Lesson 7. Modeling with Markov Chains, Revisited

1 Overview

- When is a Markov chain an appropriate model?
- More advanced Markov chain models

2 When is a Markov chain an appropriate model?

- Recall that a Markov chain must satisfy the Markov and time stationarity properties:
  - **Markov property**: only the last state influences the next state

$$\Pr\{S_{n+1} = j \mid S_n = i, S_{n-1} = a, \dots, S_0 = z\} = \Pr\{S_{n+1} = j \mid S_n = i\}$$

  - **Time stationarity property**: transition probabilities don't depend on when the transition happens

$$\Pr\{S_{n+1} = j \mid S_n = i\} \text{ is the same for all } n = 0, 1, 2, \dots$$

Example 1. For each of the following cases, discuss what assumptions need to be made in order for the Markov property and time stationarity property hold, and whether these assumptions are plausible.

a. A model of a taxi's movement defines the state of the system to be the region of the city that is the destination of the current ride, and the time index to be the number of riders the taxi has transported. When the taxi delivers a rider, it stays in the destination region until it picks up another rider.

b. A model of computer keyboard usage defines the state of the system to be the key that a person is currently typing, and the time index to be the number of keys typed.

c. A model of the weather in Annapolis defines the state of the system to be the high temperature, in whole degrees, and the time index to be the number of days.

3 Examples of more advanced Markov chain models

Example 2. Arrow Auto Insurance wants to model the likelihood that a policyholder will be in an accident in a given year. The insurance company believes that after having an accident, a policyholder is a more careful driver for the next 2 years. In particular, actuarial records show that a customer who had at least 1 accident in the past two years has a 5% chance of having an accident during the current year, while a customer who did not have an accident during the past two years has a 10% chance of having an accident during the current year. Model this setting as a Markov chain.

Example 3. Instagloat, a popular social media platform, tracks the behavior of its users based on two dimensions: the frequency of their posts and the sentiment of their posts. A user's sentiment is categorized as either positive or negative. A user's frequency is categorized as low, medium, or high. You've been hired to model and analyze the user behavior.

Based on historical data, you find that:

- If a user has positive sentiment, the probability of that user remaining positive with their next post is 0.8.
- If a user has negative sentiment, the probability of that user remaining negative with their next post is 0.6.
- If a user has low frequency, the probability of that user transitioning to medium frequency in their next post is 0.3, and the probability of transitioning to high frequency is 0.1.
- If a user has medium frequency, the probability of transitioning to low frequency in the next post is 0.3, and the probability of transitioning to high frequency is 0.2.
- If a user has high frequency, the probability of transitioning to low frequency in the next post is 0.2, and the probability of transitioning to medium frequency is 0.1.

Assume that changing sentiment types and changing frequency types are independent.

Model this setting as a Markov chain.

4 Exercises

Problem 1. Consider a model of a consumer's preferences for toothpaste brands that defines the state of the system to be the brand that the consumer currently uses, and the time index to be the number of tubes of toothpaste purchased. Discuss what assumptions need to be made in order for the Markov property and time stationarity property hold, and whether these assumptions are plausible.

Problem 2. The meteorologist at the Simplexville Observatory is trying to come up with a simple model for Simplexville's weather that can be used for some quick, preliminary analysis. According to their data, if the previous two days were sunny, then there is a 75% chance that the next day is also sunny. In a similar vein, if the previous two days were rainy, then there is a 35% chance that the next day is rainy. Otherwise, there is a 60% chance that the next day is sunny. Model Simplexville's weather as a Markov chain.

Problem 3. The Simplex Company is interested in how many software developers it needs to hire. The company has two software developer positions: "junior" and "senior". How its software developers transition between these positions depends on the number of years they have been in that position. Based on the company's historical data:

- Among junior software developers that have been in the position for 2 years or less, 10% of them get promoted to the senior position, and 10% of them leave the company.
- Among junior software developers that have been in the position for 3 years or more, 40% of them get promoted to the senior position, and 50% of them leave the company.
- Among junior software developers that have been in the position for 2 years or less, 40% of them leave the company.
- Among senior software developers that have been in the position for 3 years or more, 20% of them leave the company.

Model this setting as a Markov chain.

Problem 4 (SMAS Exercise 6.20)**.** Quality control engineers at KRN Corporation are monitoring the performance of a manufacturing system that produces an electronic component. Components are inspected in the sequence they are produced. The engineers believe that there is some dependence between successively produced components. In particular, if the previous component is good, there is a 99.5% chance that the next component is good. If the previous component is bad, then there is a 50.5% chance that the next component is bad. Furthermore, there is a chance that a component will be declared good when it is actually bad. Suppose that the probability that a bad component is misclassified as good is 0.06, independent of what happened with previous components. Model this setting as a Markov chain.