ε-Machine Estimation and Forecasting

Comparative Study of Inference Methods

D. Shemetov[1]

[1]Department of Mathematics
University of California, Davis

Natural Computation, 2014

Outline

Outline

ε-Machines and Time Series.

- ε-machines eat sequences of alphabet symbols: $\{\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots\}$

ε-Machines and Time Series.

- ε-machines eat sequences of alphabet symbols: $\{\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots\}$
- These look way too much like time series: $\{X_t | t \in \mathbb{Z}\}$

ε-Machines and Time Series.

- ε-machines eat sequences of alphabet symbols: $\{\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots\}$
- These look way too much like time series: $\{X_t | t \in \mathbb{Z}\}$
- Let's compare with time series methods?

Time Series.

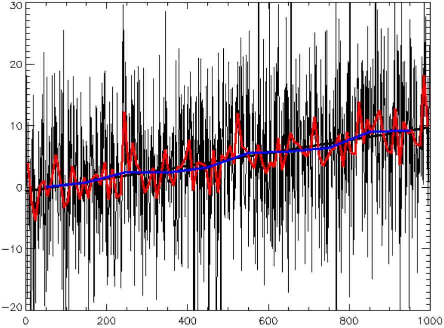
- A time series is a sequence of measurements taken from a system, spaced at regular time-intervals: $\{X_t | t \in \mathbb{Z}\}$

Time Series.

- A time series is a sequence of measurements taken from a system, spaced at regular time-intervals: $\{X_t | t \in \mathbb{Z}\}$
- Examples:
  - Atmospheric/climate data
  - Seismic activity data
  - Stock price data
  - Life

Time Series.

- A time series is a sequence of measurements taken from a system, spaced at regular time-intervals: $\{X_t | t \in \mathbb{Z}\}$
- Examples:
  - Atmospheric/climate data
  - Seismic activity data
  - Stock price data
  - Life



Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:

Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:
- Autocorrelation

Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:
- Autocorrelation
- Spectral analysis

Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:
- Autocorrelation
- Spectral analysis
- Principal component analysis

Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:
- Autocorrelation
- Spectral analysis
- Principal component analysis
- Neural networks

Time Series Method Ecosystem.

- There are a lot of methods for studying and forecasting these things:
- Autocorrelation
- Spectral analysis
- Principal component analysis
- Neural networks

Tools for investigating time-series data include:

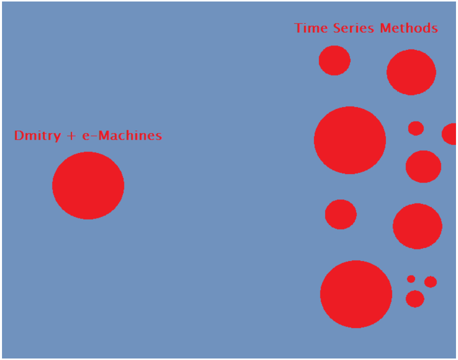
- Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions)
- Scaled cross- and auto-correlation functions to remove contributions of slow components[8]
- Performing a Fourier transform to investigate the series in the frequency domain
- Use of a filter to remove unwanted noise
- Principal component analysis (or empirical orthogonal function analysis)
- Singular spectrum analysis
- "Structural" models:
  - General State Space Models
  - Unobserved Components Models
- Machine Learning
  - Artificial neural networks
  - Support Vector Machine
  - Fuzzy Logic
- Hidden Markov model
- Control chart
  - Shewhart individuals control chart

[Dmitry + ε-machines] vs. [Time series]

- Let's compare!

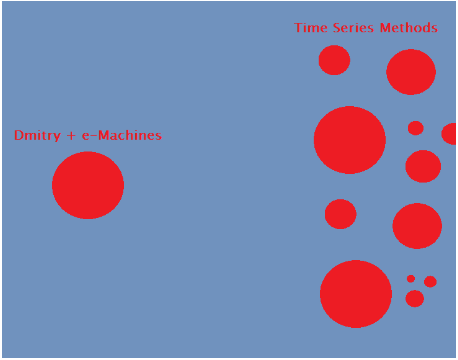
[Dmitry + ε-machines] vs. [Time series]

- Let's compare!



[Dmitry + ε-machines] vs. [Time series]

- Let's compare!



- Too many! Pick battles you can win!

Outline

Baum-Welch - What is it?

The Baum-Welch algorithm - a method for inferring the most likely parameters of a Hidden Markov Model responsible for some given string of data.

Example

Input

- Number of states - n
- Network topology and initial guess for the parameters of the HMM - θ
  - Parameters are the start distribution, state transition probabilities, and symbol emission probabilities at each state.
- A sequence of outputted symbols - $Y = \{Y_1, Y_2, ..., Y_n\}$

Example

Output

- New parameters θ' with the property that $P(Y|\theta') \geq P(Y|\theta)$.

Baum-Welch - How does it work?

- Belongs to a broad class of EM algorithms - expectation maximization.

Baum-Welch - How does it work?

- Belongs to a broad class of EM algorithms - expectation maximization.
- Too many details:
  - Makes a forward pass to compute $\alpha_i(t) = P(''y_1, \ldots, y_t'', X_t = i|\theta)$
    - ★ Probability of seeing "word" and ending up in state i at time t.
  - Makes a backwards pass to compute $\beta_i(t) = P(''y_{t+1}, \ldots, y_T''|X_t = i, \theta)$
    - ★ Probability of having the last $T - t$ symbols be the specified "word", assuming that you ended up in state i at time t.
  - Combined these in the update step to calculate $\xi_{ij}(t) = P(X_t = i, X_{t+1} = j|Y, \theta)$
    - ★ Probability of transitioning $i \to j$ at time t.
    - ★ $\xi_{ij}(t)$ can be computed as a function of α_i's and β_i's
  - Where this eventually gets us:
    - ★ Summing $\xi_{ij}(t)$ over time, we can compute the expected number of transitions from state $i \to j$, as well as $i \to *$.
    - ★ This will allow us to compute the expected transition probabilities between states.
    - ★ And! Similarly, we can compute the expected symbol emission matrix for each state.
    - ★ Starting probabilities.
    - ★ We now have θ'!

Baum-Welch - More.

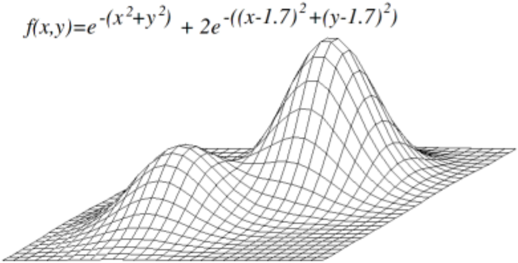
- We can run this procedure again.

Baum-Welch - More.

- We can run this procedure again.
- Thus, Baum-Welch algorithm is an iterative, gradient-based optimizer in the space of HMM parameters.

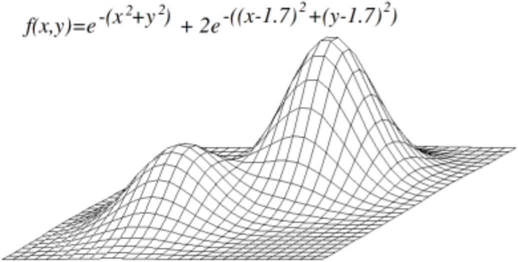
Baum-Welch - More.

- We can run this procedure again.
- Thus, Baum-Welch algorithm is an iterative, gradient-based optimizer in the space of HMM parameters.
- "Hill-climber" algorithm.



Baum-Welch - More.

- We can run this procedure again.
- Thus, Baum-Welch algorithm is an iterative, gradient-based optimizer in the space of HMM parameters.
- "Hill-climber" algorithm.



- Gets stuck in local optima, so you have to seed it and rerun it with many starting conditions.

Outline

1. Motivation
   - ε-Machines and Time Series
   - Baum-Welch Algorithm
   - The Future, The Comparison Project

ε-Machines and Baum-Welch.

- We can ask a number of questions.

ε-Machines and Baum-Welch.

- We can ask a number of questions.

Pause

- Given impoverished data [small sample], do they differ in their rate of convergence to the true HMM as sample size is increased?

ε-Machines and Baum-Welch.

- We can ask a number of questions.

Pause

- Given impoverished data [small sample], do they differ in their rate of convergence to the true HMM as sample size is increased?
- What about their error bars or confidence intervals?

ε-Machines and Baum-Welch.

- We can ask a number of questions.

Pause

- Given impoverished data [small sample], do they differ in their rate of convergence to the true HMM as sample size is increased?
- What about their error bars or confidence intervals?
- Maybe run-time speed, but that's tough - have to prove to C-language speed demons that you have a fast implementation.

ε-Machines and Baum-Welch.

- We can ask a number of questions.

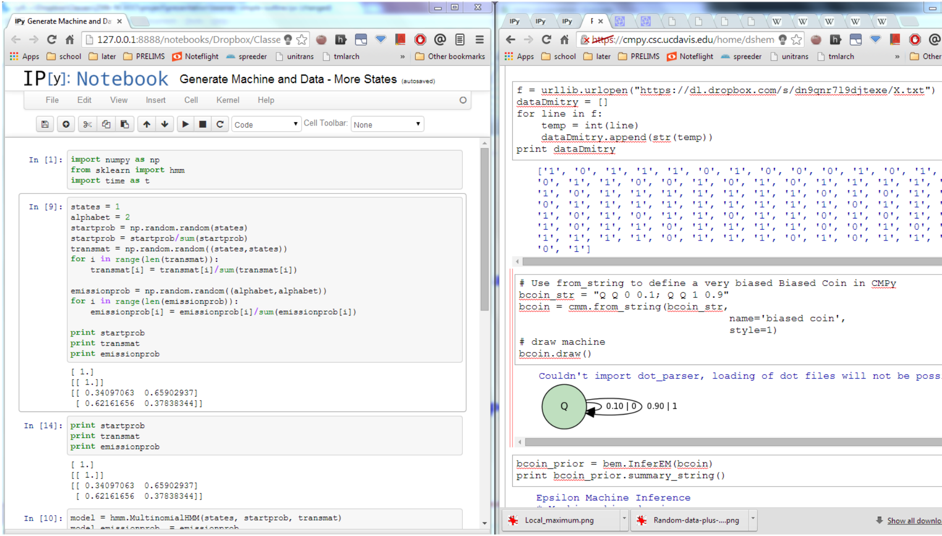
Pause

- Given impoverished data [small sample], do they differ in their rate of convergence to the true HMM as sample size is increased?
- What about their error bars or confidence intervals?
- Maybe run-time speed, but that's tough - have to prove to C-language speed demons that you have a fast implementation.
- As fun side-project / implementation, would like to translate HMMs into some audio signal [for ex. each symbol corresponds to a pitch being emitted].
  - See if the HMMs the algorithms produce "sound" different.

Sidenote: IPython.

- IPython is great for Windows machines that have run VM's.
- It is Sage-like, here is my workflow:

Sidenote: IPython.



Thank you!

Thank you! Everybody!