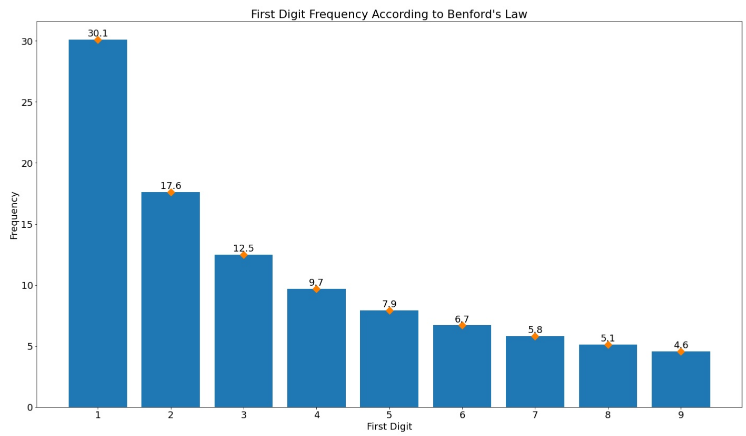
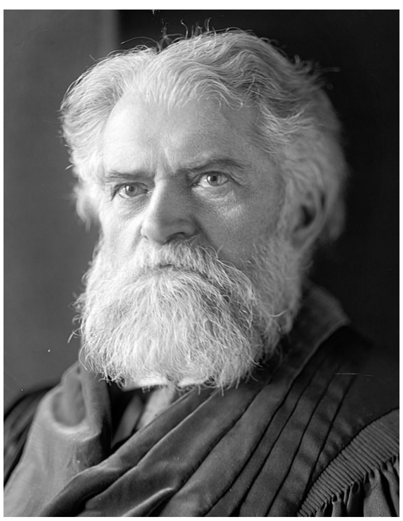
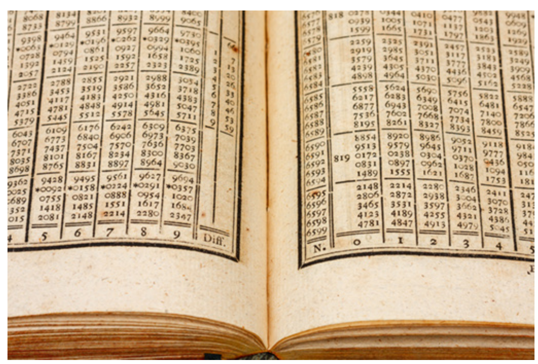
Benford's Law and irrational rotation





Simon Newcomb
1835-1909
1881



Numberphile video (on log tables)
https://www.youtube.com/watch?v=VRzH4xB0GdM

FRANK BENFORD

The frequency of first digits thus follows closely the logarithmic relation

$$F_a = \log\left(\frac{a+1}{a}\right), \tag{1}$$

where F_a is the frequency of the digit a in the first place of used numbers.

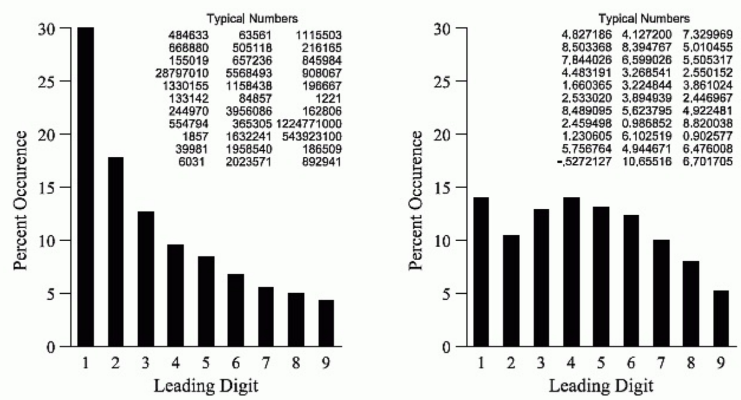
TABLE II

OBSERVED AND COMPUTED FREQUENCIES

Natural Number	Number Interval	Observed Frequency	Logarithm Interval	Observed − Computed	Prob. Error of Mean
1	1 to 2	0.306	0.301	+0.005	±0.008
2	2 to 3	0.185	0.176	+0.009	±0.004
3	3 to 4	0.124	0.125	−0.001	±0.004
4	4 to 5	0.094	0.097	−0.003	±0.003
5	5 to 6	0.080	0.079	+0.001	±0.002
6	6 to 7	0.064	0.067	−0.003	±0.002
7	7 to 8	0.051	0.058	−0.007	±0.002
8	8 to 9	0.049	0.051	−0.002	±0.002
9	9 to 10	0.047	0.046	+0.001	±0.003

Frank Benford
1883-1948
1938

also true for 2^n
(or any other a^n, $a \neq 10^k$)



a. Tax return numbers, *Benford's law*

b. RNG numbers, *not Benford's law*

FIGURE 34-2

Two examples of leading-digit histograms. The left figure shows the leading-digit distribution for 14,414 numbers taken from U.S. Federal income tax returns. The figure on the right is for numbers produced by a computer random number generator (RNG). This shows one of the longstanding mysteries of Benford's law– Why do some sets of numbers follow the law (such as tax returns), while others (such as this RNG) do not? Many have claimed that this is some sort of secret code hidden in the fabric of Nature.

Key to understanding B.L in math is through rational approximations

Dirichlet's Drawer Box principle
(A.K.A pigeonholes pr.)

Dirichlet's rational approximation thm

Gustav Lejeune Dirichlet
1805-1859
1840

Th 1
Dirichlet proved in 1840 that

for every number α
and any positive N (whole)
there's a rational p/q
s.t. $\left|\alpha - \frac{p}{q}\right| < \frac{1}{qN}$, $q \le N$

$|q\alpha - p| < \frac{1}{N}$ $\quad 0 \le \{x\} = x - \lfloor x \rfloor \le 1$

Pf (Idea:) Plot fractional values of $k\alpha$, $k = 1, 2, \dots, N+1$

$N+1$ p.

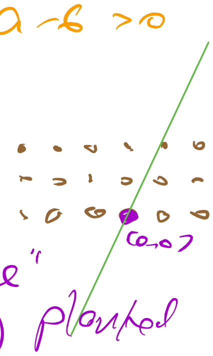
$a > b$

So
$|\{a\alpha\} - \{b\alpha\}| \le \frac{1}{N}$
$\parallel$
$|a\alpha - b\alpha - p| \le \frac{1}{N}$
whole

i.e. $|\alpha(\underbrace{a-b}_{q}) - p| \le \frac{1}{N}$

$\left|\alpha - \frac{p}{q}\right| \le \frac{1}{Nq}$, $N \ge q = a - b > 0$ //

Corollary 1 (proved last time)



A line $y = \alpha x$ intersects a "tree" of radius $\varepsilon > 0$ (for any ε) planted at some lattice point

(Dirichlet's simultaneous approximation)

Thm 2 Thm 1 also holds for several numbers $\alpha_1, \dots, \alpha_k$, i.e. for every N there's q

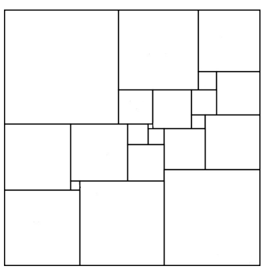
$$\left|\alpha_1 - \frac{p_1}{q}\right| \le \frac{1}{Nq}, \quad \dots, \quad \left|\alpha_k - \frac{p_k}{q}\right| \le \frac{1}{Nq}$$

(only now $q \le N^k$). **Proof**: Exercise

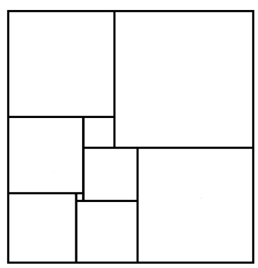
Corollary (Max Dehn's Thm on dissecting into squares)

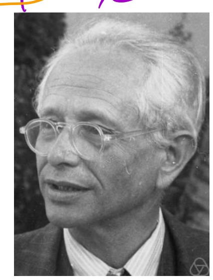
Thm An $a \times b$ rectangle can be dissected into squares if and only if a/b is rational

e.g.



or



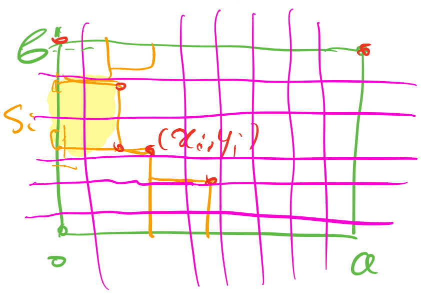


Max Dehn
1878-1952
1903

(Solved 3rd Hilbert's problem)

Proof. The if part is easy (if $a/b = m/n$ then can cut into squares $a/m \times a/m$)

Only if: Assume the rectangle is cut into k squares of sides S_i, $i = 1, 2, \dots, k$



Need to prove that a/b rational

By Dir's Simultan. Approx. Thm we can assume that all vertices (x_i, y_i) of squares are within $\frac{1}{5}$ of a lattice pt

(choose q such that

$$\begin{cases} \left|x_i - \frac{p_i}{q}\right| < \frac{1}{5q} \\ \left|y_i - \frac{p_i'}{q}\right| < \frac{1}{5q} \end{cases}$$

and mult' by all coords by q).

Now draw horizontal & vertical lines at $x = n+\frac{1}{2}$, $y = m+\frac{1}{2}$. The total length of all horiz lines inside the rectangle is

$$\boxed{L = a\cdot[b]}$$

length of each segment; $[b]$: the number of horizontal segms. (Notation $[x]$ is closest integer to x)

and for the vertical lines we have similarly the total length

$$\boxed{M = b\cdot[a]}$$

b: length of each segm; $[a]$: their number

<u>Claim</u> $L=M$. Indeed, look at the lines inside the i-th square: . They contribute exactly $s_i\cdot[s_i]$ to L and exactly the same to M, so we have

$$L = \sum_{i=1}^{k} s_i\cdot[s_i] = M.$$

But $L = a\cdot[b]$ and $M = b\cdot[a]$,

so $a\cdot[b] = b\cdot[a] \implies \frac{a}{b} = \frac{[a]}{[b]}$ = rational!

Back to Benford's Law

How to go about proving B. L. for 2^n?

E.g. what exactly it means that 2^n starts with digit d? $\boxed{d=1,2,\dots,9}$

This means: $2^n = d \cdot 10^k + r$, $0 \le r < 10^k$

$2^n = d\underbrace{\_\,\_\,\_\,\_}_{k \text{ digits}}$ i.e.:

$$\boxed{d \cdot 10^k \le 2^n < (d+1) \cdot 10^k}$$ for some k

How to get rid of k?

Take $\log_{10}$: $\log_{10}(d \cdot 10^k) \le \log_{10} 2^n < \log_{10}\left((d+1) \cdot 10^k\right)$

$$k + \underbrace{\log_{10} d}_{\ge 0,\ <1} \le n \cdot \log_{10} 2 < k + \underbrace{\log_{10}(d+1)}_{\ge 0,\ <1}$$

i.e. $0 \le \log_{10} d \le \underbrace{n \cdot \log_{10} 2 - k}_{= \{n \cdot \log_{10} 2\}} < \log_{10}(d+1) < 1$

So $k = \lfloor n \cdot \log_{10} 2 \rfloor$ ← floor

and

$\{x\} = x - \lfloor x \rfloor$ ← fractional part of x

So $\Leftrightarrow$ $\boxed{\{n \cdot \log_{10} 2\} \text{ is in } [\underbrace{\log_{10} d}_{x}, \underbrace{\log_{10}(d+1)}_{y}) = I}$

→ interval on $[0,1)$

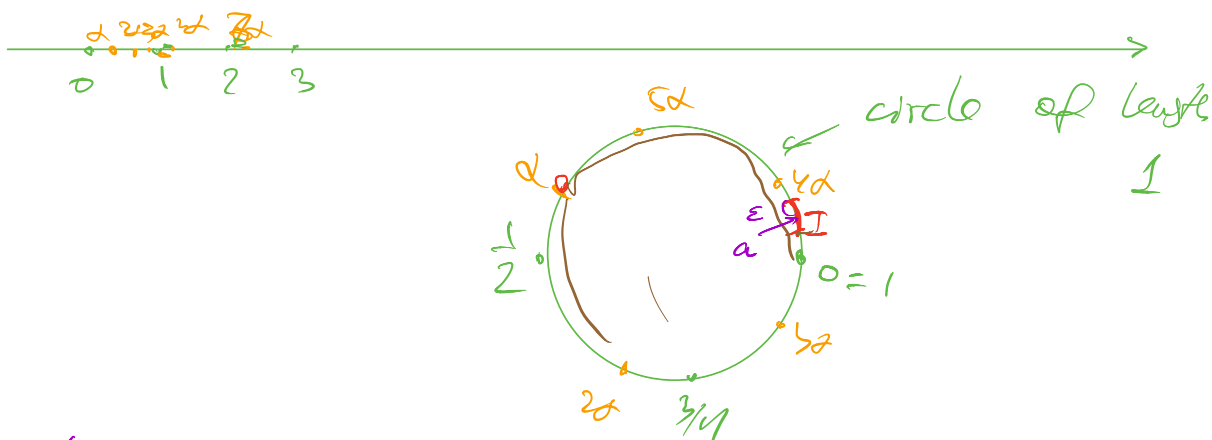
th is $\log_{10} \frac{d+1}{d}$

i.e. we want to study when

$\{n \cdot \log_{10} 2\}$ lands inside the interval I

$\quad \overset{\shortparallel}{\alpha}$ (of length $\boxed{\log_{10} \frac{d+1}{d}}$)

i.e. $\{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots$



Thm (Kronecker's Approximation Thm)

For any $\varepsilon > 0$, irrational α

and arbitrary a there exists n and m

s.t. $|n\alpha - a - m| < \varepsilon$

i.e. $\{n\alpha - a\} < \varepsilon$

fractional part

$\Leftrightarrow a - \varepsilon + m < n\alpha < a + \varepsilon + m$

<u>Corollary</u> There exist n s.t. 2^n begins with any ~~fixed~~ fixed combination of k-digits

<u>Pf</u> The only thing we need to check is that $\alpha = \log_{10^k} 2$ is irrational. Indeed if we assume $\log_{10^k} 2 = \frac{p}{q}$ is rational

then $(10^k)^{p/q} = 2$ i.e.

$$(10^k)^p = 2^q$$

$$\parallel$$

$$10^{kp} = 2^q \leftarrow \text{false!} \quad //$$

<u>Pf</u> of Kronecker's Thm.

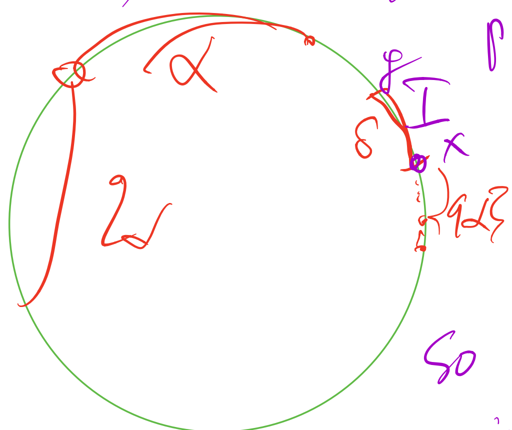
Let N be such that $\frac{1}{N} < \varepsilon$, then by Dirichlet there are $a > b > 0$

integers s.t. $|\{a\alpha\} - \{b\alpha\}| < \frac{1}{N}$

$\{\underbrace{(a-b)}_{q}\alpha\} < \varepsilon$, so by

continuing with points $q\alpha, 2q\alpha$

$3q\alpha, \ldots, Nq\alpha$ we'll bypass pt x of I but not y since $\{q\alpha\} < \varepsilon = |I|$



So we'll land in I //

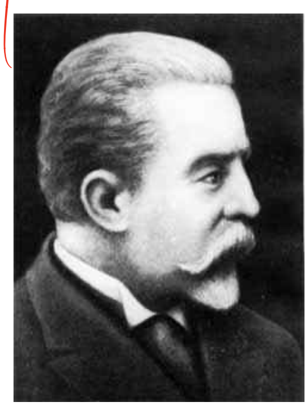
Leopold Kronecker
1823-1891
1884

"Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk"
"God made the natural numbers; all else is the work of man."

Final statement follows from the <u>Equidistribution Thm</u>

<u>Thm</u> Let α be any irrational number and I be an interval on $[0,1)$ of length δ. Then the probability that the fractional part $\{n\alpha\}$ for $n=1,2,3,\ldots$ belongs to I is equal to δ.

This was proved in 1909 and 1910 by:



Piers Bohl
1865-1921
1909

Wacław Sierpiński
1882-1969
1910

Hermann Weyl
1895-1955
1910