Benford's Law of Digit Bias

Problem: Given a sequence of data or some random variables, determine the distribution of the first (or leading) digits.

↳ data $\{a_n\}_{n=1}^{\infty}$ calculate (if it exists!)

$$\lim_{N \to \infty} \frac{\#\{n \leq N : \text{first digit of } a_n \text{ is } d\}}{N}$$

Applications: ① leads to some good, interesting math questions

↳ ex: quantified equidistribution in $\{n\alpha \bmod 1\}$

② data integrity / fraud detection (IRS story)

Needed Input: Fourier Analysis + Number Theory

Amount needed depends on application.

Number Theory: Equidistribution of $n\alpha$

Fourier Analysis:
- Fejér's Thm and a generalization:

  ↳ Lebesgue's Thm: $f \in L^1 \to T_N f$ converges in L^1 to f

  $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx$
- Poisson summation:

  Then f nice $\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$

  ↳ Why useful? Can convert a slowly converging sum to a rapidly converging one, and often small error in replacing with $\hat{f}(0)$

BENFORD'S LAW

Newcomb (1880s), Benford (1930s)

Many processes: $\text{Prob}(1^{st} \text{ digit } d) = \log_{10}\left(\frac{d+1}{d}\right)$

↳ check prob distr by adding and see sums to 1

More generally: $x > 0$ write $x = M_{10}(x) \cdot 10^k$ $k \in \mathbb{Z}$

Then $\text{Prob}(M_{10}(x) \leq s) = \log_{10} s$

Not all seqs are Benford: some don't even have a limit!

↳ ex: 1, 2, 3, ... : first dig freq oscillates b/w 1/9 and 5/9

ASIDE: Analytic Density can help for questions like this

Ex: First digit of primes (Serre's comment about Bombieri): $\text{Density}_{\text{analytic}}(A) = \lim_{s \to 1} \frac{\sum_{n \in A} 1/n^s}{\sum_{n=1}^{\infty} 1/n^s}$

Examples: Fibonacci, Recurrence relations, L-fns and RMT, financial data (IRS!), hydrology data ...

Also $3x+1$ (Tell IRS Story, writing @ Alex K..., Hawaii)

↳ Erdös + Kakutani quotes

$$a_{n+1} = \begin{cases} 3a_n + 1 & a_n \text{ odd} \\ a_n/2 & a_n \text{ even} \end{cases} \quad \text{or } a_{n+1} = \frac{3a_n + 1}{2^k}$$

where $2^k \| 3a_n + 1$

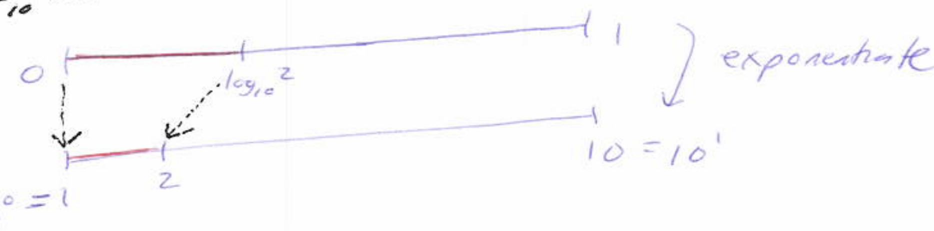
RECURRENCE RELATIONS + BENFORD'S LAW

Why is $\alpha \notin \mathbb{Q} \Rightarrow n\alpha \bmod 1$ so useful?

- 10^u and 10^v same first digits iff $u \equiv v \bmod 1$
  - Thus instead of studying seq $\{X_n\}_{n=1}^{\infty}$, study $\{Y_n\}_{n=1}^{\infty}$
    with $Y_n = \log_{10} X_n \bmod 1$
    - Note first digit of X_n is d iff $M_{10}(X_n) \in [d, d+1)$
      ie, iff $Y_n \in [\log_{10} d, \log_{10}(d+1))$

<u>THM (9.2.4)</u>: If $\log_{10} X_n \bmod 1$ is equidistr then X_n is Benford

Proof: Calculation:



<u>Example</u>: $\log_{10} r \notin \mathbb{Q} \Rightarrow X_n = ar^n$ is Benford

Proof: $\log_{10} X_n = \underbrace{n \log_{10} r}_{\text{equidistr}} + \underbrace{\log_{10} a}_{\text{constant}}$

<u>Example</u>: Fibonacci numbers are Benford

- "Many" recurrence relations Benford (Thm 9.3.1)

  $a_{n+k} = c_1 a_{n+k-1} + \dots + c_k a_n$ with $c_1, \dots, c_k$ fixed constants

  (ex: $a_{n+1} = a_n + a_{n-1}$: Fibonacci)
  - solve by divine inspiration: $a_n = u_1 \lambda_1^n + \dots + u_k \lambda_k^n$
    - Binet's formula: $n-1^{st}$ Fibonacci is $\frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right)$
      note n large main term dominates: book-keeping

<u>ASIDE</u>: Theory of difference/differential eqs

ASIDE

① Values of L-fns, RMT
② 3X+1 : need {quantified equidistribution, irrationality exponent} } Poisson Sum

③ Mod 1 CLT : need Fejér/Lebesgue

↳ Many noticed amalgamated data more Benford than constituents

Analysis of products of random variables involves an enormous amount of prob and Fourier Analysis

④ Order Statistics

Other asides

- Proof of quantified equidistribution: Erdös-Turan, irrationality exponent plays huge role.

- 3X+1 problem
  ↳ what is known...
  heuristics: "$\mathbb{E}[\log a_{n+1}]$" = "$\sum_{k=1}^{\infty} \frac{1}{2^k} \log\left(\frac{3a_n}{2^k}\right)$"
  ↳ This involves differentiating an identity as must evaluate $\sum_{k=1}^{\infty} \frac{k}{2^k}$