The Case for Base-8: A Novel Way to Apply Benford's Law to Music

Sybil Prince Nelson

Washington and Lee University

January 6, 2026

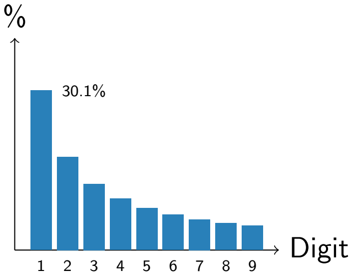
Roadmap

What is Benford's Law?

Observation: In many naturally occurring datasets, the leading digit is not uniformly distributed.

The digit 1 appears as the first digit about 30% of the time, while 9 appears less than 5% of the time.

First observed by Simon Newcomb (1881), later rediscovered by Frank Benford (1938) who tested it on 20 different datasets.



The Mathematical Formula

Benford's Law (Base-10)

The probability that the first significant digit is d is:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right) \quad \text{for } d \in \{1, 2, \ldots, 9\}$$

Digit d	1	2	3	4	5	6	7	8	9
$P(d)$	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

Key insight: Data spanning multiple orders of magnitude tends to follow this distribution.

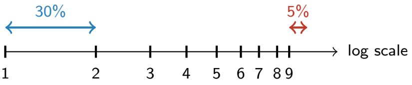
Why Does Benford's Law Work?

The key: Scale Invariance

Imagine measuring something in different units (meters vs. feet, dollars vs. euros). The leading digit distribution shouldn't change—Benford's Law is the *only* distribution with this property.

Logarithmic spacing:

- To go from 1 to 2, you increase by 100%
- To go from 8 to 9, you increase by only 12.5%
- Numbers "spend more time" with leading digit 1



The Intuition

On a logarithmic scale, the interval from 1 to 2 is *much wider* than from 8 to 9. Randomly sampled values are more likely to land in wider intervals.

Where Benford's Law Applies

Natural datasets:

- Population of cities
- River lengths and areas
- Physical constants
- Stock prices
- Electoral data

Practical applications:

- Fraud detection in financial audits
- Tax return verification
- Scientific data validation
- Election integrity analysis
- Image forensics

The Core Principle

Deviation from Benford's Law can indicate data manipulation, fabrication, or anomalies.

Benford's Law Meets Music

Three key studies have extended Benford's Law to musical data:

1. **Khosravani & Rasinariu (2018):** Note durations
2. **Nelson et al. (2022):** Note frequencies
3. **Barbancho et al. (2015):** Audio features

Let's examine each in detail...

Khosravani & Rasinariu (2018)

"Emergence of Benford's Law in Music"
Journal of Mathematical Sciences: Advances and Applications

What they studied:

- Note *durations* (how long each note is held)
- Hundreds of classical pieces
- Composers: Bach, Beethoven, Mozart, Schubert, Tchaikovsky

Key findings:

- Note durations follow Benford's Law
- The logarithmic pattern appears not just in first digits, but *all* digit positions
- First empirical evidence of Benford behavior in music

Limitation: Only examined temporal structure, not pitch/frequency

Nelson et al. (2022)

"Benford's Law and Music Note Frequencies"
Mathematics and Computation in Music (MCM), Springer LNCS

What they studied:

- Note *frequencies* in Hertz (the 88 piano keys: 27.5 Hz to 4186 Hz)
- Extended analysis beyond piano range (16.35 Hz to 7902.13 Hz)
- Compared Romantic-era classical vs. 2000s pop music

Key findings:

- Piano key frequencies are Benford distributed
- Romantic classical music conforms well to Benford
- Modern pop music shows weaker conformance
- Introduced "Naturalness" metric for measuring fit

Question raised: Why does modern music deviate?

Barbancho et al. (2015)

"Benford's Law for Music Analysis"
ISMIR 2015 (International Society for Music Information Retrieval)

What they studied:

- Audio signal features (not symbolic MIDI data)
- Proposed three new Benford-based audio features
- Practical applications in music information retrieval

Key findings:

- Can distinguish real chords from artificially assembled ones
- Useful for MIDI-to-audio conversion quality assessment
- Broader distributions (spanning more orders of magnitude) $\rightarrow$ better Benford fit

Implication: Benford conformance as a "naturalness" indicator

The Gap in the Literature

All existing music studies use Base-10

What we know:

- Benford's Law generalizes to any base
- Music has inherent mathematical structure
- Base-10 analysis shows mixed results

What hasn't been explored:

- Does the *choice of base* matter?
- Could a different base better capture musical structure?
- Is Base-10 even appropriate for music?

Generalizing Benford's Law

Benford's Law in Base b

$$P(d) = \log_b\left(1 + \frac{1}{d}\right) \quad \text{for } d \in \{1, 2, \ldots, b-1\}$$

Key theoretical results:

- **Hill (1995):** Proved base-invariance property—Benford's Law holds regardless of the base used
- **Berger & Hill (2015):** Comprehensive mathematical treatment in *An Introduction to Benford's Law*
- **Nigrini & Miller (2007):** Empirical verification across different bases

The Unexplored Question

If Benford's Law works in any base, shouldn't we choose the base that best matches our data's structure?

Why Base-10 Dominates

Base-10 is the default because:

- It's our standard numeral system
- Convenient for human interpretation
- "That's how it's always been done"

But consider:

- Computer scientists use Base-2 (binary), Base-8 (octal), Base-16 (hex)
- Time uses Base-60 (seconds/minutes) and Base-24 (hours)
- Music uses... what base?

"For computers with binary-power bases, base 8 (octal) is particularly natural." —Hill (1998)

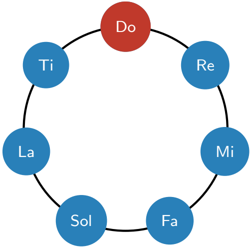
The Structure of Western Music

The Diatonic Scale:

- Western music is built on **7 distinct notes** per octave
- Do–Re–Mi–Fa–Sol–La–Ti–*(Do)*
- The 8th note is the **octave**—the cycle repeats

The Octave:

- Frequency doubles every octave
- $A4 = 440$ Hz $\rightarrow$ $A5 = 880$ Hz
- The octave is a natural "reset point"



7 notes + octave = 8

Base-8: A Natural Fit

The Hypothesis

Base-8 accommodates 7 scale degrees + the octave return.
This structural alignment may produce better Benford conformance than Base-10.

In Base-8 Benford analysis:

$$P(d) = \log_8 \left(1 + \frac{1}{d}\right) \quad \text{for } d \in \{1, 2, 3, 4, 5, 6, 7\}$$

Digit d	1	2	3	4	5	6	7
$P(d)$	33.3%	19.5%	13.8%	10.7%	8.7%	7.4%	6.5%

Scale degrees 1–7 map directly to first digits 1–7 in Base-8!

Tonal Hierarchy: A Music Theory Foundation

Not all scale degrees are created equal.

Music theory identifies a *stability hierarchy* among scale degrees:

Degree	Name	Stability	Role
1	Tonic	Most stable	Home base, resolution point
5	Dominant	Stable	Strong pull to tonic
3	Mediant	Stable	Defines major/minor quality
4	Subdominant	Less stable	Pre-dominant function
6	Submediant	Less stable	Relative minor connection
2	Supertonic	Unstable	Passing tone, approach
7	Leading tone	Least stable	Strong pull to tonic

Why Tonal Hierarchy Matches Benford

Benford predicts (Base-8):

- Digit 1: 33.3%
- Digit 2: 19.5%
- Digit 3: 13.8%
- Digit 5: 8.7%
- Digit 7: 6.5%

Compositional practice:

- Melodies start/end on tonic (1)
- Chord tones (1, 3, 5) are emphasized
- Passing tones (2, 4, 6) are brief
- Leading tone (7) resolves quickly

The Deeper Connection

Western music's tonal system naturally produces a *logarithmic-like* distribution of scale degrees. Base-8 captures this structure; Base-10 (analyzing raw Hz) does not.

Data and Approach

Dataset: 22 lead sheets (monophonic melodies)

- Jazz standards, pop classics, contemporary songs
- MIDI format with key signature metadata
- Single melodic line (avoids polyphonic complexity)

Two parallel analyses:

1. **Base-10 (Traditional):** Extract note frequencies in Hz, analyze first digit
2. **Base-8 (Novel):** Convert pitches to scale degrees (1–7), analyze distribution

Conformance metric: Mean Absolute Deviation (MAD) from expected Benford distribution

$$\text{MAD} = \frac{1}{b-1}\sum_{d=1}^{b-1}|P_{\text{observed}}(d) - P_{\text{Benford}}(d)|$$

Lower MAD = Better conformance

Why MAD? Captures how well the *entire* distribution fits, not just the worst digit.

Sample of Pieces Analyzed

- Blue Skies (Ella Fitzgerald)
- Brown Eyed Girl
- Can't Help Falling in Love
- Dancing Queen
- Danny Boy
- Dream a Little Dream of Me
- For Once in My Life
- From The Start (Laufey)
- Hallelujah
- Radioactive
- It's Been a Long, Long Time
- Jingle Bells
- Killing Me Softly
- La Bamba
- Misty
- My Way
- Perfect (Ed Sheeran)
- Que Sera, Sera
- Somewhere Beyond the Sea
- Tomorrow (Annie)
- Wonderful Tonight

Diverse genres: Jazz, Pop, Rock, Standards, Show tunes

Key Finding: Base-8 Wins

Statistical Result

Welch Two-Sample t-test:

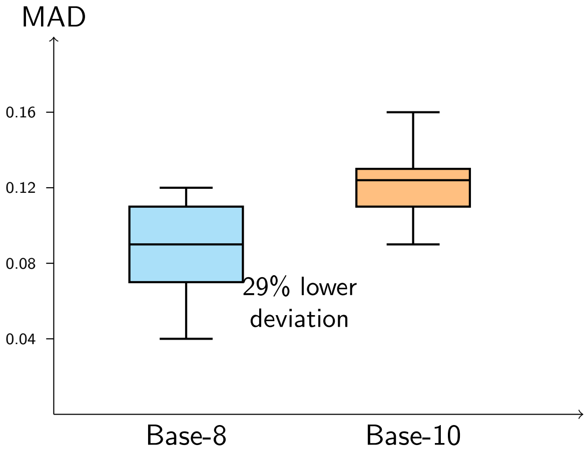
$t = -5.10$, $df = 37.4$, $p < 0.001$ ($p = 1.0 \times 10^{-5}$)

	Base-8 (Scale Degrees)	**Base-10 (Hz)**
Mean MAD	**0.088**	0.124

Consistent Pattern

Base-8 shows **29% lower** mean deviation from Benford's Law than Base-10.

Visualizing the Difference



Base-8 produces distributions that are 29% closer to Benford's Law on average.

Best and Worst Conformance

Lowest MAD (Best Fit):

1. Dancing Queen (0.037)
2. For Once in My Life (0.040)
3. Dream a Little Dream (0.051)
4. Blue Skies (0.054)
5. Perfect (0.062)

Highest MAD (Worst Fit):

1. Que Sera Sera (0.124)
2. Hallelujah (0.119)
3. Somewhere Beyond the Sea (0.118)
4. Misty (0.117)
5. Radioactive (0.112)

Notable: Radioactive has the lowest Base-10 MAD (0.074)—an interesting outlier.

Outlier Analysis: The Radioactive Puzzle

Radioactive (Imagine Dragons) behaves unusually:

	Base-8 MAD	**Base-10 MAD**
Radioactive	0.112 (5th worst)	0.074 (best!)
Dataset average	0.088	0.124

What might explain this?

- **Melodic structure:** Radioactive has a narrow, repetitive melodic range
- **Modern production:** Electronically produced, pitch-corrected vocals
- **Genre effect:** Rock/pop anthems may use different melodic patterns than jazz standards
- **Chromatic content:** May use more non-diatonic tones (excluded from Base-8 analysis)

Future work: Investigate what musical features predict Base-8 vs. Base-10 conformance.

Why Does Base-8 Work Better?

Three structural reasons:

1. **Octave alignment:** Base-8's digit range (1–7) matches the diatonic scale exactly
2. **Tonal hierarchy:** Music theory's stability hierarchy roughly follows logarithmic distribution
3. **Frequency-independent:** Scale degrees abstract away from arbitrary Hz values

Implications

For music analysis:

- More accurate "naturalness" metrics for genre classification
- Better anomaly detection (AI-generated vs. human-composed music?)
- New approach to style analysis and authenticity verification

For Benford's Law research:

- Demonstrates value of **domain-appropriate base selection**
- Challenges the "default to Base-10" assumption
- Opens questions: What other domains have natural non-decimal structure?

Broader methodological point:

The base we choose should reflect the structure of our data, not just convenience.

Summary

1. **Benford's Law** describes the surprising distribution of leading digits in natural data
2. **Previous music research** used Base-10 with mixed results
3. **The 7-note diatonic scale** suggests Base-8 as a natural alternative
4. **Empirical comparison** (n=22): Base-8 produces significantly better Benford conformance
   - $t = -5.10$, $p < 0.001$
   - 29% reduction in mean deviation
5. **Conclusion:** Base-8 better captures the mathematical structure of Western music

Future Directions

- **Expand the corpus:** More pieces, include minor keys, diverse genres
- **Non-Western music:** Do other scales suggest different bases?
  - Pentatonic (Base-6?)
  - Chromatic (Base-12?)
- **Polyphonic analysis:** Extend beyond monophonic lead sheets
- **Historical analysis:** Has Benford conformance changed over music history? - Base-10 (yes) Base-8 (?)
- **AI music detection:** Can Base-8 Benford analysis distinguish human vs. AI compositions? - yes!

Thank You

Questions?

sprincenelson@wlu.edu
Washington and Lee

References I

- Barbancho, A. M., et al. (2015). Benford's Law for Music Analysis. *ISMIR 2015*.
- Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Berger, A., & Hill, T. P. (2015). *An Introduction to Benford's Law*. Princeton University Press.
- Hill, T. P. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, 10(4), 354–363.
- Hill, T. P. (1998). The First Digit Phenomenon. *American Scientist*, 86(4), 358–363.
- Khosravani, A., & Rasinariu, C. (2018). Emergence of Benford's Law in Music. *J. Math. Sci.: Adv. & Appl.*
- Nelson, R., et al. (2022). Benford's Law and Music Note Frequencies. *MCM 2022, LNCS*. Springer.
- Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *Amer. J. Math.*, 4, 39–40.

References II

- Nigrini, M. J., & Miller, S. J. (2007). Benford's Law Applied to Hydrology Data. *J. Hydrology*, 334, 255–269.