Searching Applications of Benford’s Law to Investigate Beam Jitter

Presenter: Dominic Fallico, Geneva High School
Mentor: Dr. Arun Saini, APS-TD
TRAC
8/3/2023
Outline

• Introduction to Benford’s Law
  – Definition and Origin
  – Applications/Uses
• Application for particle accelerators
  – PIP2-IT
  – Beam jitters/noise and distribution
• Methodology
  – Generate synthetic and real distributions using Python
  – Extracting First Digits and Analyze
  – Results, Findings, & Shortcomings
• Conclusions
1. Introduction
   - Motivation
   - Benford’s Law Definition and Application
2. Methodology
3. Results
4. Conclusion
Motivation and Objective

• Explore Applications of Benford’s Law to predict jitter and noise in the beam distribution

• Apply Benford’s Law on beam measurement data of PIP2IT beamline.
Benford’s Law

Origin - Simon Newcomb and worn logarithmic tables $\rightarrow P_{10} = \log_{10} \left( 1 + \frac{1}{d} \right)$

Frank Benford redisCOVERS in 1938 and extends it to $\rightarrow P_d = \log_b \left( 1 + \frac{1}{d} \right)$

Benford tests it himself

| First Digit Distributions Expressed as Percentages for Various Physical Data Sets |
|----------------------------------------|-----------------|-----------------|
| First Digit Frequencies               | Number of Values in Each Data Set | Dynamic Range of the Data (max/min) |
| 1                                      | 2                | 3                | 4                | 5                | 6                | 7                | 8                | 9                | 10^10            | 10^3             | 10^6             | 10^9             | 10^12            |
| P_d                                    | 30.1             | 17.6             | 12.49            | 9.69             | 7.92             | 6.69             | 5.80             | 5.12             | 4.58             | 36512            | 93               | 423776           | 25917            | 401              |
| Geomagnetic Field                      | 28.9             | 17.7             | 13.3             | 9.4              | 8.1              | 6.9              | 6.1              | 5.1              | 4.5              | 1861             | 324              | 1451             | 248915           | 1.7             |
| Geomagnetic reversals                  | 32.3             | 19.4             | 13.9             | 11.8             | 5.3              | 4.3              | 3.2              | 5.4              | 4.3              | 1451             | 184              | 534              | 1527             | 326              |
| Seismic wave speeds below SW-Pacific  | 30.0             | 17.6             | 13.3             | 9.8              | 7.9              | 6.4              | 5.6              | 4.8              | 4.7              | 1451             | 184              | 534              | 1527             | 326              |
| Earth’s gravity                       | 33.0             | 16.6             | 11.2             | 8.5              | 7.5              | 6.7              | 5.9              | 5.7              | 5.0              | 1451             | 184              | 534              | 1527             | 326              |
| Exoplanet mass                        | 33.9             | 15.4             | 10.7             | 9.2              | 6.2              | 9.4              | 9.7              | 9.8              | 4.8              | 1451             | 184              | 534              | 1527             | 326              |
| Pulsars rotation freq                 | 33.9             | 20.7             | 12.7             | 7.6              | 5.3              | 5.0              | 4.9              | 4.6              | 4.8              | 1451             | 184              | 534              | 1527             | 326              |
| Fermi space telescope \(\gamma\)-ray source fluxes | 30.3             | 17.9             | 13.0             | 9.9              | 7.6              | 6.9              | 6.5              | 5.2              | 5.3              | 1451             | 184              | 534              | 1527             | 326              |
| Earthquake depths                     | 31.6             | 16.9             | 14.0             | 8.6              | 9.8              | 7.4              | 3.5              | 4.8              | 4.3              | 248915           | 10^2             | 10^3             | 10^6             | 10^9             |
| S-A seismicogram                      | 28.4             | 15.7             | 12.5             | 9.6              | 8.9              | 7.7              | 5.2              | 6.0              | 4.9              | 240000           | 10^2             | 10^3             | 10^6             | 10^9             |
| Green house gas emissions by country  | 29.9             | 17.9             | 11.4             | 7.6              | 9.2              | 8.5              | 5.9              | 4.8              | 4.8              | 184              | 10^4             | 10^3             | 10^6             | 10^9             |
| Global Temp anomalies in period 1880-2008 | 27.7             | 19.4             | 12.7             | 12.1             | 8.9              | 5.4              | 6.6              | 4.3              | 2.8              | 1527             | 10^2             | 10^3             | 10^6             | 10^9             |
| Fund. Phys. constants                 | 34.0             | 18.4             | 9.2              | 8.2              | 8.5              | 7.3              | 5.3              | 5.2              | 5.2              | 326              | 10^4             | 10^3             | 10^6             | 10^9             |
| Global Infectious disease cases       | 33.7             | 16.7             | 13.2             | 10.7             | 7.3              | 5.4              | 4.5              | 5.0              | 3.4              | 987              | 10^6             | 10^3             | 10^6             | 10^9             |
| Geometric series                      | 29.8             | 17.4             | 13.0             | 10.0             | 7.8              | 6.6              | 5.8              | 5.0              | 4.6              | 1000             | 10^21            | 10^3             | 10^6             | 10^9             |
| Fibonacci sequence                    | 30.0             | 17.7             | 12.5             | 9.6              | 8.0              | 6.7              | 5.7              | 5.3              | 4.5              | 1000             | 10^14            | 10^3             | 10^6             | 10^9             |
| Combined                               | 30.9             | 17.4             | 13.2             | 9.0              | 7.6              | 6.4              | 5.7              | 4.8              | 5.0              | 10000            | 10^23            | 10^3             | 10^6             | 10^9             |
Benford Applications

- Minimal interest except for Knuth in 1968 programming.
- More recent, prominent applications include validity testing:
  - COVID data
  - Fraud Detection & Auditing
  - Election Data
  - Statistical Physics?

- Non-Examples:
  - Height
  - IQ scores
  - Car prices
  - Phone Numbers
Benford Criteria

Something conforms to Benford’s Law if:

- \( P_d \approx \log_{10} \left( 1 + \frac{1}{d} \right) \)
- Orders of magnitude
- Scale invariance
- Data describes the same object
- No artificial max/min imposed on data
- Number cannot be assigned (e.g. phone numbers)
1. Introduction

2. Methodology
   ➢ Python Programming
   ➢ Validation of Benford’s law on synthetic distribution

3. Results

4. Conclusion
from matplotlib import pyplot as plt
import numpy as np
import math

def benford(d):
    return np.log10(1+1/d)
d_values=np.linspace(.1, 10, 100)
benford_prob=benford(d_values)

numbers = [float(n) for n in range(1, 10)]
benford = [math.log10(1 + 1 / d) for d in numbers]

digits = list(range(1,10))

#Beam Position Data X
beam_data_X = np.loadtxt('ALLBPMX.txt', unpack = False)
plt.style.use('bmh')
plt.hist(beam_data_X, bins = 200,color = 'teal',hatch = '')
plt.xlabel('Horizontal Beam Centroid (\u03BCm)')
plt.ylabel('Pulse Counts')
plt.title('All BPMs')
plt.show()

#Beam Position Data Y
beam_data_Y = np.loadtxt('ALLBPMY.txt', unpack = False)
plt.style.use('bmh')
plt.hist(beam_data_Y, bins = 200,color = 'salmon',hatch = '')
plt.xlabel('Vertical Beam Centroid (\u03BCm)')
plt.ylabel('Pulse Counts')
plt.title('ALL BPMs')
plt.show()

#First Digit Counts X
digit_counts_X = [0]*9
with open("ALLBPMX 1ST DIGIT.txt","r") as file:
    lines=file.read().splitlines()
    for line in lines:
        if line.isdigit() and int(line) in range (1,10):
            digit_counts_X[int(line)-1]+=1
print(digit_counts_X)

#First Digit Counts Y
digit_counts_Y = [0]*9
with open("ALLBPMY 1ST DIGIT.txt","r") as file:
    lines=file.read().splitlines()
    for line in lines:
        if line.isdigit() and int(line) in range (1,10):
            digit_counts_Y[int(line)-1]+=1
print(digit_counts_Y)

#Digit Proportions X
total_count = sum(digit_counts_X)
digit_proportions_X = [count/total_count for count in digit_counts_X]

#Digit Proportions Y
digit_proportions_Y=[count/total_count for count in digit_counts_Y]

#Benford Deviation
actual_X = (digit_proportions_X)
predicted = (benford)
array1 = np.array(actual_X)
array2 = np.array(predicted)
residuals_X = np.subtract(array1,array2)

summand_X = (residuals_X/predicted)**2
summation_X = sum(summand_X)
Benford Deviation_X = np.sqrt(summation_X)
print("Benford Deviation \u03b4= " + str(Benford Deviation_X))
Calculating Benford Deviation: Methods

Benford Deviation

\[ \delta = \sqrt{\sum_{d=1}^{9} \left[ \frac{F(d) - P(d)}{P(d)} \right]^2} \]

\( P(d) = \) ideal proportion
\( F(d) = \) observed proportion

\( \delta = 0 - 0.03 \) considered “good”
\( \delta = 0.6 - 0.7 \) shows “preference to small digits”

**Chi-Square

\[ \chi^2 = \sum_{1}^{n} \frac{(O_n - E_n)^2}{E_n} \]

\( O_n = \) observed frequency
\( E_n = \) expected frequency

**Not best metric for Benford’s Law

Mean Absolute Deviation

\[ MAD = \frac{1}{9} \cdot \sum_{i=1}^{9} |p_i - b_i| \]

\( p_i = \) observed proportions
\( b_i = \) expected proportions

**Not best metric for Benford’s Law

<table>
<thead>
<tr>
<th>MAD</th>
<th>Conformity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.006</td>
<td>Close</td>
</tr>
<tr>
<td>0.006 - 0.012</td>
<td>Acceptable</td>
</tr>
<tr>
<td>0.012 - 0.015</td>
<td>Marginal</td>
</tr>
<tr>
<td>&gt; 0.015</td>
<td>Non-conformity</td>
</tr>
</tbody>
</table>
Popular Number Sets Raw Data

First 100 Powers of 2

First 100 Factorials

First 100 Fibonacci Numbers
Popular Number Sets Against Benford

First 100 Powers of 2 Deviation from Benford

First 100 Factorials Deviation from Benford

First 100 Fibonacci Numbers Deviation from Benford

Dominic Fallico | Searching Applications of Benford’s Law to Investigate Beam Jitter

8/3/2023

Fermilab
Popular Number Sets Against Benford (again)

First 10000 Powers of 2

Deviation from Benford

Ideal Benford Probabilities
\( p_i = \frac{1}{\log_{10}(i+1)} \)

\( \chi^2 = 0.004539901863641374 \)

MAD = 0.002625587725811167

First 10000 Factorials

Deviation from Benford

Ideal Benford Probabilities
\( p_i = \frac{1}{i+1} \)

\( \chi^2 = 7.4560821458781999 \)

MAD = 0.00227599930350623996

First 10000 Fibonacci Numbers

Deviation from Benford

Ideal Benford Probabilities
\( p_i = \frac{1}{\log_{10}(i+1)} \)

\( \chi^2 = 0.0153302456786505566 \)

MAD = 0.005196744580451122
1. Introduction
   ➢ Motivation
   ➢ Benford’s Law Definition and Application
2. Methodology
3. Results
   ➢ Beam Measurement Data for PIP2IT beamline
4. Conclusion
**PIP2-IT**

- PIP2-IT Goals – address risks for PIP-II linac, test front-end of PIP-II
- Facility was completed & successfully commissioned in 2021

- Starts with ion source, turns Hydrogen gas → beam of $\text{H}^-$ ions
- RFQ accelerates beam to MEBT where beam is manipulated.
- Two super conducting cryomodules accelerate beam up to 25 MeV.
- Remaining are beam measurement, diagnostics, and dump station
• Beam Horizontal and Vertical positions are measured using Beam Position Monitor (BPMs)
  • BPMs are located at center of each Quadrupole Doublet and triplet.
• BPM data were used to quantify eam position jitters.
Jitter and Noise

The variation in timing or arrival of particles or beams at specific points within the accelerator system. Arises from accelerator defects, external disturbances, fluctuations in the power supply.

Affect efficiency of accelerator performance, thus affecting measurement accuracy and experimental outcomes.

If we want to take more data, it has to be accurate!
Beam Position Data

- Beam Position measured almost 12,000 times in all nine BPMs.
- Position varies in both planes
- Origin of jitters unknown, methods to resolve unsuccessful.
- Ultimately impacts were negligible.
Testing Benford on Normal Distributions

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>3000</td>
<td>30000</td>
</tr>
<tr>
<td>$N$</td>
<td>11000</td>
<td>11000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$\approx 0.739$</td>
<td>$\approx 0.743$</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>$\approx 519.258$</td>
<td>$\approx 539.031$</td>
</tr>
<tr>
<td>$MAD$</td>
<td>0.0222</td>
<td>0.0222</td>
</tr>
</tbody>
</table>

Deviation from Benford

- $\delta = 0.7386713840900823$
- $\chi^2 = 519.2582137695816$
- $MAD = 0.022245457119922117$
- $\delta = 0.7429269553810672$
- $\chi^2 = 539.030672885993$
- $MAD = 0.022296286829974008$
PIP2-IT BPM Data: First two BPMs

Deviation from Benford

- **Ideal Benford Probabilities**
  - $\delta = 1.1505079526644768$
  - $X^2 = 2279.691806071065$
  - MAD = 0.044681457934664194
  - $\gamma = 5.467344697897703$
  - $X^2 = 56118.40980568727$
  - MAD = 0.18309083132095974

- **Ideal Benford Probabilities**
  - $\delta = 0.605468506691364$
  - $X^2 = 383.9688404461003$
  - MAD = 0.017621913475026042
  - $\gamma = 1.0586922268307528$
  - $X^2 = 1157.2012006093087$
  - MAD = 0.02862741622640174
**PIP2-IT BPM Data: All BPMs**

<table>
<thead>
<tr>
<th></th>
<th>Horizontal</th>
<th>Vertical</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>22.5</td>
<td>-252.52</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>130.92</td>
<td>890.35</td>
</tr>
<tr>
<td>( N )</td>
<td>107946</td>
<td>107946</td>
</tr>
<tr>
<td>( \delta )</td>
<td>( \approx 0.825 )</td>
<td>( \approx 1.285 )</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>( \approx 8887.41 )</td>
<td>( \approx 22383.23 )</td>
</tr>
<tr>
<td>( MAD )</td>
<td>0.0289</td>
<td>0.0372</td>
</tr>
</tbody>
</table>

**Deviation from Benford**

- Ideal Benford Probabilities
  - \( G = 0.8252781854043079 \)
  - \( X^2 = 8887.405487406957 \)
  - \( MAD = 0.028939018813869233 \)

- Observed
  - \( G = 1.2851102351258459 \)
  - \( X^2 = 22383.230817971034 \)
  - \( MAD = 0.03719764427081064 \)
Using MAD, for all nine BPMs, 0/18 distributions exhibited even marginal Benford conformity.

Using $\delta$, about 4/18 showed “preference to leading digits” but no conformity.

Poor conformity – why?
Conclusions

Noise → Randomness → Benford Deviation

Experimental noise is persistent, even at low temps

Ambiguity in metrics

Benford’s law is somewhat trivial and postdictive, not predictive
Where to go from here?

Other test methods
   – Kolmogorov-Smirnov Test yielded similar results for

Examine other details:
   – Outliers
   – Skewness
   – Orders of magnitude

Analyze additional beam data
Take Home

Immersed in the Fermilab culture
   – Mentored by and interact with top scientists

Cross-disciplinary

Math, Stats, Python, Particle Accelerators, & Beam Dynamics

Every teacher’s favorite question
Acknowledgements

This work was done under the Teacher Research Associate (TRAC) Program at Fermilab.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Thank you to Harry Cheung, Don Lincoln, and the other TRAC teachers for a great experience, and a special thank you to my mentor Arun Saini for his guidance this summer.
HWR CM & SSR1 CM are first two in PIP-II
Niobium insulated cavities, cooled to 2° K
RF doubles from HWR to SSR1
More Pictures

Top Right – PIP2IT Facility
Top Left – RFQ at PIP2IT
Bottom Left – PIP2IT Chopper
Bottom Right – SSR1 at PIP2IT
References


