Estimating Dynamic Networks Using A Hidden Markov Model

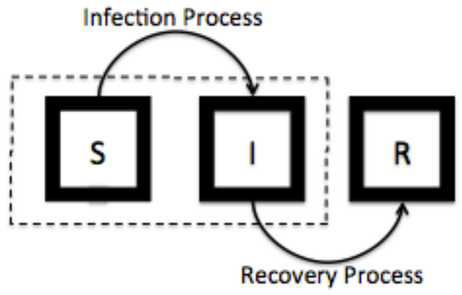
Haleh Ashki, Peter Beerli

Department of Scientific Computing, Florida State University, Tallahassee, FL

INTRODUCTION

Human life and diseases are inseparable. Diseases can be caused by our own bodies as they age and degenerate or by infectious pathogens. Our study is about infectious diseases, such as flu or sexually transmitted diseases. The prediction of the spread of a disease is paramount to establish intervention methods or procedures to curb an epidemic. There are three key parameters in modeling of epidemic diseases:

- SIR model : Susceptible, Infected, and Recovered
- Social Contact network, representing person-to-person contact: static or dynamic
- Genome sequenced data of infected host



We have developed theoretical approaches that can take into account dynamic networks and, independently, that can use genomic data of the pathogen, sampled from infected individuals, to reconstruct the path of an epidemic. By considering the location and time of the sampled pathogen sequence data we can combine the sampled infection network and the mutational history of the pathogen to reconstruct a more accurate contact network. We can reconstruct this dynamic contact networks using genetic data and epidemic parameters via a Hidden Markov Model: HMM

Method

HMM is a powerful statistical probability distribution modeling method typically used for time series data. Given plenty of data that are generated by some hidden mechanism, we create a HMM architecture and the Expectation Maximization algorithm allow us to find out the best model parameters that account for the observed data. Here we will use the Baum-Welch algorithm also known as forward-backward algorithm estimates the model parameters.

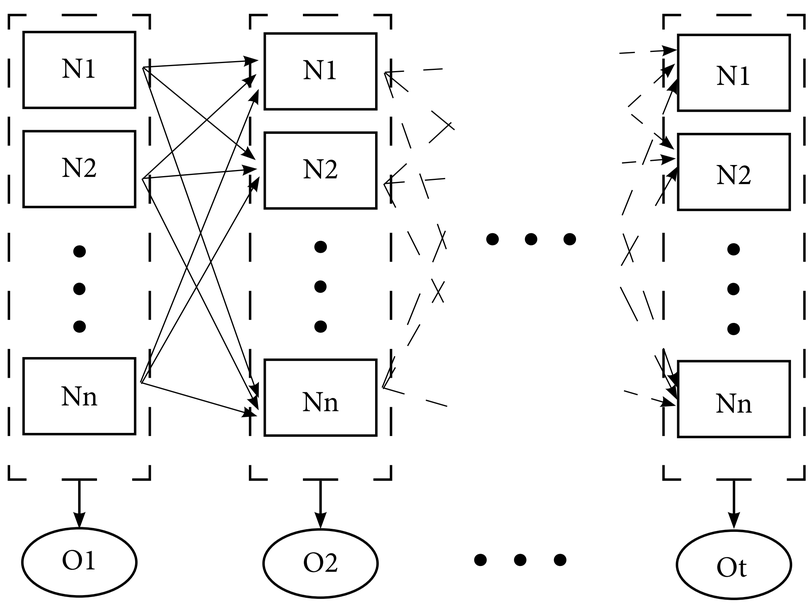
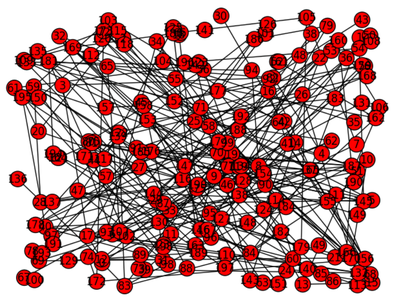


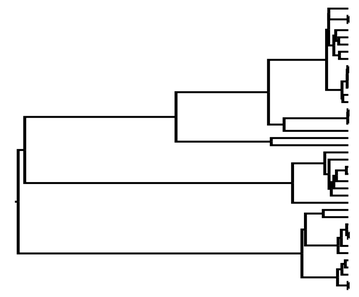
Fig1: Hidden and observed state of HMM for time series data

REFERENCES

1. Zhang, Yingjian. *Prediction of financial time series with Hidden Markov Models*. Diss. Simon Fraser University, 2004.
2. Erdős, P., Renyi, A., 1959 On the evolution of random graph. Publicationes Mathematicae 6: 290-297.
3. Welch, Lloyd R. "Hidden Markov models and the Baum-Welch algorithm." *IEEE Information Theory Society Newsletter* 53.4 (2003): 10-13.



Hidden state: $[N_1 \dots N_n]$ set of dynamic contact networks representing social contact structure changes over time. Figure shows one state: N_j



Observed data: $[O_1 \dots O_t]$ Coalescent tree constructed based on genome sequenced data of sampled infected host at different time over the course of an epidemic.

HMM for parameter maximization

Given an observation sequence, want to find the model parameters $\mu = (A,B, \boldsymbol{\pi})$ that best explains the observation sequence.
Reformulated as find the parameters that maximize $P(O|\mu)$

$$\underset{\mu}{\operatorname{argmax}}\ P(O|\mu)$$

$$\mu = (A, B, \pi)$$

$$\pi = P(N_1 = i)$$

$$A = \{a_{ij}\} = P(N_t = j | N_{t-1} = i)$$

$$B = bj(o_t) = P(O_t = o_t | N_t = j)$$

μ: Initial state
A: State Transition
B: Probability of observation given hidden state

Challenge: Likelihood function

$P(Ot|Nt)$: relates the probability of an observed coalescent tree given a particular hidden network structure. We approximate the likelihood numerically using a distance variant between the tree and each of the hidden networks. Both coalescent tree and network structure would be mapped to adjacency matrix and then the Euclidian matrix would be calculated.

Baum-Welch algorithm

This is a special case of the EM method. It works iteratively to improve the likelihood of **P(O|μ)**.

$$\alpha_j(t+1) = bj(o_{t+1}) \sum \alpha_i(t) a_{ij}$$

$$\beta_i(t) = \sum \beta_j(t+1) a_{ij} b_j(o_{t+1})$$

$$\gamma_i = \frac{\alpha_i(t)\beta_i(t)}{\sum \alpha_j(t)\beta_j(t)}$$

$$\xi_{ij}(t) = \frac{\alpha_i(t) a_{ij} \beta_j(t+1) b_j(o_{t+1})}{\sum_k \alpha_k(t)\beta_k(t)}$$