ABUAD INTERNATIONAL JOURNAL OF NATURAL AND APPLIED SCIENCES
ISSN: 2955-1021
AIJNAS 2021, Volume 1, Issue 1, pp 24–29
https://doi.org/10.53982/aijnas.2021.0101.04-j

https://journals.abuad.edu.ng/index.php/aijnas

Benford's Model as a Tool for Analysing Students Performance Pattern and Anomaly

***Adebusola M. TOPE-OKE, Oluwafemi A. SANYA and Lateefah AHMED.**
Computer Science Programme, Department of Mathematical and Physical Sciences,
Afe Babalola University, Ado-Ekiti, Nigeria
*Corresponding author: topeokea@abuad.edu.ng , +2348032349222

Abstract
Studies on existing methods of analyses are still insufficient due to the lack of investigation of students' performance pattern. Prior studies found that there is a limit in the existing systems that can be used to make a comparative analysis and monitoring of the student academic performance. To look into this problem, this study explores the possibility of analysing students' performance pattern in higher institutions. A system based on Benford's law was designed to analyse students' performance pattern and the system was tested with students Cumulative Grade Point Average (CGPA). Python programming language, Malplot library, wxpython graphical user interface and Atom text and source code editor were used for developing and testing the application. After a couple of error debugging and fixes, a well-functioning application that met the requirements of the system was achieved. This study also features a spiral model which was adopted as the methodological approach in the development of the Benford's Analysis system. The interaction between users and the Benford's Analysis system was described. Furthermore, a brief description of Benford's law, pattern recognition, students' performance and stochastic modeling was given.
Keywords: *Benford's Law, Performance Pattern, Stochastic*

INTRODUCTION

Students' academic performance was often measured more by ear in the past than present day. Tutors' comments only made up the bulk of the appraisal. The aggregation method of deciding how sound a student is performing is a fairly recent invention that is used in union with Tutors evaluations of students. The track recording of academic performance fulfils a number of purposes. Measures of success and failure in a student's school major need to be evaluated in order to foster improvement and make full use of the learning curve. The track record of students gives a structure for the analyses of performance. Performance results also allow students to be ranked and sorted on a scale that is numerically obvious thereby reducing complaints because the teachers and schools are held accountable for the workings of every grade (Bell, 2018). In higher institution of learning, the Cumulative Grade Point Average is used as a measure of student's performance. For the purpose of this study we therefore define performance as the degree to which a student has accomplished their short educational goals in terms of their Cumulative Grade Point Average (CGPA).

Evidence from literature shows the application of Benford's law in fraud detection and issues during election, with little study on student performance. This study however explores the possibility of utilizing Benford's law to detect pattern or lack thereof in a given set of CGPA. This can lead to important applications in data science such as catching anomalies in the sequence of occurrence of these numbers (CGPA).

Benford's law states *that in many naturally occurring collections of numbers, the leading significant digit is likely to be small.* A set of numbers is said to satisfy Benford's law if the leading digit d {d∈1...,9} occurs with probability

$$(d)=log_{10}(d+1) - log_{10}(d) = log_{10}(\frac{d+1}{d})$$

$$=log_{10}(1+\frac{1}{d}). \qquad (1)$$

Where d= leading digit (i.e. First digit)
(Brilliant.org 2019)

A set of CGPA is said to satisfy Benford's law if the leading digit d {d ∈0..., 5} occurs with probability.

$$P(d)=\log(1+\frac{1}{d}) \qquad (2)$$

Where $0 \geq d \leq 5$

For the purpose of this study, we would adapt the formula 1 to 2.

Benford's law is usually applied to data that span several orders of scale. As an unwritten rule, the more scale of magnitude that the data evenly covers, the more accurately Benford's law apply. It is based on the assumption that fabricated figures are most likely to be distributed homogeneously; a simple contrast of first-digit frequency distribution from the data with the expected distribution according to Benford's Law should reveal anomalous results (Fewster, 2009)

With the above discussion in mind, the objectives of this study therefore are:

i. to model and design the students' CGPA pattern using Benford's first digit law
ii. to simulate the result in (i)
iii. to compare the simulation in (ii) with the original benfords model

Benford law has been applied in detecting patterns and anomaly of numbers in various fields. Although Arno and Hill (2011) postulated that there is no integrated approach to explain the form of a dynamic system, Sarkar (2018) compared the significance of Benford's law in pattern detection of natural number set to the ubiquitous nature of Normal distribution in all kinds of natural observations. Their study revealed at profundity, the idea of using Benford's law for detecting potential forgery in data in scientific studies. Shaban *et. al*. (2015) audits the efficiency of Benford's law by identifying some numbers recurring over other numbers in a given data set. The study found that the use of the Benford law analyses as part of the computer's internal controls improved the effectiveness of internal controls and reinforce them. Aris *et al*. (2013) however revealed that a combined dosage of Benford's Law and Beneish model will allow user of accounting data and assist auditor and investigator in a more accurate findings of anomalies which can be construed into fraud occurrences.

In another study carried out by Hüllemann (2017), Benford's law was used in a case of confirmed fallacious articles against a comparison group (unfalsified). The study compared the features of 12 known fraudulent articles from a single author to the features of 13 articles in the same research field from other authors, published during the same time period and identified with a Medline database search. They found that all the known fraudulent papers debased Benford's law and 6 of the 13 articles used for comparison followed the law. Although the previous studied have applied Benford's law in other settings apart from performance, its application in predicting pattern of student CGPA cannot be overlooked. This study therefore simulates and generates a performance pattern using Benford's law, while students' Cumulative Grade Point Average (CGPA) is used as input data.

MATERIALS AND METHODS

The Spiral developmental approach was adopted for this study. This approach enabled user involvement in the Benford's Analysis System even before its final build and release. With the use of the spiral model approach significant changes expected in the system during the development cycle were detected and debugged earlier thereby reducing the run time errors. This model joins the design of iterative development with the organized, controlled aspects of the waterfall model; therefore the activities in this model can be organized like a spiral. The system development repeatedly passes through four phases in iterations which are considered the spiral. They are discussed as follows;

Requirement analysis, identifying objectives and alternatives

First, the requirement analysis (in this case performance pattern prediction), identification of objectives (as stated in section 1(i,ii,iii)) and alternatives was done. This phase starts with gathering the basic functional requirements for the system in the baseline spiral in terms of how the data is analysed (CGPA). It also includes the identification of implementation alternatives as well as the constraints imposed on these alternatives. The system was given to the users to test and feedbacks were gotten from these tests. From these feedbacks, missing requirements and functionalities were identified and added. These include the plot functionality which displays and plots graph of the analysed data.

Identifying tools for designing system and designing prototype

Next, the conceptual design of the system in the baseline spiral which involved architectural design and logical design of the system's modules was arrived at. The initial prototype of the system was designed and developed in this stage; also the modelling and simulation techniques to identify the risks and to reduce them in the actual development process were carried out in this stage.

Construct and build of the system

Thirdly, the actual system was developed. It involved the coding, testing and also the integration of modules (Python programming language, Malplot library, wxpython graphical user interface and Atom text and source code editor were used to achieve this). The build was made as result of the prototyping efforts that resolved the development process and then a waterfall approach was employed which included the concept of operations, design, development, integration, and testing of the system.

Evaluation and Modification of the System

Finally, an evaluation and modification of the System after testing the build. At the end of the first iteration, the system is evaluated. Advanced technical planning and multidisciplinary reviews at critical staging or control points were carried out in this phase.

Figure 1 describes the interaction activities between the user and the system. It shows how the user saves data in text format, the user then clicks on view data, user also clicks on either of the buttons which are; analyse data using benford's law, analyse data using cho and gaines formula, analyse data using leemis formula and plot graph, the user would be shown the result of all analysis and can save the graph.

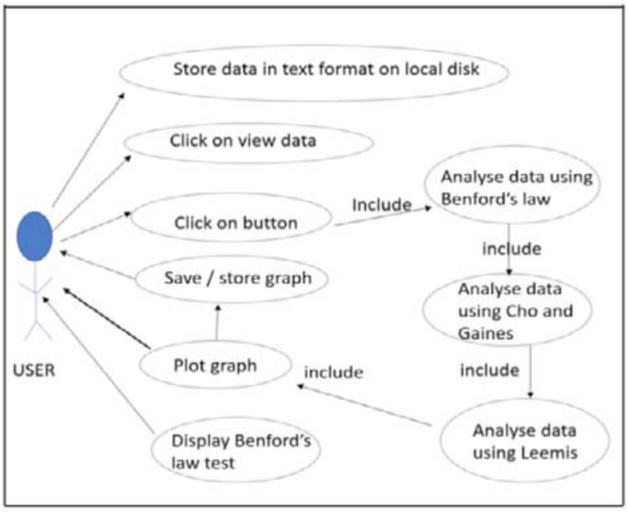

Figure 1: Use Case Diagram

Figure 2 explains the way data flows and how each state interacts with each other. The user stores the data, the system then reads the data, the system allows the user to view the data, the system then fetches the saved data, the data would be analysed using Benford's law, the result would be displayed and a graph would be plotted and saved. For the purposes of this study, 1000 cgpa data was randomly generated with a python library called NumPy

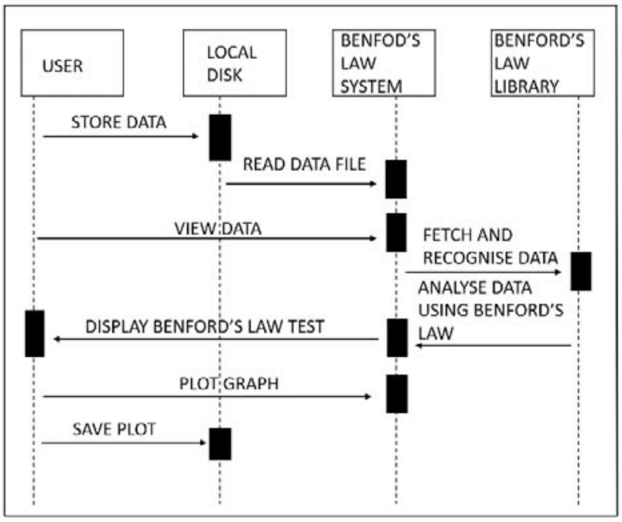

Figure 2: State Sequence Diagram

Stochastic modeling

In this study our desired result is a Benford probability distribution that actually conforms to the rule of Benford's law. What we have done was to initially calculate the Benford's probability distribution for the data we generated calculating the Benford's probability distribution for each semester CGPA compilation, we discovered that not even a single semester had a probability distribution that conforms with Benford's law and this had a range of reasons starting with the fact that there were lots of constraints in the data we used for this process, one of Benford's law major concerns is that data has to be randomly ranged with no constraint, but in our case a CGPA cannot be less than a 0.0 and cannot be more than a 5.0. then we decided to run the simulation by creating 1000 randomly generated CGPA data each for a 1000 simulation, after each randomly generated CGPA data has been generated we calculate a Benford's probability for the data in the 1000 set and see if it conforms to Benford's law which is our desired output, we do this for 1000 times (simulation). At the end of this simulation Random walk simulation was used to represent the stochastic simulation on the graph, where each time a Benford's probability distribution from any of the simulation conforms to Benford's law, there is an upward trend and vis-versa .

Then this 1000 simulation was carried out another 30 times and this was also represented on the graph so as to compare several trends.

RESULTS

Figures 3, 4 and 5 show the analysis of Benford's distribution using stochastic model in the students first, second and third semester respectively.

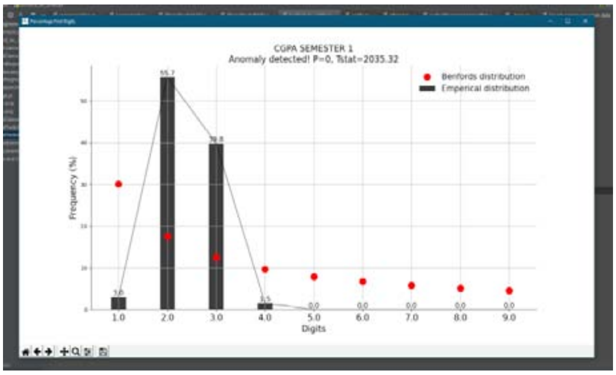

Figure 3: Frst semester stochastic model

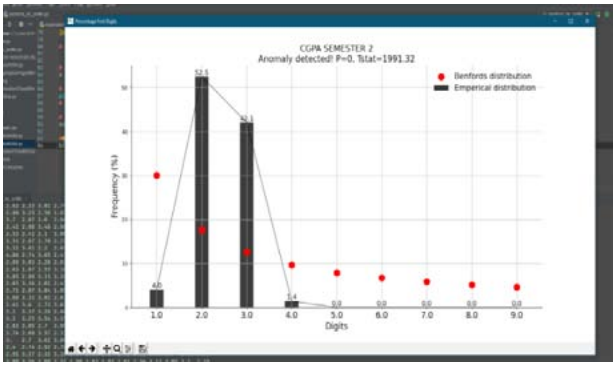

Figure 4: Second semester stochastic model

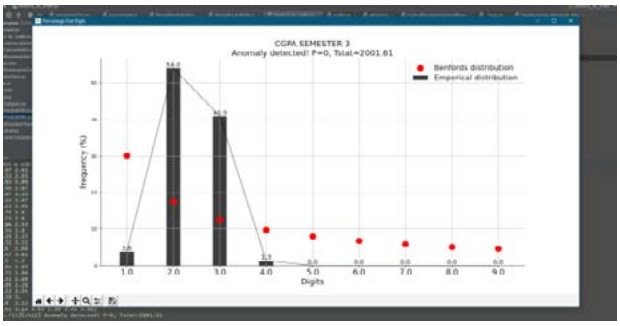

Figure 5: Third Semester stochastic model

IMPLEMENTATION

Welcome Page

This is the first window that pops up to the user when the system is launched. This window contains two buttons; the skip to start button which skips the process of using the system and moves straight to the data sampling page, and the next button which takes the user to the next window containing the necessary information about the system. This window contains very simple welcome page content, a taskbar, a background image, an icon and the system logo. The welcome page is displayed in Figure 6.

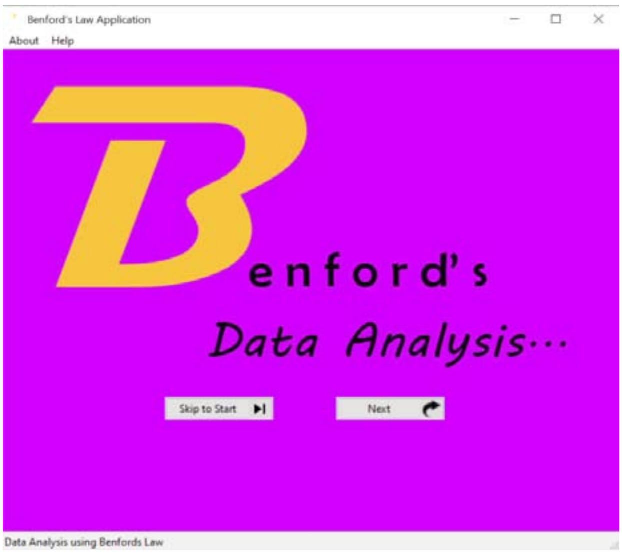

Figure 6: Welcome page and its contents.

Data Sampling

In this window the data is analysed in five ways. For the **random int** button random integers are analysed using Benford's law. For the **Benford's** button the data selected by the user is analysed using Benford's law. For the **Leemis** button the data is being analysed using Leemis formulation. For the **Cho and Gaines** button the data is being analysed using Cho and Gaines Formulation. For the **plot** button a graph of the data being analysed would be plotted. This window also has a status bar and a task bar with the about and help menu and the **close** button which takes the user to the end application page. Figure 7 shows the windows with the different data analysis function.

Figure 7: Data Sampling

DISCUSSION

In this study, a set of randomly generated variables was used. The probabilities of all the variables were forecasted under different sets of conditions. These processes were repeated until a probability value of interest was arrived at (i.e a value that is close to Benford probability distribution). The Benford's probability distribution for the generated data was initially calculated. This study found that no single semester had a probability distribution that conforms to Bedford's law. This had a range of reasons starting with the fact that there were lots of constraints in the data we used for this process: one of Benford's law major concerns is that data has to be randomly ranged with no constraint, but in our case a CGPA cannot be less than a 0.0 and cannot be more than a 5.0.

After this finding, a simulation of 1000 randomly generated CGPA data was ran, for each randomly generated CGPA, a Benford's probability was calculated for the 1000 dataset. We also check if it conforms to Benford's law. This simulation was done 1000 times. At the end of this simulation, Random walk simulation was used to show the stochastic simulation on the graph. For every time the probability distribution from any of the simulation conforms to Benford's law, there is an upward trend and vis-versa

Then this 1000 simulation was carried out another 30 times and this was also represented on the graph so as to compare several trends. The trends from most of this graphs curve tend towards the negative Y axis (Figure 8). The implication here is that even after 1000 simulations, the chances that the CGPA dataset will conforms to Benford's law was low (since the curve tend towards negative Y axis).

Finally, the stochastic simulation reveals that the use of Benford's law to analyse or find patterns in the performance of students using the respective CGPA over a number of semesters cannot be validated. The fact that the Benford's probability distribution of the CGPA dataset do not conform to Benford's law is not an indication of an anomaly in the collation of student performances in terms of the CGPA.

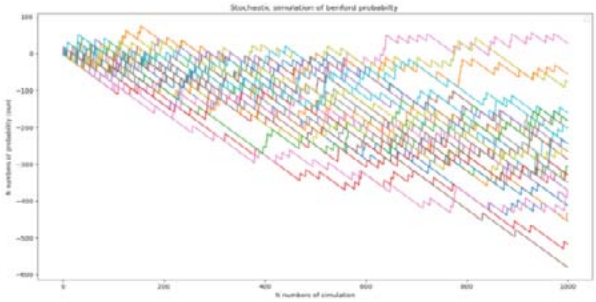

Figure 8: Random walk Simulation Result

Figure 8 shows random walk simulation to illustrate the result of stochastic simulation of the Benford's probability distribution of students CGPA over a number of semesters.

CONCLUSION

The systems using Benford's law have been researched and a couple of models are available. This research however was centered on a robust system that is modeled around Benford's law to analyse and predict the students' performance pattern. Furthermore the system is supported by Windows, MacOs and Linux operating systems. This study therefore found that analyzing student CGPA at its face value alone might not be sufficient, we are therefore looking into combining fuzzy logic with Benfords law in a follow up study to achieve a more efficient system.

List of Abbreviation:
CGPA: Cumulative Grade Point Average

Declarations
Ethical Approval: Not Applicable

Consent for Publication:
Competing Interests: There is no competing interest among the authors

Funding: There was no grant funding for this research

Authors' Contribution: Tope-Oke A.M. conceived and presented the idea.

Tope-Oke A.M., Sanya O.A. and Ahmed L. adapted the theory and performed the computations.

Tope-Oke A.M verified the analytical methods and encouraged Hammed L. to investigate Benford model.

All authors discussed the results and contributed to the final manuscript.

Acknowledgments: We thank Mr. Olajide Olanrewaju for his assistance with Python programming debugging, and all Computer Science Lecturers in Afe Babalola University, Ado Ekiti for comments that improved the manuscript.

RFERENCES

Aris, N.A.,Othman R., Arif, S.M.M., Malek, M.A.M. and Omar, N. (2013), Fraud Detection: Benford's Law Vs Beneish Model in *2013 IEEE Symposium on Humanities, Science and Engineering Research* (SHUSER), 726-731.

Arno, B. and Theodore, P. H. (2011). Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem. *Math Intelligence,* 33**:**85–91 . https://doi.org/10.1007/s00283-010-9182-3.

Bell, M.J. (2018) Define Academic Performance, *The Classroom*. Retrieved from: https://www.theclassroom.com/define-academic-performance-4740750.html.

Fewster, R. M. (2009). A simple explanation of Benford's Law. *The American Statistician,* 63 (1): 26–32. CiteSeerX 10.1.1.572.6719. *https://*doi.org/10.1198/tast.2009.0005.

Gašević, D., Zouaq, A. and Janzen, R. (2013). Choose your classmates, your GPA is at stake: The association of cross-class social ties and academic performance. *American Behavioral Scientist,* 57(10): 1460–1479. https://doi.org/10.1177/0002764213479362

Hüllemann, S., Schüpfer, G. and Mauch, J. (2017). Application of Benford's law: a valuable tool for detecting scientific papers with fabricated data? *Anaesthesist,* 66:795–802. https://doi.org/10.1007/s00101-017-0333-1

Sakar, T. (2018). What is Benford's Law and why is it important for Data Science? *Towards Data Science*.
Retrieved from: https://towardsdatascience.com/what-is-benfords-law-and-why-is-it-important-for-data-science-312cb8b61048

Shaban, M., Behrad M., Nezhad, Ali, M. and Fateme Z. (2015). Benford's Law and its Application in Auditing. International *Journal of Industrial Distribution & Business*, (6-2): 13-16. doi: http://dx.doi.org/10.13106/ijidb.2015.vol6.no2.13.