BENFORD'S LAW LECTURE

• HISTORY

- Newcomb: 1880s: log-tables
- Benford: 1930s: lots of data, physics paper next to it

• DEFINITIONS

- $X > 0$, write $X = M_{10}(x) \cdot 10^{k(x)}$ where $k(x)$ is an integer and $1 \le M_{10}(x) < 10$.
  Call $M_{10}(x)$ the mantissa (others say significand)
  See Scientific notation

- Data set satisfies Benford's Law if

$$\text{Prob}(\text{first digit base 10 is } d) = \log_{10}\left(\frac{d+1}{d}\right)$$

↳ similar result other bases

- Scale invariance of data sets
  ↳ look at powers of 2 - issues!

- More generally, strongly Benford if

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

BENFORD'S LAW LECTURE

QUESTION: What systems satisfy Benford?

Financial transactions (IRS)

Mathematical data (often)

Fibonacci numbers

3X+1 iterates

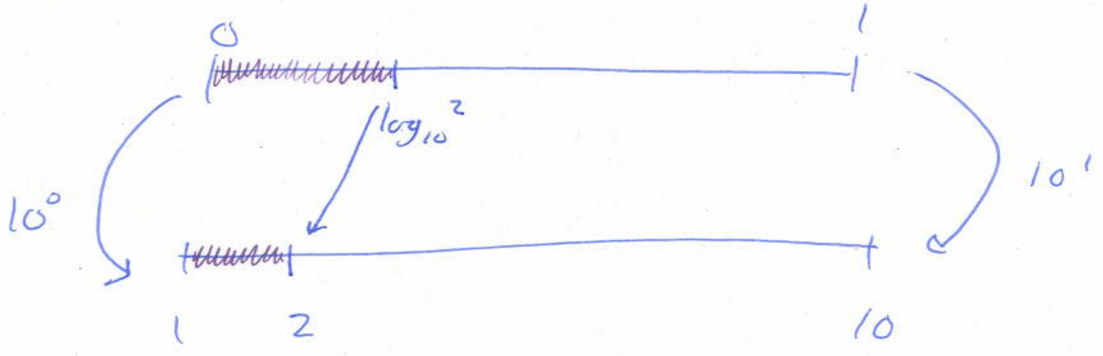
Amalgamations

NEW VIEWPOINT

↳ Clearly don't care about $k(x)$ if studying leading digits

$$Y_n = \log_{10} X_n \bmod 1 = \log_{10} M_{10}(X)$$

$\boxed{\text{\underline{Fundamental Equivalence}: } \{Y_n\} \text{ is equidistributed (uniformly distributed) if and only if } X_n \text{ is strongly Benford}}$



What does uniformly distributed mean?

$$\lim_{N \to \infty} \frac{\#\{n \le N : a \le Y_n \le b\}}{N} \to b-a$$

for any $[a,b] \subset [0,1]$.

BENFORD'S LAW LECTURE

Reduced Benford to uniformly distributed logarithm of mantissa: how do we show this?

$$\boxed{\text{Kronecker's Thm: If } \alpha \text{ is irrational then } y_n = n\alpha \bmod 1 \text{ is equidistributed.}}$$

Proof involves lot of nice adv math (Fourier Series) sines/cosines love mod 1 constraints

Clear need α irrational: $\alpha = 1/5$ and cycle: $\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 0, \frac{1}{5}, \ldots$

Beyond scope of class to prove, but can at least show

<u>dense</u>: given any $x \in [0,1]$ and any small $\varepsilon > 0$ can find an n st $n\alpha \bmod 1$ is within ε of x.

Uses <u>Dirichlet's Box Principle</u>:

↳ Pidgeon gun, fire $n+1$ pidgeons into n boxes, then at least one box gets two pidgeons.

<u>Proof of Claim</u>

Choose Q st $1/Q < \varepsilon/2$ (wiggle room!)

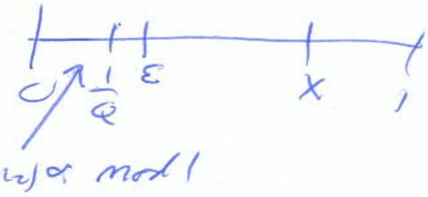
look at Q boxes $[0, 1/Q), [1/Q, 2/Q), \ldots, [\frac{Q-1}{Q}, 1)$

Consider $Q+1$ numbers $\alpha \bmod 1$, $2\alpha \bmod 1, \ldots, (Q+1)\alpha \bmod 1$

At least two in same box, say $n_1\alpha \bmod 1$, $n_2\alpha \bmod 1$

Get $|n_1\alpha \bmod 1 - n_2\alpha \bmod 1| < 1/Q$ as in same box

So $|(n_1 - n_2)\alpha \bmod 1| < 1/Q$, so



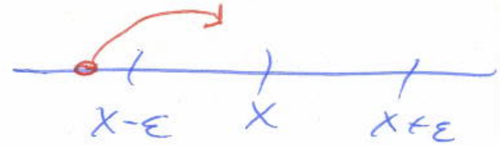
BENFORD'S LAW: CONTINUED

Proof of Claim (cont)

Now just keep walking down

Move at most $1/Q < \varepsilon$ every $n_1 - n_2$ steps

Must land within ε of x



↳ to miss requires step size $> 2\varepsilon$

Notes

- Pidgeon hole appears on many math competitions
- See in closure in group theory
- Applications to estimating irrationals with rationals

↳ had $|(n_1 - n_2)\alpha \bmod 1| < 1/Q$

so is an integer p such that

$$|(n_1 - n_2)\alpha - p| < 1/Q$$

Let $q = n_1 - n_2$; note $1 \le q < Q$. Then

$$|\alpha - p/q| < 1/qQ < 1/q^2$$

↳ little work: infinitely often rational within $1/q^2$

↳ measure cost by size of denom

↳ much better than expect: $\frac{314159}{100000}$ "expensive"

↳ "worst" number is golden mean, $\frac{1+\sqrt{5}}{2}$;
hardest to approx with rationals!

BENFORD LECTURE CONT

EXAMPLES OF BENFORD

- Long Street?

No, limit doesn't exist. Stop at $2 \cdot 10^k$ just saw a lot of 1s, stop at $10^{k+1}-1$ just had a long block without 1s. Percentage oscillates b/w $1/9$ and $5/9$.

- Powers of 2?

Yes, though strange as no leading 9s in first 30

$X_n = 2^n$, $Y_n = \log_{10}(2^n) \bmod 1 = n \log_{10} 2 \bmod 1$

↳ Just need to show $\log_{10} 2$ irrational

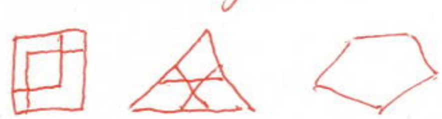
↳ assume not: $\log_{10} 2 = p/q$ in lowest terms

$$\Rightarrow 2 = 10^{p/q}$$
$$\Rightarrow 2^q = 10^p = 2^p 5^p$$
$$\Rightarrow 2^{q-p} = 5^p \Rightarrow q-p = p = 0$$

↳ Why so few lines? 1 Gb = 1024 Mb: $2^{10} = 1024 \approx 10^3$

Almost periodic

IRS cares about rates of convergence

Nice geometric irrationality proofs: 

• EXAMPLE: FIBONACCI NUMBERS

$F_{n+2} = F_{n+1} + F_n$, $F_0 = 0$, $F_1 = 1$, ...

Seq now determined, but what is F_n?

Remarkably a closed form expression exists!

METHOD OF DIVINE INSPIRATION

Guess $F_n = r^n$ and see what happens

↳ $r^{n+2} = r^{n+1} + r^n$

$r^{n+2} - r^{n+1} - r^n = 0$

$r^n(r^2 - r - 1) = 0$; $r^n = 0$ gives trivial soln

$\underbrace{r^2 - r - 1 = 0}_{\text{characteristic polynomial}} \rightarrow r = \frac{1 \pm \sqrt{5}}{2} = r_1, r_2$

↳ see r^n solves recurrence for either value

↳ little algebra: $a r_1^n + b r_2^n$ also a soln for any a, b:

↳ $a r_1^{n+2} + b r_2^{n+2} = a(r_1^{n+1} + r_1^n) + b(r_2^{n+1} + r_2^n)$

$= (a r_1^{n+1} + b b_2^{n+1}) + (a r_1^n + b r_2^n)$

↳ What are a, b? Linear algebra!

$$\begin{aligned} a r_1^0 + b r_2^0 &= 0 \\ a r_1^1 + b r_2^1 &= 1 \end{aligned} \Rightarrow \begin{pmatrix} 1 & 1 \\ r_1 & r_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

BENFORD LECTURE: CONT

Fibonacci Numbers (Cont)

Many ways to solve: linear algebra and matrix inversion or substitution.

$$a + b = 0 \quad \Longrightarrow \quad b = -a$$

$$\frac{1+\sqrt{5}}{2}a + \frac{1-\sqrt{5}}{2}b = 1 \quad \Longrightarrow \quad \left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right)a = 1 \Longrightarrow a = \frac{1}{\sqrt{5}}$$

General Soln:

$$\boxed{\text{BINET'S FORMULA: } F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n}$$

Absolutely amazing

↳ This is ALWAYS an integer, but has division by 2 and has irrational $\sqrt{5}$'s....

↳ Can obtain by generating fns (bring it on, partial fractions, geo series).

Claim: Fibonacci numbers are Benford!

If n is big, as $\frac{1+\sqrt{5}}{2} > 1$ and $\left|\frac{1-\sqrt{5}}{2}\right| < 1$, have

$F_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$ as other term very small.

(Need some technical book-keeping in following arguments, which will ignore).

Benford's Law: Cont

Fibonacci Numbers are Benford

$$X_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

$$Y_n = \log_{10} X_n \overset{\text{mod } 1}{=} \log_{10}\left[\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n\right] \text{ mod } 1$$

$$= \underbrace{n \log_{10}\left(\frac{1+\sqrt{5}}{2}\right)} \underbrace{- \frac{1}{2}\log_{10} 5}_{\text{This is indep of } n} \text{ mod } 1$$

→ if this part is equidistributed, so too is this $-\frac{1}{2}\log_{10} 5$.

Just need to show $\log_{10} \frac{1+\sqrt{5}}{2}$ is irrational.

If not: $\log_{10} \frac{1+\sqrt{5}}{2} = p/q$ in lowest terms

$$\frac{1+\sqrt{5}}{2} = 10^{p/q}$$

$$1+\sqrt{5} = 2 \cdot 10^{p/q}$$

$$(1+\sqrt{5})^q = 2^q \cdot 10^p = \text{integer}$$

→ expand, get integer + integer $\sqrt{5}$ = integer

→ impossible!

> Geometric Growth is Benford! Financial data: says goes up 2% per year... Takes more time to go from \$1 to \$2 than from \$9 to \$10.

BENFORD'S LAW: CONTINUED

Difference Equations

Relation $F_{n+2} = F_{n+1} + F_n$ just one example of a linear difference eq. Can solve with "divine inspiration" (though sometimes need to tweak method if characteristic polynomial has a repeated root). Could use generating functions.

Application: Double Plus 1 in Roulette

Assume no greens, 50% chance red, 50% chance black.

Bet \$1 red: if red up \$1 else down \$1

↳ if lose bet \$2 red: if red up \$1 else down \$3 = \$1 + \$2

↳ if lose bet \$4 red: if red up \$1 else down \$7...

Eventually get red and net \$1: lather, rinse, repeat, retire!

Okay, something must go wrong!

↳ Rich Uncle: always willing to bankroll ∞ amount of money, but won't give you money

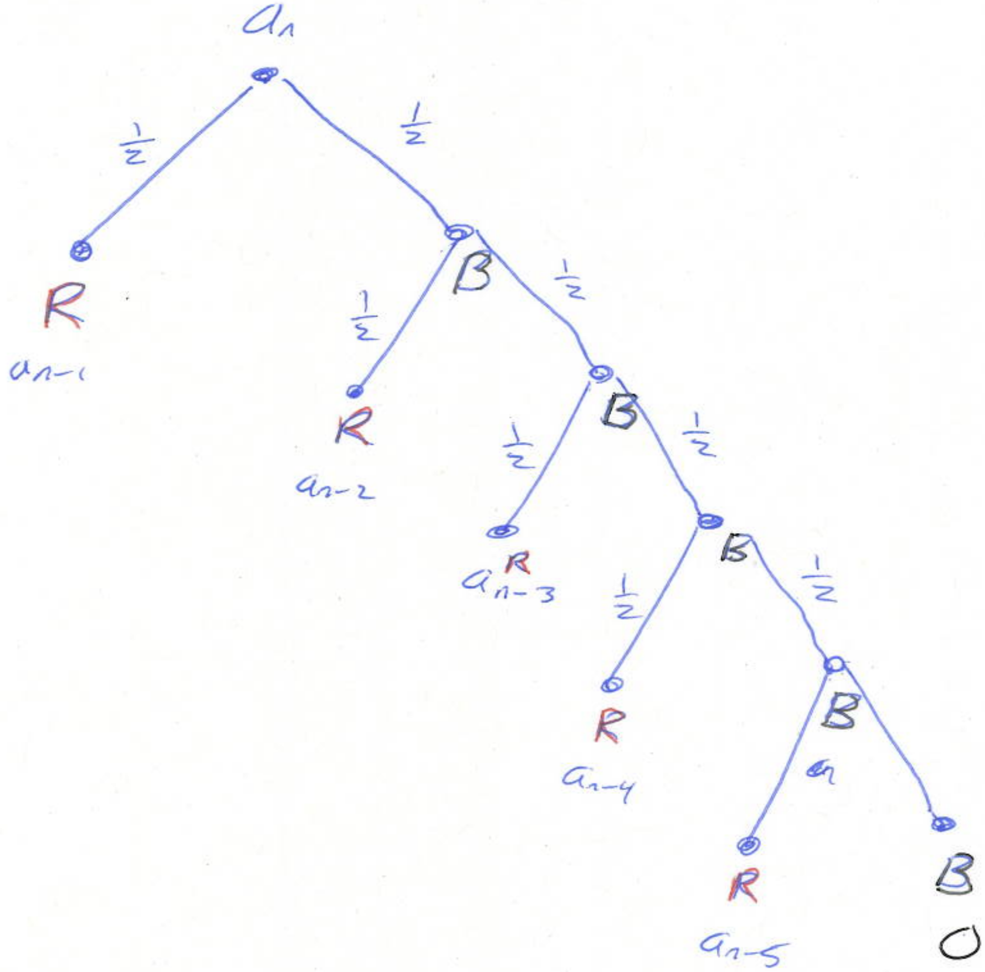
↳ House Limits: Cannot keep doubling: hit table max.

Benford's Law: Cont

Double Plus One

Let a_n = prob no 5 consec blacks in n spins

b_n = prob get 5 consec blacks in n spins $= 1 - a_n$



$$a_n = \frac{1}{2} a_{n-1} + \frac{1}{4} a_{n-2} + \frac{1}{8} a_{n-3} + \frac{1}{16} a_{n-4} + \frac{1}{32} a_{n-5}$$

Guess $a_n = r^n$ and get a quintic!

No nice soln, numerically approximate

roots $\approx .983,\ .098 \pm .424i,\ .339 \pm .229i$

$a_n \approx 1.05 * .983^n + \cdots$, $b_n \approx 1 - 1.05 * .983^n + \cdots$

$b_{100} \approx 81\%$ $b_{200} \approx 96.5\%$ $b_{300} \approx 99.4\%$

BENFORD: CONT

Can approximate: Split 100 into 20 blocks of 5

Prob has 5 blacks in 5: $\frac{1}{32}$

Prob doesn't have 5 blacks: $1-\frac{1}{32} = 1-\frac{1}{2^5}$

Prob none of the 20 blocks has 5 blacks: $\left(1-\frac{1}{2^5}\right)^{20}$

Prob don't have 5 consec blacks anywhere $\leq \left(1-\frac{1}{2^5}\right)^{20}$

↳ could have 5 cutting across two blocks

Prob have 5 consec blacks somewhere $\geq 1-\left(1-\frac{1}{2^5}\right)^{20}$

↳ using Prob(A happens) + Prob(A doesn't happen) = 1

Get Prob have 5 consec black in 100 is at least 47%

↳ Can get better lower bound by blocks of 6, ...

Can do block of length 10 without too much trouble

BBBBB *****	2^5 possibles	all distinct
RBBBBB ****	2^4 possibles	Get 96 possibles out of
*RBBBBB ***	2^4 possibles	$2^{10} = 1024$
**RBBBBB **	2^4 possibles	
***RBBBBB *	2^4 possibles	
****RBBBBB	2^4 possibles	

lower bound is $\geq 1-\left(1-\frac{63}{1024}\right)^{10} \approx 62.6\%$

IMPORTANT SKILL: GETTING SOME KIND OF ESTIMATE