IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF BENFORD'S LAW FOR IMAGE PROCESSING

[1]Mihika Bodke, [2]Devyani Mishra, [3]Kavan Vasani, [4]Juhi Janjua
[1]Student, [2]Student, [3]Student, [4]Assistant Professor
[1]Computer Engineering,
[1]Thadomal Shahani Engineering College, Mumbai, India

Abstract:

Benford's law is known as a 'first digit law', 'digit analysis' or 'Benford-Newcomb phenomenon'.This paper presents a generalization of Benford's law for analysing the first significant digit in images. It studies this law's application in the aspects of digital image modification or fraud detection. Also, it established the results for edited as well as non- edited images. The amount of deviation from the actual Benford's curve is used as an approximate indicator of forgery.

This paper uses Benford's law to analyze two real world examples of Benford's law's applications. They include firstly an image modification detection such as adding a filter to an image and secondly fraud banknote authentication. The results are plotted onto a graph by calculating the frequency of the first significant digit to check the deviation from the original Benford's curve, this shows any fraudulent data which may be present in the database.

***Index Terms* -** Benford's law, deviation, forgery, image modification, MSD (Most Significant Digits), digital image forensics, frequency

I. INTRODUCTION

In these modern times where technology has advanced by leaps and bounds, tampering with digital images is a relatively simple task. The myriad of image editing softwares available these days are freely available to be used to create altered images. Celebrities can use filters on images to enhance their beauty and give false advertising to their audience. Using advanced techniques, images can be edited to either add people in the image, completely remove them from it or make changes in it that can be completely undetectable to the human eye, making it harder for humans to detect a phony image. Many images are also used as evidence in courts. Various times, crucial decisions have to be made based on these images. This makes it very important for these images to be completely authentic and not manipulated. This need has spurred a need of exploring various image forensics techniques. When there are a huge number of images, such as in a dataset, Benford's Law can be useful to check the authenticity of these images.

Benford's law is commonly used with financial data all over the globe. The law is commonly applied to detect fraud. Natural, or unaltered data, should follow the predicted frequencies of Benford's law. A significant deviation from this law indicates that the data has been tampered with.

II. WHAT IS BENFORD'S LAW?

The discovery of Benford's Law dates back to 1881, when an astronomer named Simon Newcomb observed that the earlier pages of a logarithm table that started with one-which was used for various calculations in those days- were more worn out than the later pages. It was then that he thought of researching on the same topic. He concluded that the probability of a digit N, being the first digit of any number was (log(N+1) - log(N)). This means that the ten decimal digits are not equally common in a natural random dataset [1].

Then in 1938, a physicist Frank Benford tested this on data from different domains. He used numbers from random domains, like the surface areas of rivers, sizes of populations of cities, entries from a mathematical handbook, molecular weights, physical constants to further verify Newcomb's discovery. He also used street addresses of people listed in American Men of Science and also death rates. Due to this work, this law was named Benford's Law [2]. The following table 1 gives an account of the real-world values that Frank Benford took into consideration while creating Benford's Law.

Table 1 Benford's value observations

Title	1	2	3	4	5	6	7	8	9	Count
Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
$n^{-1}, \sqrt{n}$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
Digest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
$n, n^2, \ldots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Benford's Law	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6	

This law was further studied and developed by Hill to analyze the probability distributions of the first significant digits (1-9) obtained from natural data. According to the laws of probability, the chances of any of the digits being the first significant digit should be 1 in 9 - or 11.1%. But it was discovered that the occurrence of smaller digits is more than the larger digits in any natural dataset. The occurrence of 1 as the Most Significant Digit (MSD) was found to be around 30.1% and it tapers as we move to the larger digits.

The following table 2 shows the distribution of Benford's Law.

Table 2 Frequency Probability of most significant digits

d	P(d)
1	30.10%
2	17.60%
3	12.50%
4	9.70%
5	7.90%
6	6.70%
7	5.80%
8	5.10%
9	4.60%

The distribution for Benford's Law can be expressed mathematically as:

$$P\left\{d_0 = \frac{d}{B} = 10\right\} = (log_{10}(d+1) - log_{10}d) \ / \ (log_{10}10 - log_{10}1)$$

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

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

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

$$= \frac{\ln\left(1+\frac{1}{d}\right)}{ln10}$$

where d is the most significant digit of the number and P is the probability distribution of d [3]. According to this formula we can make a diagram such as Figure 4 for the first significant digit distribution.

The below figure is a bar chart representation of the probability figures in Table 2. If a line is drawn on the top of the bar of each significant digit, a curve will begin to form called as Benford's curve.

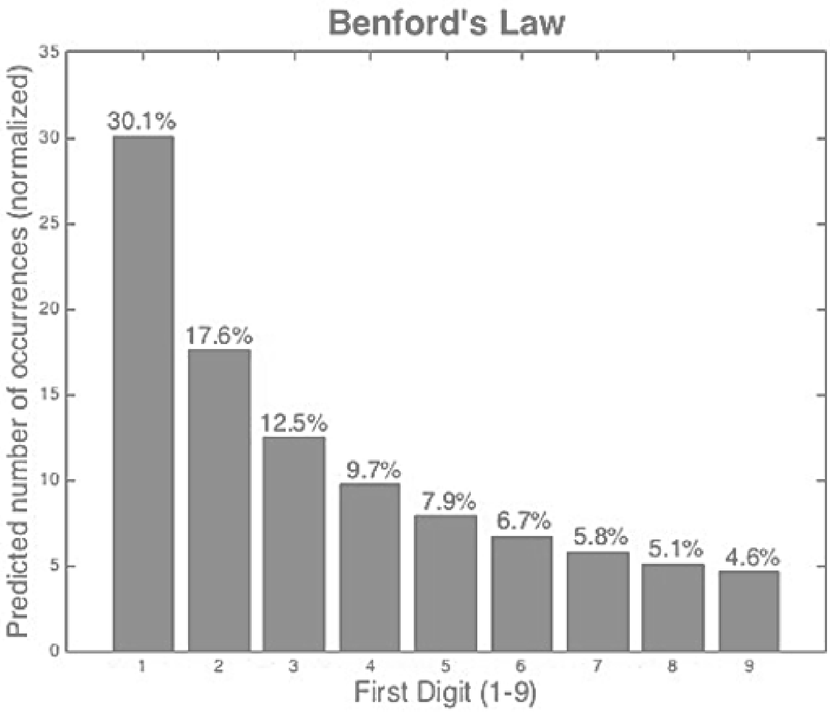

Figure 1 Graphical Representation of Benford's Law

III. IMPLEMENTATION DETAILS AND RESULTS

1. DETECTING MODIFIED IMAGE USING BENFORD'S LAW

In the following section, Benford's Law is used for detecting modified images and compares the result with the actual Benford's Curve. An untouched image shown in Figure 2 is taken as an input and its pixel data is taken from which significant values are extracted using a function. The function is applied using the math and cv2 modules of Python and the output is shown in a form of graph which is displayed using matplotlib in Data Visualization [4]. The frequency of the significant values is calculated and it is compared with the frequency of the Benford's law values. This comparison can be seen in the Figure 3 as shown below.

Figure 2 Original Image

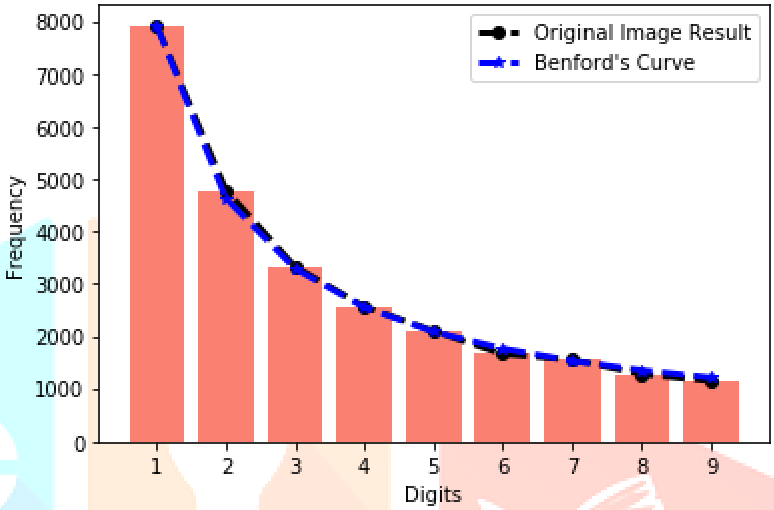

Figure 3 Comparison of Original Image with Benford's Curve

The dashed lines represent the frequency of most significant digits. Since both of the lines coincide, the original figure 2 follows Benford's Law where the frequency of the digit 1 is 30.1%.
Now the input image is tampered, as shown in Figure 4, which has been modified by adding the black and white filter.

Figure 4 Modified Image

By adding the filter, there is a change in the pixel data which results in a change in the most significant digits. On comparing with the actual Benford's curve, the result will not conform and will instead deviate from the Benford's expected curve. Thus this will show that the image in figure 4 has been tampered or edited. This result can be observed in Figure 5.

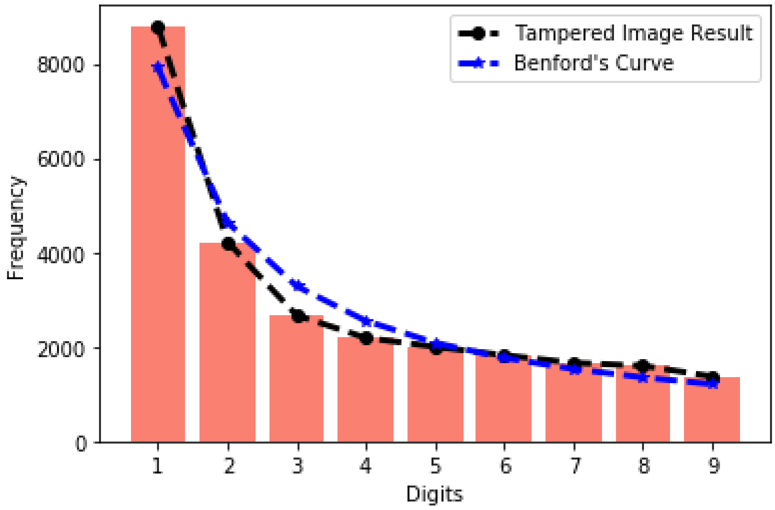

Figure 5 Comparison of Modified Image with Benford's Curve

The below Figure 6 shows the flowchart of the process to detect altered images.

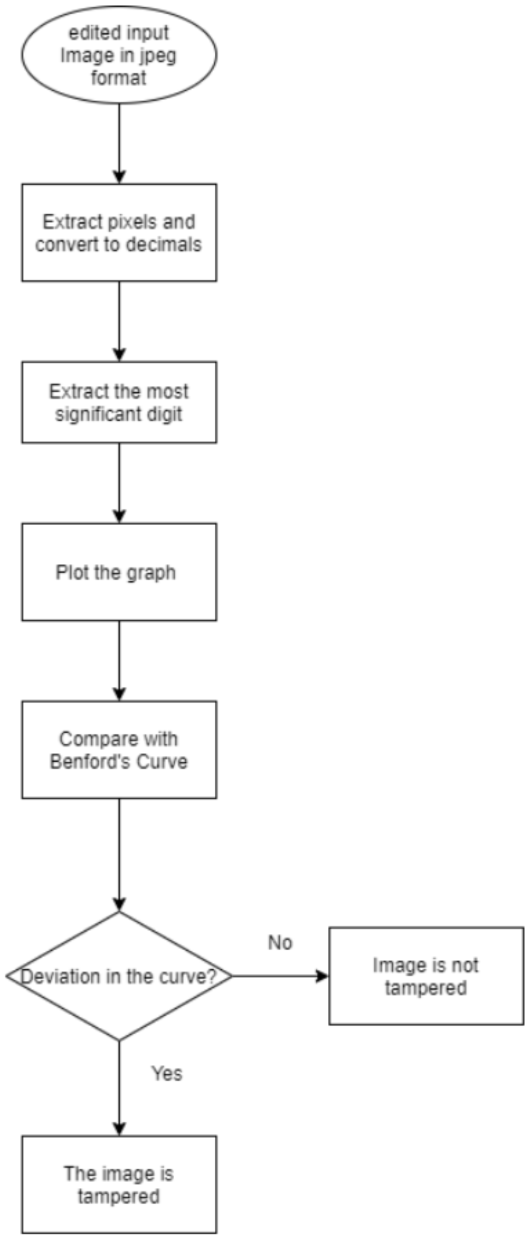

Figure 6 Flowchart for Image Processing Algorithm

2. AUTHENTICATING BANKNOTES USING BENFORD'S LAW

Building upon the previous tampered image detection, Benford's Law is used to authenticate banknotes. Forging of banknotes has been a major problem for a long time now. With the advancements in technology in recent years, there has been a huge rise in the amount of forged currency.

To validate this counterfeit currency, this paper uses a general banknote dataset [5]. In the specific dataset, data was extracted from images that were taken from genuine and forged banknote-like specimens.

From the dataset used, the image information has been classified into four categories, namely -

1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)

Based on these attributes, a banknote can be classified as real or fake based on whether it conforms with Benford's law of most significant digits.

It can be observed from Figure 7 that the attributes of an authentic banknote more or less follow Benford's Law.

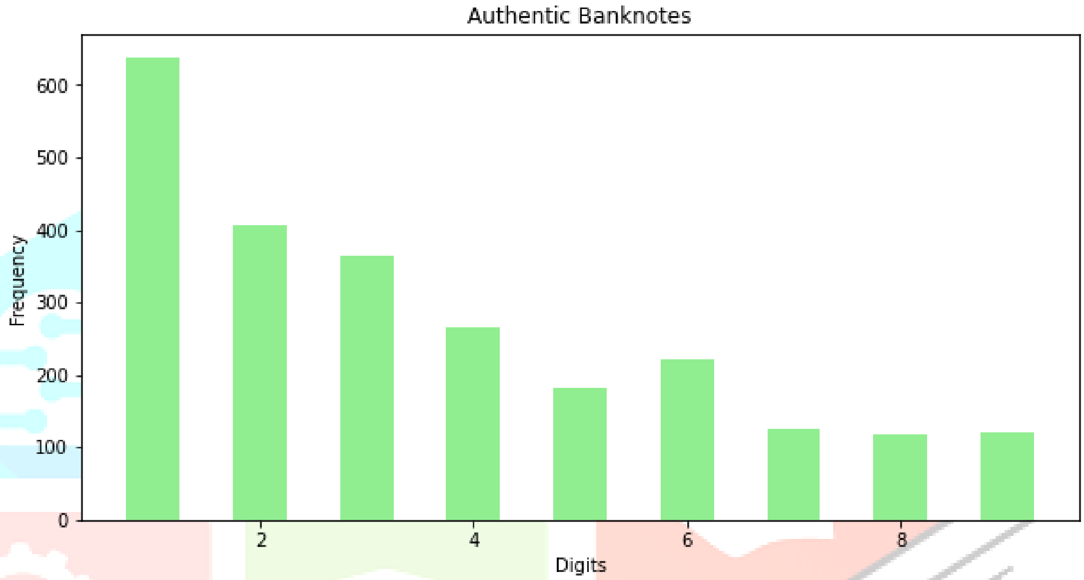

Figure 7 Graphical Representation of MSD of Authentic Banknotes

But when the attributes of a forged banknote are plotted in Figure 8, a greater deviation from Benford's Law is observed. This should raise concerns that the banknote is inauthentic.

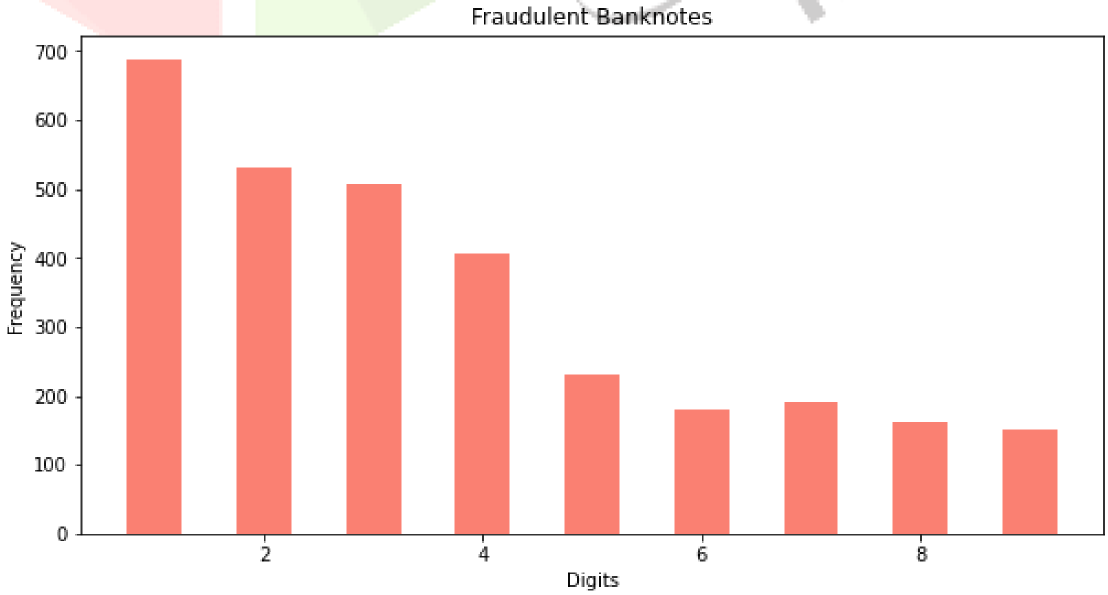

Figure 8 Graphical Representation of MSD of Fraudulent Banknotes

The below Figure 9 shows the flowchart which is used to authenticate banknotes.

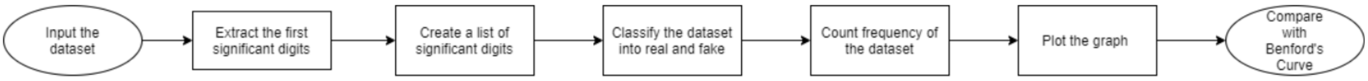

Figure 9 Flowchart for Banknote Authentication Algorithm

IV. FUTURE SCOPE

The best-known use of Benford's Law is in fraud detection. A mobile or web application can be created to verify whether an image or any natural dataset has been manipulated or edited. As this application will not use a machine learning model but instead a simple algorithm to calculate most significant digits, it will work much faster and more efficiently to provide results. This application will make identifying forged pictures much easier and portable since it will not require a high-powered computer for calculations.

V. CONCLUSION

The inexorable development towards an all-digital world comes hand-in-hand with the manipulation or even the fabrication of digital information. Images are growing vulnerable to digital forgery in today's day and age. Digital image forensics are making genuine efforts to rebuild the dwindling confidence in the easily manipulative digital content. The widely used image editing softwares such as Photoshop are primarily accountable for this lost trust in genuine images. When images are available in such a huge quantity such as large datasets, application of digital forensic methods for authentication is not feasible [6].

Due to this, there is an increase in the need for tools that ease the identifying of those manipulations. This paper has so far investigated Benford's law and its application in detecting tampered images as well as its efficiency in banknote authentication.

As the results shows, Benford's law could be an excellent option and is in fact already being used to detect various frauds in many domains.

VI. REFERENCES

[1] S. Newcomb, "Note on the frequency of use of different digits in natural numbers" Amer. J. Math. 4(1881)

[2] https://mathworld.wolfram.com/BenfordsLaw.html, 22/06/21

[3] Z. Jasak and L. Banjanovic-Mehmedovic, "Detecting Anomalies by Benford's Law," 2008 IEEE International Symposium on Signal Processing and Information Technology, 2008, pp. 453-458, doi: 10.1109/ISSPIT.2008.4775660.

[4] https://medium.com/explorations-in-python/benfords-law-in-python-64c28a560c98 , 22/06/21

[5] https://www.kaggle.com/shanks0465/banknoteauthentication

[6] Fernando Perez-Gonzalez, Greg L. Heileman and Chaouki T. Abdallah, " Benford's law in image processing," 2007 IEEE International Symposium on Signal Processing and Information Technology, 2007, doi: 10.1109/ISSPIT.2008.4775660.