Markov Chains

Antarish Rautela

October 2019

1 Introduction

The story begins in 1856, St. Petersburg, Russia. Andrei Andreyevich Markov was born and fell in love with mathematics, and became prominent in the Academy of Sciences, established in St. Petersburg by Peter the Great (1682-1725). Markov was born at a time when the study of probability was very popular in Europe. During this time Jacob Bernoulli had proved one of the first versions of the Law of Large Numbers. He proved that the proportion of heads in repeated tosses of a fair coin converged to the expected value of the process, $\frac{1}{2}$. Coin flips are independent events, meaning that the outcome of a current coin flip does not depend on any previous coin flips. Thus Bernoulli proved that independent events have a convergence property. This inspired a both a moral and mathematical argument from the Russian mathematician Pavel Nekrasov. Markov chains are helpful in many areas including the medical and technological fields. A Markov chain essentially consists of a set of transitions, which are determined by some probability distribution, that satisfy the Markov property. Formally, a Markov chain is a probabilistic automaton[1]. The probability distribution of state transitions is typically represented as the Markov chain's transition matrix. One example is from transitioning to state A to state B with probability. If the Markov chain has N possible states, the matrix will be an N x N matrix, such that entry (X, Y) is the probability of transitioning from state I to state J. Additionally, the transition matrix must be a stochastic matrix[2], a matrix whose entries in each row must add up to exactly 1. This makes complete sense, since each row represents its own probability distribution. Markov chains are really helpful to predict randomness. A Markov chain is a stochastic process that satisfies the Markov property, which means that the past and future are independent when the present is known.

[1] probabilistic automaton is a generalization of the non-deterministic finite

[2] a square matrix used to describe the transitions of a Markov chain where each of its entries is a non-negative real number that represents a probability.

2 What is a Markov Chain?

A Markov chain essentially consists of a set of transitions, which are determined by some probability distribution, that satisfy the Markov property. Observe how in the example, the probability distribution is obtained solely by observing transitions from the current day to the next. A transition matrix is said to be regular if some power of T has all positive entries. The Markov chain represented by T is called a regular Markov chain. The chain tells you the probability from transitioning to one state to another state. For example, when you have two phases that a certain object can be in there are four possible transitions. A to itself, A to B, B to itself, and B to A. To demonstrate this one can use a model, but what most data scientists do is set up a transition matrix. For example, one might make the diagram of A and B as a path from A to B and a path from B to A, and two more paths that are A to A, and B to B.

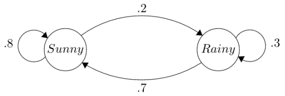
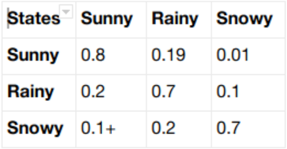


Figure 1: Model of Markov Chain

3 Markov Matrices



States	Sunny	Rainy	Snowy
Sunny	0.8	0.19	0.01
Rainy	0.2	0.7	0.1
Snowy	0.1+	0.2	0.7

Figure 2: Markov Matrix

A data scientist would make a 3 by 3 and exclude the top right corner and then the corner is the middle of the leftmost column would be A and the on below that square would be B and then the column in the middle of the top most row would A, and the one to the right of it would be B. Then the one in the middle of the 3 by 3 would be the probability of A going to itself and the one to the right of the box would be the probability of A going to B. The one in the middle of the bottom row would be the probability of B going to A, and the square to the right of that square would be the probability of B going to itself. Each one of the four squares that contain probabilities have to deal with conditional probability, which is the probability of a certain happening when something else happens. The left column is the event that happened and the ones on the top row are the probabilities of an event happening. This is represented by $P(X|Y)$ which means the probability of X happening given that Y happened. The Markov chains can have more than two phases, and with that the number of transition increase too. This causes the models to be really time consuming to make and this is why most people use Markov matrices. One example of visualizing this we will see the matrix below. The top left box that has the number 0.8 written inside of it is the probability of it being sunny tomorrow knowing that today is sunny, and the box next to it is the probability of it being rainy tomorrow knowing that today was rainy is depicted by the number 0.19. The number 0.01 is the probability of it being snowy knowing that today is sunny.

Those are a couple examples of how the Markov matrix works.A Markov chain is called homogeneous if P(V;+ = P—v; = a) is independent of i for all a and P.

4 Instances of the Usage of Markov Chains

Google is one company that uses the Markov chain. They use it in their search engine, called Pagelink, which is a Markov chain. One thing that people are using Markov chains for is the modeling of cancer metastasis. Meteorologist use Markov Chains to predict the weather. A lot of fields that require probability require Markov Chains. Google uses the Markov chains in the form given a web with n pages, construct an $n \times n$ matrix A.

5 Markov Ciphers

In this section, a class of iterated ciphers that are especially interesting for differential cryptanalysis will be considered. For such a cipher, the sequence

$$\Delta Y(O), \Delta Y(l), ..., \Delta Y(r)$$

forms a Markov chain. Recall that a sequence of discrete random variables $v_0, v_1, ..., v_r$, is a Markov chain if, for $0 \leq i < r$ (where $r = \infty$ is included),

$$P(v_{i+l} = \beta_{i+1}|v_i = \beta_i, v_{i-1} = \beta_{i-1}, ..., v_0 = \beta_0) = P(v_{i+l} = \beta i + 1|v_i = \beta_i)$$

. A Markov chain is called homogeneous if $P(v_{i+l} = \beta|v_i = \alpha)$ is independent of i for all α and all β. [In what follows, we always assume that the plain text X is independent of the sub keys $Z^1, ..., Z^r$.

6 Definition

An iterated cipher with round function Y = f(X, Z) is a Markov cipher if there is a group operation @ for defining differences such that, for all choices of $\alpha(\alpha \neq e)$ and $\beta(\beta \neq e)$

$$P(\Delta Y = \beta|\Delta X = \alpha, X = \gamma)$$

and P $(P\#e)$, is independent of γ when the sub key Z is uniformly random, or, equivalently, if

$$P(\Delta Y = P|\Delta X = a, X = 7) = P(\Delta Y) = P||AUX = a)$$

for all choices of γ when the sub key Z is uniformly random. The following crucial theorem explains the terminology Markov cipher.

7 Theorem

If an r-round iterated cipher is a Markov cipher and the r round keys are independent and uniformly random, then the sequence of differences

$$\Delta X = \Delta Y(O)$$

,

$$\Delta Y(l), ..., \Delta Y(r)$$

is a homogeneous Markov chain. Moreover, this Markov chain is stationary if ΔX is uniformly distributed over the non-neutral elements of the group.

8 References

`https://link.springer.com/content/pdf/101007%2F3-540-46416-6_2.pdf`

9 Citations

Molina, Alessandro. “Markov Chains with Python.” Medium, Medium, 26 Nov. 2018, `https://medium.com/@__amol__/markov-chains-with-python-1109663f3678`.

Home, `bookdown.org/probability/beta/markov-chains.html`.

“Nondeterministic Finite Automaton.” Wikipedia, Wikimedia Foundation, 25 Nov. 2019, `en.wikipedia.org/wiki/Nondeterministic_finite_automaton`.