Monte Carlo Methods

Cryptography and Substitution Ciphers

Anand Kamble

Department of Scientific Computing
Florida State University

1 Introduction

This project is based on an application of Markov Chain Monte Carlo (MCMC) to cryptography, introduced by Persi Diaconis (2008). Here, MCMC was used to decrypt coded messages from a state prison.

The prison's coded messages (or ciphertext) were assumed to be a simple substitution cipher. In these ciphers, the original message m (plaintext) is scrambled using a permutation f, one letter at a time, yielding the ciphertext c.

For example, if the plaintext alphabet is 'ABC' and the ciphertext alphabet is 'XYZ', then the message $m =$ 'AABBCAA' is encrypted as $c =$ 'XXYYZXX'.

2 Goal

This project aims to decode the given secret phrase using a unique 1-1 substitution mapping among the 26 letters of the alphabet.

3 Methodology

Markov Chain Monte Carlo (MCMC) methods are widely used to obtain samples from complicated, high-dimensional probability distributions.

For the problem of decryption, we seek a bijection f (also called a *key*) that maps characters in the *code space* to those in our *regular alphabet.*

$$f : \{\text{code space}\} \longrightarrow \{\text{regular alphabet}\}$$

A common approach to guess this mapping is for the cryptanalyst to look at the frequency distribution of letters in the ciphertext and compare that to the frequency distribution [4] of the natural language.

This project's method models text as a string of characters from a first-order Markov process. Using the bigram[5] character frequency from a reference text, a first-order transition matrix from character-to-character can be constructed. For example, a transition from 'q' to 'a' might rarely occur compared to going from 'q' to 'u'.[3]

3.1 Initial Mapping

We start by guessing an initial mapping f'. Which is generated by using the numpy function `np.random.permutation` to shuffle the alphabets and create a mapping.

3.2 Compute the Plausibility

Using the mapping generated in the last step, we are updating the text and calculating the plausibility.

We are using the following function to calculate the plausibility of the given text based on the provided transition matrix.

$$Pl(f) = \prod_{i=1}^{N-1} M(f(s_i), f(s_{i+1}))$$

where s_i is the i^{th} letter of the encrypted message.

3.3 Proposing a new mapping

Using the previous mapping, we are creating a new map by randomly interchanging two alphabets. This mapping can be denoted by f'

3.4 Compute the Plausibility of new mapping

The plausibility of the new mapping is calculated the same way as described above. Which will be denoted as $Pl(f')$

3.5 Accept or Reject the new mapping

The decision to accept the new mapping is based on the likelihood ratio, given by

$$\frac{Pl(f')}{Pl(f)} > 1$$

If this ratio is less than 1, a random number $U[0, 1]$ is generated. If this randomly generated number is less than the ratio, the new mapping is accepted. This step introduces a probabilistic element, allowing the Markov Chain Monte Carlo (MCMC) algorithm to explore alternative mappings. The randomness in accepting or rejecting proposals aids in navigating the solution space and converging to a deciphered message.

3.6 Simulation

We are then repeating these steps until the text doesn't change or it starts making sense.

3.7 Sample Run

Below is an example of how the text evolved during the simulation.

```
Iteration    1: W AMVY XYEFYIJWVY PG JAY BJMGG JAWJ C JYIN JP HY ...
Iteration  500: U HOGE YERCENTUGE IF THE LTOFF THUT A TEND TI BE ...
Iteration 1000: A HUGE BERPENTAGE OF THE STUFF THAT I TEND TO CE ...
Iteration 1500: A HUGE BERCENTAGE OF THE STUFF THAT I TEND TO KE ...
Iteration 2000: A HUGE BERPENTAGE OF THE STUFF THAT I TEND TO CE ...
Iteration 2500: A HUGE PERCENTAGE OF THE STUFF THAT I TEND TO BE ...
```

Iteration 3000: A HUGE PERCENTAGE OF THE STUFF THAT I TEND TO BE ...

4 Results

We ran the MCMC simulation for 5,000 iterations and successfully decoded the secret phrase. The final mapping and the decoded message are presented below.

Final Mapping:

$$\begin{aligned}
A &: O \\
B &: Y \\
C &: Q \\
D &: V \\
E &: T \\
F &: A \\
G &: K \\
H &: D \\
I &: P \\
J &: M \\
K &: S \\
L &: K \\
M &: E \\
N &: N \\
O &: B \\
P &: I \\
Q &: H \\
R &: F \\
S &: T \\
T &: L \\
U &: R \\
V &: D \\
W &: G \\
X &: J \\
Y &: U \\
Z &: O \\
\text{Space} &: \text{Space}
\end{aligned}$$

Decoded Message:

A HUGE PERCENTAGE OF THE STUFF THAT I TEND TO BE AUTOMATICALLY CERTAIN OF IS IT TURNS OUT TOTALLY WRONG AND DELUDED HERES ONE EXAMPLE OF THE UTTER WRONGNESS OF SOMETHING I TEND TO BE AUTOMATICALLY SURE OF EVERYTHING IN MY OWN IMMEDIATE EXPERIENCE SUPPORTS MY DEEP BELIEF THAT I AM THE ABSOLUTE CENTER OF

THE UNIVERSE THE REALEST MOST VIVID AND IMPORTANT PERSON IN EXISTENCE WE RARELY TALK ABOUT THIS SORT OF NATURAL BASIC SELF CENTEREDNESS BECAUSE ITS SO SOCIALLY REPULSIVE BUT ITS PRETTY MUCH THE SAME FOR ALL OF US DEEP DOWN IT IS OUR DEFAULTSETTING HARDWIRED INTO OUR BOARDS AT BIRTH THINK ABOUT IT THERE IS NO EXPERIENCE YOUVE HAD THAT YOU WERE NOT AT THE ABSOLUTE CENTER OF THE WORLD AS YOU EXPERIENCE IT IS RIGHT THERE IN FRONT OF YOU OR BEHIND YOU TO THE LEFT OR RIGHT OF YOU ON YOUR TV OR YOUR MONITOR OR WHATEVER OTHER PEOPLES THOUGHTS AND FEELINGS HAVE TO BE COMMUNICATED TO YOU SOMEHOW BUT YOUR OWN ARE SO IMMEDIATE URGENT REALYOU GET THE IDEA BUT PLEASE DONT WORRY THAT IM GETTING READY TO PREACH TO YOU ABOUT COMPASSION OR OTHERDIRECTEDNESS OR THE SOCALLED VIRTUES THIS IS NOT A MATTER OF VIRTUEITS A MATTER OF MY CHOOSING TO DO THE WORK OF SOMEHOW ALTERING OR GETTING FREE OF MY NATURAL HARDWIRED DEFAULTSETTING WHICH IS TO BE DEEPLY AND LITERALLY SELFCENTERED AND TO SEE AND INTERPRET EVERYTHING THROUGH THIS LENS OF SELF

Below is the graph which shows the evolution of $logP(f')$

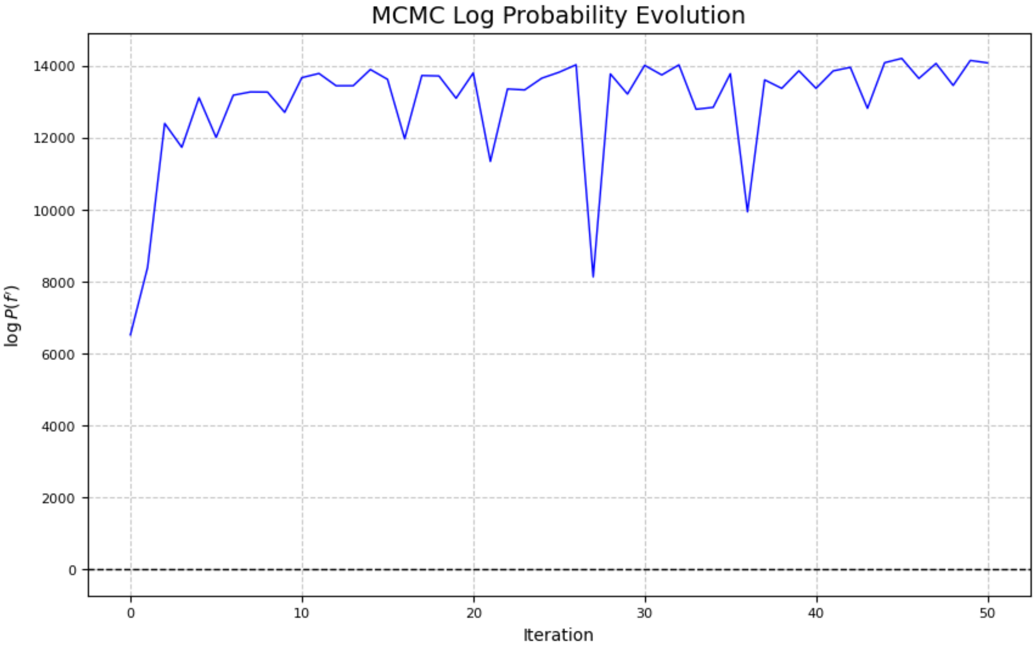

Figure 1: Log Probability Evolution

5 Conclusion

In conclusion, the Markov Chain Monte Carlo approach proved effective in decoding the secret phrase encoded with a substitution cipher. The project demonstrates the application of MCMC techniques in cryptanalysis.

References

[1] Diaconis, P. (2009). ”Markov Chain Monte Carlo revolution,” Bull. Amer. Math. Soc., 46, 179-205.

[2] andrew. (2013). ”Text Decryption Using MCMC” https://www.r-bloggers.com/2013/01/text-decryption-using-mcmc/

[3] ”Explain Substitution cipher with python” prompt. ChatGPT, OpenAI, 9 Dec. 2023. https://chat.openai.com.

[4] ”Frequency analysis”. Wikipedia. https://en.wikipedia.org/wiki/Frequency_analysis.

[5] ”Bigram”. Wikipedia. https://en.wikipedia.org/wiki/Bigram.