Benford's Law: Applications in Astronomical Data

Team : 1-Digit

Members :

1. Parth Kumar Singh (parth.singh@adypy.edu.in)
2. Veer Shah (veer.shah@adypy.edu.in)
3. Ganesh Wayal (ganesh.wayal@adypy.edu.in)
4. Bhavishya Sharma (bhavishya.sharma@adypy.edu.in)

Introduction to Benford's Law

Definition

- Also known as the "First-Digit Law" or the "Law of Anomalous Numbers"
- A mathematical phenomenon where in many naturally occurring datasets, the leading digit is more likely to be small
- Specifically, digit 1 appears as the first digit about 30% of the time, while 9 appears less than 5% of the time

Mathematical Formula

The probability of digit d is $P(d) = \log_{10}(1 + 1/d)$.

- 1: 30.1%
- 2: 17.6%
- 3: 12.5%
- 4-9: decreasing probabilities

Historical Background

Origins of Benford's Law:

- First noticed by astronomer Simon Newcomb in 1881
  - Observed that pages of logarithm tables were more worn at the beginning
  - Published in the American Journal of Mathematics
- Rediscovered by physicist Frank Benford in 1938
  - Analyzed 20,000+ numbers from diverse sources (river lengths, populations, physical constants)
  - Published "The Law of Anomalous Numbers" showing the pattern held across different datasets
- Mathematical proof provided by Theodore Hill in 1995
  - Showed why naturally occurring numbers tend to follow this distribution

Applications of Benford's Law

Data Validation

- Quality control for scientific measurements
- Testing computer models and simulations
- Identifying errors in datasets

Fraud Detection

- Financial auditing and accounting
- Detecting manipulated data in tax returns
- Election fraud analysis

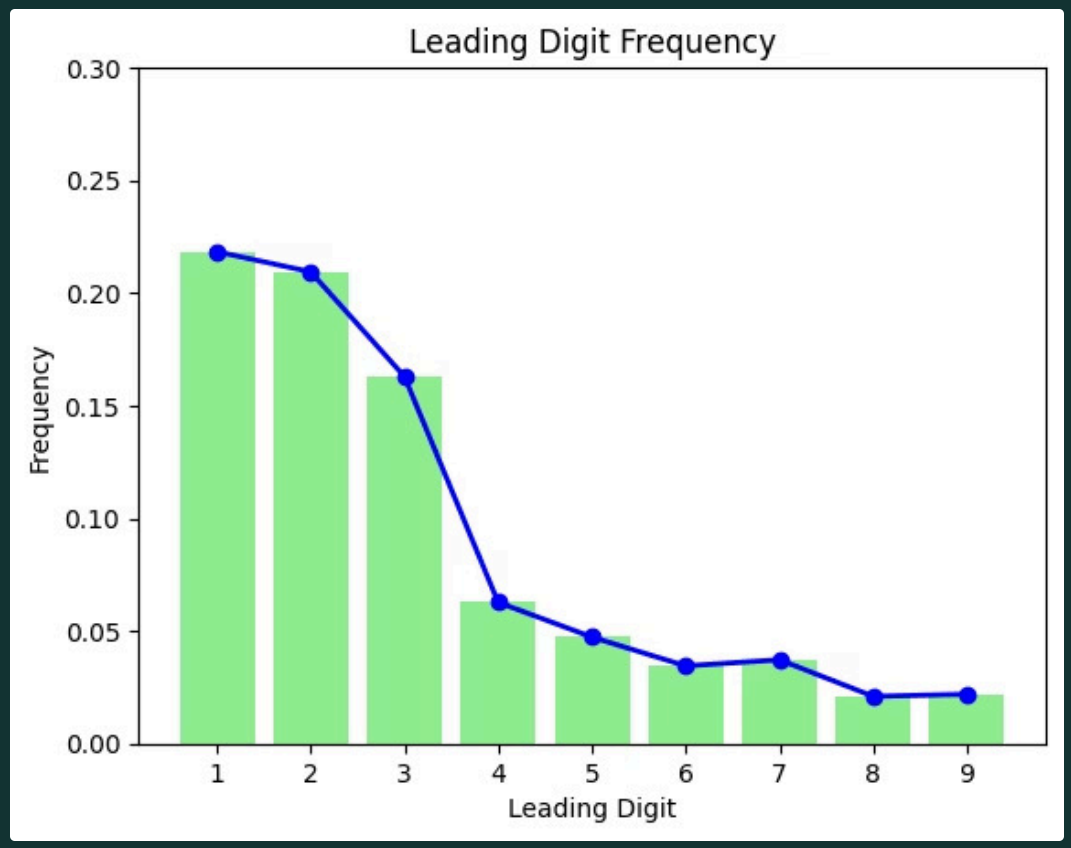
Natural Phenomena

- Population figures
- Physical constants
- Geographic measurements
- Astronomical data

Key Characteristics of Benford-compliant Data:

- Spans multiple orders of magnitude
- Not artificially constrained or truncated
- Generated by natural processes

Applying Benford's Law to the Stars Dataset



- **Goal:**
  - To apply Benford's Law to the stars dataset and assess whether the numerical attributes conform to the expected distribution.
- **Dataset Overview:**

1. Contains data on 113 stars from the night sky
2. Includes various astronomical measurements:
   - Visual magnitude
   - Absolute magnitude
   - Parallax
   - Distance (light years)
   - Galactic coordinates
   - Spectral information

- **Methodology:**

1. Extract different numerical columns from the dataset
2. Analyze the first digits of each measurement type
3. Compare observed frequencies with Benford's expected distribution
4. Calculate deviation to assess compliance with Benford's Law

Observations and Conclusions

Observations:

1. **Natural astronomical measurements that span orders of magnitude follow Benford's Law**
    - Distance and parallax measurements show strong conformity
    - These parameters represent naturally occurring physical properties
2. **Constrained or artificially selected measurements deviate from Benford's Law**
    - Visual magnitude is constrained by observational limitations
    - The dataset itself contains selection bias (mostly bright, visible stars)

3. **Dominance of Lower Digits:**

- The digit 1 appears most frequently, followed by 2, 3, and so on — this is consistent with Benford's Law. • This suggests the dataset generally follows a natural distribution.

4. **Overall Alignment:**

- The bars for actual frequencies closely match the expected Benford distribution, especially for digits 1 through 5.
- Minor fluctuations (such as slightly higher/lower bars for certain digits) are normal in real-world data.

5. **No Major Anomalies Detected:**

- There is no significant deviation in the frequency of any leading digit.
- This implies the data is likely authentic, not manipulated, and was collected from naturally occurring phenomena (like stellar attributes).

Conclusion:

The application of Benford's Law on the star dataset reveals a distribution that aligns closely with expected natural patterns. This indicates that the dataset is likely:

- Genuine and unaltered.
- Reflective of real-world astronomical measurements.
- Suitable for further scientific or statistical analysis.

Benford's Law proved effective in validating the integrity of the dataset, which is critical in fields like astronomy, finance, and forensics.

Thank You

We appreciate your attention and look forward to your feedback.