Pattern Identification in Random Numbers

Megha Bharti (24/AFI/20)
Computer Science and Engineering
Delhi Technological University
New Delhi, India
meghabharti_24afi20@dtu.ac.in

Pooja Verma(24/AFI/22)
Computer Science and Engineering
Delhi Technological University
New Delhi, India
poojaverma_24afi22@dtu.ac.in

Abstract— **This report presents a methodology to identify patterns in human-generated random number sequences using statistical models and pattern recognition techniques. The study utilizes Damerau-Levenshtein distance to calculate sequence similarity and applies a Markov Chain-based predictive model for next-sequence estimation. The analysis demonstrates the ability to identify recurring patterns and achieve high prediction accuracy. Furthermore, the study evaluates the uniqueness of human-generated randomness by computing identification rates and exploring overlaps between patterns. Comprehensive visualizations and metrics, such as prediction accuracy and pattern frequency, highlight the effectiveness of the proposed methodology. These results provide significant insights into the cognitive processes behind human attempts at generating randomness and have implications for fields such as cryptography, psychology, and data analysis.**

Keywords- **Pattern Recognition, Markov Chains, Damerau-Levenshtein Distance, Sequence Analysis, Random Numbers, Prediction Accuracy.**

I. INTRODUCTION

Random number sequences are often used in psychological experiments, cryptography, and simulations. Understanding patterns in such sequences is crucial for applications requiring unbiased randomness. This research aims to identify recurring patterns in human-generated random sequences and develop a predictive model for future sequence prediction.

The study utilizes Damerau-Levenshtein distance for pattern similarity analysis and Markov Chains for sequence prediction. The primary objectives are:

- To identify frequently occurring patterns in the dataset.
- To evaluate prediction accuracy based on historical data.
- To compute identification rates, highlighting sequence uniqueness.

II. RELATED WORK

Extensive research has been conducted in pattern recognition and randomness evaluation. Damerau-Levenshtein distance is a well-known metric for sequence similarity. Markov Chains have been widely used for predictive modeling. However, limited work has focused on human-generated random number sequences. This study integrates these techniques to analyze the unique characteristics of human-generated randomness.

III. METHOD

3.1 DATASET

The dataset, titled "Can humans really be random?", comprises human-generated random number sequences collected through a survey. Each row represents a unique sequence provided by a participant. The dataset contains 50 rows and 20 columns, where each row represents a sequence of 20 numbers generated by a human participant.

3.2 Pattern Analysis

Damerau-Levenshtein Distance: Measures the similarity between sequences by calculating the minimum number of operations (insertions, deletions, substitutions, and transpositions) to transform one sequence into another.

The distance between two strings **a** and **b** can be defined by using a function $f_{a,b}(i,j)$ where **i** and **j** represent the prefix length of string **a** and **b** respectively which can be defined as follows:

$$f_{a,b}(i,j) = \min \begin{cases} 0 & \text{if } i = j = 0 \\ f_{a,b}(i-1, j) + 1 & \text{if } i > 0 \\ f_{a,b}(i-1, j) + 1 & \text{if } j > 0 \\ f_{a,b}(i-1, j-1) + 1 \ (a_i \text{ and } b_j \text{ are not equal}) & \text{if } i,j > 0 \\ f_{a,b}(i-2, j-2) + 1 & \text{if } i,j > 1 \text{ and } a_{i-1} = b_j \text{ and } a_i = b_{j-1} \end{cases}$$

Damerau – Levenshtein distance

Pattern Identification: Recurring patterns of fixed lengths are extracted and ranked by frequency to highlight common cognitive tendencies in generating randomness.

$$s(m,z) = \frac{1}{l} \sum_{i=n}^{l} \frac{1}{di+1}$$

3.3 Predictive Model

A Markov Chain is trained on historical data to predict the next number in a sequence. The transition probability for a number is defined as:

3.4 Evaluation Metrics

- **Prediction Accuracy**: Proportion of correctly predicted numbers:
- **Identification Rate**: Measures the uniqueness of patterns across sequences by comparing intra- and inter-sequence prediction accuracy.

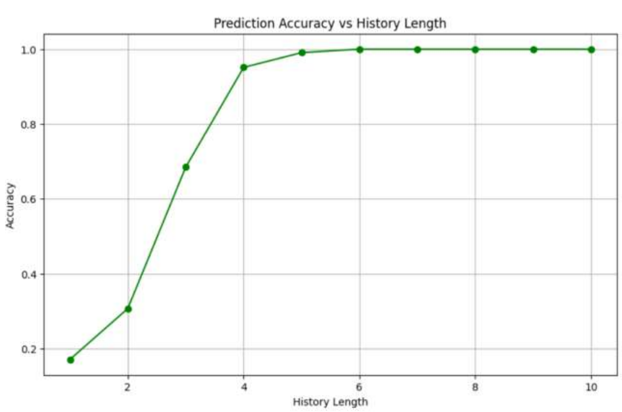
IV. RESULT

4.1 Top Patterns

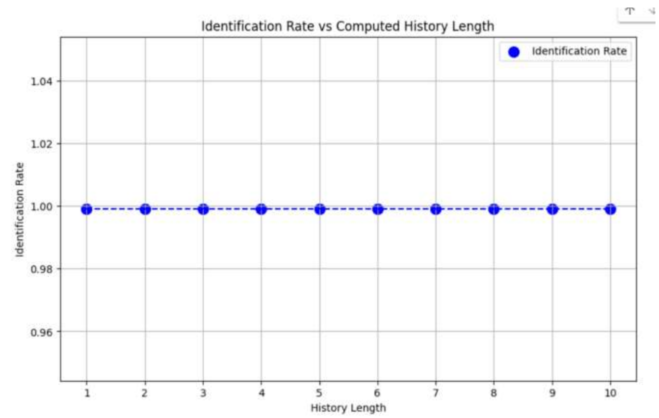
The most frequent patterns identified in the dataset are listed below:

Pattern	Frequency
(1, 10, 9, 7)	4
(9, 7, 8, 3)	3
(6, 9, 4, 3)	3
(5, 6, 3, 2)	3
(7, 5, 8, 3)	3
(1, 9, 7, 5)	3
(6, 1, 10, 9)	3

4.2 Prediction Rate



4.3 Identification Rate



4.4 Visualizations

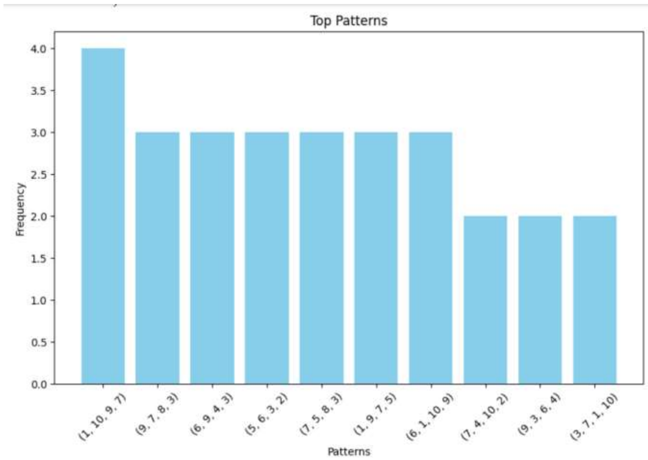


Figure 1: Top Patterns by Frequency

V. DISCUSSION

The high prediction rate indicates that human-generated random number sequences often exhibit recurring patterns, influenced by cognitive biases. However, the relatively low identification rate suggests significant overlaps in patterns across participants. This implies that humans share common tendencies in generating randomness. Future work could explore longer history lengths, alternative models such as recurrent neural networks, and strategies to differentiate patterns more effectively.

CONCLUSION

This study successfully identifies and analyzes patterns in human-generated random number sequences using Damerau-Levenshtein distance and Markov Chains. The methodology achieves high prediction accuracy and provides insights into the randomness of human-generated sequences. These findings could benefit applications in cryptography, psychology, and human-computer interaction.

ACKNOWLEDGMENT

We would like to express my deepest gratitude to Prof. Anurag Goel, for his invaluable guidance, encouragement, and continuous support throughout this course project. We gained

deep insights in the area of Pattern Identification and Recognition in randomly generated numbers by humans.

REFERENCES

1. TUNE GS (1964) RESPONSE PREFERENCES: A REVIEW OF SOME RELEVANT LITERATURE. PSYCHOLOGICAL BULLETIN 61.

2. BADDELEY A (1998) RANDOM GENERATION AND THE EXECUTIVE CONTROL OF WORKING MEMORY. QUART J EXPER PSYCH 51.

3. Daniels C, Witt K, Wolff S, Jansen O, Deuschl G (2003) Rate dependency of the human cortical network subserving executive functions during generation of random number series–a functional magnetic resonance imaging study. Neurosci Lett 345.

4. Dirnberger G, Frith CD, Jahanshahi M (2005) Executive dysfunction in Parkinson9s disease is associated with altered pallidal-frontal processing. Neuroimage 25.

5. Jahanshahi M, Dirnberger G, Fuller R, Frith CD (2000) The role of the dorsolateral prefrontal cortex in random number generation: a study with positron emission tomography. Neuroimage 12: 713–725.

6. Spatt J, Goldenberg G (1993) Components of random generation by normal subjects and patients with dysexecutive syndrome. Brain Cogn 23: 231–242.

7. Brugger P, Monsch AU, Salmon DP, Butters N (1996) Random number generation in dementia of the Alzheimer type: a test of frontal executive functions. Neuropsychologia 34: 97–103.

8. Brown RG, Soliveri P, Jahanshahi M (1998) Executive processes in Parkinson9s disease–random number generation and response suppression. Neuropsycholo gia 36: 1355–1362.

9. G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

10. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892.