

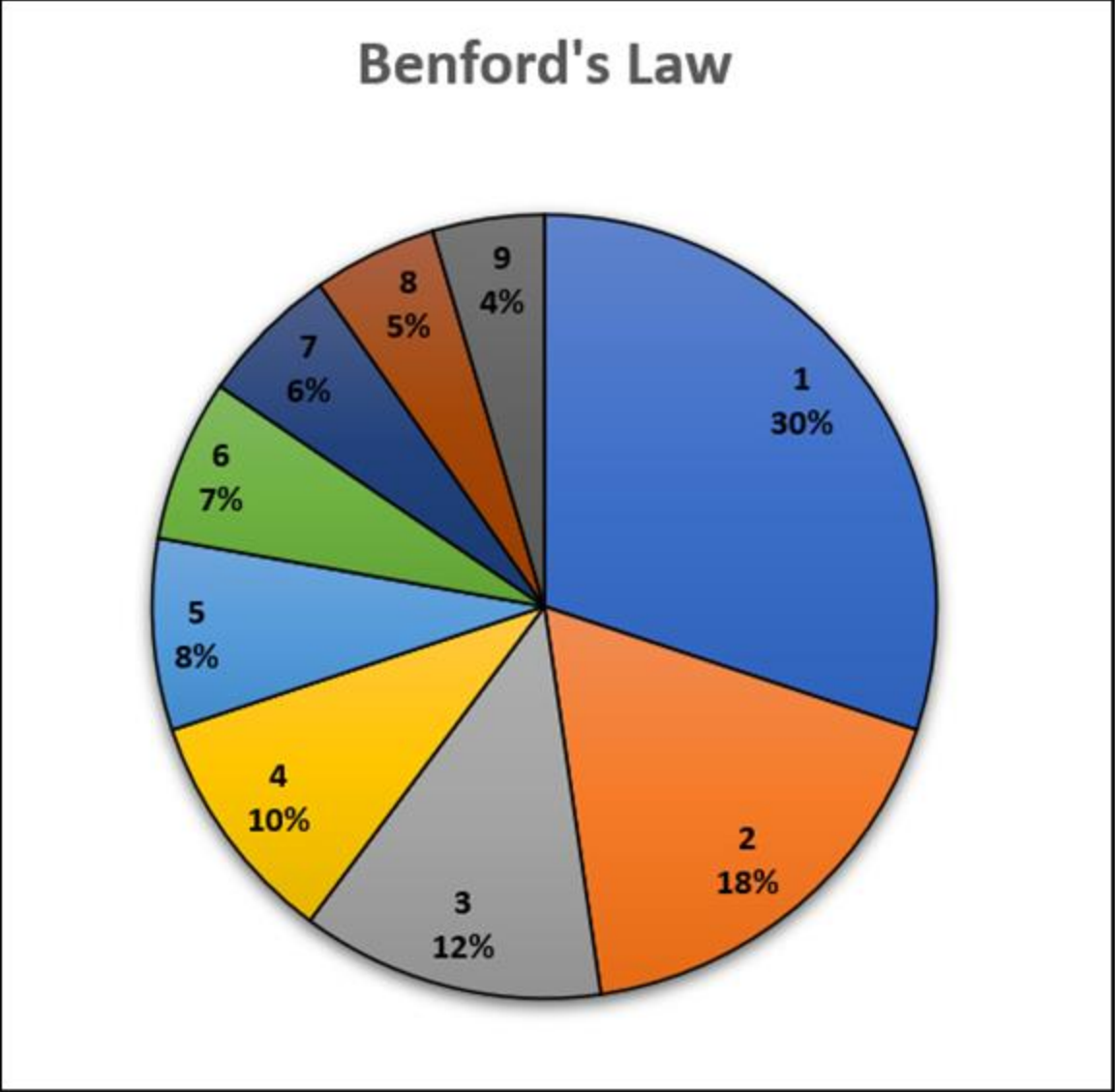
The Digit Files: Uncovering the patterns

Sujata B Sharma

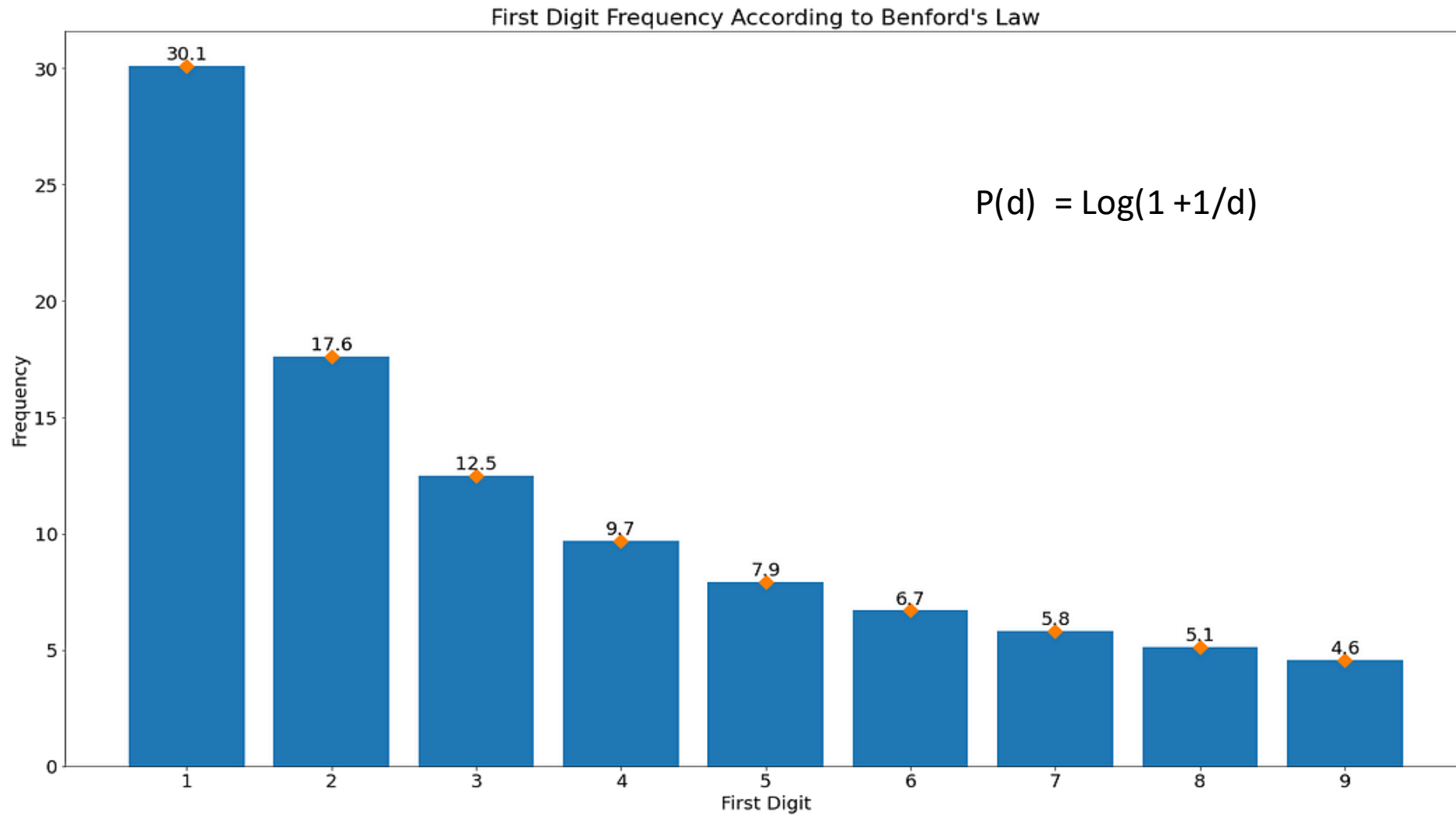
Quiz : Can there be a pattern in these data?

- ***Financial Data:*** Like income tax returns and expenses.
- ***Population Statistics:*** City and country populations
- ***Stock Market Prices:*** Daily closing values.
- ***Scientific Data:*** Constants and physical measurements.
- ***Natural entities:*** Lengths of rivers, Intergalactic distances
- ***Sports:*** Statistics, like scores and player achievements.
- ***Geological Data:*** Like earthquake magnitudes.
- ***Social Media:*** Engagement metrics

The Frequencies of Leading Digits in the Dataset

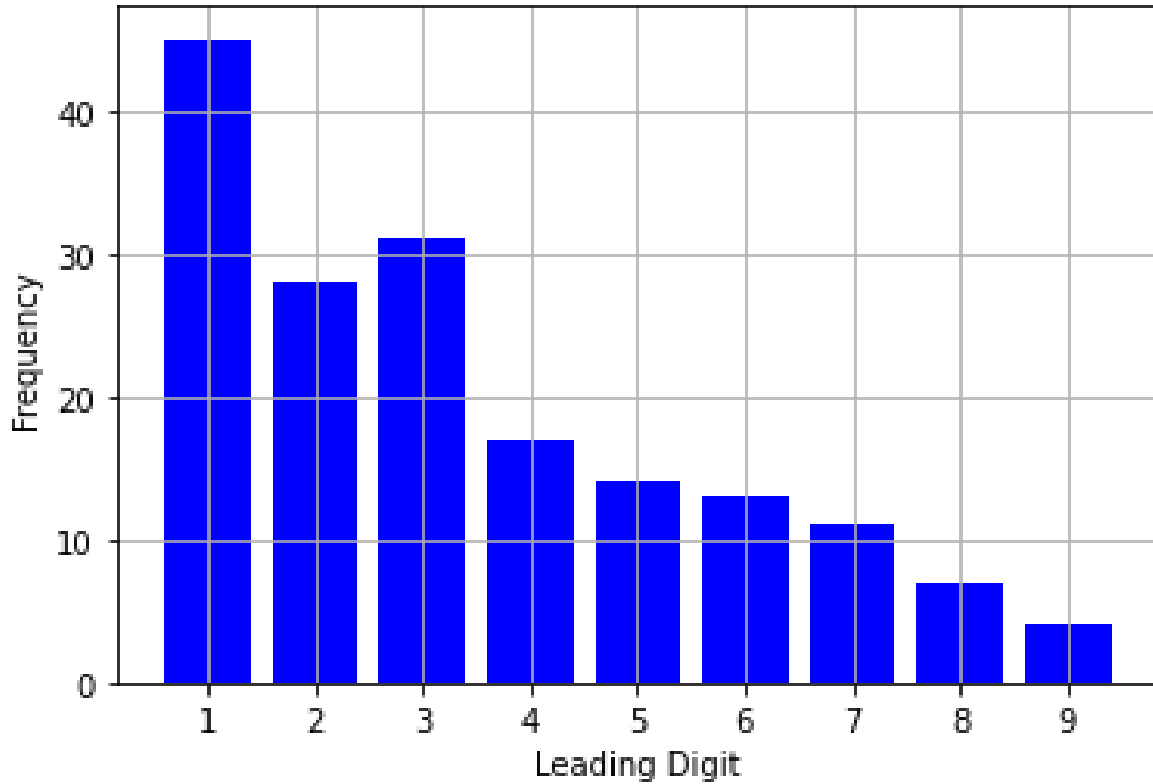


First Digit Frequency : Benford's law

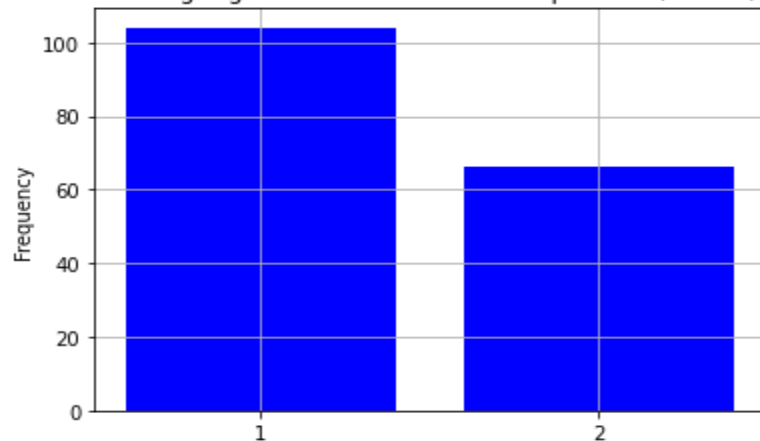


Leading Digit Distribution of Global Populations Across Number Systems

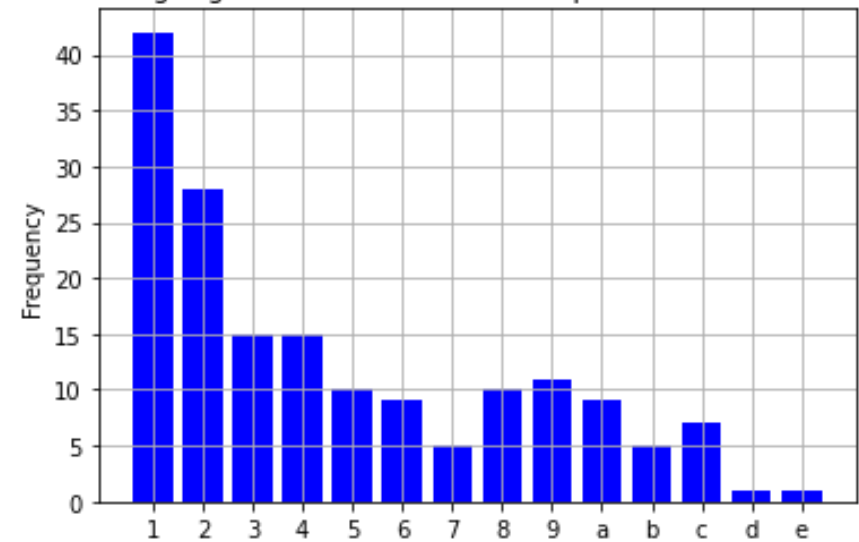
Leading Digit Distribution of World Population (Decimal)



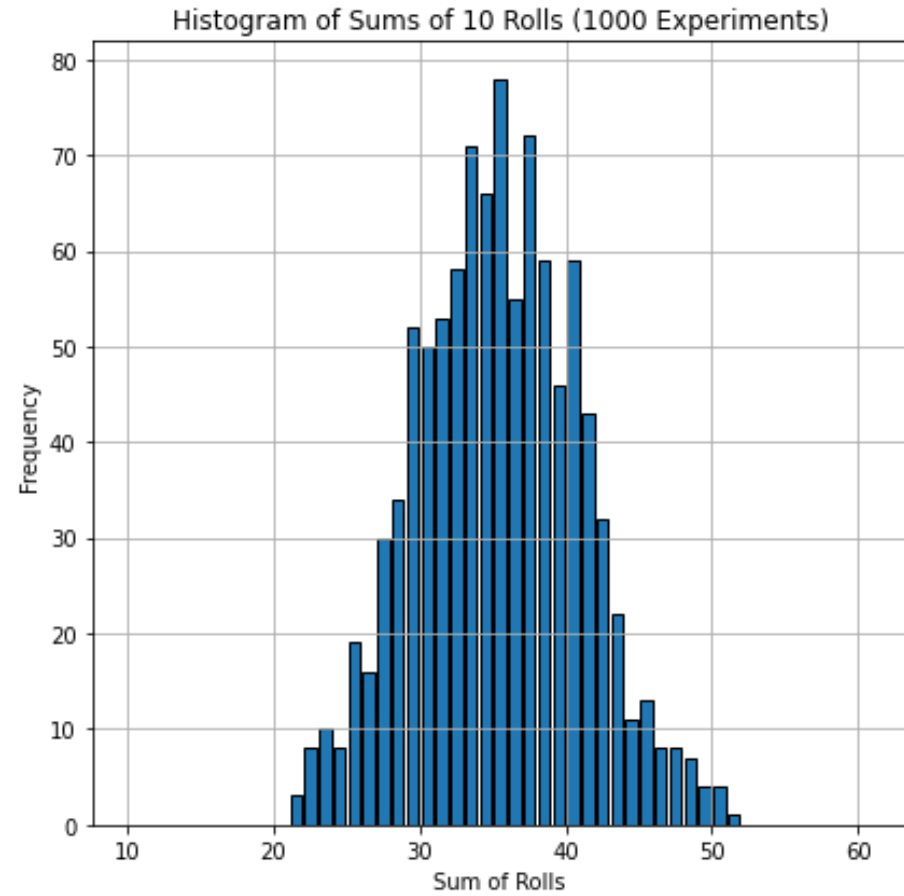
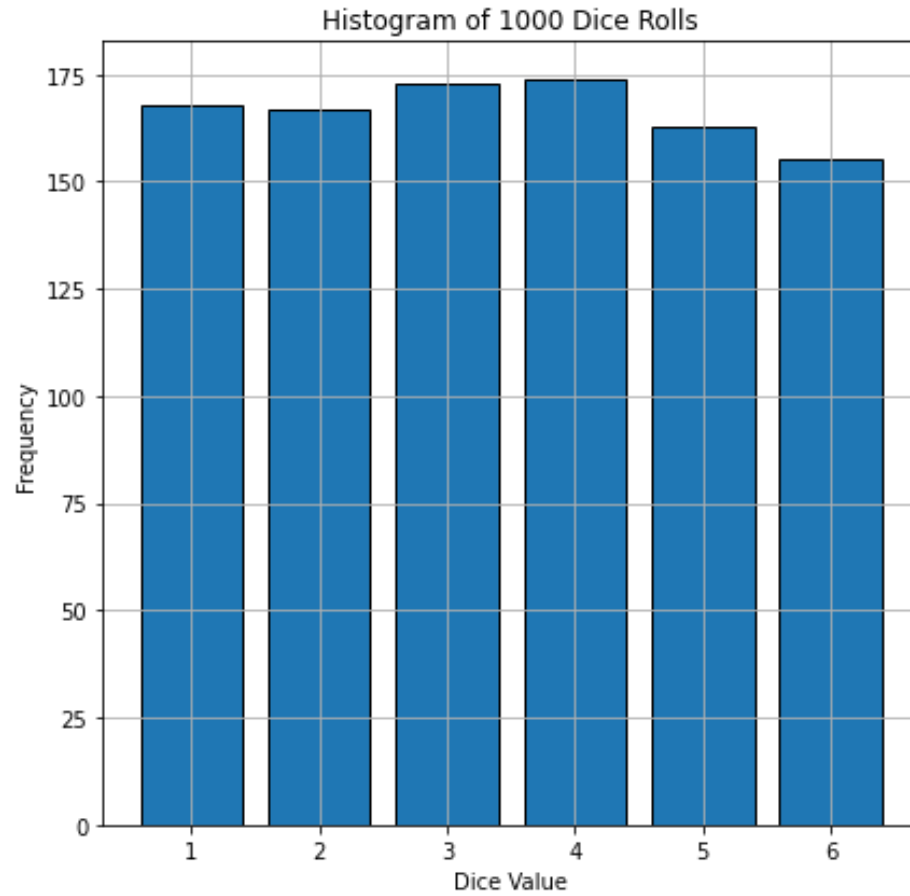
Leading Digit Distribution of World Population (Base 3)



Leading Digit Distribution of World Population (Hexadecimal)

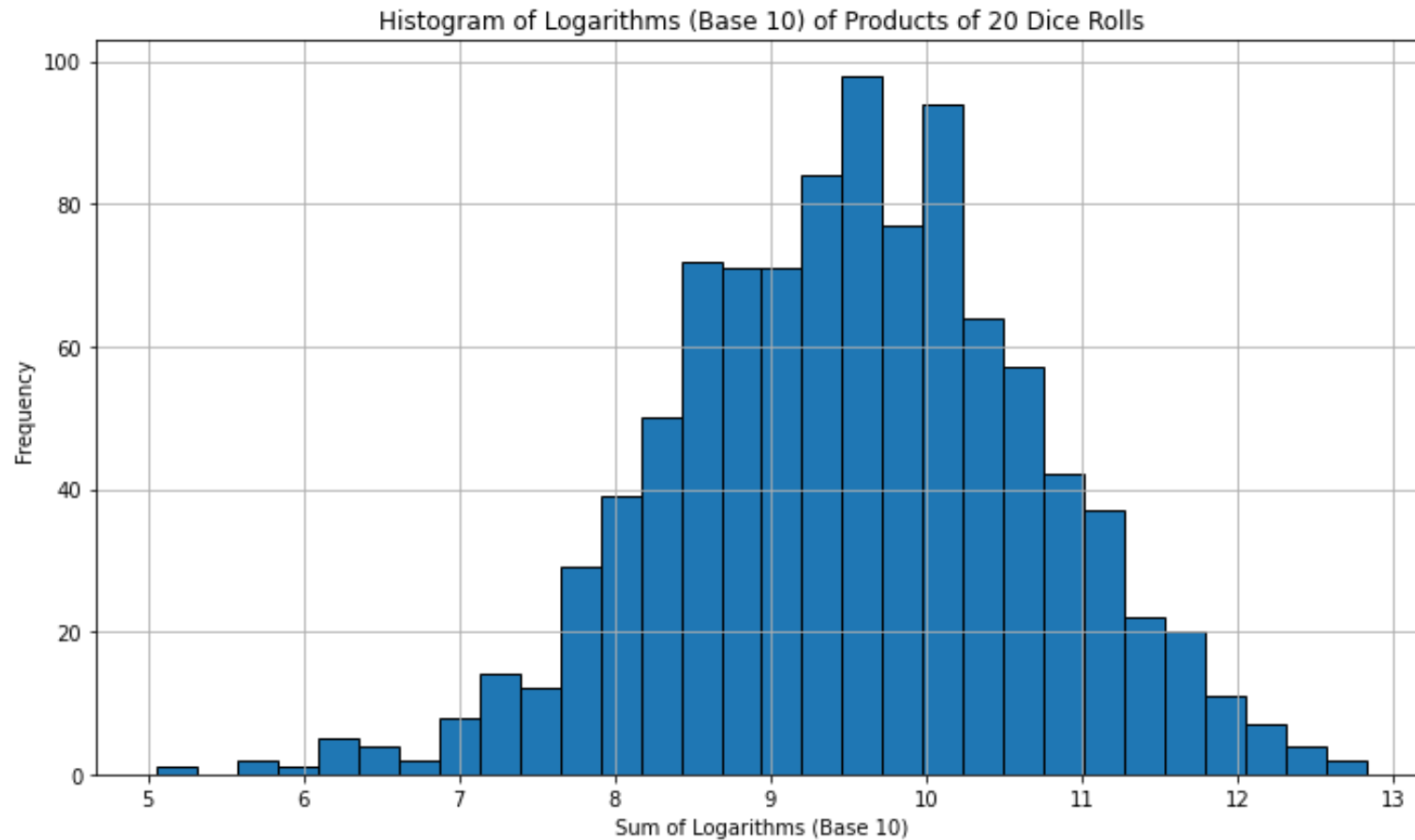


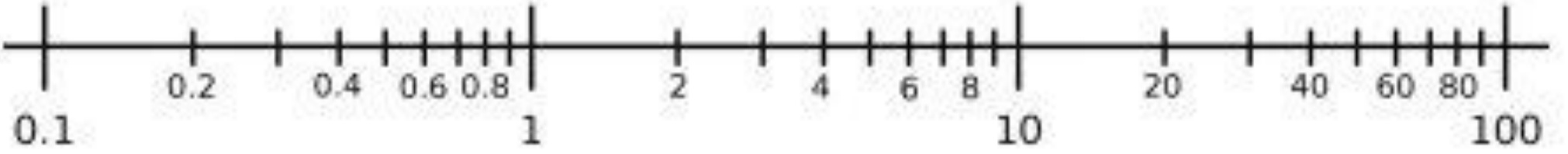
Experiment :1000 dice rolls



Sum of the outputs of 10 rolls

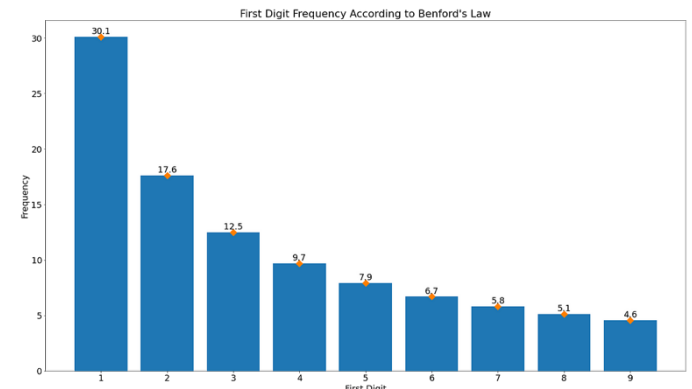
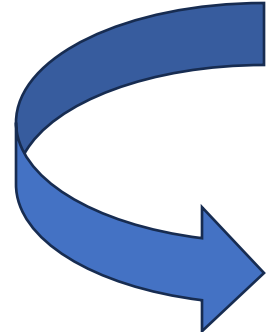
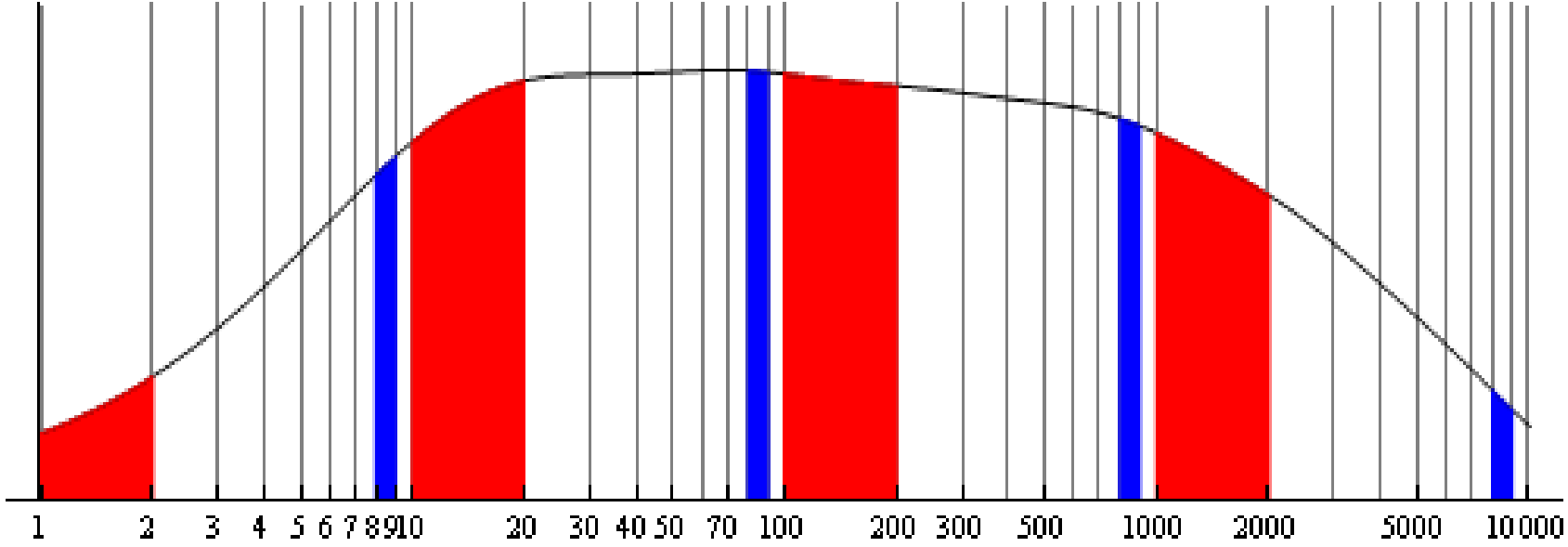
Logarithmic Distribution of Dice Roll Product





A logarithmic scale bar. Picking a random x position uniformly on this number line, roughly 30% of the time the first digit of the number will be 1.

- $10^{0.0} = 1.00000$
- $10^{0.1} = 1.25893$
- $10^{0.2} = 1.58489$
- $10^{0.3} = 1.99526$
- $10^{0.4} = 2.51189$
- $10^{0.5} = 3.16228$
- $10^{0.6} = 3.98107$
- $10^{0.7} = 5.01187$
- $10^{0.8} = 6.30957$
- $10^{0.9} = 7.94328$
- $10^{1.0} = 10.00000$
- $10^{1.1} = 12.58925$
- $10^{1.2} = 15.84893$
- $10^{1.3} = 19.95262$
- $10^{1.4} = 25.11886$
- $10^{1.5} = 31.62278$
- $10^{1.6} = 39.81072$
- $10^{1.7} = 50.11872$
- $10^{1.8} = 63.09573$
- $10^{1.9} = 79.43282$
- $10^{2.0} = 100.00000$



Lognormal distribution in Nature

Length of rivers : Average rainfall, geological features, winds etc

Population: Areas, housing unit density, migration etc.

Finance books: Profit/loss, exponential growth, taxes etc

Conclusion

- Why Benford's law ?

Due to cumulative effects of numerous multiplicative random factors, resulting in log-normal distribution

We use linear systems to describe and measure exponentially dynamic world.

Thank you.