Journal of Humanistic Mathematics

Volume 13 | Issue 2 July 2023

The "Benfordness" of Bach Music

Chadrack Bantange
Washington and Lee University

Darby Burgett
Washington and Lee University

Luke Haws
Washington and Lee University

Sybil Prince Nelson
Washington and Lee University

Follow this and additional works at: https://scholarship.claremont.edu/jhm

Part of the Applied Statistics Commons, Mathematics Commons, Musicology Commons, and the Probability Commons

Recommended Citation

Chadrack Bantange, Darby Burgett, Luke Haws & Sybil Prince Nelson, "The "Benfordness" of Bach Music," *Journal of Humanistic Mathematics*, Volume 13 Issue 2 (July 2023), pages 389-397. DOI: 10.5642/jhummath.SGFV8169. Available at: https://scholarship.claremont.edu/jhm/vol13/iss2/21

©2023 by the authors. This work is licensed under a Creative Commons License.
JHM is an open access bi-annual journal sponsored by the Claremont Center for the Mathematical Sciences and published by the Claremont Colleges Library | ISSN 2159-8118 | http://scholarship.claremont.edu/jhm/

The editorial staff of JHM works hard to make sure the scholarship disseminated in JHM is accurate and upholds professional ethical guidelines. However the views and opinions expressed in each published manuscript belong exclusively to the individual contributor(s). The publisher and the editors do not endorse or accept responsibility for them. See https://scholarship.claremont.edu/jhm/policies.html for more information.

The "Benfordness" of Bach

Chadrack Bantange

Washington and Lee University, Lexington, Virginia, US
bantangec25@mail.wlu.edu

Darby Burgett

Washington and Lee University, Lexington, Virginia, US
dburgett@mail.wlu.edu

Luke Haws

Washington and Lee University, Lexington, Virginia, UStry
hawsl23@mail.wlu.edu

Sybil Prince Nelson

Mathematics Department, Washington and Lee University, Lexington, Virginia, USA
sprincenelson@wlu.edu

Synopsis

In this paper we analyze the distribution of musical note frequencies in Hertz to see whether they follow the logarithmic Benford distribution. Our results show that the music of Johann Sebastian Bach and Johann Christian Bach is Benford distributed while the computer-generated music is not. We also find that computer-generated music is statistically less Benford distributed than human-composed music.

Keywords. Johann Sebastian Bach, Johann Christian Bach, Benford's Law, David Cope, Emmy Cope.

1. Introduction

Initially discovered in 1881 by American astronomer and mathematician Simon Newcomb, Benford’s law is a distribution of the first leading digits in a given numerical data set. Noticing how much faster the first pages of logarithmic tables wear out than the last ones, Newcomb pointed out that in a data set, the leading digits of numbers do not appear at equal frequency [1]. This discovery did not, however, attract much attention from the scientific world until 1938 when Frank Albert Benford, an American electrical engineer and physicist, proposed the Law of First Digits, also known as Benford’s Law. After researching over 20,000 entries from 20 different tables, Benford showed that in many naturally occurring tables of numerical data, the leading digits are not uniformly distributed but rather follow a logarithmic distribution with subsequent numbers in decreasing frequency.

Benford’s Law has been applied in fields such as economics, sociology, physics, computer science, and biology [3]. This paper investigates Benford’s Law by looking into music note frequencies. Previous work of the fourth author with collaborators has shown that as time progressed throughout the major musical time periods (Medieval, Renaissance, Baroque, Classical, Romantic), European music became more and more Benford distributed [4].

This paper continues this research by analyzing works of the German composers Johann Sebastian Bach (1685–1750) and Johann Christian Bach (1735–1782), and “Emmy” Cope, a computer algorithm designed to emulate Johann Sebastian Bach. Interestingly, while Johann Christian was the eighteenth child of Johann Sebastian, their compositions fall in different musical time periods. Johann Sebastian’s compositions are considered Baroque, while Johann Christian’s compositions fall into the Classical period.

David Cope is an American author, composer, and Professor Emeritus at the University of California Santa Cruz. Commissioned to compose music for an opera, Cope began to experiment with a musical algorithm as a form of procrastination. And so the Experiments for Musical Intelligence was born. “EMI” or “Emmy” is a patented algorithm designed by Cope to replicate the compositional styles of any selected composer. We found that while music generated through artificial intelligence to sound like Bach indeed sounds like Bach to the human ear, upon analysis, it is statistically less Benford than both Johann Christian and Johann Sebastian Bach. This suggests that there is a necessary human component needed in the composition of music.

2. Background

Benford's Law states that in many naturally occurring data sets, the probability of the first leading digit is

$$\text{Prob(d)} = \log_{10}\left(1 + \frac{1}{\text{d}}\right), \tag{1}$$

where $d = 1, 2, ..., 9$. Using Equation (1), we calculate that the probability of having 1 as a leading digit is .30103; 2 as a leading digit is .17609 and so on, logarithmically leading to the distribution presented in Figure 1 below.

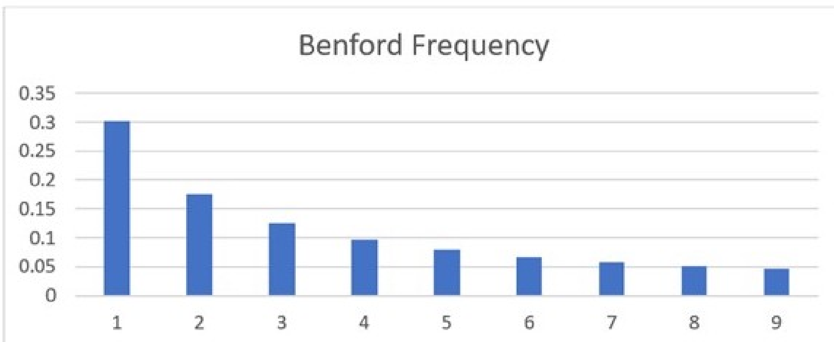

Figure 1: Visualization of the Benford Distribution.

The purpose of our research is to further our understanding of Benford's Law in music. Specifically, we endeavor to look at the distribution of music note frequencies in songs composed by the Bachs.

A music note frequency is a representation of pitch measured in Hertz (Hz). *Harvard Concise Dictionary of Music* [7] explains that sounds are produced by systems for vibrating that send their vibrations through a medium such as air to the ear. In music, the most common systems used are built with strings (such as violins and pianos) or columns of air (such as clarinets or trumpets). For example, consider a string stretched between points X and Y with midpoint A that is stretched to point B and released causing it to vibrate in the pattern A-B-A-C-A, as shown in Figure 2. The distance l which encompasses one complete vibration or wave (where the string has traversed A-B-A-C-A) is called the *wavelength*. The number of complete waves occurring per unit of time is the *frequency* of vibration and is measured in cycles per second [7].

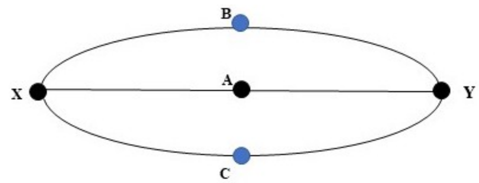

Figure 2: Visualization of string vibration.

We call a cycle per second a *Hertz* (Hz) after German physicist Heinrich R. Hertz [7]. To change notes or hear a different frequency, each new pitch can be derived by multiplying (ascending) or dividing (descending) the frequency of the previous pitch by the twelfth root of two. For example, to go from A_4 (Tuning A, with a frequency of 440) to C_5 (the C an octave above middle C, see Figure 3), we multiply 440 three times by the twelfth root of two (this is a minor third).

Figure 3: A_4 and C_5.

We already know that the musical note frequencies of most of the classical works studied are Benford distributed [4]. We wanted to answer another question: Are certain artists more Benford than others? We also wondered whether computer generated music would be more or less Benford.

To determine whether a song was more or less Benford distributed, we calculated Delta (Δ), which is the maximum deviation a distribution D of notes in a song has from the Benford distribution, using the following formula:

$$\Delta = 100 \cdot \max_{1 \leq d \leq 9} \left| \mathrm{Prob}(\mathrm{D}_1 = \mathrm{d}) - \log_{10}(1 + \frac{1}{\mathrm{d}}) \right|. \tag{2}$$

Here, D_1 stands for the distribution of the first digits in distribution D. Figure 4 shows two examples of Delta values for piano note frequencies using

$$f(k) = 440 \cdot (2)^{\frac{k-49}{12}}. \tag{3}$$

In Equation (3), the k corresponds to the key on a piano. The 49^{th} key is tuning and thus corresponds to 440Hz.

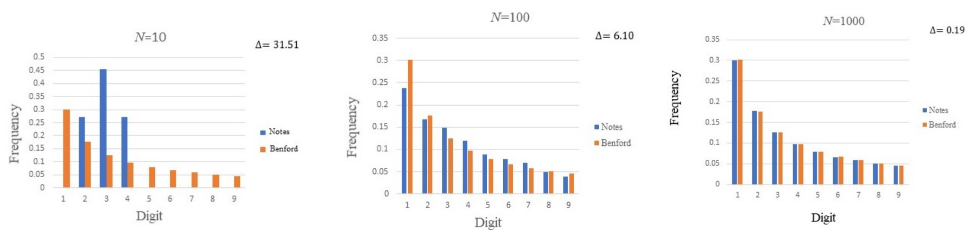

Figure 4: (a) On the left: First ten values of piano note frequencies in Hertz. (b) In the middle: First hundred values of piano note frequencies in Hertz. (c) On the right: First thousand values of piano note frequencies in Hertz.

If we let $k = 1, \cdots, 10$ in Equation (3), we are considering the first ten keys of the piano. We use this formula to convert the first ten keys to frequencies; then we apply Equation (2) to calculate a delta value. Figure 4a shows that for these first ten keys, the sequence of numbers is not very Benford, with a delta value of 31.51. Though a typical piano only contains 88 keys, we consider what would happen to this sequence if we could increase the number to 100. So, we let $k = 1, \cdots, 100$ in Equation (3), convert the key values to Hertz, and consider the distribution of these hundred notes. The delta value becomes 6.10 indicating that it is Benford (Figure 4b). In fact, as you increase the number of theoretical notes from a hundred to a thousand and beyond, the distribution of the frequencies created by Equation (3) becomes more and more Benford. We can see this in Figure 4c, where the delta value shrinks to 0.19. Thus, musical note frequencies in Hertz are Benford distributed [5].

A characteristic of sequences that are Benford distributed is that their deviation (Δ) from Benford goes to 0 as the number N of data points involved goes to ∞. That is, for each N, let f_N be a sequence of N numbers selected from the data set D. If D is Benford distributed, then we have:

$$\lim_{N \mapsto \infty} \frac{\text{total number of elements of } f_N \text{ whose first digit is } d_j}{N}$$

$$= \log(1 + \frac{1}{d_j}).$$

Our Equation (3) satisfies this property.

3. Methods

To collect music to analyze, we used the website http://www.kunstderfuge.com/, which has a wide collection of music from various composers throughout the various musical eras, all in MIDI format, a type of musical file that can be analyzed easily by a computer.[1]

Due to the significantly larger sample size of songs for Johann Sebastian Bach, and because the AI Bach was trained on Bach fugues, we ended up limiting our data set for the human Bachs to make things more comparable. In the end we worked with thirteen fugues for Johann Sebastian Bach and thirteen piano sonatas for Johann Christian Bach.

To analyze the MIDI files, we used the programming software `R` [6]. First, we imported the package *tuneR* to read the MIDI files and extracted the note values for each song. The note values were then converted to Hertz using Equation (3). After converting each of the notes to frequencies in Hertz, we then found the distribution of the first digits of those frequencies. We compared the distribution of the first digits of each song to the Benford distribution using Equation (2).

4. Results

To see which among the music of Emma (AI), Johann Christian Bach (JC), and Johann Sebastian Bach (JS) is more Benford, we used the Kruskal-Wallis test, typically used to determine if the medians of two or more independent groups are significantly different. With a $p < 0.001$, we were able to reject the hypothesis that the medians of all the groups are the same.

To determine which of the group medians differ, we performed a pairwise t-test with a Bonferroni-Holm adjustment. From the boxplot in Figure 5, the AI music composed by Emmy appears to have the highest delta with an approximate value of 30, indicating that it is the farthest from the Benford distribution. Both Johann Sebastian Bach and Johann Christian Bach have relatively low median delta values of 7.99 and 11.33, respectively.

[1]The songs for Emmy were taken from YouTube in MP3 form and subsequently converted to MIDI.

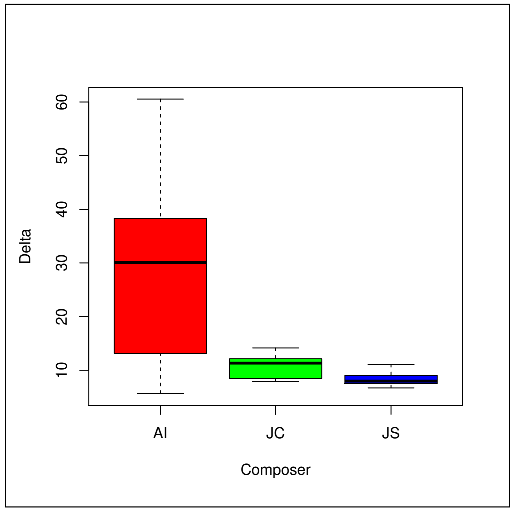

Figure 5: Box plot of the AI, JS, and JC music delta values.

To test for statistical significance, we used a post hoc analysis, the results of which are summarized below in Table 1:

Table 1: Post-hoc Analysis

	AI	JC	JS
AI	-	0.00049	0.000031
JC	0.00049	-	0.000194
JS	0.000031	0.000194	-

Table 1 shows how a group statistically differs from another. For example, the median delta value of AI Bach music is statistically different from the median delta value of the music of Johann Sebastian Bach with a p-value of 0.000031. So, we would reject the null assumption that AI Bach music and the music of Johann Sebastian Bach have the same median delta value. Likewise, with a p-value of 0.00049, we would reject the the hypothesis that AI Bach music and the music of Johann Christian Bach have the same median delta value. With the values for the music of Johann Sebastian Bach and that of Johann Christian Bach being significantly different from each other with a p-value of 0.000194, we reject the hypothesis that they have the same delta value. Combining this result with the lowest median delta value from the boxplot in Figure 5, we conclude that Johann Sebastian Bach's music is more Benford than AI music and the music of Johann Christian Bach.

5. Conclusion

The connection between Benford's Law and music has already been established through prior research [2, 4]. The purpose of this research was to investigate whether this connection can differentiate between artists and between human composed music compared to computer generated music. To do this, we collected samples of music from Johann Sebastian Bach, Johann Christian Bach, and AI-generated Bach music and looked at the distribution of musical note frequencies for all the songs. We computed the deviation Δ of the distribution of those frequencies using Equation (2).

Looking at the median delta for each group, we notice that the AI-generated Bach music had the highest Delta, followed by Johann Christian Bach, followed by Johann Sebastian Bach with the lowest. A Kruskal Wallis test and a post hoc analysis showed that each group's median delta value was significantly different from the others'. Thus we conclude that the music of Johann Sebastian Bach is the most Benford while computer-generated music is not Benford at all.

This finding is slightly different than previous findings that Classical music is more Benford than Baroque music [4]. This may be accounted for by the fact that we limited our analysis to thirteen Johann Sebastian Bach fugues and thirteen Johann Christian Bach piano sonatas since the latter did not write fugues. Because the AI Bach was trained on Johann Sebastian Bach fugues, we felt this was the closest comparison to make. Analysis of more songs for each of the Bachs may confirm previous findings that Classical music is more Benford than Baroque music.

Further investigation is needed to determine how the "Benfordness" of a song affects a listener's enjoyment. We hypothesize that though the computer-generated music does sound like Bach, it is less enjoyable as it sounds a little repetitive and formulaic. Another theory is that the "Benfordness" of a song reflects in its mainstream appeal. Thus far, the research done in this context has focused on popular musicians from each time period. We next wish to explore whether artists with less recognition or that are not as mainstream will subsequently have a higher Delta value.

Software. Software in the form of `R` code, together with a sample input data set and complete documentation is available upon request from the fourth author (sprincenelson@wlu.edu) and on Github.

Acknowledgements.
We would like to acknowledge the Washington and Lee Summer Research Scholars (SRS) and Advanced Immersion and Mentorship programs for their funding. A special thanks to 2021 SRS students Brian Wickman, Eric Gazin, and Jack Null.

References

[1] Berger, A. and Hill, T., *An Introduction to Benford's Law*, Princeton University Press, 2015.

[2] Khosravani, A., "Emergence of Benford's law in music," *Journal of Mathematical Sciences: Advances and Applications*, Volume **54** Issue 1 (2018), pages 11–24.

[3] Li, F., Han, S., Zhang, H., Ding, J., Zhang, J., and Wu, J., "Application of Benford's law in data analysis," *Journal of Physics: Conference Series*, Volume **1168** (2019), 032133. doi:10.1088/1742-6596/1168/3/032133

[4] Prince Nelson, S., Wickman, B., Null, J., and Gazin, E., "An evaluation of music's adherence to Benford's law throughout history," *Journal of Mathematical Sciences: Advances and Applications*, Volume **67** Issue 1 (2021), pages 73–84.

[5] Prince Nelson, S., Wickman, B., Null, J., and Gazin, E. "Benford's law and music note frequencies," pages 383–389 in *Mathematics and Computation in Music – MCM 2022*, edited by Mariana Montiel, Octavio A. Agustín-Aquino, Francisco Gómez, Jeremy Kastine, Emilio Lluis-Puebla, and Brent Milam (Lecture Notes in Computer Science, Volume 13267, Springer, Cham, 2022). doi:10.1007/978-3-031-07015-0_34

[6] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[7] Randel, D., *Harvard Concise Dictionary of Music*, Belknap Press, Cambridge, 1978.