Journal of Mathematical Sciences: Advances and Applications
Volume 67, 2021, Pages 73-84
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jmsaa_7100122222

AN EVALUATION OF MUSIC'S ADHERENCE TO BENFORD'S LAW THROUGHOUT HISTORY

SYBIL PRINCE NELSON, BRIAN WICKMAN, JACK NULL and ERIC GAZIN

Department of Mathematics and Data Science
Washington and Lee University
Lexington, VA 24450
204 W. Washington St.
USA
e-mail: sprincenelson@wlu.edu

Abstract

Western Music history can be divided into six major categories: Medieval, Renaissance, Baroque, Classical, Romantic, and Post-War. We analyzed a large collection of music from each time period and discovered a clear mathematical connection. Within each time period, we found that the note frequencies measured in hertz (Hz) and note durations are all Benford distributed. We also found that as music progressed through time, note lengths adhered closer and closer to the Benford distribution with the exception of the Post-War time period.

2020 Mathematics Subject Classification: Primary 00A65; Secondary 60E05.
Keywords and phrases: Benford distribution, classical music, logarithmic distribution of first digits, box plots, Kruskal-Wallis test.
Received July 27, 2021

1. Introduction

The paper “Emergence of Benford’s Law in Music” by Khosravani et al. showed that there is a clear connection between time intervals of each note in classical works and the Benford distribution. They considered the works of Bach, Beethoven, Mozart, Tchaikovsky, and Schubert. This collection of works included composers from several musical eras leaving one to question: Is there a connection between the musical era and how well music adheres to the Benford distribution ?

Historians commonly divide Western music history into six major categories: Medieval, Renaissance, Baroque, Classical, Romantic, and Post-War. These eras of music are characterized by differences in style, inspiration, emotion, and creative genius of the performers (Rosentiel [5]). In addition, wars, social and technological advances, religious beliefs, and philosophical trends contributed to these contrasts in musical eras. After investigating a collection of songs from each era, we found that each category was relatively Benford. Looking at the progression through time, we discovered that the time periods adhered more and more close to the Benford distribution.

To collect MIDI files for analysis, we used the website http://www.kunstderfuge.com, which is a large resource of thousands of songs from hundreds of composers. We used R to find the note durations and musical note frequencies of each song. To measure how closely each song adhered to the Benford distribution in note duration and musical note frequency, we calculated a delta value as follows:

$$\Delta = 100 \cdot \max_{d=1}^{9}\left|\mathrm{Prob}(D_1 = d) - \log_{10}\left(1 + \frac{1}{d}\right)\right|.$$

This value, noted by Berger and Hill, implies that the smaller a delta value is, the closer the data adheres to a Benford distribution. We aggregated the delta values for each time period and then performed an ANOVA to determine if there is a statistical difference between median delta values for each time period.

For this paper, we will first provide a brief summary of Benford's law and give an overview of the six time periods of Western music. Next, we will provide our numerical analysis of how each time period adheres to Benford's law and the trend we discovered. Finally, we will give a brief discussion of how we can interpret our results and what future steps should be taken.

2. Benford's Law

What is known today as Benford's law was first discovered in 1881 by Canadian-American astronomer and mathematician Simon Newcomb. He noticed that the pages of the logarithmic tables did not wear out at equal frequency. Namely, the first page being more used, the second page used less often, the third page used less often than the second and so diminishing up to page 9 (Newcomb [1]). This discovery did not attract further interest until Frank Benford rediscovered this natural phenomenon in 1938. After researching over 20,000 digits, he formally proposed the "First Digit Law", which was later named "Benford's Law" in his honor. He showed that $\mathrm{Prob}(d) = \log_{10}\left(1 + \frac{1}{d}\right)$, where $d = 1, 2, \ldots, 9$.

Over the next several decades after the First Digit Law was formalized by Benford, development of this research was relatively slow; it was not until 1995 that Theodore Hill provided a theoretical proof that first digits obeyed Benford's law (Hill [3]).

One might think that the probability of a first digit being 1 to 9 in a set of naturally occurring data is equal or constant. Yet Benford's law shows instead that naturally occurring data sets follow a uniform logarithmic distribution. Using the formula, we see that the probability that $d = 1$ is $\mathrm{Prob}(1) = \log_{10}\left(1 + \frac{1}{1}\right) = \log_{10}(2) = 0.30103$ and the probability that $d = 2$ is $\mathrm{Prob}(2) = \log_{10}\left(1 + \frac{1}{2}\right) = \log_{10}(1.5) = 0.17609$.

Table 1 shows all the probabilities from 1 through 9 and Figure 1 shows a Benford distribution.

Table 1. First digit probabilities

Digit	1	2	3	4	5	6	7	8	9
Prob	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

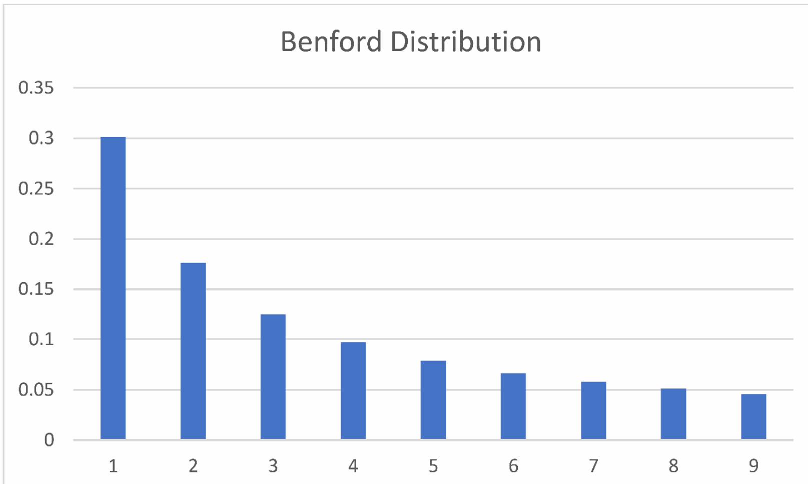

Figure 1. Benford probability distribution.

The application of Benford's law has proven quite ubiquitous. Today, research is being conducted in the fields of economics, sociology, physics, computer science, and biology (Li et al. [4]). This paper explores an unsuspected yet interesting field for the application of Benford's law: music.

3. Data and Methods

Using the timeline of music history from www.naxos.com, we categorized our sample of musical composers into their respective time period. Table 2 shows a sample of composers from each time period that we used for this analysis. Our data set consisted of 320 songs from the

Medieval through the Post-War musical era. For the sake of continuity, we considered only Neo-Classical artists from the Post-War era. Otherwise, this which includes everything from 1920 to today would be too broadly defined, as it would include diverse genres such as rag time, rap and country music. Table 2 also shows the number of songs we used and the number of total notes analyzed from each respective time period.

Table 2. Number of songs per time period

Descriptive statistics

Time period	Years	Number of songs	Total notes analyzed	Composers
Medieval	1150-1400	31	12,276	Unknown*
Renaissance	1400-1600	72	58,200	King Henry VII, William Byrd, Turlough O'Carolan
Baroque	1600-1750	89	152,600	J. S. Bach, Bivaldi, Monteverdi
Classical	1750-1830	44	176,000	Beethoven, Mozart, Chopin
Romantic	1830-1920	29	207,000	Tchaikovsky, Mendelssohn, Schumann
Post War	1920-present	55	217,200	Medtner, Prokofiev, Satie

*Medieval artist often did not take credit for their work for religious reasons.

We used the "tuneR" package in R to analyze our 320 songs. In particular, we obtained each note played in the song and the time duration in seconds. We then used the formula for frequency (Hz)$f(n) = 440\left(\sqrt[12]{2}\right)^{n-69}$ to convert each note to a specific wavelength. Next, we extracted the leading digit for the cumulative number of seconds played for every note and the leading digit of the wavelength frequencies. The cumulative length durations for Felix Mendelssohn's Symphony No. 3 are shown below in Table 3 in German musical notation. As an example, tuning A, or A4, is written as a' in Table 3 and has a first digit of 8.

Table 3. Cumulative note length durations for Mendelssohn's Symphony No. 3

Cumulative note length durations for Mendelssohn's Symphony No. 3

German note	Total length	German note	Total length	German note	Total length	German note	Total length
C,	0.06	A	45.53	f'	23.94	c#'''	1.92
D,	0.33	A#	4.30	f#'	25.89	d'''	9.33
D#,	0.19	B	18.10	g'	32.73	d#'''	4.96
E,	1.16	c	14.06	g#'	18.23	c'''	14.49
F,	0.76	c#	6.57	a'	81.30	f'''	4.36
F#,	0.15	d	13.69	a#'	9.82	f#'''	3.91
G,	1.38	d#	11.25	b'	38.77	g'''	2.99
G#,	0.84	c	72.14	c''	37.45	g#'''	0.03
A,	13.49	f	11.17	c#''	9.13	a'''	1.86
A#,	0.79	f#	8.80	d''	35.11		
B,	5.02	g	21.86	d#''	19.28		
C,	6.56	g#	14.02	c''	70.41		
C#	1.62	a	59.40	f''	16.05		
D	7.79	a#	7.64	f#''	21.83		
D#	3.53	b	50.19	g''	25.76		
E	17.19	c'	55.53	g#''	12.68		
F	6.16	c#'	6.86	a''	36.15		
F#	3.30	d'	44.70	a#''	4.29		
G	7.63	d#'	19.82	b''	19.67		
G#	4.43	c'	95.57	c'''	17.42		

We analyze the distribution of the first digits of the total length column.

4. Methodology

Figure 2 shows the distribution of leading digits for wavelength frequencies and cumulative note lengths compared to the Benford distribution for all 29 songs in our data set from the Romantic period (c. 1800-1920).

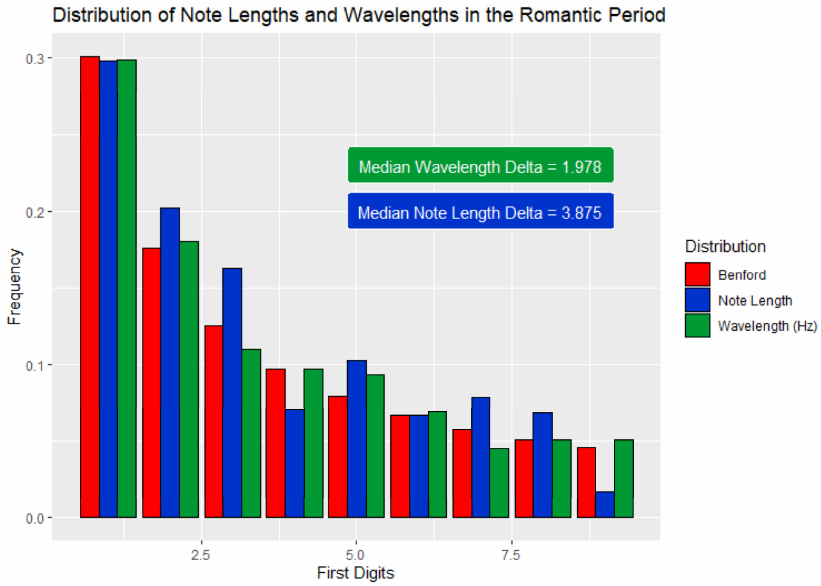

Figure 2. Distribution of leading digits for the Romantic period compared to Benford's law.

To compare how well a data set fits the Benford distribution, we use the following formula to calculate delta:

$$\Delta = 100 \cdot \max_{d=1}^{9}\left|\text{Prob}(D_1 = d) - \log_{10}\left(1 + \frac{1}{d}\right)\right|. \quad [5]$$

The higher the delta value, the less the data set fits the Benford distribution. In our songs from the Romantic period, for example, the median delta value is 1.978 for note frequency and 3.875 for cumulative note length (Figure 2). Figures 3 and 4 show the delta values for cumulative note length and wavelength frequency, respectively.

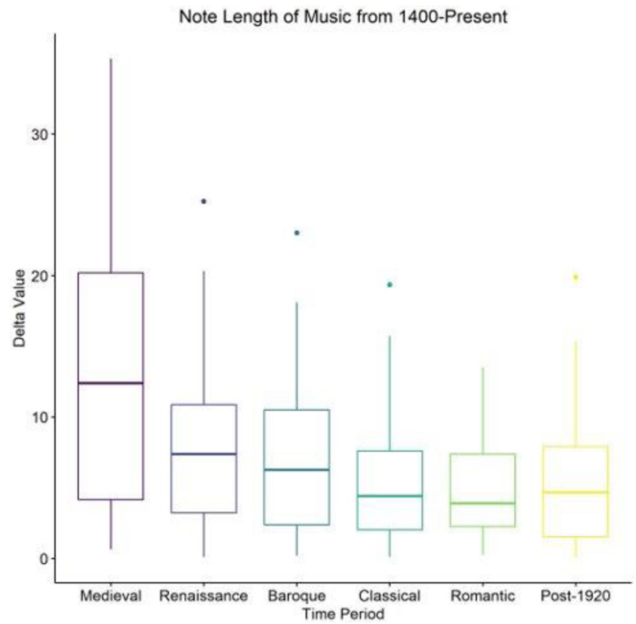

Figure 3. Box plot of note length delta values across all time periods.

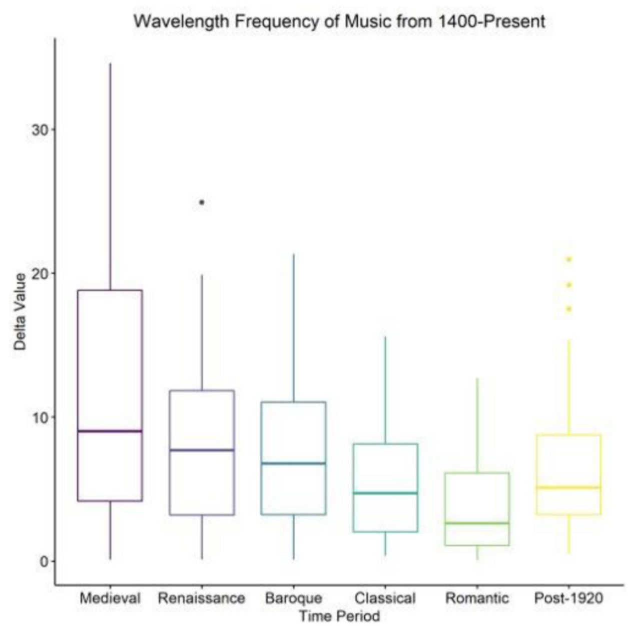

Figure 4. Box plot of wavelength frequencies delta values across all time periods.

A Kruskal-Wallis test revealed statistically-significant differences between groups for both wavelength frequency and cumulative note length. We then used a Mann-Whitney post-hoc test with a Bonferroni adjustment to identify which time periods had significantly different distributions of deltas. The results of the post-hoc tests are captured for note length and wavelength in Tables 4 and 5, respectively.

Table 4. Post hoc analysis of median delta values for note lengths

	Baroque	Classical	Medieval	Post-1920	Renaissance
Classical	1.0000	-	-	-	-
Medieval	0.0459	0.0095	-	-	-
Post-1920	0.6179	1.0000	0.0013	-	-
Renaissance	1.0000	0.7778	0.1677	0.1486	-
Romantic	1.0000	1.0000	0.0134	1.0000	0.4977

Table 5. Post hoc analysis of differences in median delta value for wavelength frequency

	Baroque	Classical	Medieval	Post-1920	Renaissance
Classical	0.7785	-	-	-	-
Medieval	0.3300	0.0221	-	-	-
Post-1920	1.0000	1.0000	0.0962	-	-
Renaissance	1.0000	0.2550	1.0000	1.0000	-
Romantic	0.0087	0.7502	0.0011	0.1737	0.0022

5. Results

Figure 3 illustrates that as time progressed, note lengths adhered more closely to a Benford distribution with the exception of the Post-War period. As seen in the figure, the median delta value for each period, represented as a line on the box plot, decreases as time progresses. Table 4 identifies the following statistically significant differences between time periods: Baroque-Medieval, Classical-Medieval, Post-War-Medieval, and Romantic-Medieval. Essentially, the Medieval time period is statistically different at the $\alpha = 0.1$ level from all of the other time periods with the exception of the Renaissance period.

Similarly, Figure 4 shows that as time progressed, the wavelength frequencies adhered more closely to the Benford distribution with the exception of the Post-War period. Table 5 reveals that there were significant differences at the $\alpha = 0.1$ level between the following periods for wavelength delta values: Romantic-Baroque, Classical-Medieval, Medieval-Post-War, Romantic-Medieval, and Romantic-Renaissance.

6. Discussion and Conclusion

At first glance, one might reasonably conclude that the varying median delta values for the different eras of Western music can simply be attributed to advances in tuning mechanisms for instruments. However, this would be a drastic oversimplification for two major reasons. One, many of the time periods used the same tuning system. The tuning systems went from Just Intonation, Pythagorean tuning, Meantone tuning, Well temperament, to Equal temperament which is currently used. Some time periods used several tuning systems and in other cases, one tuning method was used for several eras. For example, well temperament was firmly established in the 16th century and became widely used during the Baroque period and used until the 20th century when equal temperament which is based on the 12th root of two became the standard. Thus, well temperament was used during both the Baroque and Classical periods yet, they have different median delta values, indicating different levels of adherence to Benford's law.

Another reason that differing delta values cannot solely be attributed to tuning methods is that, in most cases, the different tuning methods only produced minor differences in the hertz measurement. For example, musicians in the medieval era used a tuning method attributed to Pythagoras; it was a 12-tone temperament that made all 5ths perfectly consonant and tuned in a ratio of 3.2. The note A was tuned such that its frequency equalled 3/2 the frequency of D. So, if D was 288Hz, then A was 432Hz. In equal and well temperament, A is 440Hz. Notice the first digit has not changed meaning the change in tuning standards could not have affected conformity to Benford's law. There are many different tuning methods available and widely used. While they may produce pitches slightly different to the trained ear, these differences clearly do not affect the first digit of frequency value in hertz.

A difference in tuning methods also would not explain why there was a significant difference in note durations between the time periods. A more plausible and less simplistic explanation involves the different styles in music through the time periods as influenced by social, religious, technological, and political variations. For example, during the medieval era, most music was religious and involved monophonic chants (Rosentiel [5]). In general, the music was simple, most often sung, and used primarily for sacred purposes. Also, the Pythagorean system of tuning did not allow for easy harmonic development. For these reasons, each song of that time period most likely used fewer notes. This would also explain the large variance for the era. On the other end of the spectrum, the Romantic music is known for its passion, imagination and the virtuosity of the composers. Thus, there were probably more varied notes used during this period.

The Post-War era noticeably does not follow the trend of the other time periods. Nevertheless, it does reveal an interesting observation. The songs selected for the other eras were of similar style. Even though there were different composers used, each time period in Western music is very much identifiable by a certain feeling it evokes. Even some untrained ears can tell the difference between a Bach fugue from the Baroque era and a Chopin Nocturne from the Romantic era. The Post-War era on the other hand includes everything from Stravinsky to Nine Inch Nails. So, because this time period focused on similar styles of music, we chose to look at only Neoclassical songs of the modern era. It is interesting to note that the median delta values for both wavelength and note duration are similar to that of the classical period instead of continuing the trend and being smaller than the Romantic era. Further investigation is needed to determine the best way to delineate groups in the Post-War era.

In conclusion, this paper has shown that there is a clear mathematical connection between note durations and wavelength frequencies of music in the major time periods of Western music. This connection, demonstrated by an adherence to Benford's law, also has a significant time dependence as shown by the results of our ANOVA analysis. With the exception of the Post-War era, music has adhered more and more closely to the Benford distribution. Further analysis is currently underway to discover connections of modern musical genres with Benford's law.

References

[1] Simon Newcomb, Note on the frequency of use of the different digits in natural numbers, American Journal of Mathematics 4(1) (1881), 39-40.

DOI: https://doi.org/10.2307/2369148

[2] Frank Benford, The law of anomalous numbers, Proceedings of the American Philosophical Society 78(4) (1938), 551-572.

[3] Theodore P. Hill, The significant-digit phenomenon, The American Mathematical Monthly 102(4) (1995), 322-327.

DOI: https://doi.org/10.1080/00029890.1995.11990578

[4] Feifei Li, Shuqing Han, Hongyv Zhang, Jiaojiao Ding, Jianhua Zhang and Jianzhai Wu, Application of Benford's law in data analysis, Journal of Physics: Conference Series 1168 (2019), 032133.

DOI: https://doi.org/10.1088/1742-6596/1168/3/032133

[5] Rosenstiel Léonie, Schirmer History of Music, Schirmer Books, 1982.

[6] Mathematics and Music, American Mathematical Society.

www.ams.org/publicoutreach/math-and-music

[7] Arno Berger and Theodore Preston Hill, An Introduction to Benford's Law, Princeton University Press, 2015.

[8] Azar Khosravani and Constantin Rasinariu, Emergence of Benford's law in music, Journal of Mathematical Sciences: Advances and Applications 54(1) (2018), 11-24.

DOI: https://doi.org/10.18642/jmsaa_7100122017

■