Hidden Markov Models & their Applications to Statistical Genetics

Freddy Barragan, Spring 2021, Macalester College

Contents

Welcome

This is bookdown on discrete Hidden Markov Models & their applications to statistical genetics is my capstone! I made this for my Mathematical Statistics course taught by Kelsey Grinde— big thanks to her for her continued guidance & for looking this over— with the goal of learning some advanced statistical method.

Content was written and gathered by Freddy Barragan with appropriate citations for referenced materials.

Note: I assume no mathematical background, but you shouldn't need much! If any equations look scary, feel free to skip over to the written content :)

Chapter 1

Mathematical Intuition

Markov Chains are cool! And Hidden Markov Models are even cooler! So, let's make sure you can get at the cool-stuff by starting off with the necessary basics. In this section, we'll review conditional probabilities & set up the basis for Hidden Markov Models by getting comfortable with chains first.

1.1 Conditional Probability & Bayes Rule

Probability, as a field, formalizes how we predict events with some equally beautiful & ugly notation, sharp concepts, and complex mathematical principles. But the premise is simple: by ascribing a numeric value to the outcomes of an event happening, we can abstract the real-world and study it with math. This process of ascribing numeric values to the outcome of an event is called a mapping & by mapping all possible outcomes of an event we create a *random variable*.

> **NOTE:** This can be confusing! The "random" part of the word doesn't mean all outcomes have an equal chance of happening, really it means that within an event there are multiple possible outcomes.

For example, let's say that the weather can either be sunny, rainy, or cloudy. We can then say that the weather is an event, W, with the outcomes w_s, w_r, w_c, respectively. However, by specifying the probabilities of each event we can turn W into a random variable:

$$p_{(W)}(w) = \begin{cases} 0.5, & \text{for } w_s \\ 0.3, & \text{for } w_r \\ 0.2, & \text{for } w_c \\ 0.0, & \text{otherwise} \end{cases}$$

But what happens when the weather changes because of temperature, implying that weather is conditional on the outcome of temperature. Let's formalize

- Definition of Conditional Probability
- Definition of Bayes rule

1.2 Markov Chains & The Markov Property

Markov Property

Markov Chains

- State Space
- Transition Probabilities & Matrices

Visualizing a Markov Chain - Will Hipson's Dot Visualization

- embedding Victor Powell's Website.

1.3 Video Resources

References

Hipson, Will. 2020. "Visualizing A Markov Chain." Will Hipson's Personal Website. March 23, 2020. https://willhipson.netlify.app/post/markov-sim/markov_chain/.

Johnson, Alicia, Miles Ott, and Mine Dogucu. 2020. "Bayes' Rule." In Bayes Rules! An Introduction to Bayesian Modeling with R . https://www.bayesrulesbook.com.

Kang, Eugine. 2017. "Hidden Markov Model." Medium. August 31, 2017. https://medium.com/@kangeugine/hidden-markov-model-7681c22f5b9.

Lay, David C. 2012. "Applications to Markov Chains." In Linear Algebra and Its Applications, 4th ed., 253–62. Boston: Pearson College Division.

Powell, Victor, and Lewis Lehe. 2014. "Markov Chains Explained Visually." Explained Visually. November 7, 2014. https://setosa.io/ev/markov-chains/.

Chapter 2

Hidden Markov Models

Motivations for HIdd

2.1 Hidden Markov Models

- Definition of **Hidden Markov Model** Definition of Hidden States
  - Definition of Observation Likelihoods
  - Visual Intuition with family data and `seqHMM`

2.2 The Forward-Backward Algorithm

- The Forward-Backward Algorithm
  - Motivation: Given hidden states, find the likelihood of the observations
  - Define connection between Bayes rule
  - Define Joint Probability
  - Walkthrough of Algorithm

2.3 Limitations

References

Degirmenci, Alperen. 2014. "Introduction to Hidden Markov Models." Harvard University. https://scholar.harvard.edu/files/adegirmenci/files/hmm_adegirmenci_2014.pdf.

Eddy, Sean. 2004. "What Is a Hidden Markov Model?," October, 1315–1316. https://www.nature.com/articles/nbt1004-1315#citeas.

Helske, Satu, and Jouni Helske. 2019. “Mixture Hidden Markov Models for Sequence Data: The SeqHMM Package in R” 88: 1–32. http://dx.doi.org/10.18637/jss.v088.i03.

Jurafsky, Dan, and James Martin. 2020. “Hidden Markov Models.” In Speech and Language Processing, 3rd ed. https://web.stanford.edu/~jurafsky/slp3/A.pdf.

Kang, Eugine. 2017. “Hidden Markov Model.” Medium. August 31, 2017. https://medium.com/@kangeugine/hidden-markov-model-7681c22f5b9.

Chapter 3

Application to Statistical Genetics

3.1 Local Ancestry Inference

3.1.1 Motivation

Definition of Admixture Definition of Local Ancestry

3.1.2 Specific Application

- Definition of Smoother Analysis
- HAPMIX
- Toy Transition Matrix
- Visualizing the Smoothing Process
  - 23andMe
  - Re-using the Hipson Visualization (?)
  - Re-using

References

Eddy, Sean. 2004. "What Is a Hidden Markov Model?," October, 1315–1316. https://www.nature.com/articles/nbt1004-1315#citeas.

Helske, Satu, and Jouni Helske. 2019. "Mixture Hidden Markov Models for Sequence Data: The SeqHMM Package in R" 88: 1–32. http://dx.doi.org/10.18637/jss.v088.i03.

Price, Alkes L., Arti Tandon, Nick Patterson, Kathleen C. Barnes, Nicholas Rafaels, Ingo Ruczinski, Terri H. Beaty, Rasika Mathias, David Reich, and

Simon Myers. 2009. “Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations.” PLoS Genet 5 (6): e1000519. https://doi.org/10.1371/journal.pgen.1000519.