BENFORD'S LAW, VALUES OF L-FUNCTIONS AND THE $3x+1$ PROBLEM

STEVEN J. MILLER

Abstract. Below are slides for talks at Boston College (10/19/04), the University of Michigan (11/15/04) and the University of Arizona (1/11/06). Many systems exhibit a digit bias. For example, the first digit (base 10) of the Fibonacci numbers or 2^n equals 1 about 30% of the time. This phenomena was first noticed by observing which pages of log tables were most worn with age – it's a good thing there were no calculators 100 years ago! We show that the first digit of values of L-functions near the critical line also exhibit this bias. A similar bias exists (in a certain sense) for the first digit of terms in the $3x+1$ problem, provided the base is not a power of two. For L-functions the main tool is the Log-Normal law; for $3x+1$ it is the rate of equidistribution of $n \log_B 2$ mod 1 and understanding the irrationality measure of $\log_B 2$. This work is joint with Alex Kontorovich.

Date: January 12, 2006.

2000 *Mathematics Subject Classification.* (primary), (secondary).

Key words and phrases. Benford's Law, Poisson Summation, L-Function, $3x$+1, Equidistribution, Irrationality Type.

BENFORD'S LAW, VALUES OF L-FNS AND 3X+1

Following the [illegible]

(1) HISTORY

- Newcomb: 1881 $\quad \text{Prob}(d) = \log_B\left(\frac{d+1}{d}\right)$
- Benford: 1938
- Long Street, invariant under rescaling: define PROB Mantissa

(2) LOGS + BENFORD

$X_n, \quad Y_n = \log_B X_n$

Y_n ED mod 1 $\Longleftrightarrow$ X_n Benford (B)

Ex: • $X_n = \alpha^n, \quad \log_B \alpha \notin \mathbb{Q}$

• Diff Eq (Fibonacci)

↳ Cauchy: $a_n = 2a_{n-1} - a_{n-2}$

$a_0 = a_1 = 1$

$a_0 = 0 \quad a_1 = 1$

(3) Poisson Sum + Benford

↳ Pinkham (1961), Feller : errors

Mantissa: $x>0$: $M_B(x) \cdot B^k$, extend to $x \in \mathbb{C}$

Sets: $\mathbb{P}(A) = \lim_{T\to\infty} \frac{\#\{n\in A : n \le T\}}{T}$ or $\lim_{T\to\infty} \frac{\mu(0\le t\le T : t\in A)}{T}$

Setup: $\vec{X}_T$ & $\vec{Y}_{T,B} (= \log_B \vec{X}_T) \to \vec{Y}_B$

Say $\exists$ nice density f st $\vec{Y}_{T,B}$'s spread f plus errors

$$CDF_{\vec{Y}_{T,B}}(x) = \int_{-\infty}^{x} \frac{1}{T} f\left(\frac{t}{T}\right) dt + E_T(x)$$

$$= F\left(\frac{x}{T}\right) + E_T(x)$$

often Gaussian

Assume $\exists$ mono incr $h(T) \to \infty$ st $\leftarrow P_T(a,b) = \sum_k \mathrm{Prob}(a+k \le \vec{Y}_{T,B} \le b+k)$

Cond 1: $F_T(\infty) - F_T(Th(T)) = o(1)$

$F_T(-Th(T)) - F_T(-\infty) = o(1)$

Cond 2: $\frac{1}{T} \sum_{|k| > Th(T)} \int_a^b f\left(\frac{x+k}{T}\right) dx = o(1)$

Cond 3: $\sum_{k\ne 0} \left| \frac{\hat{f}(Tk)}{k} \right| = o(1)$

☆ Cond 4: $\sum_{|k| \le Th(T)} \left[E_T(b+k) - E_T(a+k) \right] = o(1)$

THM (M–, ?)

Conds 1-4 imply Benford (if f is sq-integrable prob dist)

$$P_T(a,b) = \mathbb{P}(\vec{Y}_{T,B} \bmod 1 \in [a,b])$$
$$= \sum_k \mathbb{P}(a+k \le \vec{Y}_{T,B} \le b+k)$$

↓ Poisson Sum

$$= \hat{f}(0)(b-a) + \sum_{k\neq 0} \hat{f}(Tk) \frac{e^{2\pi i b k} - e^{2\pi i a k}}{2\pi i k} + o(1)$$

but $\hat{f}(0) = 1$ as prob dist

CORR: GEO. BROWNIAN MOTIONS ARE BENFORD

GENERAL IDEA

- Structure Thm of sorts
  ↳ main term spread out of something nice
  apply Poisson Sum
- Control of errors

VALUES OF L-FNS + BENFORD

Unconditionally ζ, Dirichlet, holo Hecke cusp forms level 1, even k

(Density Conj replaces GRH) $\left(N(\sigma,T) = O\left(T^{1-\beta(\sigma-\frac{1}{2})}\log T\right), \beta>0\right)$

- STRUCTURE THM

SELBERG'S LOG-NORMAL LAW

$$\frac{\mu\left(t\in[T,2T] : a \le \log|\zeta(\tfrac{1}{2}+it)| \le b\right)}{T} \to \frac{1}{\sqrt{2\pi\sigma_T^2}}\int_a^b e^{-\frac{u^2}{2\sigma_T^2}}du + O\left(\frac{\log^2\sigma_T}{\sigma_T}\right)$$

$$\sigma_T = \sqrt{\tfrac{1}{2}\log\log T} + O(\log\log\log T)$$

Error term too large for pointwise summation

Need Hejhal's refinement

Ingredients of proof

(1) Approx $\log L(\sigma+it)$ with $\sum_{n\le X}\frac{c(n)\Lambda(n)}{\log n}n^{-\sigma-it}$

(2) look at moments $\int_T^{2T}|\quad|^{2k}$, $k \le \log^{1-\delta} T$

(3) Mont-Vaughan:

$$\int_T^{T+H}\left(\sum a_n n^{-it}\right)\overline{\left(\sum b_n n^{-it}\right)}dt = H\sum a_n\overline{b_n} + O(1)\sqrt{\sum n|a_n|^2\cdot\sum n|b_n|^2}$$

(4) work @ test fns: char fns $\chi_{a,b}$

Do for $\sigma = \frac{1}{2} + \frac{1}{\log^\delta T}$

-4-

3X+1 AND BENFORD

- Kakutani: Conspiracy
- Erdos : not ready

X odd: $T(x) = \frac{3X+1}{2^k}$ $\quad 2^k \| 3X+1$

Conj: eventually 1 $\quad 7 \xrightarrow[1]{} 11 \xrightarrow[1]{} 17 \xrightarrow[2]{} 13 \xrightarrow[3]{} 5 \xrightarrow[4]{} 1$

STRUCTURE THM (S, K-S)

Given pos ints $(k_1, \dots, k_m)$: two arithm prog of form $X, X + 6 \cdot 2^{k_1 + \dots + k_m}$ full (start initially)

$\Longrightarrow$ get natural density

$$\mathbb{P}(A) = \lim_{N \to \infty} \frac{\#\{n \le N, n \equiv 1,5\ (6), n \in A\}}{\#\{n \le N, n \equiv 1,5\ (6)\}}$$

$\Longrightarrow$ Geo Brownian Motion in a sense $\quad P(n) = \left(\frac{1}{2}\right)^n, n = 1,2,3,\dots$

- k_j are iidrv @ exp distr @ param $\frac{1}{2}$
- $\mathbb{P}\left(\log_2 \frac{X_m}{\left(\frac{3}{4}\right)^m X_0} \le a \right) = \mathbb{P}\mathrm{rob}\left(S_m - 2m \le a \right)$

where S_m is sum m exp-dist rv $\left(\frac{1}{2}\right)$

THM (K-M)

As $m \to \infty$, $\frac{X_m}{\left(\frac{3}{4}\right)^m X_0}$ is Benford

- Failed Proof: lattice, bad errors (polygonal approx)

- Proof: CLT: $S_m - 2m \to N(0, \sqrt{2m})$

(1) $\text{Prob}\left(\frac{\overline{S}_m}{\sqrt{m}} = \frac{k}{\sqrt{m}}\right) = \frac{\eta\left(\frac{k}{\sqrt{m}}\right)}{\sqrt{m}} + o\left(\frac{1}{\sqrt{m}}\right)$, η is std normal

$\hookrightarrow O\left(\frac{1}{g(m)\sqrt{m}}\right)$

(2) $I_\ell = \{\ell M, \ell M+1, \dots, (\ell+1)M-1\}$

$M = m^c$, $c < \frac{1}{2}$

- $k_1, k_2 \in I_\ell \implies \left|\frac{1}{\sqrt{m}}\eta\left(\frac{k_1}{\sqrt{m}}\right) - \frac{1}{\sqrt{m}}\eta\left(\frac{k_2}{\sqrt{m}}\right)\right|$ is manageable

$\hookrightarrow$ allows us to use just left endpoints

- assume $C = \log_B 2$ is irrational of type $k < \infty$

$\forall \varepsilon > 0$ $\#\{k \in I_\ell : kC \bmod 1 \in [a,b]\} = M(b-a) + O\left(M^{1+\varepsilon-\frac{1}{k}}\right)$

(quantified equidistr + irrationality measure)

(3) Poisson Sum: $\frac{1}{\sigma}\sum_n e^{-n^2\pi/\sigma^2} = \sum_n e^{-n^2\pi\sigma^2}$

$Y_m = \log_B \frac{X_m}{\left(\frac{3}{4}\right)^m X_0}$; mult by $\frac{1}{\log_2 B} = \log_B 2$

Study $\overline{S}_m \cdot \log_B 2 \bmod 1$ in $[a,b]$

$$P_m(a,b) = \sum_{|\ell| \leq \frac{\sqrt{m}\, h(m)}{M}} \mathrm{Prob}\left(\overline{S_m} = t \in I_\ell : t \bmod 1 \in [a,b]\right)$$

$$+ \text{ Sum @ large } \ell$$

Rate of Equid

Given seq $x_1, x_2, \ldots$

$$D_N = \frac{1}{N} \sup_{[a,b] \subset [0,1]} \left| N(b-a) - \#\{n \leq N : x_n \in [a,b]\} \right|$$

Erdos-Turan: $\exists C$ st $\forall m$

$$D_N \leq C \cdot \left(\frac{1}{m} + \sum_{h=1}^{m} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right| \right)$$

Say $x_n = n\alpha \bmod 1$

Exp sum is $\leq \frac{1}{|\sin \pi h\alpha|} \leq \frac{1}{2\langle h\alpha \rangle}$ ← distance to nearest int

Must Control

$$\sum_{h=1}^{m} \frac{1}{h \langle h\alpha \rangle}$$ Now see why can't be too close to a rational

Say of type κ if κ the sup of all γ with $\underline{\lim}_{q \to \infty} q^{\gamma} \langle q\alpha \rangle = 0$

↳ Roth: alg #s of type 1: $\left|\alpha - \frac{p}{q}\right| \geq \frac{c}{q^{2+\epsilon}}$

↳ Gives $\sum_{h=1}^{m} \frac{1}{h\langle h\alpha \rangle} = O\left(m^{\kappa - 1 + \epsilon}\right)$, take $m = \lfloor N^{1/\kappa} \rfloor$

$\log_{10}$

$$\left|\log_{10} 2 - \frac{p}{q}\right| = \left|\frac{\log 2}{\log 10} - \frac{p}{q}\right| = \left|\frac{q \log 2 - p \log 10}{q \cdot \log 10}\right|$$

Enough to show $|q \log 2 - p \log 10| \gg \frac{1}{q^c}$

(Literature always works with log not $\log_{10}$ — otherwise get probs @ integer powers)

THM (Baker)

$\alpha_1, \dots, \alpha_n$ alg numbers of height at most $A_j \ (\geq 4)$
$\beta_1, \dots, \beta_n$ rational ints " " " $B \ (\geq 4)$

$\Lambda = \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n$

If $\Lambda \neq 0$, $|\Lambda| > \frac{1}{B^{C \Omega \log \Omega'}}$

where $C = (16nd)^{200n}$

$\mathbb{K} = \mathbb{Q}(\alpha_i, \beta_j)$ of deg d

$\Omega = \log A_1 \cdots \log A_n$

$\Omega' = \Omega / \log A_n$

(Consider special fns, know poles, contour $\int, \dots$)

For us: $d = 1$, $n = 2$, $C = 2^{2000}$

$\Omega = \log 4 \cdot \log 10 \quad \Omega' = \log 4$

$K = 1 + C \Omega \log \Omega' = 2^{2000} (\log 4)(\log 10)(\log \log 4) + 1 = 1.197082 \cdot 10^{602}$

E-mail address: sjmiller@math.brown.edu

Department of Mathematics, Brown University, Providence, RI 02912