

# How low can we go? Lower order terms in CLTs from Benford's Law to Elliptic Curves

Steven J Miller  
Williams College

Steven.J.Miller@williams.edu  
<http://www.williams.edu/go/math/sjmilller>

NES MAA Fall 2009 Meeting, November 21, 2009

## Outline

- Review of Central Limit Theorem results.
- Applications to Benford's Law.
- Applications to Elliptic Curves.

## Central Limit Theorems

## General Statement

### Central Limit Theorem

Let  $X_1, \dots, X_N$  be **nice** iidrv with mean  $\mu$  and variance  $\sigma^2$ .  
Then

$$Y_N = \frac{X_1 + \dots + X_N}{N}, \quad \lim_{N \rightarrow \infty} \frac{Y_N - \mathbb{E}[Y_N]}{\text{StDev}(Y_N)} \rightarrow N(0, 1).$$

- What is nice? Finite higher moments:

$$k^{\text{th}} \text{ centered moment} = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx.$$

- Speed of convergence controlled by higher moments (especially third).

## Overview of Benford's Law

## Benford's Law: Newcomb (1881), Benford (1938)

### Statement

For many data sets, probability of observing a first digit of  $d$  base  $B$  is  $\log_B \left( \frac{d+1}{d} \right)$ ; base 10 about 30% are 1s.

- Not all data sets satisfy Benford's Law.
  - ◇ Long street  $[1, L]$ :  $L = 199$  versus  $L = 999$ .
  - ◇ Oscillates between  $1/9$  and  $5/9$  with first digit 1.
  - ◇ **Many streets of different sizes: close to Benford.**

## Examples

- recurrence relations
- special functions (such as  $n!$ )
- iterates of power, exponential, rational maps
- products of random variables
- $L$ -functions, characteristic polynomials
- iterates of the  $3x + 1$  map
- differences of order statistics
- hydrology and financial data
- many hierarchical Bayesian models

## Applications

- analyzing round-off errors
- determining the optimal way to store numbers
- detecting tax and image fraud, and data integrity

## Mantissas

Mantissa:  $x = M_{10}(x) \cdot 10^k$ ,  $k$  integer.

$M_{10}(x) = M_{10}(\tilde{x})$  if and only if  $x$  and  $\tilde{x}$  have the same leading digits.

**Key observation:**  $\log_{10}(x) = \log_{10}(\tilde{x}) \pmod{1}$  if and only if  $x$  and  $\tilde{x}$  have the same leading digits. Thus often study  $y = \log_{10} x$ .

## Equidistribution and Benford's Law

### Equidistribution

$\{y_n\}_{n=1}^{\infty}$  is equidistributed modulo 1 if probability  $y_n \bmod 1 \in [a, b]$  tends to  $b - a$ :

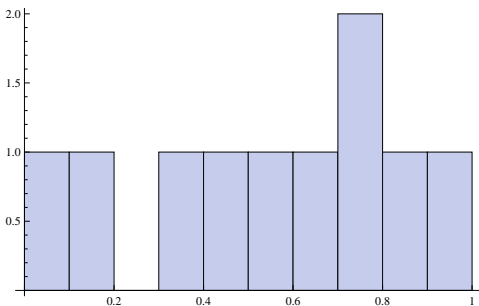
$$\frac{\#\{n \leq N : y_n \bmod 1 \in [a, b]\}}{N} \rightarrow b - a.$$

### Theorem: Kronecker, Weyl

$\beta \notin \mathbb{Q}$ ,  $n\beta$  is equidistributed mod 1.

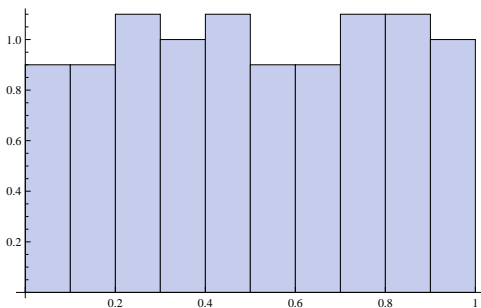


## Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$  for  $n \leq 10$

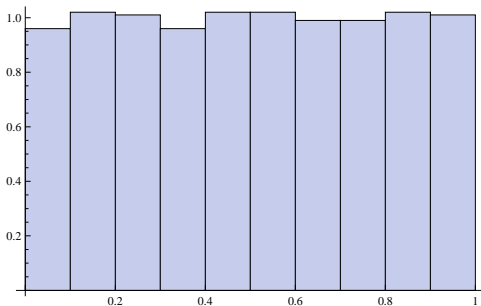
## Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$  for  $n \leq 100$



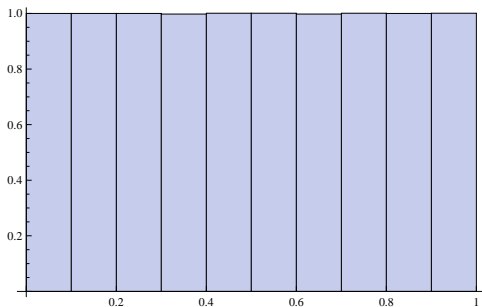
## Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$  for  $n \leq 1000$



## Example of Equidistribution: $n\sqrt{\pi} \bmod 1$



$n\sqrt{\pi} \bmod 1$  for  $n \leq 10,000$

## Logarithms and Benford's Law

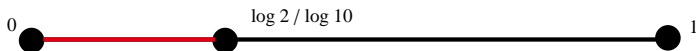
### Fundamental Equivalence

Data set  $\{x_i\}$  is Benford base  $B$  if  $\{y_i\}$  is equidistributed mod 1, where  $y_i = \log_B x_i$ .

# Logarithms and Benford's Law

## Fundamental Equivalence

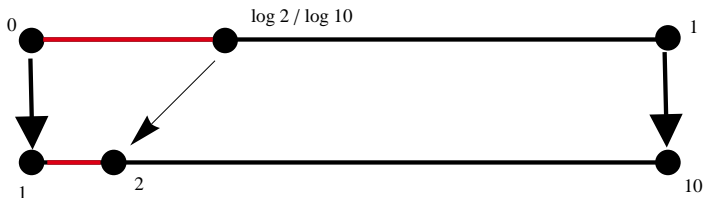
Data set  $\{x_i\}$  is Benford base  $B$  if  $\{y_i\}$  is equidistributed mod 1, where  $y_i = \log_B x_i$ .



# Logarithms and Benford's Law

## Fundamental Equivalence

Data set  $\{x_i\}$  is Benford base  $B$  if  $\{y_i\}$  is equidistributed mod 1, where  $y_i = \log_B x_i$ .



Benford's Law and the  
 $3x + 1$  Problem

## $3x + 1$ Problem

- Kakutani (conspiracy), Erdős (not ready).
- $x$  odd,  $T(x) = \frac{3x+1}{2^k}$ ,  $2^k \parallel 3x + 1$ .
- Conjecture: for some  $n = n(x)$ ,  $T^n(x) = 1$ .
- $7 \rightarrow_1 11 \rightarrow_1 17 \rightarrow_2 13 \rightarrow_3 5 \rightarrow_4 1 \rightarrow_2 1$ ,  
 2-path  $(1, 1)$ , 5-path  $(1, 1, 2, 3, 4)$ .  
*m*-path:  $(k_1, \dots, k_m)$ .

## Structure Theorem: Sinai, Kontorovich-Sinai

### Theorem (Sinai, Kontorovich-Sinai)

$k_i$ -values are i.i.d.r.v. (geometric,  $1/2$ ):

$$\mathbb{P} \left( \frac{\log_2 \left[ \frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{\sqrt{2m}} \leq a \right) = \mathbb{P} \left( \frac{S_m - 2m}{\sqrt{2m}} \leq a \right)$$

## Structure Theorem: Sinai, Kontorovich-Sinai

### Theorem (Sinai, Kontorovich-Sinai)

$k_i$ -values are i.i.d.r.v. (geometric,  $1/2$ ):

$$\mathbb{P} \left( \frac{\log_2 \left[ \frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{(\log_2 B) \sqrt{2m}} \leq a \right) = \mathbb{P} \left( \frac{S_m - 2m}{(\log_2 B) \sqrt{2m}} \leq a \right)$$

## Structure Theorem: Sinai, Kontorovich-Sinai

### Theorem (Sinai, Kontorovich-Sinai)

$k_i$ -values are i.i.d.r.v. (geometric,  $1/2$ ):

$$\mathbb{P} \left( \frac{\log_B \left[ \frac{x_m}{\left(\frac{3}{4}\right)^m x_0} \right]}{\sqrt{2m}} \leq a \right) = \mathbb{P} \left( \frac{\frac{(S_m - 2m)}{\log_2 B}}{\sqrt{2m}} \leq a \right)$$

## $3x + 1$ and Benford

### Theorem (Kontorovich and M–, 2005)

*As  $m \rightarrow \infty$ ,  $x_m/(3/4)^m x_0$  is Benford.*

### Theorem (Lagarias-Soundararajan 2006)

*$X \geq 2^N$ , for all but at most  $c(B)N^{-1/36} X$  initial seeds the distribution of the first  $N$  iterates of the  $3x + 1$  map are within  $2N^{-1/36}$  of the Benford probabilities.*

## Sketch of the proof

- Failed Proof: lattices, bad errors.
- CLT:  $(S_m - 2m)/\sqrt{2m} \rightarrow N(0, 1)$ .
- Quantified Equidistribution:  $I_\ell = \{\ell M, \dots, (\ell + 1)M - 1\}$ ,  
 $M \ll m^{1/2}$

$\log_B 2$  of irrationality type  $\kappa < \infty$ :

$$\#\{k \in I_\ell : k \log_B 2 \bmod 1 \in [a, b]\} = M(b - a) + O(M^r),$$

$$r = 1 + \epsilon - 1/\kappa < 1.$$

## Irrationality Type

### Irrationality type

$\alpha$  has irrationality type  $\kappa$  if  $\kappa$  is the supremum of all  $\gamma$  with

$$\liminf_{q \rightarrow \infty} q^{\gamma+1} \min_p \left| \alpha - \frac{p}{q} \right| = 0.$$

- Algebraic irrationals: type 1 (Roth's Thm).
- Theory of Linear Forms:  $\log_B 2$  of finite type.

## Linear Forms

### Theorem (Baker)

$\alpha_1, \dots, \alpha_n$  algebraic numbers height  $A_j \geq 4$ ,  $\beta_1, \dots, \beta_n \in \mathbb{Q}$  with height at most  $B \geq 4$ ,

$$\Lambda = \beta_1 \log \alpha_1 + \dots + \beta_n \log \alpha_n.$$

If  $\Lambda \neq 0$  then  $|\Lambda| > B^{-C\Omega \log \Omega'}$ , with  $d = [\mathbb{Q}(\alpha_i, \beta_j) : \mathbb{Q}]$ ,  
 $C = (16nd)^{200n}$ ,  $\Omega = \prod_j \log A_j$ ,  $\Omega' = \Omega / \log A_n$ .

Gives  $\log_{10} 2$  of finite type, with  $\kappa < 1.2 \cdot 10^{602}$ :

$$|\log_{10} 2 - p/q| = |q \log 2 - p \log 10| / q \log 10.$$

## $3x + 1$ Data: random 10,000 digit number, $2^k \parallel 3x + 1$

80,514 iterations ( $((4/3)^n = a_0$  predicts 80,319);  
 $\chi^2 = 13.5$  (5% 15.5).

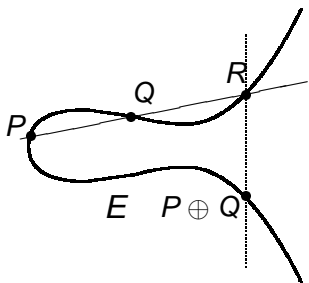
Digit	Number	Observed	Benford
1	24251	0.301	0.301
2	14156	0.176	0.176
3	10227	0.127	0.125
4	7931	0.099	0.097
5	6359	0.079	0.079
6	5372	0.067	0.067
7	4476	0.056	0.058
8	4092	0.051	0.051
9	3650	0.045	0.046

## Elliptic Curves

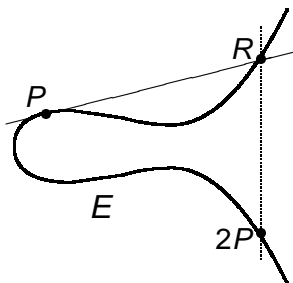
## Mordell-Weil Group

Elliptic curve  $y^2 = x^3 + ax + b$  with rational solutions

$P = (x_1, y_1)$  and  $Q = (x_2, y_2)$  and connecting line  $y = mx + b$ .



Addition of distinct points  $P$  and  $Q$



Adding a point  $P$  to itself

$$E(\mathbb{Q}) \approx E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^r$$

## Riemann Zeta Function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}, \quad \operatorname{Re}(s) > 1.$$

### Functional Equation:

$$\xi(s) = \Gamma\left(\frac{s}{2}\right) \pi^{-\frac{s}{2}} \zeta(s) = \xi(1-s).$$

### Riemann Hypothesis (RH):

All non-trivial zeros have  $\operatorname{Re}(s) = \frac{1}{2}$ ; can write zeros as  $\frac{1}{2} + i\gamma$ .

## General $L$ -functions

$$L(s, f) = \sum_{n=1}^{\infty} \frac{a_f(n)}{n^s} = \prod_{p \text{ prime}} L_p(s, f)^{-1}, \quad \operatorname{Re}(s) > 1.$$

### Functional Equation:

$$\Lambda(s, f) = \Lambda_{\infty}(s, f)L(s, f) = \Lambda(1 - s, f).$$

### Generalized Riemann Hypothesis (GRH):

All non-trivial zeros have  $\operatorname{Re}(s) = \frac{1}{2}$ ; can write zeros as  $\frac{1}{2} + i\gamma$ .

## Elliptic curve $L$ -function

$E : y^2 = x^3 + ax + b$ , associate  $L$ -function

$$L(s, E) = \sum_{n=1}^{\infty} \frac{a_E(n)}{n^s} = \prod_{p \text{ prime}} L_E(p^{-s}),$$

where

$$a_E(p) = p - \#\{(x, y) \in (\mathbb{Z}/p\mathbb{Z})^2 : y^2 \equiv x^3 + ax + b \pmod{p}\}.$$

### Birch and Swinnerton-Dyer Conjecture

Rank of group of rational solutions equals order of vanishing of  $L(s, E)$  at  $s = 1/2$ .

## 1-Level Density

$L$ -function  $L(s, f)$ : by RH non-trivial zeros  $\frac{1}{2} + i\gamma_{f,j}$ .

$N_f$ : analytic conductor.

$\varphi(x)$ : compactly supported even Schwartz function.

$$D_{1,f}(\varphi) = \sum_j \varphi\left(\frac{\log N_f}{2\pi} \gamma_{f,j}\right)$$

- individual zeros contribute in limit
- most of contribution is from low zeros

### Katz-Sarnak Conjecture:

$$\begin{aligned} D_{1,\mathcal{F}}(\varphi) &= \lim_{N \rightarrow \infty} \frac{1}{|\mathcal{F}_N|} \sum_{f \in \mathcal{F}_N} D_{1,f}(\varphi) = \int \varphi(x) \rho_{G(\mathcal{F})}(x) dx \\ &= \int \widehat{\varphi}(u) \widehat{\rho}_{G(\mathcal{F})}(u) du. \end{aligned}$$

## Explicit Formula (Contour Integration)

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds} \log \zeta(s) = -\frac{d}{ds} \log \prod_p (1 - p^{-s})^{-1} \\ &= \sum_p \frac{\log p \cdot p^{-s}}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s} + \text{Good}(s). \end{aligned}$$

## Explicit Formula (Contour Integration)

$$\begin{aligned}
 -\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds} \log \zeta(s) = -\frac{d}{ds} \log \prod_p (1 - p^{-s})^{-1} \\
 &= \sum_p \frac{\log p \cdot p^{-s}}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s} + \text{Good}(s).
 \end{aligned}$$

## Explicit Formula (Contour Integration)

$$\begin{aligned}
 -\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds} \log \zeta(s) = -\frac{d}{ds} \log \prod_p (1 - p^{-s})^{-1} \\
 &= \sum_p \frac{\log p \cdot p^{-s}}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s} + \text{Good}(s).
 \end{aligned}$$

Contour Integration:

$$\int -\frac{\zeta'(s)}{\zeta(s)} \frac{x^s}{s} ds \quad \text{vs} \quad \sum_p \log p \int \left(\frac{x}{p}\right)^s \frac{ds}{s}.$$

## Explicit Formula (Contour Integration)

$$\begin{aligned}
 -\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds} \log \zeta(s) = -\frac{d}{ds} \log \prod_p (1 - p^{-s})^{-1} \\
 &= \sum_p \frac{\log p \cdot p^{-s}}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s} + \text{Good}(s).
 \end{aligned}$$

Contour Integration:

$$\int -\frac{\zeta'(s)}{\zeta(s)} \phi(s) ds \quad \text{vs} \quad \sum_p \log p \int \phi(s) p^{-s} ds.$$

## Explicit Formula (Contour Integration)

$$\begin{aligned}
 -\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds} \log \zeta(s) = -\frac{d}{ds} \log \prod_p (1 - p^{-s})^{-1} \\
 &= \sum_p \frac{\log p \cdot p^{-s}}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s} + \text{Good}(s).
 \end{aligned}$$

Contour Integration (see Fourier Transform arising):

$$\int -\frac{\zeta'(s)}{\zeta(s)} \phi(s) ds \quad \text{vs} \quad \sum_p \log p \int \phi(s) e^{-\sigma \log p} e^{-it \log p} ds.$$

**Interplay b/w zeros and coefficients**

# 1-Level Expansion

$$\begin{aligned}
 D_{1,\mathcal{F}}(\phi) &= \frac{1}{|\mathcal{F}|} \sum_{E \in \mathcal{F}} \sum_j \phi \left( \frac{\log N_E}{2\pi} \gamma_E^{(j)} \right) \\
 &= -\frac{2}{|\mathcal{F}|} \sum_{E \in \mathcal{F}} \sum_p \frac{\log p}{\log N_E} \frac{1}{p} \hat{\phi} \left( \frac{\log p}{\log N_E} \right) a_E(p) \\
 &\quad - \frac{2}{|\mathcal{F}|} \sum_{E \in \mathcal{F}} \sum_p \frac{\log p}{\log N_E} \frac{1}{p^2} \hat{\phi} \left( 2 \frac{\log p}{\log N_E} \right) a_E^2(p) \\
 &\quad + \hat{\phi}(0) + \phi(0) + o \left( \frac{\log \log N_E}{\log N_E} \right)
 \end{aligned}$$

## Inputs

One-parameter family:

$$\mathcal{E} : y^2 = x^3 + A(T)x + B(T), \quad A(T), B(T) \in \mathbb{Z}[T],$$

take  $t \in \mathbb{Z}$ .

Let

$$A_{r,\mathcal{F}}(p) = \sum_{t(p)} a_t^r(p), \quad r = 1 \text{ or } 2.$$

For many families

$$(1) : A_{1,\mathcal{F}}(p) = -rp + O(1)$$

$$(2) : A_{2,\mathcal{F}}(p) = p^2 + O(p^{3/2})$$

## Main Term

**Theorem: M- '04**

For small support, one-param family of rank  $r$  over  $\mathbb{Q}(T)$ :

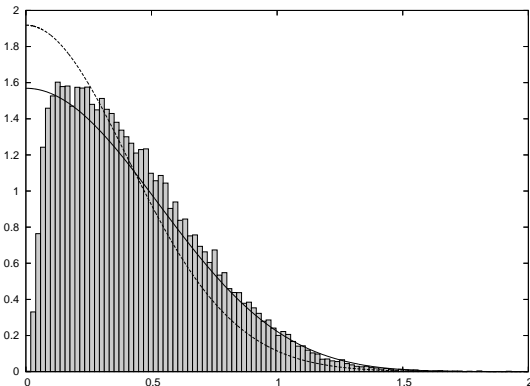
$$\lim_{N \rightarrow \infty} \frac{1}{|\mathcal{F}_N|} \sum_{E_t \in \mathcal{F}_N} \sum_j \varphi \left( \frac{\log N_{E_t}}{2\pi} \gamma_{E_t, j} \right) = \int \varphi(\mathbf{x}) \rho_{\mathcal{G}}(\mathbf{x}) d\mathbf{x} + r\varphi(0)$$

where

$$\mathcal{G} = \begin{cases} \text{SO} & \text{if half odd} \\ \text{SO}(\text{even}) & \text{if all even} \\ \text{SO}(\text{odd}) & \text{if all odd} \end{cases}$$

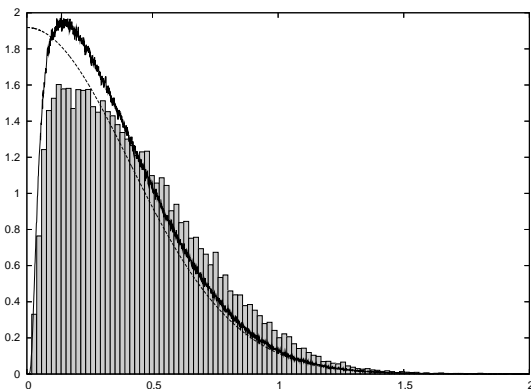
**Confirm Katz-Sarnak for main term**

# Modeling lowest zero (data & calculations from Duc Khiem Huynh)



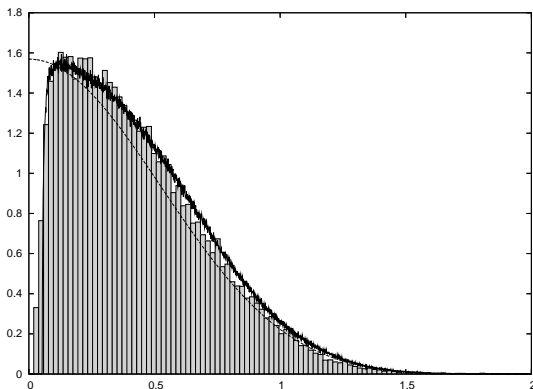
Lowest zero  $L_{E_{11}}(s, \chi_d)$ ,  $0 < d < 400,000$  (bar chart), lowest eigenvalue  $SO(2N)$  w'  $N_{\text{eff}}$  (solid), standard  $N_0$  (dashed).

# Modeling lowest zero (data & calculations from Duc Khiem Huynh)



Lowest zero  $L_{E_{11}}(s, \chi_d)$ ,  $0 < d < 400,000$  (bar chart), lowest eigenvalue  $SO(2N)$  w'  $N_0 = 12$  (solid) w' discretisation and w' standard  $N_0 = 12.26$  (dashed) w/o discretisation.

# Modeling lowest zero (data & calculations from Duc Khiem Huynh)



Lowest zero  $L_{E_{11}}(s, \chi_d)$ ,  $0 < d < 400,000$  (bar chart), lowest eigenvalue  $SO(2N)$  w'  $N_{\text{eff}} = 2$  (solid) w' discretisation and w'  $N_{\text{eff}} = 2.32$  (dashed) w/o discretisation.

## Conclusions

- Analysis of lower order terms crucial for many applications.
- Pure math can have unexpected applications.

