1. Introduction

Benford’s Law A set is Benford if probability first digit is $a$ is $\log_{10}(\frac{a+1}{a})$. 30% start with 1.

Many data sets exhibit Benford behavior: Fibonacci Sequence, Financial data (stocks, bonds, etc.), products of random independent variables

Why do we observe Benford distribution of first digits in “real world” data sets?

1.2 Conserved Quantities

First Proposed model Partition $N$ into $X$: $X = \sum^{n}_{j=1} r_{j}$, Issues: what possible $s \leq N$ is fixed?

Results

• For (small) finite $N$, brute force calculation shows $\mathbb{E}(a_{j}) = \frac{1}{\log_{10}(2)}$; Benford density is proportional to $1/x$.

• For general $N$, approximate: $\mathbb{E}(a_{j}) = \frac{N}{\sum^{n}_{j=1} r_{j}} = \frac{1}{\log_{10}(n)}$, then evaluate $N$-dimensional integral.

• $\mathbb{E}(a_{j}) = \frac{N}{\sum^{n}_{j=1} r_{j}} \rightarrow a \rightarrow 0$.

Second Proposed Model Consider $I$ sticks of lengths $l_{i}$, each $I$ drawn from the random variable L. Break each by cutting at $k_{i}$ with $k_{i} \sim \text{Unif}(0,1)$. Repeat cutting $N$ times. Theorem 1:\n
If $\mathbb{E}(a_{j})$ is on $(1,0)$ and $N \rightarrow 1$, then as $N \rightarrow \infty$, the distribution of lengths of pieces is Benford's Law.

Theorem 2: Let $L$ be some fixed constant such that $l_{1} \rightarrow l_{2} \rightarrow \cdots \rightarrow l_{N} \rightarrow \infty$. Then, as $N \rightarrow \infty$ and $N \rightarrow \infty$, the resulting first digit distribution of the lengths of the broken pieces will conform to Benford’s Law.

Conjecture: Let $I$ be fixed and consider one stick $(M = 1)$. As $N \rightarrow \infty$, the resulting first digit distribution of the lengths of the broken pieces will conform to Benford’s Law.

1.3 Copulas

Copula A form of joint CDF between multiple variables with given uniform marginals on the d-dimensional unit cube.

Sklar’s Theorem Let $X$ and $Y$ be random variables with joint distribution function $H$ and marginal distribution functions $F$ and $G$ respectively. There exists a copula, $C$, such that $C(x,y) = \mathbb{P}(F(x) \land G(y))$, Archimedean Copulas A commonly used family of copulas is of the form $C(u,v) = -\log\left(\phi(\alpha u) + \phi(\alpha v)\right)$ where $\phi$ is the generator and $\alpha$ is the inverse generator of the copula. Investigating the Benfordness of the product of random variables arising from copulas.

Clayton Copula $C(u,v) = \left(\theta + \theta - 1\right) \left\{ u^{-\theta} + v^{-\theta} - 1 \right\}^{-1/\theta}$

PDF (bivariate) $-\frac{\theta u^{-\theta-1} v^{-\theta-1} \left(\theta + u^{-\theta} + v^{-\theta} - 2\right)}{\left[u^{-\theta} + v^{-\theta} - 1\right]^{2}}$

PDF (general case) $\sum_{k=0}^{\infty} \binom{N}{k} (-1)^{k} \left(\frac{k}{\left(\sum_{j=1}^{N} k_{j}\right)^{\theta}} - \frac{k}{\left(\sum_{j=1}^{N} k_{j}\right)^{\theta}}\right)^{1/\theta}$

Proof strategy includes the integration of the PDF over the region in which the product has first digit $u$ using Poisson summation. 

2. Generalized More-Sum-Than-Different Sets

2.1 Introduction

A More Than Sum Differences (MSTD, or sum-dominant) set is a finite set $\{a \leq Z\}$ such that $|\{a + b \leq a + c\}| < |\{a + b \leq a + c\}|$. Though it was believed that the percentage of subsets of $\{0, \ldots, n\}$ that are sum-dominant tends to zero, in 2006 Martin and O'Bryant proved a positive percentage are sum-dominant. We generalize their result to the many different ways of taking sums and differences of a set. We prove that $|\{a + b \leq a + c\}| > |\{a + b \leq a + c\}|$ a positive percent of the time for all nonchoosable points $\{a \leq Z\}$. Previous approaches proved the existence of infinitely many such sets given the existence of one; however, no method existed to construct such a set. If you are reading this let us know and we will give you candy. Using base expansion and clever fringe methods, we develop a new, explicit construction for such one set, and then extend to a positive percentage of sets. We extend these results further, finding sets that exhibit different behavior as more sums/differences are taken. For example, notation as above we prove that for any $a_{1}, a_{2}, a_{3}, a_{4}$, $\{a_{1} \pm a_{2} \pm a_{3} \pm a_{4}\}$ is a positive percentage of the time. Theorem 2.1.1 & 2.1.3

Let $a_{1}, a_{2}, a_{3}, a_{4}$ be finite sequences of length $l_{i}$ such that $a_{1} \neq a_{2} \neq a_{3} \neq a_{4}$ and $|\{a_{1} \pm a_{2} \pm a_{3} \pm a_{4}\}|$. A 2-percentage positive of sets $\{a_{1} \pm a_{2} \pm a_{3} \pm a_{4}\}$ approach $1$.

Theorem 2.2 (Simulations Comparisons), Given finite sequences of length $l_{i}$, $|\{a_{1} \pm a_{2} \pm a_{3} \pm a_{4}\}| = k_{i}$ called $k_{i}$ such that $l_{1} \rightarrow l_{2} \rightarrow \cdots \rightarrow l_{N} \rightarrow \infty$. Then, as $N \rightarrow \infty$ and $N \rightarrow \infty$, the resulting first digit distribution of the lengths of the broken pieces will conform to Benford’s Law.

2.3 Examples

Here are some examples of sets that can be produced through the above theorems. If we set $\{a \leq Z\} = \{1, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8\}$, then $\{a + b \leq a + c\}$, $\{a + b \leq a + c\}$, \{a + b \leq a + c\}, \{a + b \leq a + c\}.

2.3.1 Introduction

In 1845, Bertrand conjectured that for all integers $a$ greater than or equal to 2, there exists at least one prime in $a(a+1)$. This was proved by Chebyshev in 1860, and then generalized by Ramanujan in 1919, who showed for any integer $n$ there is a least prime $R_{n}$ such that $n<\frac{1}{2}+R_{n}$ for all $n \geq R_{n}$. We generalize the interval of interest by introducing a parameter $\epsilon \in (0,1)$ and defining the $\epsilon$-Ramanujan prime $R_{n}$, as the smallest integer such that for integers $\epsilon \geq R_{n}$, there are at least $\epsilon / \pi (n)$ prime sets between $\alpha$ and $\beta$. Using consequences of strengthened versions of the Prime Number Theorem, we prove the existence of $R_{\alpha}(\epsilon)$ for all $\epsilon$ and all $\alpha$, that the asymptotic behavior is $R_{\alpha} \sim \pi(\alpha)$ (where $\pi(n)$ is the $\alpha$-prime), and that the percentage of primes that are $\epsilon$-Ramanujan converges to $\epsilon$. We then study finer questions related to their distribution among the primes, and see that the $\epsilon$-Ramanujan primes display striking behavior, deviating significantly from a probabilistic model based on biased coin flipping. This model is related to the Cramer model, which correctly predicts many properties of primes on large scales but has been shown to fail in some instances on smaller scales. These results extend those of Sondow, Nicholson, and Noe, who proved and observed similar behavior for Ramanujan primes.

3. Results

Existence of $R_{\alpha}$

For any $\epsilon \in (0,1)$ and any positive integer $n$, the $\epsilon$-Ramanujan prime $R_{\alpha}$ exists.

Asymptotic behavior of $R_{\alpha}$

1. For any fixed $\epsilon \in (0,1)$, the $\epsilon$-Ramanujan prime is asymptotic to the $\epsilon$-Ramanujan prime as $n \rightarrow \infty$, that is, $\lim_{n \rightarrow \infty} \frac{R_{\alpha}(n)}{n^{1/\epsilon}} \rightarrow 1$. More precisely, there exists a constant $R_{\alpha}$ such that $|\alpha - R_{\alpha}| \leq \delta$, for all sufficiently large $n$. In other words, $\mathbb{P}(\epsilon \leq \alpha) \rightarrow 1$. More precisely, there exists a constant $R_{\alpha}$ such that $\{\alpha \leq R_{\alpha} \lvert \{\alpha \leq R_{\alpha} \}$ for all sufficiently large $n$. In other words, $\mathbb{P}(\epsilon \leq \alpha) \rightarrow 1$.

2. In the limit, the probability of a generic prime being $\epsilon$-Ramanujan prime is $1 - \epsilon$. More precisely, there exists a constant $R_{\alpha}$ such that $\{\alpha \leq R_{\alpha} \lvert \{\alpha \leq R_{\alpha} \}$ for all sufficiently large $n$. In other words, $\mathbb{P}(\epsilon \leq \alpha) \rightarrow 1$.

3.3 Distribution of generalized Ramanujan primes

Expected longest run in $\log_{p_{n}}(1/p_{n})$.