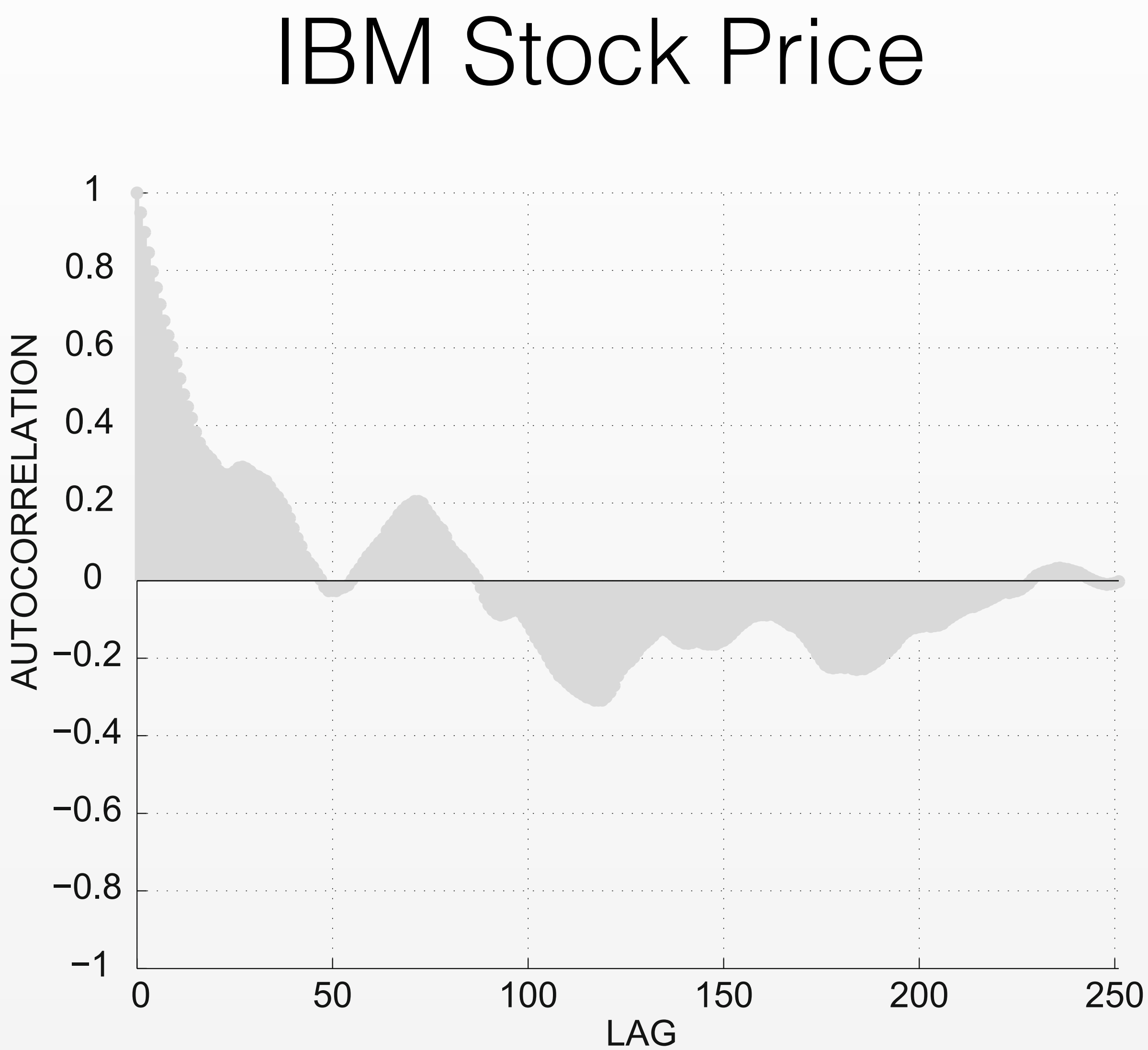
Differencing y_t - y_{t-1}

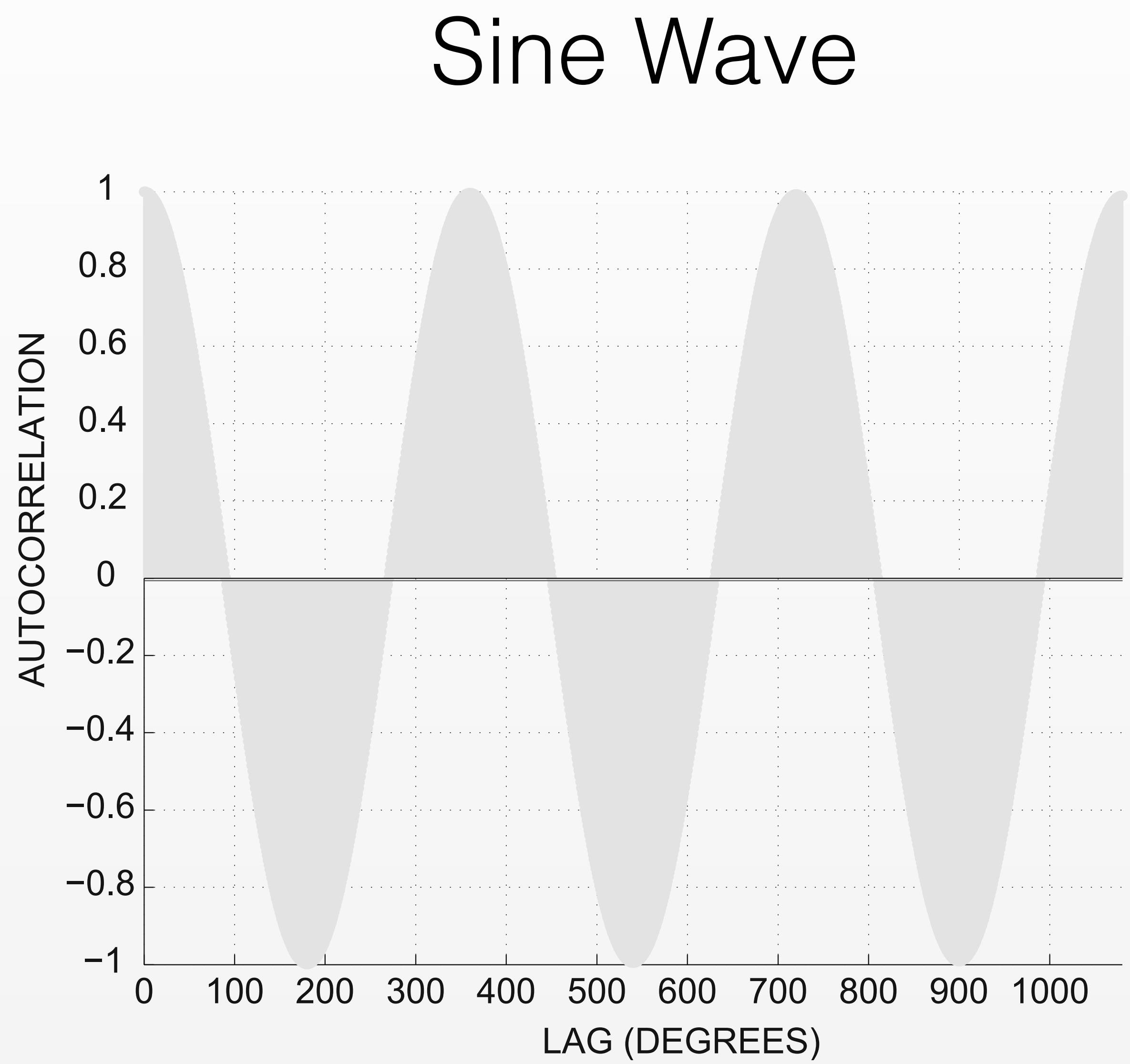


Log differencing log y_t - log y_{t-1}



Auto-correlation





$$\text{Autocorrelation}(L) = \frac{\text{Covariance}_t(y_t, y_{t+L})}{\text{Variance}_t(y_t)}.$$

Autoregressive Models

Autoregressive: AR(p)

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + c + \epsilon_t$$

Moving-Average: MA(q)

$$y_t = \sum_{i=1}^{q} b_i \cdot \epsilon_{t-i} + c + \epsilon_t$$

Autoregressive moving-average: ARMA(p,q)

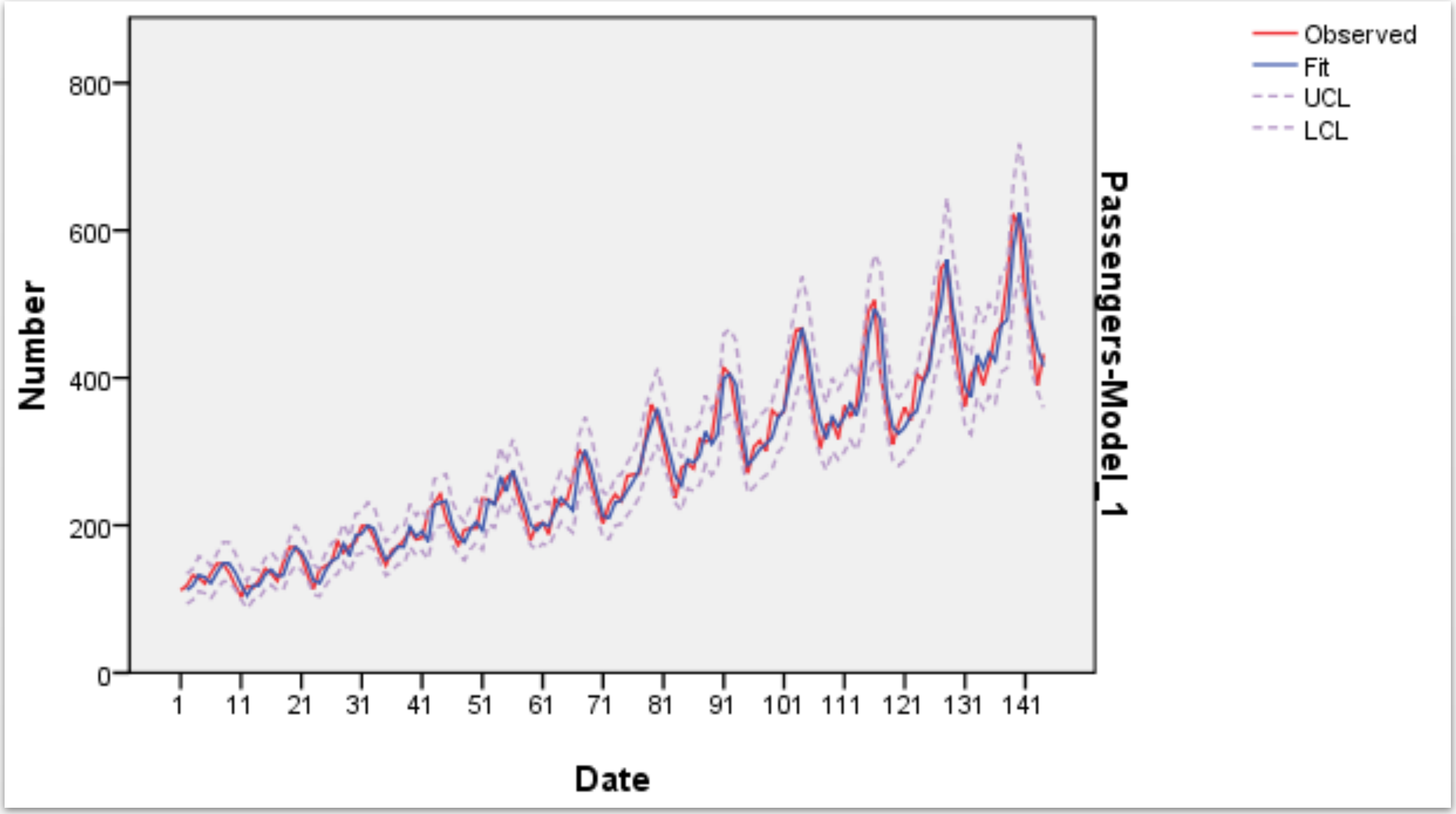
$$y_t = \sum_{i=1}^{p} a_i \cdot y_{t-i} + \sum_{i=1}^{q} b_i \cdot \epsilon_{t-i} + c + \epsilon_t$$

Autoregressive integrated moving-average: ARIMA(p,d,q)

$$y_t^{(d)} = \sum_{i=1}^{p} a_i y_{t-i}^{(d)} + \sum_{i=1}^{q} b_i \epsilon_{t-i} + c + \epsilon_t$$

Do least-squares regression to estimate a,b,c

ARIMA on Airline Data

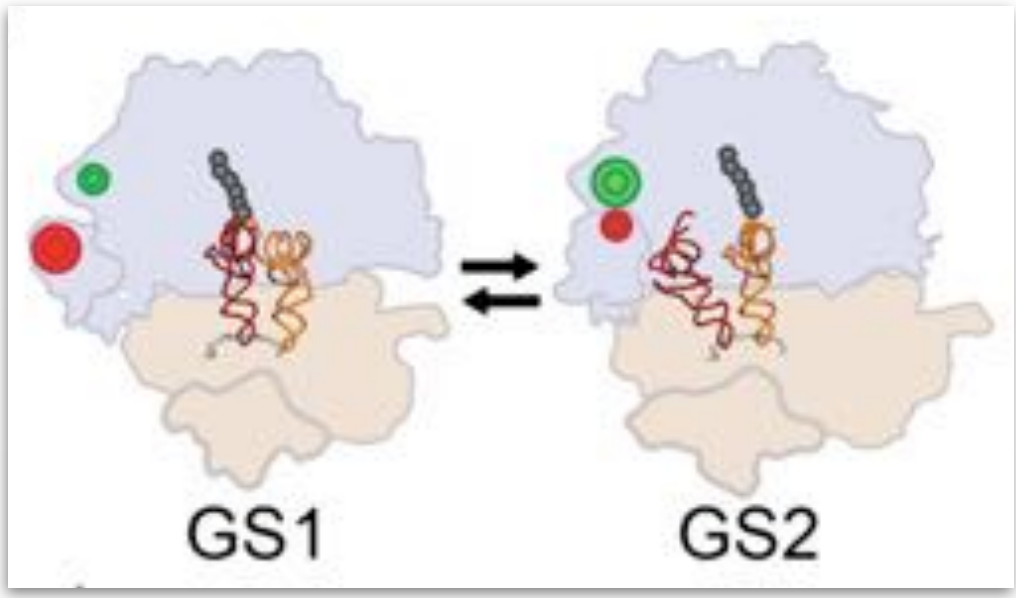


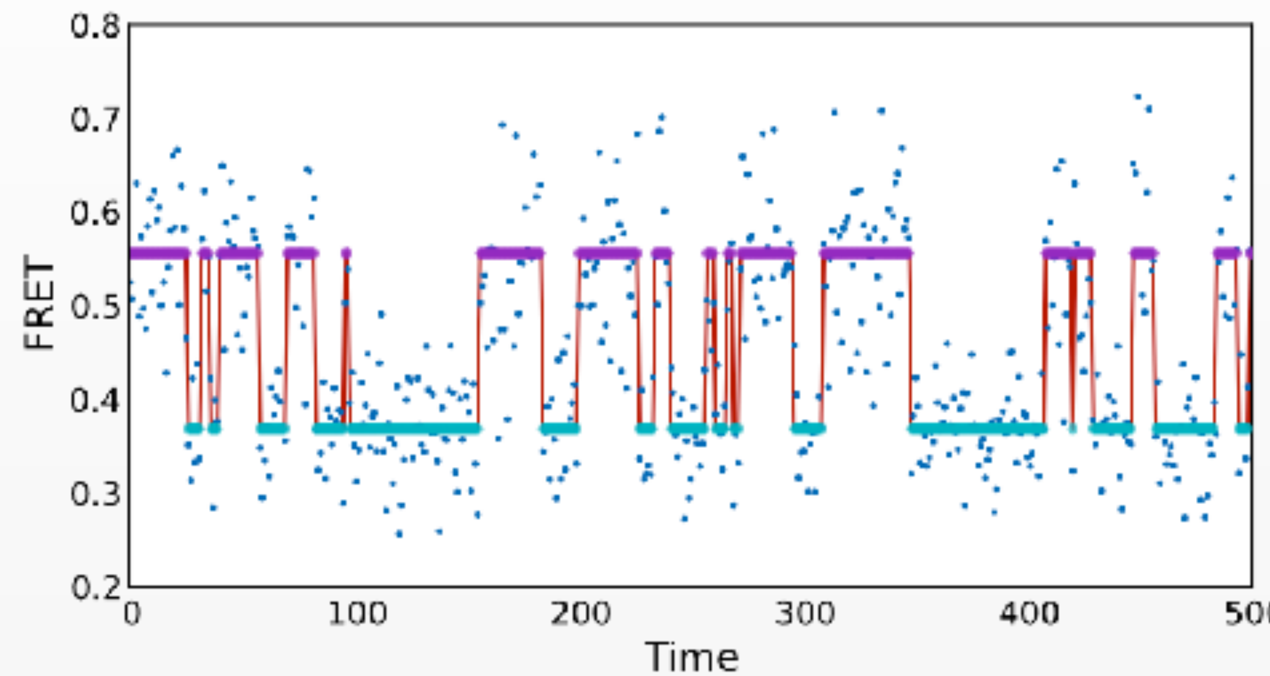
$(p,d,q) = (0,1,12)$

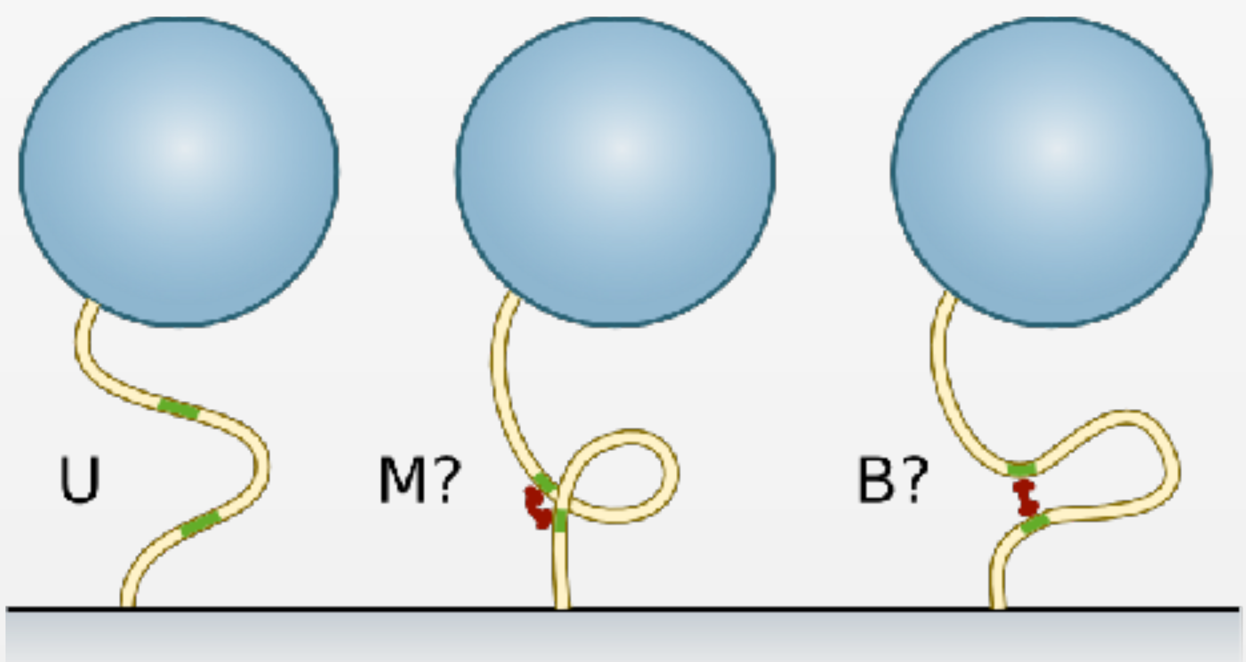
source: http://www.statsref.com/HTML/index.html?arima.html

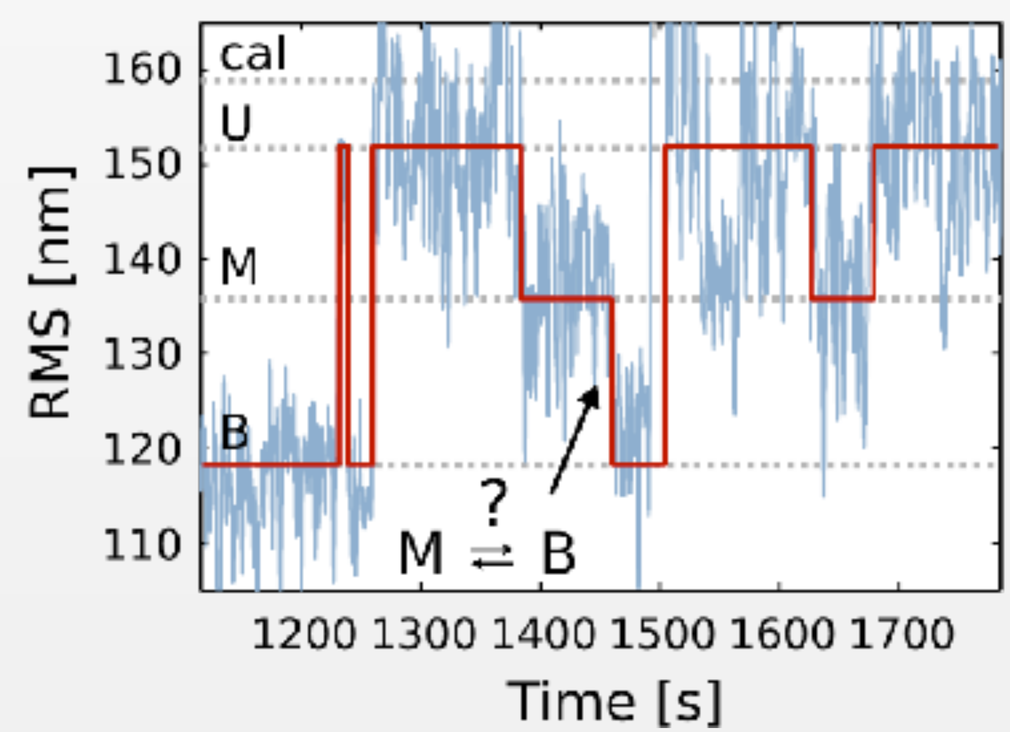
Hidden Markov Models

Time Series with Distinct States



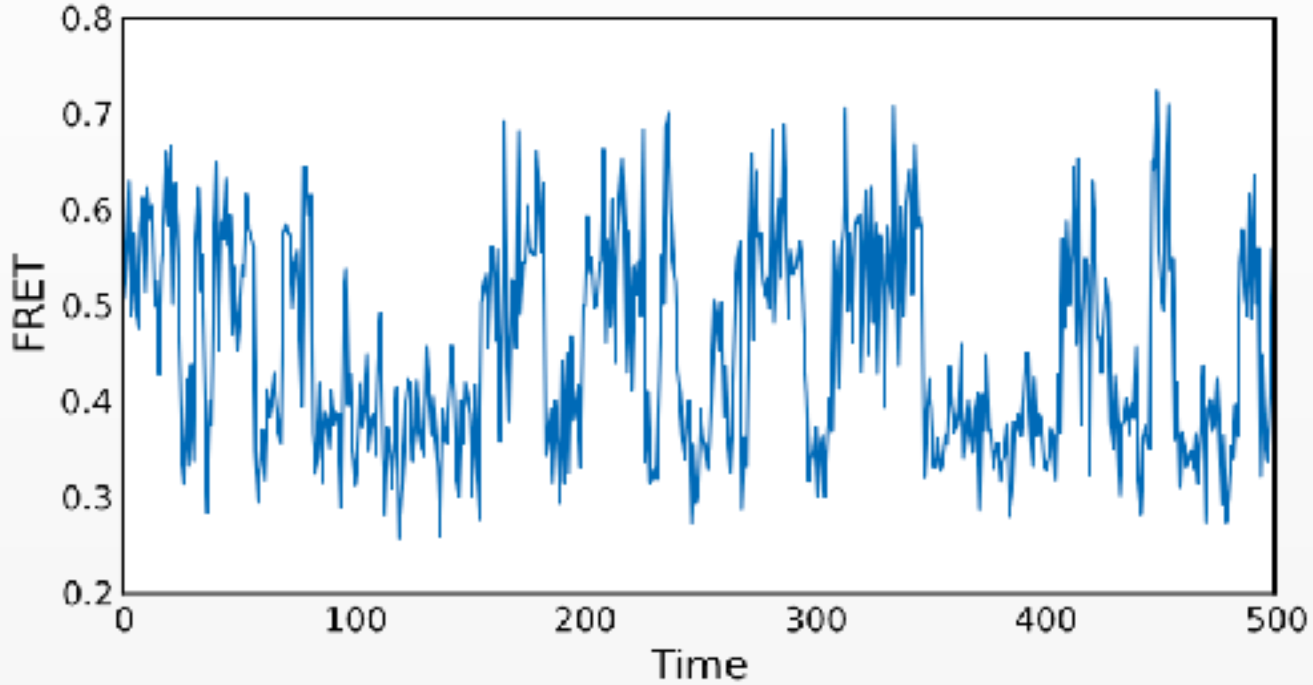




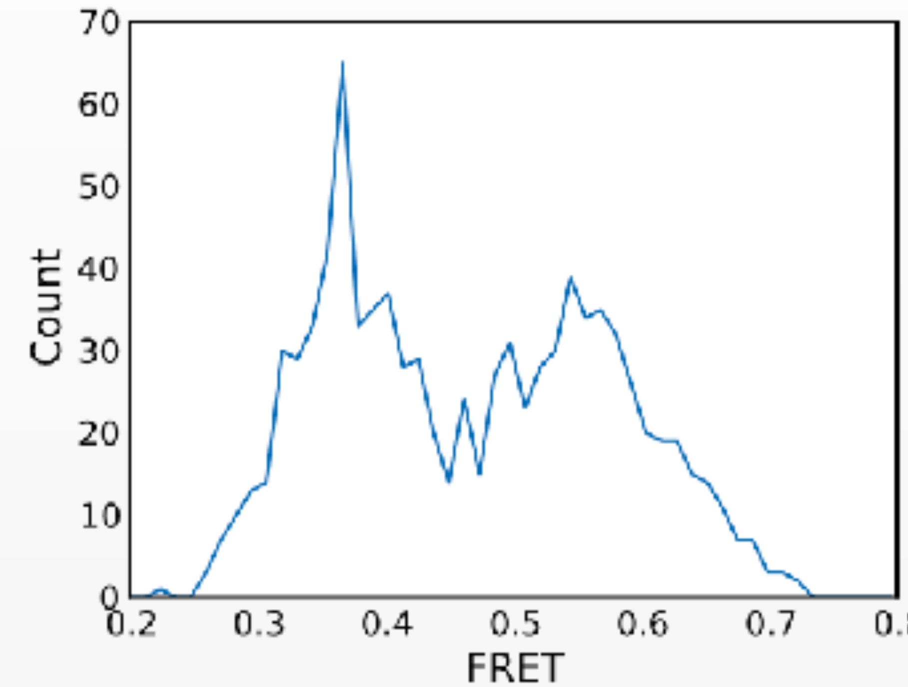


Can we use a Gaussian Mixture Model?

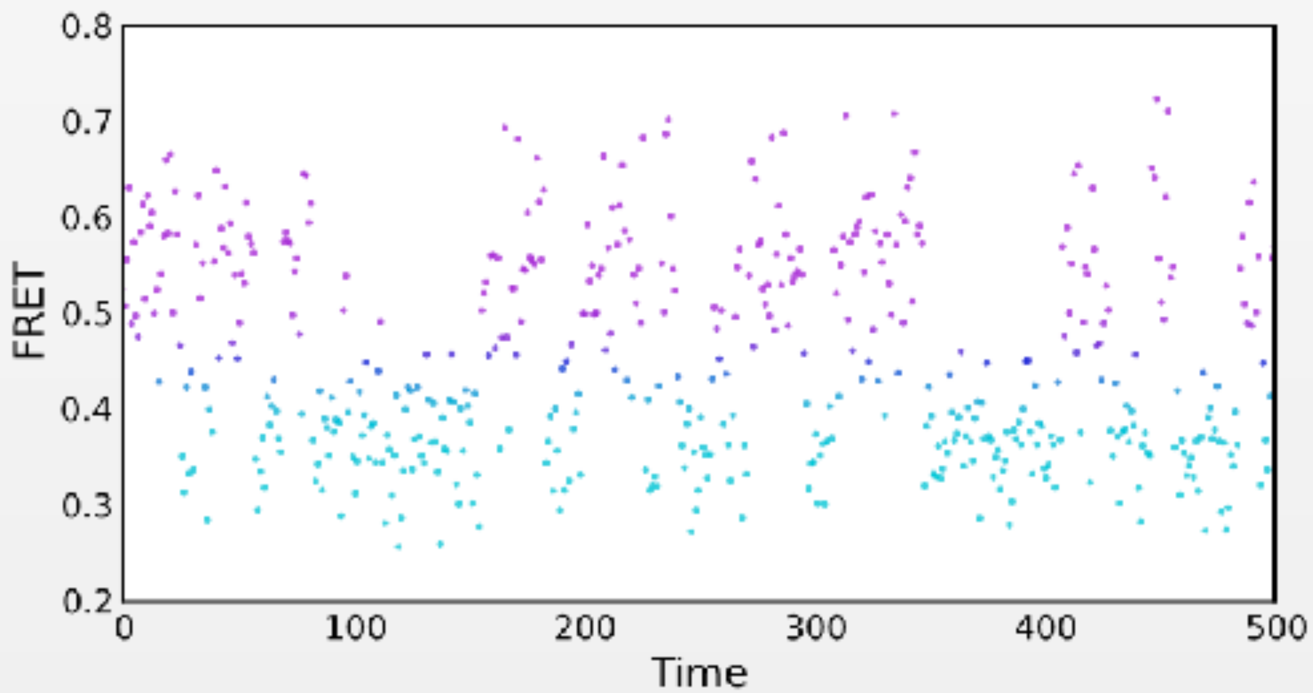
Time Series



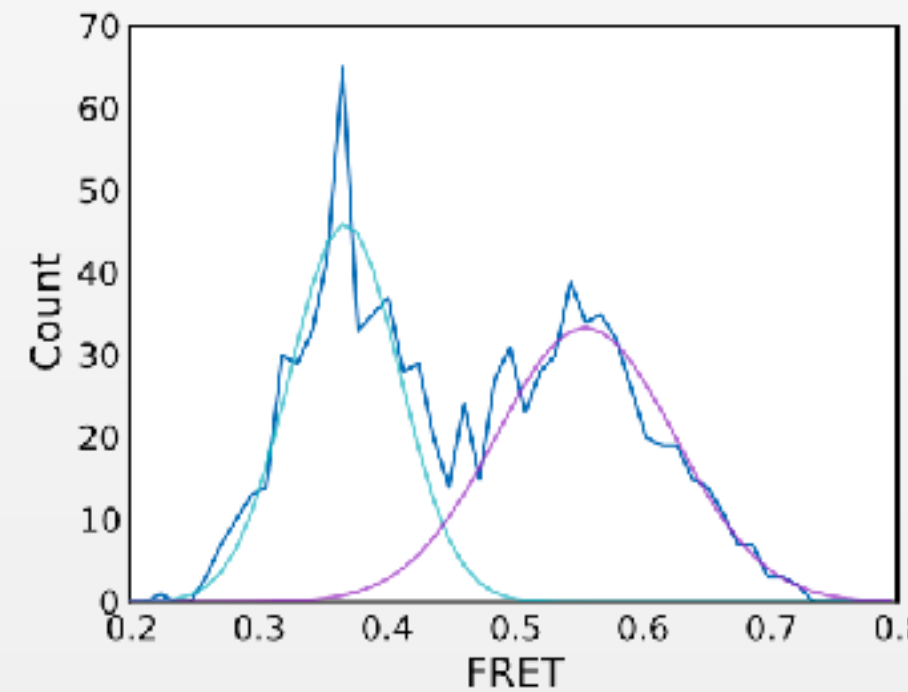
Histogram



Posterior on states

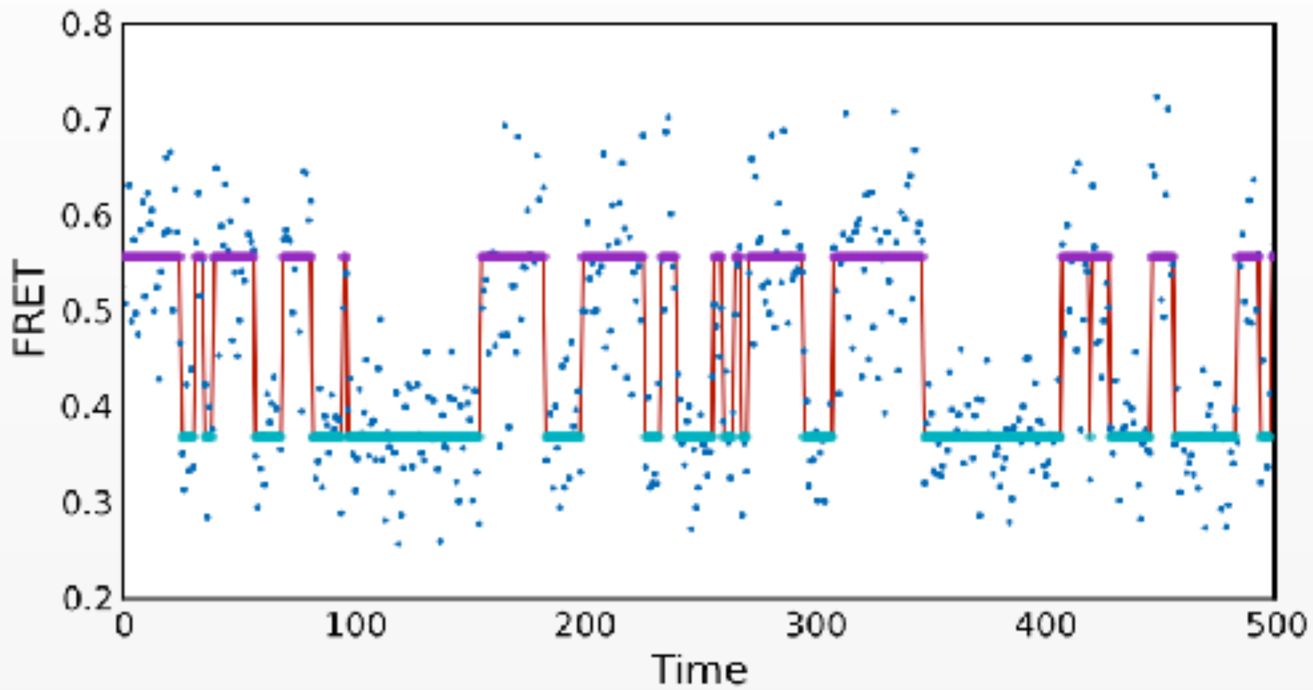


Mixture

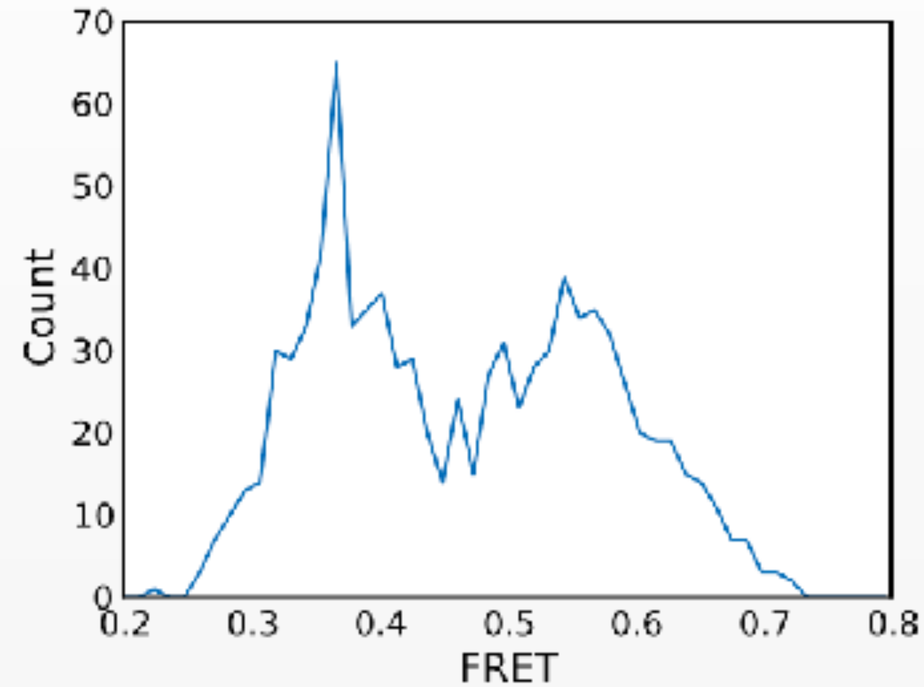


Can we use a Gaussian Mixture Model?

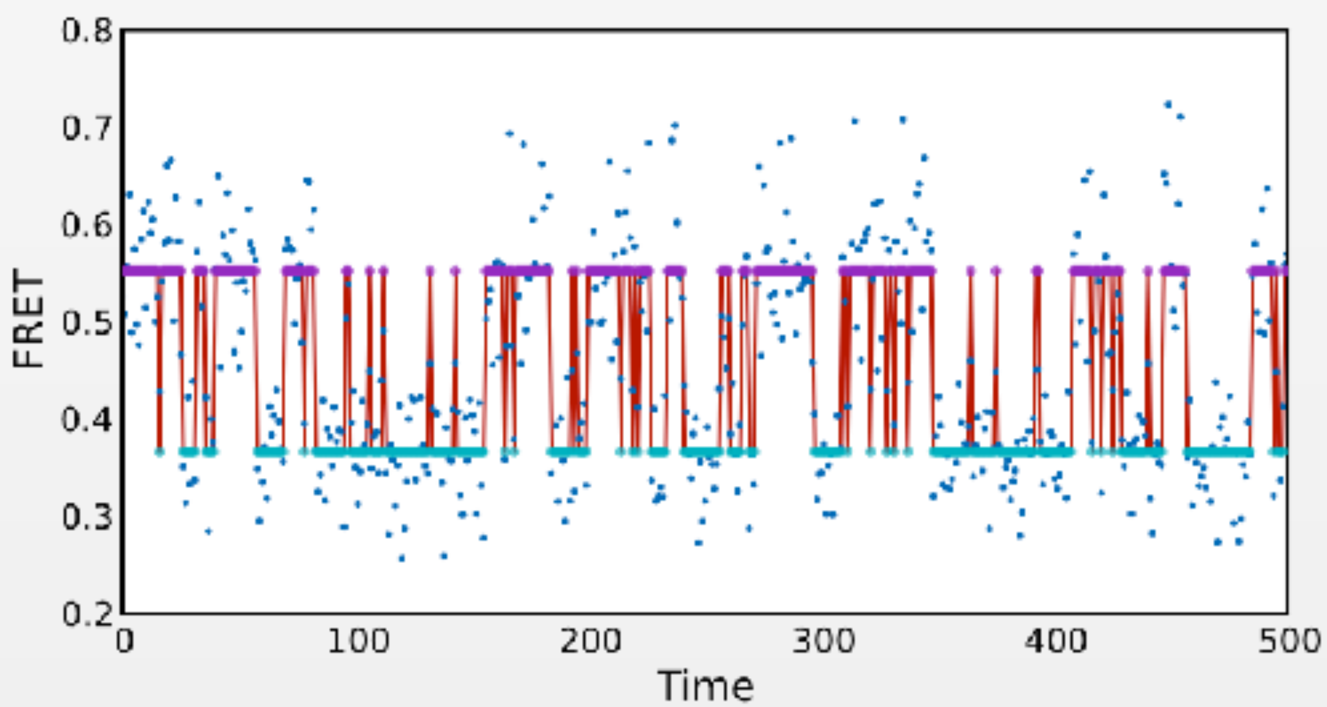
Time Series



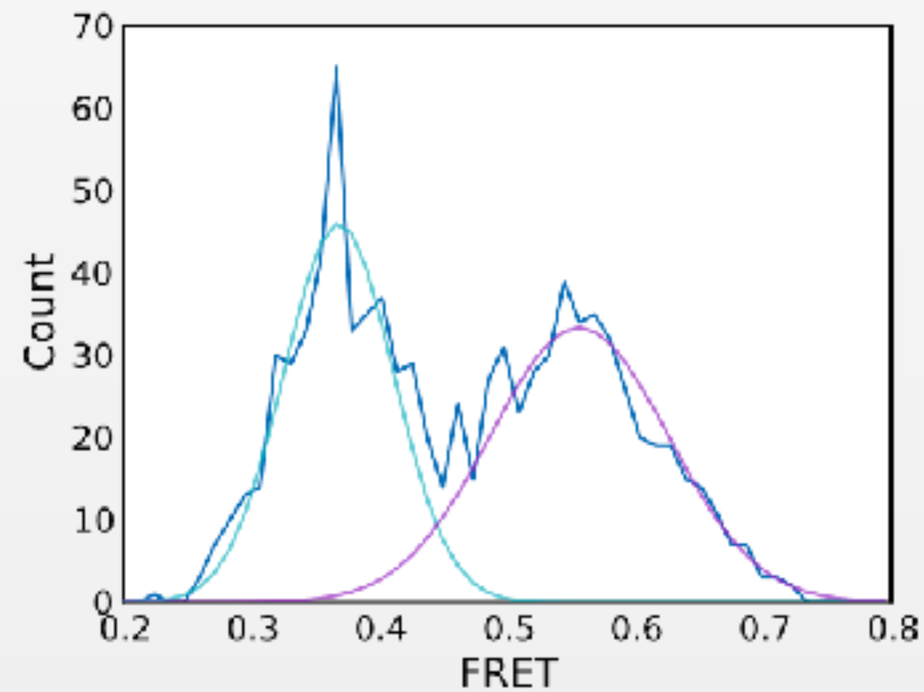
Histogram



Posterior on states

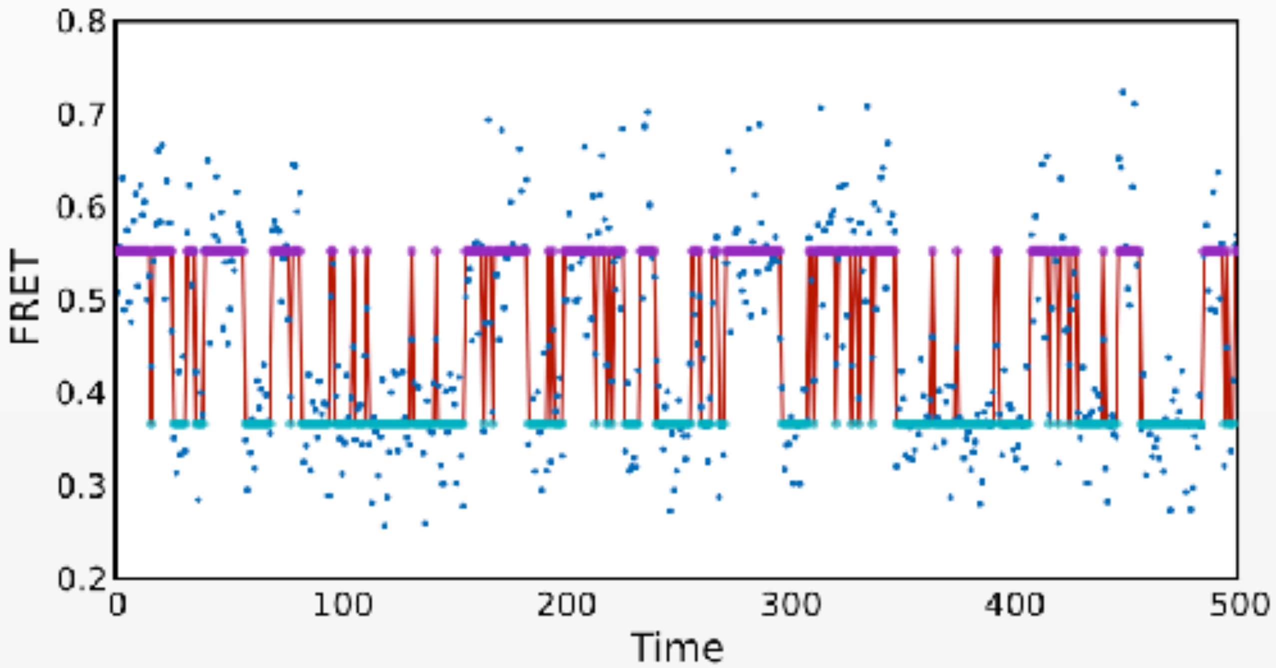


Mixture

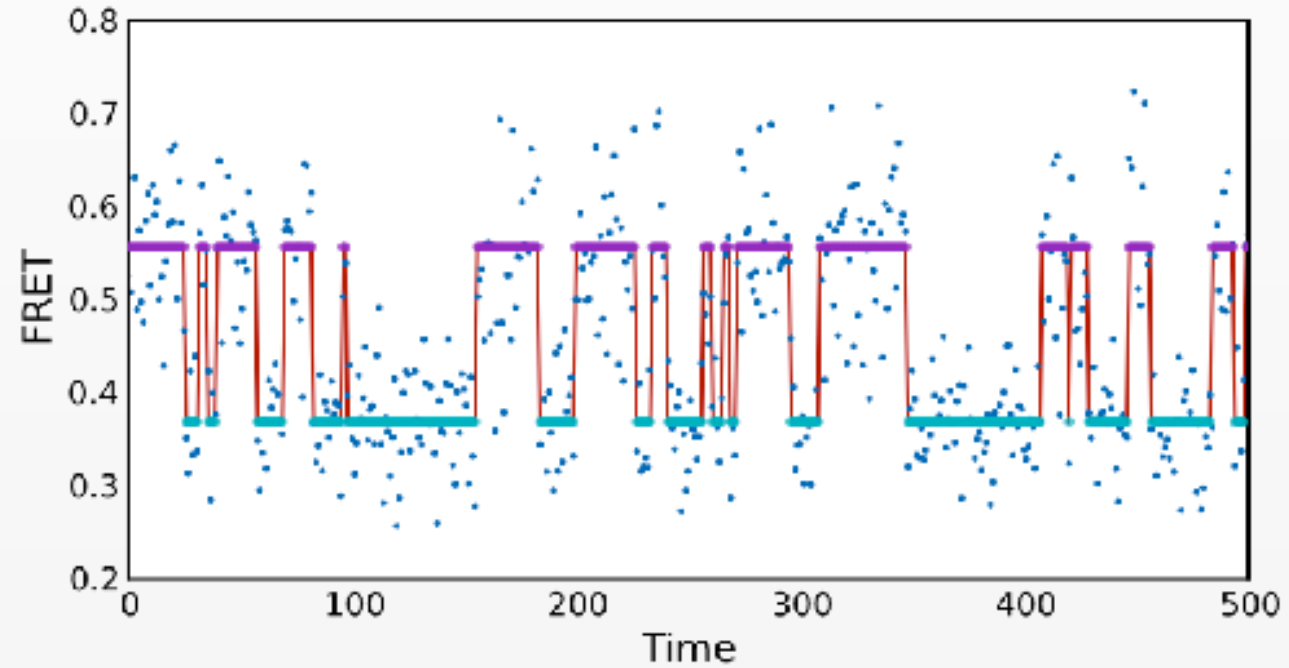


Hidden Markov Models

Estimate from GMM

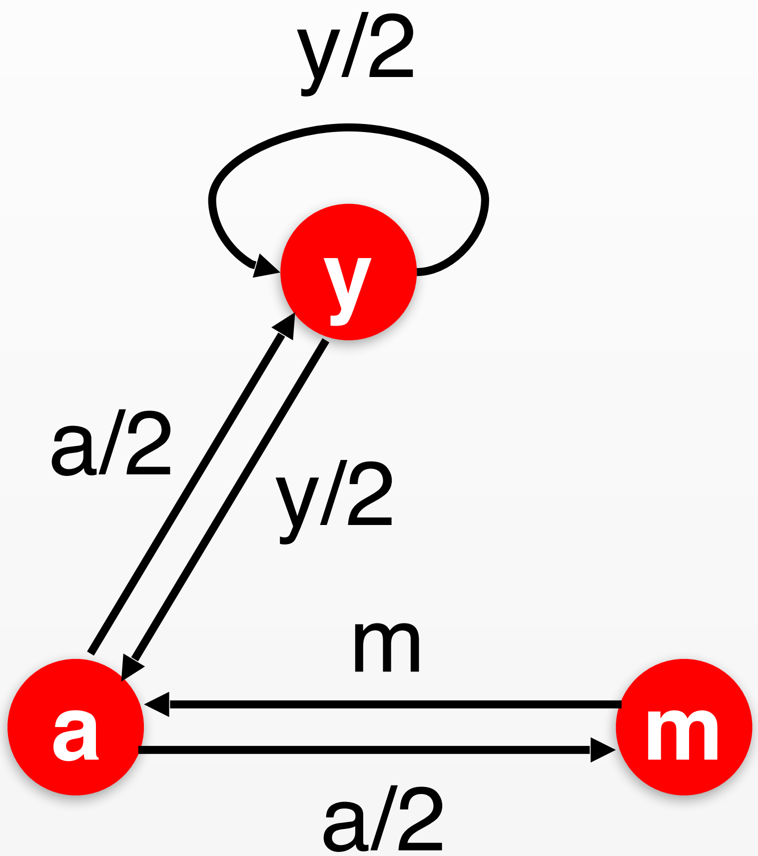


Estimate from HMM



- Idea: Mixture model + Markov chain for states
- Can model correlation between subsequent states (more likely to be in same state than different state)

Reminder: Random Surfers in PageRank



$$\boldsymbol{p}^t = M\boldsymbol{p}^{t-1} = M^t\boldsymbol{p}^0$$

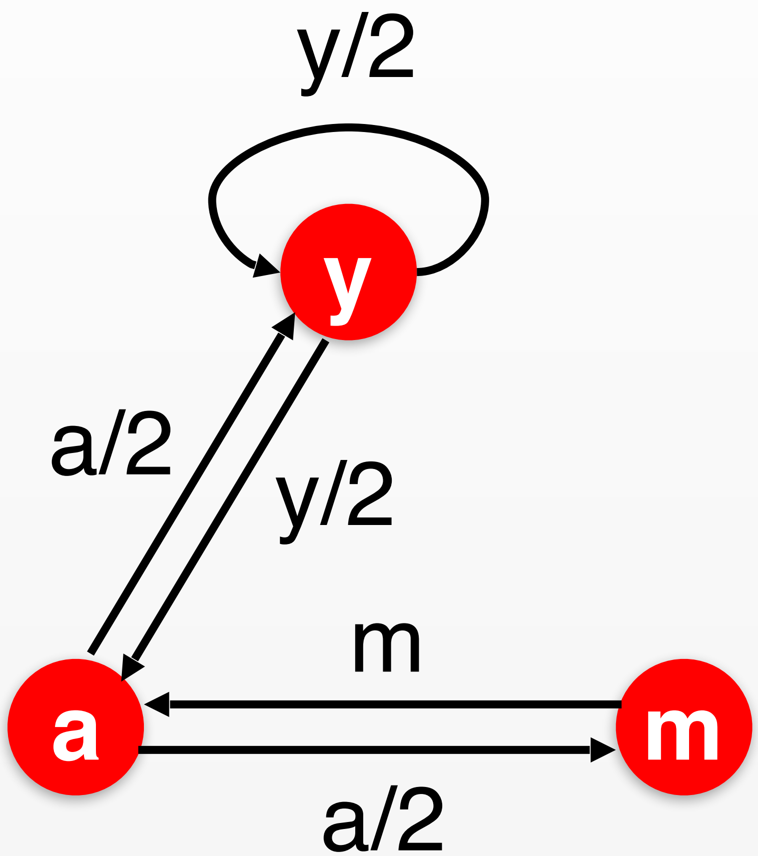
$$\boldsymbol{p}^0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \quad M = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Model for random Surfer:

- At time $t = 0$ pick a page at random
- At each subsequent time t follow an outgoing link at random

(adapted from:: Mining of Massive Datasets, http://www.mmds.org)

Reminder: Random Surfers in PageRank



$$\boldsymbol{p}^t = M\boldsymbol{p}^{t-1} = M^t\boldsymbol{p}^0$$

$$\boldsymbol{p}^0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \quad M = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

$$p(z_0 = i) = 1/N$$

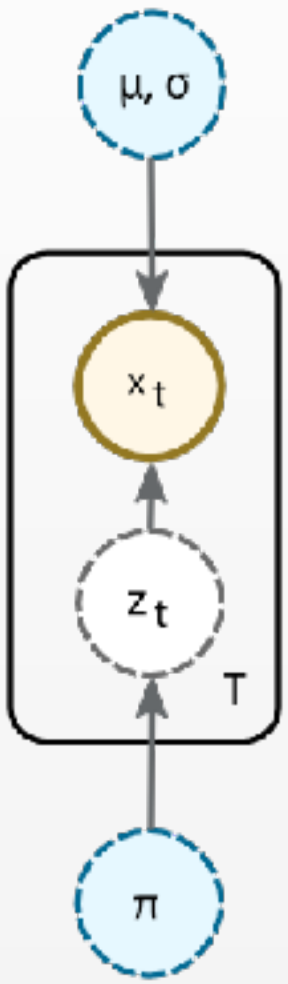
$$p(z_t = i \,|\, z_{t-1} = j) = M_{ij}$$

$$\begin{aligned} p(z_t = i) &= \textstyle\sum_j p(z_t = i, z_{t-1} = j) \\ &= \textstyle\sum_j M_{ij} p(z_{t-1} = j) \end{aligned}$$

(adapted from:: Mining of Massive Datasets, http://www.mmds.org)

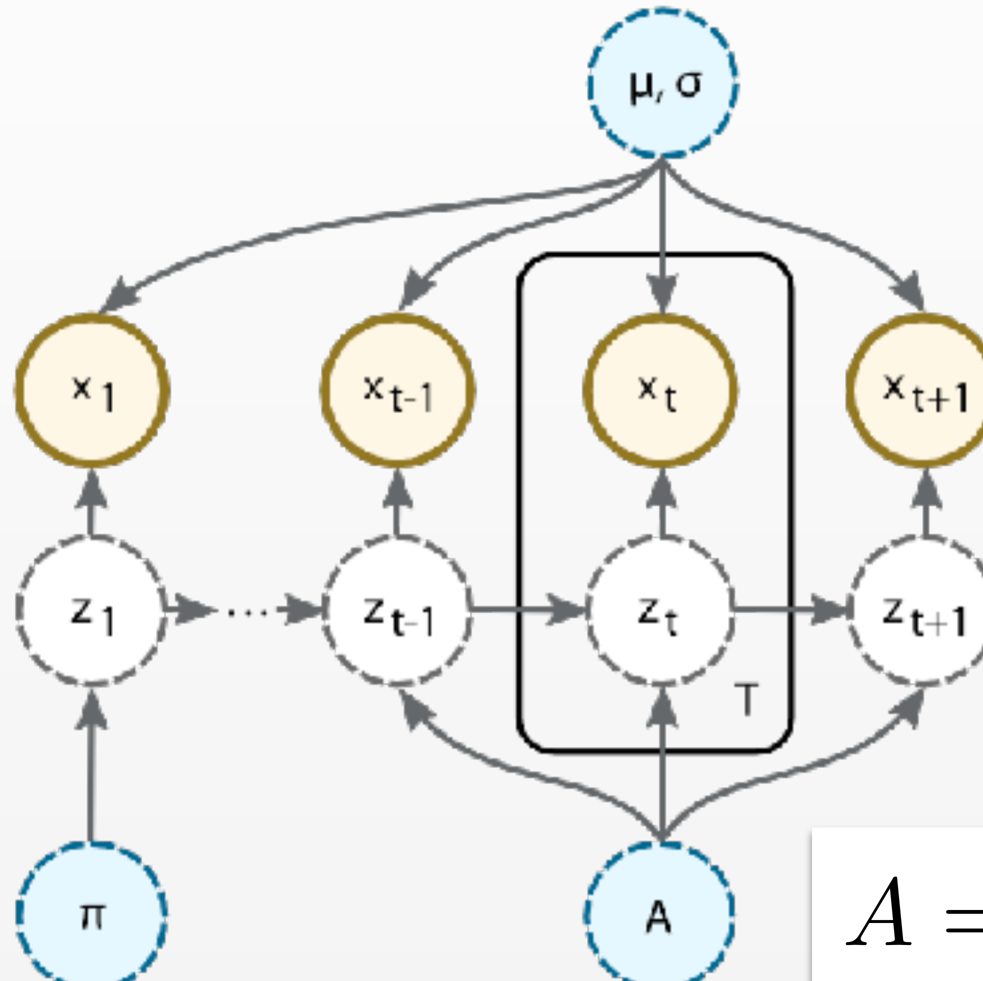
Hidden Markov Models

Gaussian Mixture



$$z_n \sim \text{Discrete}(\boldsymbol{\pi})$$
$$x_n | z_n = k \sim \text{Normal}(\mu_k, \sigma_k)$$

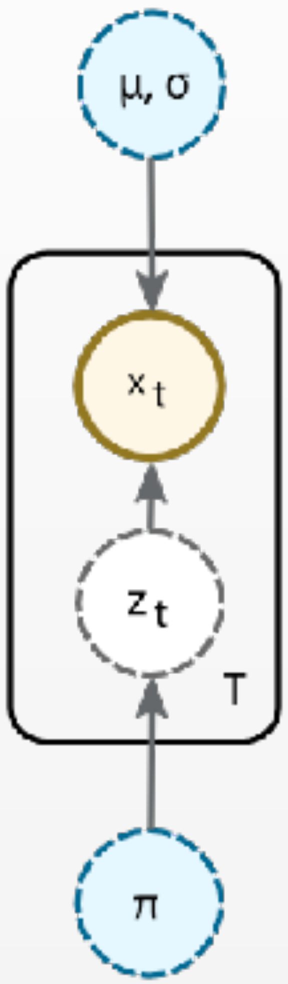
Gaussian HMM



$$A = M^\top$$

$$z_1 \sim \text{Discrete}(\boldsymbol{\pi})$$
$$z_{t+1} | z_t = k \sim \text{Discrete}(A_k)$$
$$x_t | z_t = k \sim \text{Normal}(\mu_k, \sigma_k)$$

Review: Gaussian Mixtures



$$z_n \sim \text{Discrete}(\boldsymbol{\pi})$$
$$x_n | z_n = k \sim \text{Normal}(\mu_k, \sigma_k)$$

Expectation Maximization

1. Update cluster probabilities

$$\gamma_{tk}^i = p(z_t = k \,|\, x_t, \boldsymbol{\theta}^{i-1}) = \frac{p(x_t, z_t = k \,|\, \boldsymbol{\theta}^{i-1})}{\sum_l p(x_t, z_t = l \,|\, \boldsymbol{\theta}^{i-1})}$$

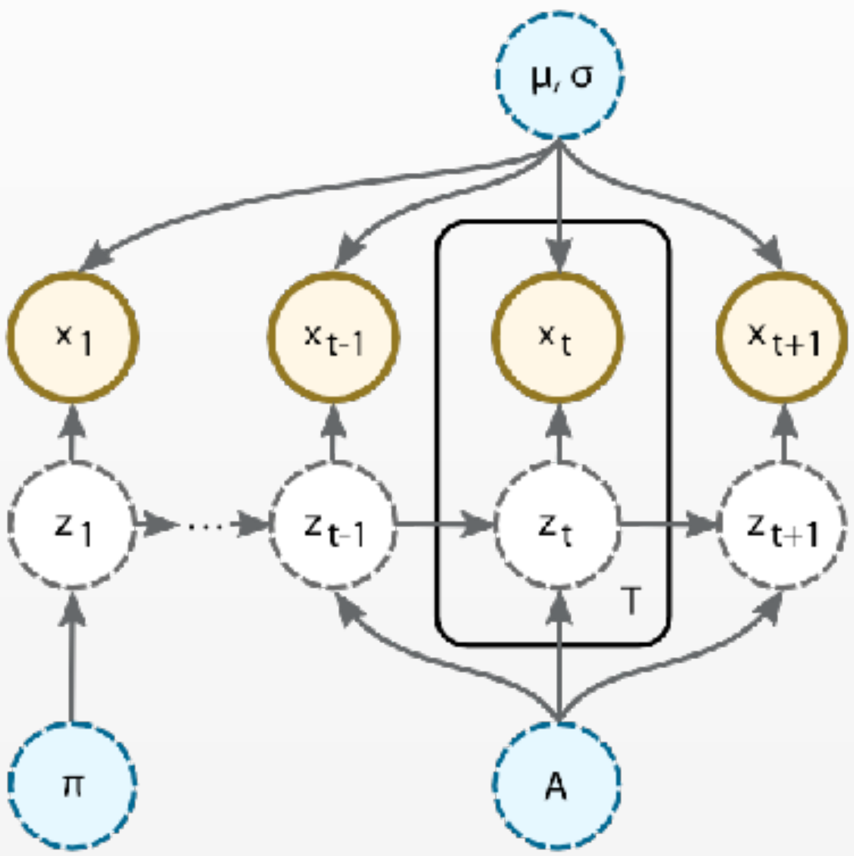
2. Update parameters

$$\mu_k^i = \tfrac{1}{N_k^i} \textstyle\sum_{t=1}^T \gamma_{tk}^i x_t$$

$$\sigma_k^i = \left(\tfrac{1}{N_k^i} \textstyle\sum_{t=1}^T \gamma_{tk}^i (x_t^i - \mu_k^i)^2 \right)^{1/2}$$

$$\pi_k^i = N_k^i / N \qquad N_k^i = \textstyle\sum_{t=1}^T \gamma_{tk}^i$$

Forward-backward Algorithm



$$z_1 \sim \text{Discrete}(\boldsymbol{\pi})$$
$$z_{t+1}|z_t = k \sim \text{Discrete}(A_k)$$
$$x_t|z_t = k \sim \text{Normal}(\mu_k, \sigma_k)$$

Expectation step for HMM

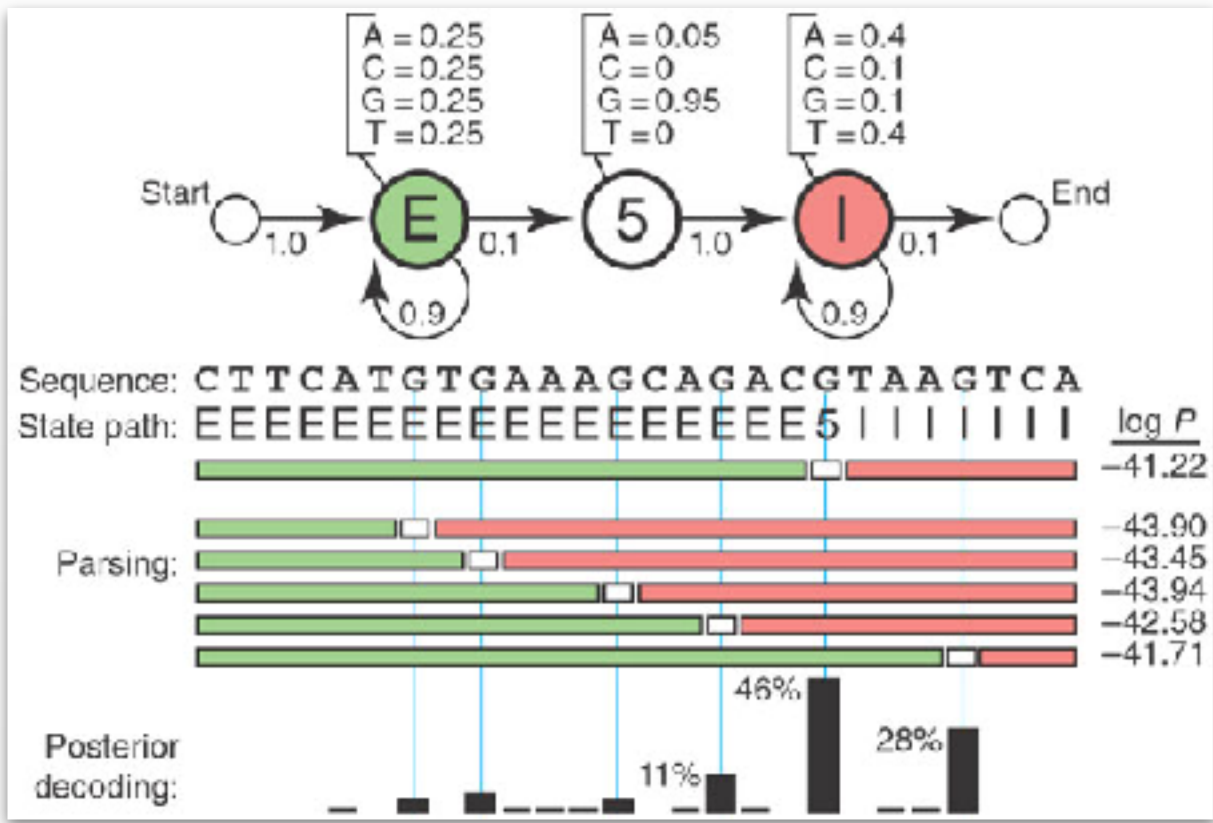
$$\gamma_{t,k} = p(z_t = k \,|\, x_{1:T}, \boldsymbol{\theta})$$
$$= \frac{p(x_{1:t}, z_t)p(x_{t+1:T}|z_t)}{p(x_{1:T})}$$
$$\propto \alpha_{t,k}\beta_{t,k}$$

$$\alpha_{t,l} := p(x_{1:t}, z_t)$$
$$= \sum_k p(x_t|\mu_l, \sigma_l) A_{kl} \alpha_{t-1,k}$$

$$\beta_{t,k} := p(x_{t+1:T} \,|\, z_t)$$
$$= \sum_l \beta_{t+1,l}\, p(x_{t+1}|\mu_l, \sigma_l)\, A_{kl}$$

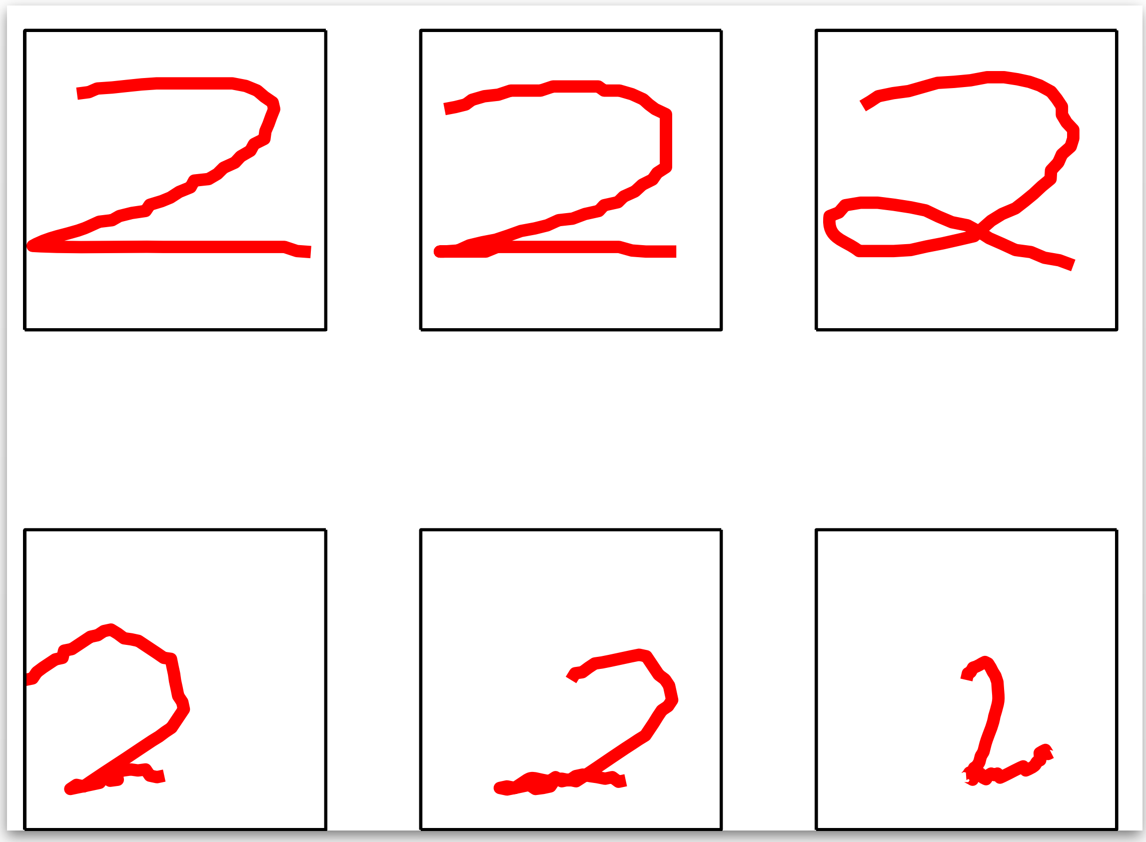
Other Examples for HMMs

RNA splicing



- State 1: Exon (relevant)
- State 2: Splice site
- State 3: Intron (ignored)

Handwritten Digits



- State 1: Sweeping arc
- State 2: Horizontal line