Title
THE RIEMANN ZETA DISTRIBUTION

Author
Peltzer, Adrien Remi

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Title of the Thesis

THE RIEMANN ZETA DISTRIBUTION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Adrien Peltzer

The Riemann Zeta Distribution Committee:
Professor Michael C. Cranston, Chair
Professor Nathan Kaplan
Professor Abel Klein

2019

DEDICATION

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, and everyone in the Math dapartment at UCI. I'd also like to specifically thank Jiancheng Lyu for helping me prepare for my advancement talk, as well as John Peca-Medlin for proof-reading this manuscript.

CURRICULUM VITAE

Adrien Peltzer

EDUCATION

Doctor of Philosophy in Mathematics **2019**
University of California, Irvine *Irvine, California*

Bachelor of Science in Mathematics **2011**
University of California, Riverside *Riverside, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2013–2019**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2013–2019**
University of California, Irvine *Irvine, California*

SOFTWARE

Python

Experience using Numpy, Pandas, Regular Expressions, parsing websites, simulating random variables, and more.

ABSTRACT OF THE THE RIEMANN ZETA DISTRIBUTION

Title of the Thesis

By

Adrien Peltzer

Doctor of Philosophy in Mathematics

University of California, Irvine, 2019

Professor Michael C. Cranston, Chair

The Riemann Zeta distribution is one of many ways to sample a positive integer at random. Many properties of such a random integer X, like the number of distinct prime factors it has, are of concern. The zeta distribution facilitates the calculation of the probability of X having such and such property. For example, for any distinct primes p and q, the events $\{p \text{ divides } X\}$ and $\{q \text{ divides } X\}$ are independent. One cannot say this if instead, X were chosen randomly according to a geometric, poisson, or uniform distribution on the discrete interval $[n] = \{1, ...n\}$ for some n. Taking advantage of such facilities, we find a formula for the moment generating function of $\omega(X)$, and $\Omega(X)$, where $\omega(n)$ and $\Omega(n)$ are the usual prime counting functions in number theory. We use this to prove an Erdos-Kac like CLT for the number of distinct and total prime factors of such a random variable. Furthermore, we obtain Large Deviation results for these random variables, as well as a CLT for the number of prime factors in any arithmetic progression. We also investigate some divisibility properties of a Poisson random variable, as the rate parameter λ goes to infinity. We see that the limiting behavior of these divisibility properties is the same as in the case of a uniformly chosen positive integer from $\{1, .., n\}$, as $n \to \infty$.

Chapter 1

Introduction

The reader familiar with fundamental concepts in probability and number theory may skip directly to Section 1.3.

1.1 Background in Probability

Suppose you have an experiment you can repeat as often as you like. Suppose further that each outcome has a certain value, a real number, attached to it (e.g. measurement of some length, or counting number of items). For example, you go fish everyday for one year in Alaska and count the number of catches per day. If you observe the experiment n times and let X_1 be the value on the first day, X_2 that on the second, and so on, then you have a sequence $X_1, ..., X_n$ of *random variables*.

Suppose that every trial is independent and identically distributed (you catch the same number of fish in January and September). Suppose, furthermore, the *expectation*, μ, and the *standard deviation*, σ, of a given X_i are known. Let

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

be the *sample mean*, or average, of your observed data. The expected value of $\overline{X}$ must also be μ. In fact, $\overline{X}$ should fall closer to μ than any one data point X_i. It should also get better as n gets larger.

By independence, it is easy to see that the variance is

$$\text{var}(\overline{X}) = (\frac{1}{n})^2\text{var}(\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 n\text{var}(X_1) = \frac{\sigma^2}{n}. \tag{1.1}$$

If we standardize $\overline{X}$, we get a new random variable $Z = \frac{\overline{X}-E\overline{X}}{\sqrt{\text{var}\overline{X}}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. This random variable now has mean 0 and variance 1.

So we have the average, $\overline{X}$, the expected value, μ, and variance, σ^2/n, of $\overline{X}$, and a rescaled version of it, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$.

The (strong) law of large numbers simply states that the average, $\overline{X}$, converges to the mean, μ, with probability 1 (see Theorem 1.1). The Central Limit Theorem (see Theorem 1.2), on the other hand, states that Z tends to be normally distributed with mean 0 and variance 1. A key point is that we assume we know the underlying mean and standard deviation of our data. Otherwise, we'd have to use statistical methods to estimate them. Figure 1.1 shows the density function of Z.

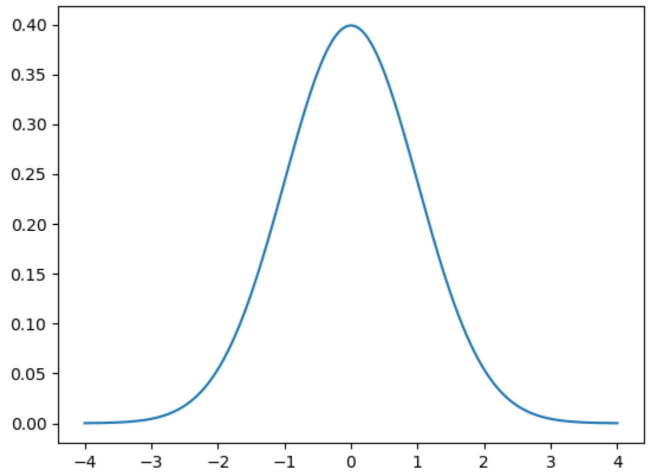

Figure 1.1: The standard Bell Curve

This is the density function $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. A well known property of $f(z)$ is that about 68% of the area under the curve lies between -1 and 1, and about 95% lies between -2 and 2, or within two standard deviations of the mean.

1.1.1 Basic Definitions

We review here some basic definitions. Most of the definitions will follow those found in Durrett ([6]). We assume the reader has some background in probability. Let X be any random variable in any probability space. Then the *expected value* of X is defined as

$EX = \sum_i iP(X = i)$

if X takes on only countably many values, and

$EX = \int_{-\infty}^{\infty} xf(x)dx$

if X is continuous with density function $f(x)$. The *variance* of X is

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2.$$

The nth *moment* of X is EX^n. The *moment generating function* (MGF) of X is defined as

$$M(t) = Ee^{tX}, \ t \in \mathbb{R}.$$

It is called "moment generating" because it generates its moments. It is the exponential generating function for the sequence $a_n = EX^n$:

$$Ee^{tX} = \sum_{n=0}^{\infty} EX^n \frac{t^n}{n!}.$$

If $M(t)$ exists for some interval (a, b) in $\mathbb{R}$, then it completely characterizes the distribution of X. We also mention two commonly used random variables throughout this paper.

Definition 1.1. *Let $\lambda > 0$. Then X is said to be Poisson distributed with parameter λ, if*

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \ k = 0, 1, 2, ...$$

Definition 1.2. *Suppose that $N \sim$ Poisson (λ), and that $X_1, X_2, ...$ are identically distributed random variables that are mutually independent and also independent of N. Let Y denote the sum*

$$Y = \sum_{i=1}^{N} X_i.$$

Then Y is said to follow a compound Poisson distribution.

1.1.2 Classical Theorems

We state a version of the weak law of large numbers given in [6], Theorem 2.2.3.

Theorem 1.1. *(Weak Law of Large Numbers) Let $X_1, X_2, ...$ be uncorrelated random variables with $EX_i = \mu$ and $var(X_i) \leq C < \infty$. If $S_n = X_1 + ... + X_n$ then as $n \to \infty$, $P\{|S_n/n - \mu| \geq \epsilon\} \to 0$ for every $\epsilon > 0$.*

We state a version of the Central Limit Theorem which allows for the random variables X_i to have different distributions (see [6], Theorem 3.4.5).

Theorem 1.2 (Lindeberg-Feller)**.** *For each n, let $X_{n,m},\ 1 \leq m \leq n$, be independent random variables with $EX_{n,m} = 0$. Suppose*

$$(i)\ \sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$$

$$(ii)\ For\ all\ \epsilon > 0,\ \lim_{n\to\infty} \sum_{m=1}^{n} E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0.$$

Then $S_n = X_{n,1} + ... + X_{n,n} \to \sigma\chi$ as $n \to \infty$. Here, $\chi \sim N(0,1)$.

One common approach to proving the central limit theorem is Levy's Continuity Theorem (see, for example [6], Theorem 3.3.6), which relates the convergence of the characteristic functions of the distributions to the convergence of the random variables themselves. It states the following:

Theorem 1.3. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables. Define $\phi_n(t) = Ee^{itX_n}$ to be their corresponding characteristic functions. If the characteristic functions converge point-wise to a function ϕ, that is*

$$\phi_n(t) \to \phi(t) \quad \forall\ t \in \mathbb{R}$$

then X_n converges in distribution to X, a random variable with ϕ as its characteristic function.

The normal distribution has characteristic function $\phi(t) = e^{-t^2/2}$. In a more general sense of the word, a central limit theorem is a theorem that identifes the limiting distribution of the running average of a sequence of random variables. The limiting distribution isn't always normal.

1.1.3 Large Deviations and Cramér's Theorem

Suppose $\{X_1, X_2, ...\}$ are taken to be independent mean 0, variance 1 normal random variables. Then $\overline{X}$ is also normal, with mean 0 and by (1.1), the variance is $1/n$. This implies, for any $x > 0$,

$$P(|\overline{X}| \geq x) \to 0$$

as $n \to \infty$.

Since $\overline{X} \sim \mathcal{N}(0, 1/n)$, we can standardize it $\sqrt{n}\overline{X} \sim \mathcal{N}(0, 1)$. Therefore,

$$P(\sqrt{n}\overline{X} \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-t^2/2} dt \tag{1.2}$$

for any measurable set $A \subseteq \mathbb{R}$.

Note now that

$$P(|\overline{X}| \geq x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-x\sqrt{n}}^{x\sqrt{n}} e^{-t^2/2} dt; \tag{1.3}$$

therefore, by taking limits and using l'hopital's rule,

$$\frac{1}{n} \log P(|\overline{X}| \geq x) \to -x^2/2. \tag{1.4}$$

Equation (1.4) is an example of a large deviations statement: The "typical" value of $\overline{X}$ is, by (1.2), of the order $1/\sqrt{n}$, but the probability it is bigger than any $x > 0$ is of the order $e^{-nx^2/2}$.

(1.2) and (1.3) above hold for any underlying distribution of the X_i, not necessarily normal, but with the equality in (1.2) replaced by $\to_{n\to\infty}$. This is what leads to the notion of a large deviation principle.

Definition 1.3. *We say that a sequence of random variables* (X_n) *with state space* X *satisfies a* large deviation principle *(LDP) with speed* a_n *and rate function* $I : X \to \mathbb{R}^+$ *if*

(1) I *is lower semi-continuous and has compact level sets* $N_L := \{x \in X : I(x) \leq L\}$ *for every* $L \in \mathbb{R}^+$

(2) for every open set $G \subseteq X$,

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in G) \le -\inf_{x\in G} I(x)$$

(3) for every closed set $A \subseteq X$,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P(X_n \in G) \ge -\inf_{x\in G} I(x)$$

EXAMPLE 1.1. *Let $\{X_1, X_2, ...\}$ be iid Poisson random variables with parameter 1. If we take the speed $a_n = 1/n$, the corresponding rate function is $I(x) = x \ln x + 1 - x$.*

Let $x > 1$ and $S_n = \sum_{i=1}^{n} X_i$. Then for any $t > 0$.

$$\begin{aligned} P(S_n \ge nx) = P(e^{tS_n} \ge e^{tnx}) \le & \frac{Ee^{tS_n}}{e^{tnx}} \\ = & e^{-tnx}(e^{e^t-1})^n \end{aligned}$$

since the X_i have mgf e^{e^t-1} and are indepedent. Taking logarithms and dividing by n yields

$$\frac{1}{n} \log P(S_n \ge nx) \le -tx + e^t - 1$$

which is true for all $t > 0$. Taking derivative to minimize the right hand side over t yields $0 = -x + e^t$ and so

$$\frac{1}{n} \log P(S_n \ge nx) \le -x \ln x + x - 1.$$

One can then show that both

$$\overline{\lim}_{n\to\infty}\frac{1}{n}\log P(S_n \geq nx) = \underline{\lim}_{n\to\infty}\frac{1}{n}\log P(S_n \geq nx) = -x\ln x + x - 1$$

and so

$$\lim_{n\to\infty}\frac{1}{n}\log P(S_n \geq nx) = -x\ln x + x - 1 \tag{1.5}$$

So the rate function is thus $I(x) = x\ln x + 1 - x$.

1.1.4 Moderate Deviations

A *moderate deviations principle* (MDP) is a LDP with the following conditions:

$$\frac{1}{b_n}\sum_{i=1}^{n}(X_i - EX_i) \qquad \frac{b_n}{n}\downarrow 0,\ \frac{b_n}{\sqrt{n}}\uparrow\infty$$

$$\overline{\lim}_{n\to\infty}\frac{n}{b_n^2}\log(nP(|X_1| > b_n)) = -\infty$$

$$I(t) = \frac{1}{2EX_1^2}t^2$$

A moderate deviation is a large deviation where the speed function a_n is of the form $a_n \sim n^\alpha$ for some $1/2 < \alpha < 1$. Formally, there is no distinction between a MDP and a large deviation principle. Usually a large deviation principle gives a sort of rate for which the average of a sequence of iid random variables converges to its mean. This is done by providing a probability of the form $e^{-nI(x)}$ for the deviation from the mean M_n. MDPs, on the other hand, describe the probabilities on a scale *between* a law of large numbers and some sort of central limit theorem. For both, large deviations principles and MDPs the three points listed under the bullets serve as a definition. However, there are differences.

Typically, the rate function in a large deviation principle will depend on the distribution of the underlying random variables, while an MDP inherits properties of both the central limit behavior as well as of the large deviation principle.

For example, one often sees the exponential decay of moderate deviation probabilities which is typical of the large deviations. On the other hand the rate function in an MDP quite often is "non-parametric" in the sense that it only depends on the limiting density in the CLT for these variables but not on individual characteristics of their distributions.

Often even the rate function of an MDP interpolates between the logarithmic probabilities that can be expected from the central limit theorem and the large deviations rate function - even if the limit is not normal.

The large deviation principle is more intuitively explained in the following simple setting. Suppose $X_1, X_2, ...$ are iid random variables with mean μ. Let $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The (strong) law of large numbers tells us that $P(\lim_{n\to\infty} M_n = \mu) = 1$, but does not tell us how fast this rate of convergence occurs. Now suppose the sequence $X_1, X_2, ...$ satisfies a LDP with speed $a_n = n$ and rate function $I(x)$. By Cramér's theorem, this rate function always exists and equals the lengendre transform of the logarithm of the moment generating function of X_1. The LDP

$$\lim_{n\to\infty} \frac{1}{n} \ln P(M_n > x) = -I(x)$$

means essentially that, for large n,

$$P(M_n > x) \approx e^{-nI(x)}$$

for all $x > \mu$.

So it gives a rate at which $P(\lim M_n = \mu)$ goes to 1, in a sense. Let us move on to Cramér's Theorem.

Theorem 1.4. *(Cramér). Suppose $X_1, X_2, ...$ are independent and identically distributed random variables with common mean μ. Suppose also that the logarithm of the moment generating function for the X_i is finite for all $t \in \mathbb{R}$. That is,*

$$\Lambda(t) := \log Ee^{tX_1} < \infty \text{ for all } t.$$

Let Λ^* be the Legendre transform of Λ:

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}}(tx - \Lambda(t))$$

Then

$$\lim_{n \to \infty} \frac{1}{n} \log(P(\sum X_i \geq nx)) = -\Lambda^*(x).$$

In other words, the random variables satisfy a LDP with speed $a_n = n$ and rate function $I(x) = \Lambda^*(x)$. This result was discovered by Harald Cramér in 1938.

1.2 Background in Number Theory

1.2.1 The counting functions $\omega(n)$ and $\Omega(n)$

On average, how many distinct prime factors does a positive integer have? This amounts to calculating the *average*, or *normal* order of the counting function $\omega(n)$.

DEFINITION 1.4. *Let n be a positive integer. Then we define $\Omega(n)$ as the number of* total *prime factors of n, and $\omega(n)$ the number of* distinct *prime factors.*

For example, $n = 12 = 2^2 * 3$ gives $\Omega(12) = 3$ and $\omega(12) = 2$, whereas for $n = 23$, a prime, both equal 1.

DEFINITION 1.5. *Let $f : \mathbb{N} \to \mathbb{C}$. The average, or normal order of $f(n)$ is an elementary function $g(n)$ such that, as $n \to \infty$, $g(n) \sim \frac{1}{n}\sum_{i=1}^{n} f(i)$.*

EXAMPLE 1.2. *The average order of $\omega(X_s)$ is* $\log\log n$ *([11], Theorem 431). That is,*

$$\frac{1}{x}\sum_{n\leq x}\omega(n) \sim \log\log x.$$

This is what is meant by the *normal* order of the counting function $\omega(n)$. It means that, a number on the order of $e^{e^3} \approx 5.28 * 10^8$ usually has 3 distinct prime factors. This astonishing result means that numbers like $2^6 * 3^4 * 104707 \approx 5.42 * 10^8$ are "normal", in the sense that it is on the same order as e^{e^3}, *and* has exactly 3 distinct prime divisors (Side note: $e^{e^3} = 528,491,311.485....$ The two nearest integers are 528491311 and 528491312. They have three and five distinct prime divisors respectively. This was checked using sage's integer factorization function).

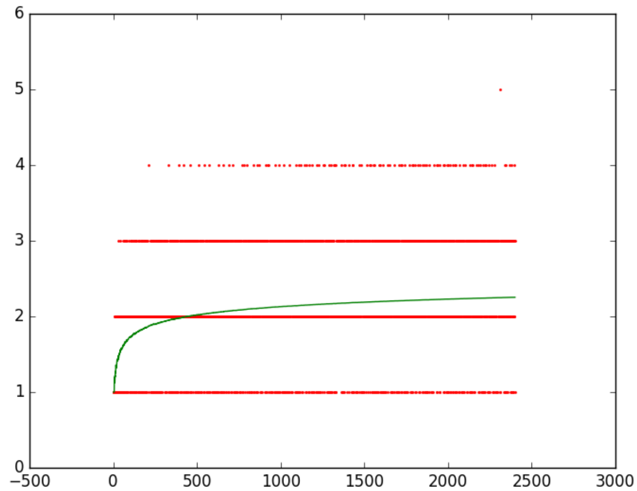

Figure 1.2: Plot of $\omega(n)$ and the cumulative average of $\omega(n)$

The scatterplot shows values of $\omega(n)$ for $n = 2, ..., 2400$. The single point at the top right is for $n = 2 \times 3 \times 5 \times 7 \times 11 = 2310$. The green plot is the cumulative average of the function: $\frac{1}{n}\sum_{k \le n} \omega(k)$. It is asymptotic to $\log\log n$.

Since there are infinitely many prime numbers, $\Omega(n) = \omega(n) = 1$ infinitely often. On the other hand, every time we pass a new primorial number (numbers of the form 2, $2*3$, $2*3*5$, $2*3*5*7$, ... see `http://oeis.org/A002110`) the function $\omega(n)$ attains a new running maximum. That is, for such a number n in this list, $\omega(k) < \omega(n)$ for all $k < n$. Both functions $\omega(n)$ and $\Omega(n)$ are oscillatory and *slowly* growing. In fact, they grow asymptotically at the order of $\log\log n$. Plots of the first few values of the two counting functions are shown in Figure 1.3.

1.2.2 Hardy-Ramunajan Theorem

In 1917, Hardy and Ramunajan proved in a precise way the statement

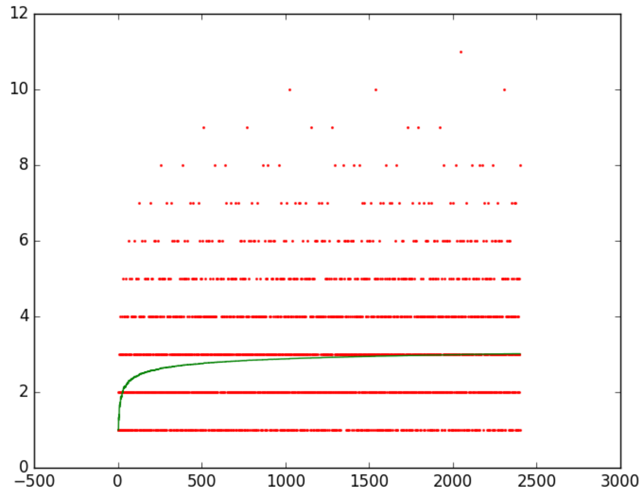

Figure 1.3: Plot of $\Omega(n)$ and the cumulative average $\Omega(n)$

The scatterplot shows values of $\Omega(n)$ for $n = 2, ..., 2400$. The maximum of these values occurs at $n = 2^{11} = 2048$, with $\Omega(2^{11}) = 11$. The next best is $\Omega(n) = 10$, which occurs at three places: $n = 1024, 1536$, and 2304 (respectively $2^{10}, 2^9 \times 3$, and $2^8 \times 3^2$). The green plot is the cumulative average of the function: $\frac{1}{n}\sum_{k\leq n}\Omega(k)$

"Almost every integer m has approximately $\log\log m$ prime divisors" ([12], Section 4.4)

In direct analogue to the law of large numbers, they proved the following

Theorem 1.5. *Let g_n be any sequence of real numbers such that $\lim_{n\to\infty} g_n = \infty$, and let l_n denote the number of integers, $1 \leq m \leq n$, for which either*

$$\omega(m) < \log\log m - g_m\sqrt{\log\log m}$$

or

$$\omega(m) > \log\log m + g_m\sqrt{\log\log m}.$$

Then

$$\lim_{m\to\infty}\frac{l_m}{m} = 0.$$

This is an analogue of the law of large numbers because it says that the asymptotic, or *relative density* of the set $\{l_n : n \in \mathbb{N}\}$, those numbers that deviate from the mean, is zero. The complement of $\{l_n : n \in \mathbb{N}\}$ in $\mathbb{N}$ has *relative density* 1(see definition 1.9).

1.2.3 Erdős-Kac Theorem

Just as the Hardy-Ramunajan is the analogue of the weak law of large numbers, the Erdős-Kac theorem plays the role of Central Limit Theorem. In fact, it was what led Kac to believe a central limit theorem should hold for the number of prime factors of a large positive integer. Specifically, this theorem states that the average, or normal order of the counting function $\omega(n)$ converges to a normal distribution (See [7] for the original paper).

Theorem 1.6. *Let a, b be real numbers with $a < b$, then as $x \to \infty$,*

$$\frac{1}{x}\#\{n \le x : \ a \le \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \le b\} \to \phi(b) - \phi(a),$$

where ϕ is the standard normal distribution.

Note that Theorem 1.6 implies Theorem 1.5, since, if $g_n \to \infty$ as $n \to \infty$, then

$$\frac{n - \log n}{n} = \frac{1}{n}\#\{m \le n : |\frac{\omega(m) - \log\log m}{\sqrt{\log\log m}}| < g_n\} \to \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz = 1.$$

1.2.4 The Prime Number Theorem

The prime number theorem states that, if $\pi(x)$ counts the number of primes less than or equal to x, then as $x \to \infty$,

$$\pi(x) \sim \frac{x}{\log x}.$$

This theorem was first proved in 1896 by Jacques Hadamard and Charles Jean de la Vallee Poussin, both of whom made use of the Riemann Zeta function $\zeta(s)$ (see `https://en.wikipedia.org/wiki/Prime_number_theorem`). Many mathematicians had previously conjectured this asymptotic growth rate, including Peter Gustav Dirichlet. During the 20th century, several different proofs were found, including some that relied only on "elementary" methods, such as those given by Selberg and Erdős in 1949 [1].

The prime number theorem is equivalent to the statement that

$$p_n \sim n \log n,$$

as $n \to \infty$, where p_n is the nth largest prime number [11].

1.2.5 Dirichlet's Theorem on Primes in Arithmetic Progressions

An arithmetic progression with first term h and common difference k consists of numbers of the form

$$h + nk, \quad n = 0, 1, 2, ...$$

If h and k are *not* coprime, then the corresponding arithmetic progession will consist of at most one prime, depending on whether h is prime or not. Therefore, we restrict our attention to the case where the greatest common divisor, $(h, k) = 1$. In this case, Dirichlet proved that the set $\{h+nk \mid n=0,1,2,...\}$ contains infinitely many primes (see [1], Theorem 7.3).

Theorem 1.7. *If $k>0$ and $(h,k)=1$, then for all $x>1$,*

$$\sum_{p\leq x,\quad p\equiv h \mod k} \frac{\log p}{p} = \frac{1}{\phi(k)}\log x + O(1),$$

where $\phi(k)$ is Euler's phi function. The sum is taken over all primes less than or equal to x that are congruent to $h \mod k$.

The first thing to note (trivially) is that this implies there are infinitely many primes in this general congruence class $\{h+nk \mid n=0,1,2,...\}$. For if there were a finite number, then the series would converge and hence not go to infinity with $\log x$. The other thing to note is the right hand side is independent of h. That is, for any h relatively prime to k, the sum is asymptotic to $\frac{1}{\phi(k)}\log x$ as $x\to\infty$. Since there are exactly $\phi(k)$ such numbers h, this theorem tells us the prime numbers are *equidistributed* among the $\phi(k)$ congruence classes. They all contain the same proportion of prime numbers. A more probabilistic interpretation of this result would be the following. Pick a prime number p "at random". What is the probability it is congruent to $1,3$, or $5 \mod 6$? The answer would be $1/3$ in each case. That is, p is uniformly distributed over the $\phi(6)=3$ (reduced) congruence classes $\mod 6$.

1.2.6 Dirichlet Series and the Riemann Zeta function

A *Dirichlet series* is a series of the form

$$\sum_{n=1}^{\infty}\frac{f(n)}{n^s},$$

where $f(n)$ is any arithmetical function, and s is any comlex number. The simplest example of a Dirichlet series is

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

This is the *Riemann Zeta function.* It is convergent for $s > 1$. In particular,

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$

a result due to Euler in 1735. The Riemann Zeta function has been studied in great detail because of its close connection with the distribution of prime numbers. A remarkable identity due to Euler is

$$\zeta(s) = \prod_{p} \frac{1}{1 - p^{-s}} \tag{1.6}$$

for all $s > 1$.

Let us determine the behavior of $\zeta(s)$ as s approaches 1 from the right. We can write it in the form

$$\zeta(s) = \sum_{1}^{\infty} n^{-s} = \int_{1}^{\infty} x^{-s} dx + \sum_{1}^{\infty} \int_{n}^{n+1} (n^{-s} - x^{-s}) dx.$$

Here,

$$\int_1^\infty x^{-s} dx = \frac{1}{s-1},$$

since $s > 1$. Also

$$0 < n^{-s} - x^{-s} = \int_n^x st^{-s-1} dt < \frac{s}{n^2},$$

if $n < x < n+1$, and so

$$0 < \int_n^{n+1} (n^{-s} - x^{-s}) dx < \frac{s}{n^2}.$$

This shows

$$\zeta(s) = \frac{1}{s-1} + O(1), \tag{1.7}$$

as $s \downarrow 1$.

1.2.7 The Prime Zeta function

Definition 1.6. *The prime zeta function is defined as*

$$P(s) = \sum_p \frac{1}{p^s}, \ s > 1.$$

Certain values of $P(s)$ are

s	Approximate value for P(s)
1	$\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\cdots\to\infty.$
2	0.45224 74200 41065 49850 . . .
3	0.17476 26392 99443 53642 . . .
4	0.07699 31397 64246 84494 . . .
5	0.03575 50174 83924 25713 . . .
9	0.00200 44675 74962 45066 . . .

Even though there are far fewer terms in the sum for $P(s)$ (for every n terms in the sum for the riemann zeta function, there are approximately $1/\log n$ terms in the sum for the Prime zeta function, by the prime number theorem), the series is still divergent for $s = 1$. Using equation 1.6, we can write the log of $\zeta(s)$ as

$$\log\zeta(s) = \sum_p \sum_{m=1}^{\infty} \frac{p^{-ms}}{m} \tag{1.8}$$

In terms of the Prime Zeta function, this gives the relation

$$\log\zeta(s) = \sum_{j\geq 1} \frac{P(js)}{j}, \tag{1.9}$$

and by mobius inversion

$$P(s) = \sum_{j\geq 1} \mu(j) \frac{\log\zeta(js)}{j}. \tag{1.10}$$

We also define a useful function below.

Definition 1.7. *The* Von Mangoldt *function is defined as* $\Delta(n) = \log p$ *if* $n = p^m$ *for some* m*, and* 0 *otherwise.*

We can rewrite equation (1.9) in terms of the Von Mangoldt function. When $n = p^m$, we have $\frac{1}{n^s} = p^{-ms}$ and $\frac{\Delta(n)}{\log n} = \frac{\log p}{\log p^m} = \frac{1}{m}$. When n is not a prime power, $\Delta(n) = 0$. These details imply

$$\log \zeta(s) = \sum_{n=2}^{\infty} \frac{\Delta(n)}{n^s \log n} \tag{1.11}$$

We will use the above in Section 1.4.1. For now, we move onto an important part of the thesis. Below, I describe the main reasons for studying these number theoretic functions in a probabilistic setting.

1.3 Distributions defined on the natural numbers

How can one sample a positive integer at random so that all outcomes are equally likely? In general, there is no way to do this, because, if it were possible, there would be a common probability c of each outcome. But $\sum_{n=1}^{\infty} c = \infty$ doesn't allow for that. We must deal with the divergence of the series somehow. In general, any convergent series $\sum a_i$ whose coefficients satisfy $0 \leq a_i \leq 1$ defines a probability distribution on $\mathbb{N}$, given by

$$P(X = i) = \frac{1}{\sum_{i=1}^{\infty} a_i} a_i, \quad i = 1, 2, ...$$

Definition 1.8. *Let n be a positive integer and define X_n such that*

$$P(X_n = i) = \frac{1}{n}, \quad i = 1, ..., n.$$

Then X_n is said to follow a uniform distribution.

The uniform distribution allows for equally likely outcomes, but cannot be extended to be supported on all of the natural numbers. For that reason, one main way of studying the distribution of numbers in a probabilistic way is to take a uniform distribution up to n, and let n go to infinity. Done in this way, the definition of natural density of a set (defined below) transfers directly to the limit of the probability of an event.

Definition 1.9. *Let A be a subset of the natural numbers. We say A has natural density $\delta(A)$, $0 \leq \delta \leq 1$, if*

$$\frac{1}{n}|A \cap [n]| \to \delta(A)$$

as $n \to \infty$.

With this definition in place, it is immediately clear that the set of positive even integers has natural density 1/2, or the set of positive integers not divisible by 3 has natural density 2/3.

The notion of natural density is very important because, one, it is intuitive, and, two, it gives us a measure of how often a number has such and such property. It also gives merit to studying uniformly distributed integers from 1 to n and letting $n \to \infty$. This is because for any A with natural density $\delta(A)$,

$$P(X_n \in A) \to \delta(A)$$

as $n \to \infty$. Another distribution defined on the natural numbers is the harmonic distribution.

Definition 1.10. *Let Y_n be a random variable such that*

$$P(Y_n = i) = \frac{1}{h_n i},$$

where $h_n = \sum_{i=1}^{n} \frac{1}{i}$. Then Y_n is said to follow a harmonic distribution.

The harmonic distribution differs from the uniform in that it does not have equally likely outcomes, and puts more weight on smaller values. Also, for any subset of the natural numbers A, sending $n \to \infty$ and looking at the limiting probability $\lim_{n\to\infty} P(Y_n \in A)$ gives rise to, not the *natural* density of A, but the *logarithmic* density.

Definition 1.11. *We say that $A \subseteq \mathbb{N}$ has logarithmic density $\delta(A)$ if the limit $\lim_{n\to\infty} \frac{1}{\log n} \sum_{i \le n:\ i \in A} \frac{1}{i}$ exists and equals $\delta(A)$.*

If the natural density of a set A exists, then so does it's logarithmic density and both are the same. The converse is not necessarily true. Depending on the set of numbers in question, logarithmic density may be more practical to calculate than natural density. One example of this is the density of primitive subsets of $\mathbb{N}$ (see Behrend's Theorem `https://en.wikipedia.org/wiki/Behrend%27s_theorem`). Finally, we define another distribution on the natural numbers, this time with the property that each natural number occurs with positive probability.

Definition 1.12. *Let $s > 1$ be a real parameter. The random variable X_s follows a zeta distribution if the probability mass function is given by*

$$P(X_s = n) = \frac{1}{\zeta(s)} \frac{1}{n^s}, \quad n = 1, 2, ...$$

This distribution doesn't have a cutoff. Every $n \in \mathbb{N}$ has positive probability. However, it is not uniformly distributed. As in the harmonic case, smaller numbers are weighted more heavily.

An overarching theme in this thesis is that, in each of these three cases, X_n uniform, Y_n harmonic, and X_s zeta, limiting probabilities seem always converge to the same value.

1.3.1 Some properties of the zeta distribution

The Riemann Zeta function $\zeta(s)$ is a meromorphic function on the complex plane with a simple pole at $s = 1$. It can be analyzed using techniques in complex analysis and also give us theorems about the integers. It is a highly scrutinized function, due in part to the elusiveness of the Riemann Hypothesis. However, in this thesis we only concern ourselves with $\zeta(s)$ on the open interval $s \in (1, \infty)$. It is for this reason that we don't discuss such extraordinary topics as the Riemann Hypothesis, and integral formulas for $\zeta(s)$. We don't make use of it here. Several authors have made use of the Riemann Zeta Distribution. Lin and Hu [15] show more generally that any dirichlet series corresponding to a completely multiplicative function defines an infinitely divisible distribution. The main articles we are concerned about in this thesis are those of Arratia, Lloyd, and Gut discussed shortly. First, let us mention and prove some basic properties of the Riemann Zeta Distribution.

1.3.2 The prime exponents are geometrically distributed

Lemma 1.3.1. *For any n, the probability of the event $\{n|X_s\}$ (this is the event $\{n$ divides $X_s\}$, not to be confused with conditional probability) is given by summing over all multiples* *of n:*

$$P(n|X_s) = \zeta^{-1}(s)\sum_{k=1}^{\infty}\frac{1}{(nk)^s} = \zeta^{-1}(s)\frac{1}{n^s}\sum_{k=1}^{\infty}\frac{1}{k^s} = \frac{1}{n^s}.$$

Furthermore, for m and n coprime, the events $\{m|X_s\}$ and $\{n|X_s\}$ are independent.

Lemma 1.3.2. *Let m, n be positive integers such that $gcd(m,n) = 1$. Then*

$$P(m|X_s, n|X_s) = P(m|X_s)P(n|X_s).$$

Proof. For m and n coprime, $\{m|X_s$ and $n|X_s\}$ if and only if $\{mn|X_s\}$. Therefore, by lemma 1.3.1,

$$P(\{m|X_s\} \cap \{n|X_s\}) = P(\{mn|X_s\}) = \frac{1}{(mn)^s} = \frac{1}{m^s}\frac{1}{n^s} = P(\{m|X_s\})P(\{n|X_s\})$$

□

Theorem 1.8. *Denote the exponents in the prime factorization of X_s by $c_p(X_s)$, so that $X_s = \prod_p p^{c_p(X_s)}$. Then the exponents are independent and geometrically distributed, with*

$$P\{c_p(X_s) \geq k\} = \frac{1}{p^{ks}},\ k = 0, 1, 2, ... \tag{1.12}$$

Proof. Let n be a positive integer. Denote the exponents in its prime factorization by $n = \prod_p p^{c_p(n)}$, where all but finitely many of the $c_p(n)$ are nonzero. Using equation (1.6),

$$\begin{aligned} P\{c_p(X_s) = c_p(n) \text{ for all } p\} = P\{X_s = n\} &= \prod_p (1 - \frac{1}{p^s})(\frac{1}{\prod_p p^{c_p(n)s}}) \\ &= \prod_{p|n}(1 - \frac{1}{p^s})\frac{1}{p^{c_p(n)s}}\prod_{p\nmid n}(1 - \frac{1}{p^s}) \\ &= \prod_p h_p(n), \end{aligned}$$

where $h_p(n) = (1-\frac{1}{p^s})\frac{1}{p^{c_p(n)s}}$. We have written the joint probability $P\{c_p(X_s) = c_p(n) \text{ for all } p\}$ in factored form $\prod_p P\{c_p(X_s) = c_p(n)\}$. Therefore, the exponents $\{c_p(X_s)\}$ form an independent set, and have respective probability mass functions given by

$$P\{c_p(X_s) = k\} = (1 - \frac{1}{p^s})\frac{1}{p^{ks}}, \ k = 0, 1, 2, ... \tag{1.13}$$

□

Remark 1.3.1. *The fact that the exponents of X_s are independent and geometrically distributed is no surprise. It is kind of "baked in" due to the euler product formula (1.6) for the zeta function.*

1.3.3 Examples using a Theorem by Diaconis

The Theorem below due gives us a direct route to connect harmonic limits with zeta limits.

Theorem 1.9. *(Diaconis, [4], Theorem 1). Let $x = (x_1, x_2, ...)$ be any bounded sequence of numbers. Then*

$$\lim_n \frac{1}{\log n}\sum_{i=1}^{n}\frac{x_i}{i} = c \iff \lim_{s\to 1^+}(s-1)\sum_{i=1}^{\infty}\frac{x_i}{i^s} = c.$$

Corollary 1.3.1. *Let A be a subset of the natural numbers whose logarithmic density exists and equals $\delta(A)$. Define the sequence $x = (x_1, x_2, ...)$ such that $x_i = 1$ when $i \in A$, zero otherwise. Then,*

$$\lim_{s\downarrow 1} P\{X_s \in A\} = \delta(A).$$

Let us look at some examples.

Definition 1.13. *An integer n is said to be squarefree if its prime factorization consists* *only of distinct primes. That is, no square number divides n.*

Example 1.3. *The probability X_n is squarefree converges to $\frac{6}{\pi^2}$ (See [11], Theorem 333). This limit also holds for X_s as $s \to 1^+$.*

Proof. By Theorem 1.8 below, we have

$$\begin{aligned}
P\{X_s \text{ squarefree }\} =& \prod_p P\{c_p(X_s) = 0 \text{ or } 1\} \\
=& \prod_p (1 - P\{c_p(X_s) \geq 2\}) \\
=& \prod_p (1 - \frac{1}{p^{2s}}) \\
=& \frac{1}{\zeta(2s)} \\
\to& \frac{1}{\zeta(2)} = \frac{6}{\pi^2}
\end{aligned}$$

as $s \downarrow 1$. □

Remark 1.3.2. *In view of Theorem 1.9, we get the same limit for Y_n.*

Definition 1.14. *(Eulers' Totient function) Let n be any positive integer. Define $\phi(n)$ to be the number of positive integers $k \leq n$ that are relatively prime to n.*

Example 1.4. *The average value of $\phi(X_n)/n$ also converges to $6/\pi^2$. We review a proof of this in the appendix in Section A.2. Here, we show this holds for the zeta distribution X_s as $s \to 1^+$.*

Proof. Using a well known equation $\phi(n) = n \prod_{p|n} (1 - \frac{1}{p})$ ([1], Theorem 2.4), we have

$$\phi(X_s) = X_s \prod_{p|X_s} (1 - \frac{1}{p}) = X_s \prod_p (1 - \frac{1_{c_p(X_s)>0}}{p})$$

Taking expected value and using Theorem 1.8 ,

$$\begin{aligned} E\frac{\phi(X_s)}{X_s} &= \prod_p E(1 - \frac{1_{c_p(X_s)>0}}{p}) \\ &= \prod_p (1 - \frac{P(c_p(X_s) > 0)}{p}) \\ &= \prod_p (1 - \frac{1}{p^{s+1}}) \\ &\to \frac{1}{\zeta(2)} = \frac{6}{\pi^2} \end{aligned}$$

as $s \downarrow 1$. □

Remark 1.3.3. *In view of Theorem 1.9, we get the same limit for Y_n.*

We generalize definition 1.13 and the corresponding example 1.3.

Definition 1.15. *We say the n is k-th power free if no prime power of the form p^k divides n.*

Example 1.5. *Let k be any positive integer greater than 1. Then as $s \to 1^+$ the probability X_s is not divisible by any k-th power converges to $\frac{1}{\zeta(k)}$.*

Proof. By Theorem 1.8,

$$\begin{aligned} P\{X_s \text{ is } k\text{-th power free }\} &= \prod_p P\{c_p(X_s) < k\} \\ &= \prod_p (1 - P\{c_p(X_s) \geq k\}) \\ &= \prod_p (1 - \frac{1}{p^{ks}}) \\ &\to \frac{1}{\zeta(k)} \end{aligned}$$

as $s \downarrow 1$. $\square$

Now let $x_i = 1$ if i is k-th power free and zero otherwise. Then the sequence $x = (x_1, x_2, ...)$ is clearly uniformly bounded (by $M = 1$). We apply Corollary 1.3.1 and see that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{i\le n:\ p^k \nmid i\ \forall p} \frac{1}{i} = \frac{1}{\zeta(k)}.$$

Example 1.6. *Let $k \geq 0$. For each $i \geq 1$, set $x_i = 1$ if $\Omega(i) - \omega(i) = k$. Otherwise, set x_i equal to 0. By Corollary 1.3.1, we see that the logarithmic density, d_k, of the set $\{i \in \mathbb{N} :\ \Omega(i) - \omega(i) = k\}$ is the same as $\lim_{s\downarrow 1} P\{\Omega(X_s) - \omega(X_s) = k\}$, if the limit exists. We will show that the limit d_k exists in Section 3.1.1.*

1.4 The Poisson approximation to the count of prime factors

It is clear that a Poisson approximation to the number of prime factors of X_s is available. For the uniform case X_n, the following theorem already hints at Poisson (see [11], Theorem 437 or [18], Section 6.1).

Theorem 1.10. *(Landau) Let k be a positive integer, and let $\pi_k(x)$ denote the number of integers $n \leq x$ with exactly k prime factors. Then*

$$\pi_k(x) \sim e^{-\log\log x} \frac{(\log\log x)^{k-1}}{(k-1)!}$$

A Poisson random variable with parameter $\log\log x$ fits nicely as an approximation (at least for an approximation to the first order) of the number of prime factors of an integer. We can also see this in the case for X_s (see Theorem 2.1). Furthermore, several authors, including Lloyd, Arratia, and Gut, make use of a Poisson approximation to X_s to simplify calculations. We shall now mention how they do so.

Both Arratia ([2], Section 3.4.2.) and Lloyd ([16], equation (4)) use some form of the Poisson distribution to approximate the prime exponents in the random variable X_s. In the case of Arratia, he looks at geometric random variables Z_p with parameter $1/p$. Arratia uses that each geometric random variable $c_p(X_s)$ (with $s = 1$) can be written in the form $\sum_{k\geq 1} kA_{p,k}$, where the A_k are independent Poisson random variables with expectation $EA_k = \frac{1}{kp^k}$. This shows stochastic domination of the geometric $Z_p \geq_d A_1$. That is, $Z_p = A_1 + R$, where $Z_p \sim Geo(\frac{1}{p})$, $A_1 \sim Poisson(\frac{1}{p})$, and R is a non-negative random remainder term $\sum\limits_{k\geq 2} kA_k$ which is much smaller (in expectation) than A_1. Notice that $P(R = 0) = P(A_k = 0\ \forall\ k \geq 2)$. Notice also that $P(R = 1) = 0$ since $2A_2 = 0$ or is greater than or equal to 2. We will explain a connection with Lloyd's decomposition in Section 1.4.2.

Gut notices that the log of X_s is *compound Poisson* [10] .

1.4.1 The logarithm is compound Poisson

In [10], Gut defines a random variable V, taking values in the set of prime powers, such that

$$P(V = p^m) = \frac{1}{\log \zeta(s)} \frac{1}{mp^{ms}} \tag{1.14}$$

He then shows that

$$\log X_s \stackrel{d}{=} \sum_{i=1}^{N} \log V_i,$$

the random variables $\log V_i$ being *iid* with common distribution that of $\log V$, and N, independent of the $\log V_i$'s, is Poisson with parameter $\lambda = \log \zeta(s)$. That is, Gut shows that $\log X_s$ is compound poisson (see definition 1.2).

1.4.2 A useful factorization of the zeta distribution ($X_s = X_s' X_s''$)

Let Y_n follow a harmonic distribution as in 1.10. Lloyd uses an approximation of Y_n by the zeta random variable X_s. If we write $Y_n = q_1(Y_n)q_2(Y_n)...$,where $q_r(Y_n)$ is the rth largest prime factor of Y_n, then Lloyd showed that the random variable $\frac{\log q_r(Y_n)}{\log Y_n}$ converges to a random variable with distribution F_r, whose moments are given by

$$\int_0^{1/r} x^m dF_r(x) = \int_0^{\infty} \frac{[\xi(t)]^m}{m!} e^{-t} \frac{t^{r-1}}{(r-1)!} dt,$$

where $\xi(t)$ is the functional inverse of the exponential integral $E(x) = \int_x^{\infty} \frac{e^{-y}}{y} dy$.

This F_r is exactly the marginal of the rth component in the Poisson-Dirichlet distribution discussed in Section 1.4.4 below. In proving this, Lloyd used an approximation of the geometric random variable by a Poisson with the same parameter. He takes the quotient of the probability generating function of a geometric with parameter ρ, $f(x) = \frac{1-\rho}{1-\rho x}$, with the probability generating function of a Poisson random variable with the same parameter, $g(x) = e^{\rho(x-1)}$. The quotient equals

$$\frac{(1-\rho)e^{-(x-1)\rho}}{1-\rho x} = (1-\rho)e^{\rho}\sum_{m=0}^{\infty}\rho^m\sigma(m)x^m$$

where $\sigma(m) = \sum_{j=0}^{m}\frac{(-1)^j}{j!}$. If α is geometric rnadom variable with parameter ρ, then so is $\alpha' + \alpha''$, where α' is Poisson distributed with the same parameter ρ, and α'' has the distribution given by the above generating function. That is,

$$P(\alpha'' = m) = (1-\rho)e^{\rho}\rho^m\sigma(m),\ m = 0, 1, 2, ... \tag{1.15}$$

The utility of this decomposition is that, as $\rho \to 0$, $P\{\alpha'' > 0\} = P\{\alpha'' \geq 2\} = O(\rho^2)$, stemming from $\sigma(1) = 0$. We state Lloyd's decomposition as a theorem.

Theorem 1.11. *(Lloyd) Let N denote a zeta random variable with parameter s, and write $N = \prod_p p^{\alpha_p}$. Define two independent variables X_s' and X_s'' such that if $X_s' = \prod_p p^{\alpha_p'}$, and $X_s'' = \prod_p p^{\alpha_p''}$, then for each prime p,*

$$\begin{aligned} P\{\alpha_p' = k\} =& e^{-1/p^s}\frac{p^{-sk}}{k!}, \quad k = 0, 1, 2, ... \\ P\{\alpha_p'' = m\} =& (1-\frac{1}{p^s})e^{1/p^s}p^{-sm}\sigma(m),\ m = 0, 1, 2, ... \end{aligned}$$

where $\sigma(m)$ is defined as in (1.15).

Then $X_s = X_s'X_s''$ in distribution.

If $X_s'' = \prod_p p^{\alpha_p''}$, where the α_p'' have the distribution in Theorem 1.11, then X_s'' will be finite as $s \to 1^+$. Essentially this means that divergence of the zeta random variable X_s is carried by the random variable X'_s, whose exponents α'_p are Poisson distributed. Once again, we see the connection with Poisson.

1.4.3 Lloyd and Arratia use the same decomposition

Lloyd's decomposition of the geometric random variable $\alpha = \alpha' + \alpha''$ is similar to Arratia's. The only difference is that Lloyd's α'' equals $\sum_{j\geq 2} jA_j$. He only decomposes the geometric into a sum of *two* variables, one Poisson and one remainder term. This suffices for first order approximation of a geometric with parameter $\rho = \frac{1}{p^s}$ as $s \to 1^+$. But if we want a better approximation, we should analyze the behavior of the variables jA_j, as well.

1.4.4 The Poisson-Dirichlet Distribution

Let $C = [0,1]^\infty$ be the infinite-dimensional unit cube, and let U_i be *i.i.d.* uniform random variables in the interval $[0,1]$. Then the vector $(U_1, U_2, ...)$ is uniformly distributed on C. If we make the transformation

$$X_1 = U_1,\ X_2 = (1-U_1)U_2,\ X_3 = (1-U_1)(1-U_2)U_3,\ ...$$

(the general term is $X_i = U_i \prod_{j<i}(1-U_j)$) then $\sum_i X_i = 1$ almost surely [5] (see Figure 1.4) , and so the vector $(X_1, X_2, ...)$ lives on the infinite dimensional simplex $\Delta = \{x \in C | \sum x_i = 1\}$. The probability measure on Δ induced by this transformation of uniform measure on C is called the GEM distribution. If we order the $X_i's$, namely we write $X = (X_{(1)}, X_{(2)}, X_{(3)}, ...)$ where $X_{(i)} \geq X_{(i+1)}$ for each i, then we get what is called the Poisson-Dirichlet distribution on $T = \{x \in \Delta | x_1 \geq x_2 \geq x_3 ...\}$.

The result that the whole sequence $(\frac{\log q_1(Y_n)}{\log Y_n}, \frac{\log q_2(Y_n)}{\log Y_n}, \frac{\log q_3(Y_n)}{\log Y_n}, ...)$ converges to the Poisson

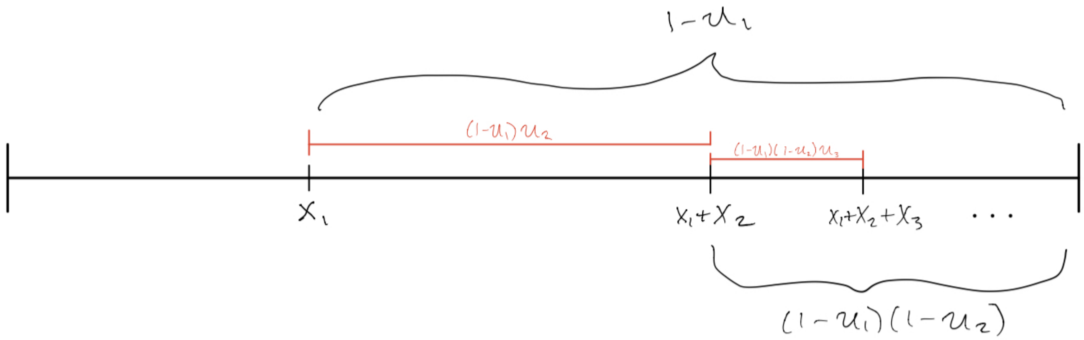

Figure 1.4: The sum of the variables X_i

This explains intuitively why $\sum_i X_i = 1$ a.s.

Dirichlet distribution was proven for the zeta, harmonic, and uniform case (in the appropriate limits). Many authors worked on this, including Billingsley, Lloyd, Knuth and Pardo, and Grimmett and Donnelly, and Arratia ([5],[14],[13]). This distribution also shows up as the limiting distribution for the fractional lengths of the cycles, in a uniformly random chosen permutation of n. In fact, there is more of a universality result at play. The component sizes of logarithmic combinatorial structures converge to the Poisson Dirichlet limit.

Chapter 2

Main Results

In this chapter, we derive formulas for the moment generating functions of $\omega(X_s)$ and $\Omega(X_s)$. We proceed to prove an Erdős-Kac like Central Limit Theorem for the number of prime factors of X_s as $s \to 1^+$, followed by a large deviations result for $\omega(X_s)$.

2.1 The moment generating functions of $\Omega(X_s)$ and $\omega(X_s)$

Theorem 2.1. *Let $A_j,\ j = 1, 2, ...$ be independent and Poisson distributed with expectation $EA_j = P(js)/j$, where $P(s)$ is as in (1.6). Then $\Omega(X_s)$ is equal in distribution to $\sum_{j\geq 1} jA_j$. Furthermore, for all $s > 1$, the moment generating functions of $\omega(X_s)$ and $\Omega(X_s)$ exist and are given by*

$$Ee^{t\Omega(X_s)} = \exp\{\sum_{j=1}^{\infty} \frac{P(js)}{j}(e^{tj}-1)\} \tag{2.1}$$

$$Ee^{t\omega(X_s)} = \exp\{\sum_{j=1}^{\infty}(-1)^{j+1}\frac{P(js)}{j}(e^{t}-1)^{j}\} \tag{2.2}$$

Proof. Note that $\omega(X_s)$ is a sum of bernoulli random variables $\sum_p 1_{c_p(X)>0}$. Write $X_s = \prod_p p^{c_p(X_s)}$ in its prime factorization, and define $P(s)$ as in (1.6).

Then

$$\begin{aligned}
Ee^{t\omega(X_s)} &= \prod_p Ee^{t1_{c_p(X)>0}} \\
&= \prod_p (e^t \tfrac{1}{p^s} + (1 - \tfrac{1}{p^s})) \\
&= \prod_p (1 + (e^t - 1)/p^s) \\
&= \exp\{\sum_p \log(1 + (e^t - 1)/p^s)\} \\
&= \exp\{\sum_p - \sum_{j=1}^{\infty}(-1)^j \tfrac{(e^t-1)^j}{jp^{js}}\} \\
&= \exp\{\sum_{j=1}^{\infty}(-1)^{j+1}\tfrac{P(js)}{j}(e^t - 1)^j\},
\end{aligned}$$

So we have (2.2). Equation (2.1) is derived similarly. Note that $\Omega(X_s)$ is a sum of independent geometric random variables $\sum_p c_p(X_s)$, whose resepctive moment generating functions are $Ee^{tc_p(X)} = \frac{1-1/p^s}{1-e^t/p^s}$. Using these and Theorem 1.8, we get

$$
\begin{aligned}
Ee^{t\Omega(X_s)} &= \prod_p \tfrac{1-1/p^s}{1-e^t/p^s} \\
&= \exp\{\sum_p \log(1-1/p^s) - \log(1-e^t/p^s)\} \\
&= \exp\{\sum_p -\sum_{j=1}^{\infty} \tfrac{1}{jp^{js}} + \sum_{j=1}^{\infty} \tfrac{e^{tj}}{jp^{js}}\} \\
&= \exp\{\sum_p \sum_{j=1}^{\infty} \tfrac{1}{jp^{js}}(e^{tj}-1)\} \\
&= \exp\{\sum_{j=1}^{\infty} \tfrac{P(js)}{j}(e^{tj}-1)\}.
\end{aligned}
$$

This proves (2.1). We proceed to show that $\Omega(X_s) \stackrel{d}{=} \sum_{j\geq 1} jA_j$. Using Arratia's fact that "every baby should know..." ([2], Section 3.4.2.), we show that $\Omega(X_s)$ is compound Poisson. This follows from the fact that (2.1) can be written in factored form

$$\prod_{j=1}^{\infty} \exp\{\frac{P(js)}{j}(e^{tj}-1)\}$$

and noticing that, if $A_j \sim Poisson(P(js)/j)$, then $Ee^{tjA_j} = e^{\frac{P(js)}{j}(e^{tj}-1)}$. □

Let us comment on a result from Theorem 2.1. Factoring the $j = 1$ term out in (2.1), we have that

$$Ee^{t\Omega(X_s)} = e^{P(s)(e^t-1)} \exp\{\sum_{j\geq 2} \frac{P(js)}{j}(e^{tj}-1)\}. \tag{2.3}$$

The first exponent on the right hand side corresponds to the moment generating function of a Poisson distribution with parameter $P(s)$, whereas the second exponent is the moment generating function of a random variable that stays finite as $s \downarrow 1$. So it is clear that $\Omega(X_s)$ is well approximated by a Poisson distribution with parameter $P(s)$ as $s \downarrow 1$.

Corollary 2.1.1. *Let $s > 1$. We can decompose $\Omega(X_s) = A_s + B_s$ into sums of independent random variables, where A_s is Poisson distributed with parameter $P(s)$, while B_s stays finite as $s \to 1^+$. The divergence of $\Omega(X_s)$ is carried by the Poisson variable A_s.*

Remark 2.1.1. *The decomposition in Corollary 2.1.1 is essentially the same as the one exhibited by Lloyd in Theorem 1.11. This is why Poissonian statistics arise in the results about $\Omega(X_s)$.*

The next section derives a normal law in the limiting behavior of the number of prime factors of X_s. Due to the similarities between equations (2.1),(2.2) and those in Flajolet (equations (1.2a),(1.2b) in [9]), it may be possible to deduce the gaussian limit from one of his theorems. There seems to be a more universal gaussian limit law for component counts of combinatorial structures, the number of "components" in the prime factorization of an integer just being a special case. We do not investigate the matter.

2.2 A Central Limit Theorem for the number of prime factors

Theorem 2.2. *Let $\hat{\omega}(X_s) = \frac{\omega(X_s) - P(s)}{\sqrt{P(s) - P(2s)}}$. Then*

$$\hat{\omega}(X_s) \xrightarrow{d} Z \sim \mathcal{N}(0,1),$$

as $s \to 1^+$. Here, $\hat{\omega}(X_s)$ is rescaled to have mean 0 and variance 1.

Proof. One first calculates that $E\omega(X_s) = \sum_p E1_{c_p(X_s)>0} = \sum_p P\{p|X_s\} = \sum_p 1/p^s = P(s)$. Also, by independence of the $c_p(X_s)$'s, we have $var\{\omega(X_s)\} = \sum_p var\{1_{c_p(X_s)}\} = \sum_p 1/p^s(1 - 1/p^s) = P(s) - P(2s)$.

Define $\hat{\omega}(X_s) = \frac{\omega(X_s)-P(s)}{\sqrt{P(s)-P(2s)}}$. Then

$$Ee^{t(\frac{\omega(X_s)-P(s)}{\sqrt{P(s)-P(2s)}})} = e^{-\frac{P(s)}{\sqrt{P(s)-P(2s)}}t} Ee^{\frac{t}{\sqrt{P(s)-P(2s)}}\omega(X_s)}$$

$$= e^{-\frac{P(s)}{\sqrt{P(s)-P(2s)}}t} \exp\{-\sum_{j\geq 1} \frac{P(js)}{j}(1 - e^{\frac{t}{\sqrt{P(s)-P(2s)}}})^j\}$$

$$= \exp\{-\frac{P(s)}{\sqrt{P(s)-P(2s)}}t - P(s)(1 - e^{\frac{t}{\sqrt{P(s)-P(2s)}}}) - \sum_{j\geq 2} \frac{P(js)}{j}(1 - e^{\frac{t}{\sqrt{P(s)-P(2s)}}})^j\}$$

$$= \exp\{P(s)(e^{\frac{t}{\sqrt{P(s)-P(2s)}}} - (1 + \frac{t}{\sqrt{P(s)-P(2s)}})) - \sum_{j\geq 2} \frac{P(js)}{j}(1 - e^{\frac{t}{\sqrt{P(s)-P(2s)}}})^j\}$$

Now let us take the limit as $s \to 1^+$ of the above.

The first two terms of the power series for $e^{\frac{t}{\sqrt{P(s)-P(2s)}}}$ are $(1 + \frac{t}{\sqrt{P(s)-P(2s)}})$. The next term is the leading term, $\frac{t^2}{2(P(s)-P(2s))}$. With the extra factor of $P(s)$ on the outside, that goes to $\frac{t^2}{2}$ as $s \to 1^+$, or $P(s) \to \infty$. The following terms all vanish. This shows that

$$Ee^{t\hat{\omega}(X_s)} \to e^{t^2/2},$$

the moment generating function of a standard normal random variable. By Levy's continuity theorem (Theorem 1.3) , we have $\hat{\omega}(X_s) \xrightarrow{d} Z \sim \mathcal{N}(0, 1)$. □

We use a different method of proof for the number of *total* prime factors.

Theorem 2.3. *Let* $\hat{\Omega}(X_s) = \frac{\omega(X_s)-P(s)}{\sqrt{Q(s)}}$*, where* $Q(s) = var\{\Omega(X_s)\} = \sum_p \frac{p^s}{(p^s-1)^2}$*. Then*

$$\hat{\Omega}(X_s) \xrightarrow{d} Z \sim \mathcal{N}(0,1),$$

as $s \to 1^+$*. Here,* $\hat{\Omega}(X_s)$ *is rescaled to have mean* 0 *and variance* 1*.*

Proof. Using Corollary 2.1.1, write $\Omega(X_s) = A_s + B_s$. Then

$$\hat{\Omega}(X_s) = \frac{A_s - P(s)}{\sqrt{Q(s)}} + \frac{B_s - P(s)}{\sqrt{Q(s)}}.$$

Since A_s is Poisson with parameter $P(s)$, its variance is $P(s)$. Using that $P(s) \sim Q(s)$ as $s \to 1^+$, the first part in the sum converges to a standard normal distribution by Theorem 1.2. The second sum goes to zero, since B_s stays finite as $s \to 1^+$. □

Once again, we see similarities between the limiting distribution of the zeta random variable X_s, and the uniform case X_n as $n \to \infty$. Theorem 2.2 is the analog of Theorem 1.6.

2.3 On the number of prime factors of X_s in an arithmetic progression

What if we consider only the number of prime factors of X_s in a given congruence class mod k for any k? Do we still get a central limit theorem for this new random variable? A corollary to Theorem 1.7 tells us that the primes are equidistributed among the $\phi(k)$ congruence classes $\{n : n \equiv h \mod k\}$, where $gcd(h,k) = 1$. Since the total number of prime factors, properly rescaled, converges to a standard normal random variable, the sum of the total prime factors in each congruence class converges to a standard normal. As the sum of normals is normal, we should expect that the limiting law of the rescaled $\omega(X_s)$ is a sum of $\phi(k)$ iid mean 0 variance $1/\phi(k)$ normals.

Let h, k be any coprime positive integers, and let $A_h = \{h + nk : n \in \mathbb{Z}^+\}$. Let

$$P_{A_h}(s) = \sum_{p \equiv h \mod k} \frac{1}{p^s}$$

be the restriction of the prime zeta function to the sum over primes congruent to $h \mod k$. Notice that if we enumerate the $\phi(k)$ numbers less than k relatively prime to k by $h_1, ..h_{\phi(k)}$, then

$$P(s) = P_{A_{h_1}}(s) + ... + P_{A_{h_{\phi(k)}}}(s).$$

It is reasonable to suspect that the prime counting function which counts the number of distinct prime divisors of n that are congruent to $h \mod k$ (let us denote this by $\omega_{A_h}(n)$) should *also* converge to a normal distribution.

Theorem 2.4. *Let $\omega_{A_h}(X_s)$ be the number of (distinct) primes congruent to $h \mod k$ that divide X_s. Then*

$$\hat{\omega}_{A_h}(X_s) = \frac{\omega_{A_h}(X_s) - P_{A_h}(s)}{\sqrt{P_{A_h}(s) - P_{A_h}(2s)}}$$

converges to a normal random variable with mean 0 *and variance* $\frac{1}{\phi(k)}$*. Furthermore, the random variables* $\hat{\omega}_{A_{h_i}}(X_s)$ *are iid, and* $\hat{\omega}(X_s) = \hat{\omega}_{A_{h_1}}(X_s) + ... + \hat{\omega}_{A_{h_{\phi(k)}}}(X_s)$.

Proof. This falls out directly from the formula for the moment generating function $Ee^{t\omega(X_s)} = \exp\{\sum_{m=1}^{\infty}(-1)^{m+1}P(ms)\frac{(e^t-1)^m}{m}\}$. Split the sum up as follows.

$$Ee^{t\omega(X_s)} = \exp\{\sum_{m=1}^{\infty}(-1)^{m+1}\sum_{i=1}^{\phi(k)}P_{A_{h_i}}(ms)\frac{(e^t-1)^m}{m}\}.$$

Then the exponential turns the sum into a product of mgf's,

$$Ee^{t\omega(X_s)} = \prod_{i=1}^{\phi(k)}\exp\{\sum_{m=1}^{\infty}(-1)^{m+1}P_{A_{h_i}}(ms)\frac{(e^t-1)^m}{m}\},$$

so we see that the $\omega_{A_{h_i}}(X_s)$'s are *independent* with respective moment generating functions given as the terms of the product above. Now by Dirichlet (Theorem 1.7), the primes are equidistributed among the $\phi(k)$ congruence classes corresponding to the h with $gcd(h,k) = 1$. Therefore,

$$P(s) = \sum_{h:\ (h,k)=1} P_{A_h}(s) = \phi(k)P_{A_1}(s)$$

as $s \downarrow 1$. So we have, for any h relatively prime to k, the expected value of $E\omega_{A_h}(X_s) \sim \frac{1}{\phi(k)}P(s)$. And since the variance of the sum of independent random variables is the sum of the variances,

$$P(s) - P(2s) = \sum_{h:(h,k)=1} P_{A_h}(s) - P_{A_h}(2s).$$

By Theorem 2.2,

$$\frac{\omega(X_s)-P(s)}{\sqrt{P(s)-P(2s)}} = \frac{\sum\limits_{h:\ (h,k)=1} \omega_{A_h}(X_s) - P_{A_h}(s)}{\sqrt{\phi(k)[P_{A_1}(s)-P_{A_1}(2s)]}} = \frac{1}{\sqrt{\phi(k)}} \sum_{h:\ (h,k)=1} \hat{\omega}_{A_h}(X_s) \to Z \sim \mathcal{N}(0,1).$$

Z is a sum of $\phi(k)$ identically distributed random variables $Z_1, ..., Z_{\phi(k)}$ which are the limiting distributions of $\frac{1}{\sqrt{\phi(k)}}\hat{\omega}_{A_h}(X_s)$. This implies for each h relatively prime to k,

$$\hat{\omega}_{A_h}(X_s) \to Z_h \sim \mathcal{N}(0, \frac{1}{\phi(k)}).$$

□

Remark 2.3.1. *A similar theorem to Theorem 2.4 but for $\Omega(X_s)$ should also be provable using the same methods as in the proof above.*

2.4 Large Deviations

From the formulas for the moment generating functions for $\omega(X_s)$ and $\Omega(X_s)$, we also obtain a large deviation principle. Recall the definition of a Large Deviation Principle (1.3). We will identify the rate function $I(x)$ for the $\omega(X_s)$ when our speed is $P(s)$.

In the classical definition, we have a sequence of *iid* random variables $X_1, X_2, ...$ Here, the $X_i's$ will instead be indicators $1_{c_p(X_s)>0}$, since $\omega(X_s)$ is a sum of these indicators. The expected value of the sum $\sum\limits_p 1_{c_p(X_s)>0}$ is just $P(s)$. So we want to find a rate function $I(x)$ such that for each $x \geq 1$,

$$\lim_{s \downarrow 1} \frac{1}{P(s)} \ln P\{\frac{\omega(X_s)}{P(s)} \geq x\} = -I(x).$$

We find that the rate function exactly matches the rate function in Example 1.1.

THEOREM 2.5. *Let $x > 1$. Then*

$$\lim_{s \downarrow 1} \frac{1}{P(s)} \ln P(\frac{\omega(X_s)}{P(s)} \geq x) = x \ln x + 1 - x.$$

Proof. We start with the upper bound.

By Chebyshev's inequality, we have

$$\begin{aligned} P\{\frac{\omega(X_s)}{P(s)} \geq x\} &= P\{\omega(X_s) \geq xP(s)\} \\ &\leq e^{-txP(s)} E e^{t\omega(X_s)} \\ &= e^{-(txP(s) + \sum_{j=1}^{\infty} P(js) \frac{(1-e^t)^j}{j})} \\ &= e^{-P(s)(tx + 1 - e^t + \sum_{j=2}^{\infty} \frac{P(js)}{P(s)} \frac{(1-e^t)^j}{j})} \end{aligned}$$

This is true for all t. The $\max_t (tx + 1 - e^t)$ occurs when $\frac{d}{dt}(tx + 1 - e^t) = x - e^t = 0$, so when $\ln x = t$.

Therefore,

$$P(\frac{\omega(X_s)}{P(s)} \geq x) \leq \exp\{-P(s)(x\ln x + 1 - x + \sum_{j=2}^{\infty} \frac{P(js)}{P(s)} \frac{(1-x)^j}{j})\}$$

Which implies

$$\frac{1}{P(s)} \ln P\{\frac{\omega(X_s)}{P(s)} \geq x) \leq -(x\ln x + 1 - x) + \sum_{j=2}^{\infty} \frac{P(js)}{P(s)} \frac{(1-x)^j}{j}$$

Therefore

$$\overline{\lim}_{s\downarrow 1} \frac{1}{P(s)} \ln P(\frac{\omega(X_s)}{P(s)} \geq x) \leq -(x\ln x + 1 - x)$$

One should then have equality, with $\overline{\lim}_{s\downarrow 1}$ replaced by $\lim_{s\downarrow 1}$ using Cramér's Theorem (Theorem 1.4). We proceed to show the lower bound is the same.

Let $Y_s = \frac{\omega(X_s)}{P(s)}$. Then

$$P(Y_s \geq x - \delta) \geq E[e^{\lambda Y_s}; Y_s \geq x - \delta] e^{-\lambda(x-\delta)}, \tag{2.4}$$

where λ solves $\Lambda_s'(\lambda) = x, \Lambda_s(\lambda) = -\sum_{j=1}^{\infty} P(js)\frac{(1-e^{\lambda/P(s)})^j}{j}$ (Here, $\Lambda_s(\lambda)$ is just the log of the moment generating function evaluated at $\lambda/P(s)$).

Define $\mu_s(A) = P(Y_s \in A), \tilde{\mu_s}(A) = \frac{E[e^{\lambda Y_s}; A]}{E[e^{\lambda Y_s}]}$. Then

$$\int y\tilde{\mu_s}(dy) = \frac{\int ye^{\lambda y}\mu_s(dy)}{\int e^{\lambda y}\mu_s(dy)} = \frac{\frac{d}{d\lambda}e^{\Lambda_s(\lambda)}}{e^{\Lambda_s(\lambda)}} = \Lambda_s'(\lambda) = x,$$

and

$$\int y^2\tilde{\mu_s}(dy) = \frac{\int y^2 e^{\lambda y}\mu_s(dy)}{\int e^{\lambda y}\mu_s(dy)} = \frac{\frac{d^2}{d\lambda^2}e^{\Lambda_s(\lambda)}}{e^{\Lambda_s(\lambda)}} = \Lambda_s''(\lambda) + \Lambda_s'^2(\lambda).$$

Also,

$$var(\tilde{Y_s}) = \Lambda_s''(\lambda) = \frac{1}{P(s)}e^{\lambda/P(s)}[1 - \sum_{j=2}^{\infty}\frac{P(js)}{j-1}(1-e^{\lambda/P(s)})^{j-1}].$$

From this last expression we see that $\Lambda_s''(\lambda) \to 0$. As this is true for any measurable A,

$$\tilde{\mu_s}((x-\delta, x+\delta)) = 1 - P(|\tilde{Y_s} - x| \geq \delta) \geq 1 - \frac{var(\tilde{Y_s})}{\delta^2} \to 1$$

Following from (2.4) with $Y_s = \frac{\omega(X_s)}{P(s)}$,

$$\begin{aligned} P(Y_s \geq x-\delta) \geq & \quad \frac{E[e^{\lambda Y_s}; Y_s \geq x-\delta]e^{-\lambda(x-\delta)}}{E[e^{\lambda Y_s}]}E[e^{\lambda Y_s}] \\ \geq & \quad \tilde{\mu_s}((x-\delta, x+\delta))e^{-\lambda(x-\delta)+\Lambda_s(\lambda)} \end{aligned}$$

Now

$$\lim_{s \downarrow 1} \frac{1}{P(s)} \ln \tilde{\mu_s}((x-\delta, x+\delta)) = 0$$

since $\tilde{\mu_s}((x-\delta, x+\delta)) \to 1$ as $s \downarrow 1$. So we have

$$\begin{aligned}
\underline{\lim}_{s \downarrow 1} \frac{1}{P(s)} \ln P(Y_s \geq x - \delta) &\geq \underline{\lim}_{s \downarrow 1} \tfrac{-\lambda(x-\delta) + \Lambda_s(\lambda)}{P(s)} \\
&\geq \underline{\lim}_{s \downarrow 1} \tfrac{\inf_\lambda(-\lambda(x-\delta) + \Lambda_s(\lambda))}{P(s)} \\
&= (x-\delta)\ln(x-\delta) + 1 - (x-\delta)
\end{aligned}$$

Now this is true for all $\delta > 0$. So we have

$$\underline{\lim}_{s \downarrow 1} \frac{1}{P(s)} \ln P(Y_s \geq x) = x \ln x + 1 - x.$$

We have shown both $\overline{\lim}$ and $\underline{\lim}$ converge to the same value. Thus,

$$\lim_{s \downarrow 1} \frac{1}{P(s)} \ln P(\frac{\omega(X_s)}{P(s)} \geq x) = x \ln x + 1 - x.$$

□

This result shows the relationship between $\omega(X_s)$ and the Poisson distribution $Z \sim Po(\lambda)$, $\lambda = P(s)$, as $s \downarrow 1$.

2.4.1 Moderate Deviations

We follow the notation in Section 1.1.4. In our case, $n \sim P(s)$, $b_n \sim P(s)^\alpha$, $0 < \alpha < 1/2$, and $b_n/n = P(s)^{\alpha-1}$. Then using (2.2), we have

$$\log Ee^{t\omega(X_s)} = -\sum_{j=1}^{\infty} P(js)\frac{(1-e^t)^j}{j}$$

$$\begin{aligned} P(s)^{1-2\alpha}\log Ee^{\frac{t}{P(s)^{1-\alpha}}\omega(X_s)} &= -P(s)^{1-2\alpha}\textstyle\sum_{j=1}^{\infty} P(js)\frac{(1-e^{t/P(s)^{1-2\alpha}})^j}{j} \\ &= -P(s)^{1-2\alpha}(P(s)(1-e^{t/P(s)^{1-\alpha}}) + \tfrac{1}{2}P(2s)(1-e^{t/P(s)^{1-\alpha}})^2 + ... \\ &= -P(s)^{1-2\alpha}(tP(s)^\alpha + \tfrac{t^2}{2}P(s)^{2\alpha} + ...) \to \tfrac{t^2}{2} \end{aligned}$$

This *suggests* that for any α between 0 and 1/2, the random variable $\omega(X_s)$ satisfies a LDP with speed $a_s = P(s)^\alpha$ and rate function $I(t) = t^2/2$. This result should be expected from the central limit theorem, as the $\omega(X_s)$ converge to a normal distribution, and this is the rate function for a sequence of *iid* $\mathcal{N}(0,1)$ rv's. We conclude with a summary of the main results.

2.5 Summary

We have exhibited formulas for the moment generating functions of $\omega(X_s), \Omega(X_s)$, and used them to prove respective central limit theorems for both of these counting functions. These results show the similarity between the zeta random variable X_s as $s \to 1^+$, and the uniform random variable X_n, as $n \to \infty$. In Theorem 2.1, we decompose $\Omega(X_s) = \sum_{j\geq 1} jA_j$ into a sum of independent random variables where $A_j \sim Poisson(P(js)/j)$. We see the utility of this in the proof of Theorem 2.3, using Corollary 2.1.1. Furthermore, if we want a more accurate approximation of $\Omega(X_s)$, we can peel off more factors $2A_2$, $3A_3$, and measure their contribution. Finally, the large deviations results once again displays the Poissonian nature of the limiting distribution. Theorem 2.5 exhibits the same rate function $I(x)$ as we would get with a sequence of *iid* Poisson random variables.

Chapter 3

Other work

3.1 The probability generating function of $\Omega(X_s) - \omega(X_s)$

3.1.1 Rényi's formula for X_s

Let us recover an elegant result by Rényi (see [17], Pg.65). Rényi showed the natural density (see definition 1.9) of the sets $\{n \in \mathbb{N} : \Omega(n) - \omega(n) = k\}$ are given by the coefficients d_k of the generating function

$$\sum_{k=0}^{\infty} d_k z^k = \prod_p (1 - \frac{1}{p})(1 + \frac{1}{p-z}).$$

Theorem 3.1. *The random variable $\Omega(X_s) - \omega(X_s)$ has probability generating function given by the infinite product $\prod\limits_p (1 - \frac{1}{p^s})(1 + \frac{1}{p^s - z})$. Moreover, this generating fuction exists for $s = 1$.*

Proof. The random variable $\Omega(X_s) - \omega(X_s)$ is equal to a sum of independent random variables

$$\Omega(X_s) - \omega(X_s) = \sum_p (c_p(X_s) - 1)^+,$$

where $f^+ = \max\{f, 0\}$. That is,

$$(c_p(X_s) - 1)^+ = \begin{cases} 0 & \text{if } c_p(X_s) = 0 \text{ or } 1 \\ c_p(X_s) - 1 & \text{if } c_p(X_s) > 1. \end{cases}$$

Writing $p(i)$ for $P(c_p(X_s) = i)$, we have the probability generating function of any one of them is

$$\begin{aligned} Ez^{(c_p(X_s)-1)^+} &= p(0) + p(1) + \textstyle\sum_{k=1}^{\infty} P\{(c_p(X_s) - 1)^+ = k\} z^k \\ &= (1 - \tfrac{1}{p^s}) + \tfrac{1}{p^s}(1 - \tfrac{1}{p^s}) + \textstyle\sum_{k=1}^{\infty} P\{c_p(X_s) = k + 1\} z^k \\ &= (1 - \tfrac{1}{p^s})(1 + \tfrac{1}{p^s}) + z^{-1} \textstyle\sum_{k=2}^{\infty} P\{c_p(X_s) = k\} z^k \\ &= (1 - \tfrac{1}{p^s})(1 + \tfrac{1}{p^s}) + z^{-1}[Ez^{c_p(X_s)} - p(0) - p(1)z]. \end{aligned}$$

Using that $Ez^{c_p(X_s)} = \frac{1-1/p^s}{1-z/p^s}$, and simplifying, one arises at

$$Ez^{(c_p(X_s)-1)^+} = (1 - \frac{1}{p^s})(1 + \frac{1}{p^s - z})$$

By Theorem 1.8, we get

$$Ez^{\Omega(X_s)-\omega(X_s)} = \prod_p Ez^{(c_p(X_s)-1)^+} = \prod_p (1-\frac{1}{p^s})(1+\frac{1}{p^s-z}).$$

Taking the limit as $s \to 1^+$, we arrive at the same infinite product as Rényi for the generating function of the d_k.

□

Once again, we see that the random variables X_s and X_n behave similarly in their respective limits. In lieu of Theorem 1.9, the same limit arises in the harmonic case Y_n (see example 1.6).

3.1.2 Some probabilities involving $Ez^{\Omega(X_s)-\omega(X_s)}$

One can then use the above generating function to calculate $d_k = \lim_{s\downarrow 1} P\{\Omega(X_s)-\omega(X_s) = k\}$ for any k. Let us rewrite the probability generating function as

$$\begin{aligned} Ez^{\Omega(X)-\omega(X)} &= \prod_p (1-\tfrac{1}{p^s})(1+\tfrac{1}{p^s}\tfrac{1}{1-z/p^s}) \\ &= \prod_p (1-\tfrac{1}{p^s})(1+\tfrac{1}{p^s}+\textstyle\sum_{k=1}^{\infty}\tfrac{z^k}{p^{s(k+1)}}). \end{aligned}$$

Let us denote the coefficient of z^k in the above as $d_{s,k} = P\{\Omega(X_s)-\omega(X_s) = k\}$. Then we see that $d_{s,0} = \prod_p (1-\frac{1}{p^s})(1+\frac{1}{p^s}) = \zeta^{-1}(2s)$, and that

$$
\begin{aligned}
d_{s,1} = &\ \sum_p (1-\tfrac{1}{p^s})\tfrac{1}{p^{2s}} \textstyle\prod_{q\neq p}(1-\tfrac{1}{q^{2s}}) \\
= &\ \sum_p \tfrac{1}{1+\frac{1}{p^s}}\tfrac{1}{p^{2s}} \textstyle\prod_q (1-\tfrac{1}{q^{2s}}) \\
= &\ \tfrac{1}{\zeta(2s)} \sum_p \tfrac{1}{p^s(p^s+1)}
\end{aligned}
$$

Using these results, we get $d_0 \approx .6079..$, and $d_1 = \frac{1}{\zeta(2)}(\sum \frac{1}{p(p+1)}) \approx .2007...$ This gives approximately

$$\lim_{s\to 1} P\{\Omega(X_s) - \omega(X_s) > 1\} \approx .1915...$$

It is also interesting to note that

$$\lim_{s\to 1} E[\Omega(X_s) - \omega(X_s)] = \sum_p \frac{1}{p(p-1)} \approx .73...$$

(Note: the above sum involving reciprocals of primes was calculated using sage)

The asymptotic density of "nth-power free" numbers is $\frac{1}{\zeta(n)}$ (see example 1.5). That is, the density of square free numbers is $\frac{1}{\zeta(2)}$, cube free is $\frac{1}{\zeta(3)}$, and so on... Since $1 - \frac{1}{\zeta(3)} \approx .16809..$, we have that most numbers ($\frac{.168}{.192} = .875$) that have $\Omega(n) - \omega(n) > 1$ also are not cube free.

3.2 Some calculations

3.2.1 Probability n iid zeta random variables are pairwise coprime

What is the probability that n iid zeta random variables are pairwise coprime? For the natural density of these coordinates in $\mathbb{Z}^n$, the answer was given by Cai & Bach 2003 ([3], Theorem 3.3 or [17], pg. 69) and is

$$\prod_p [(1-\frac{1}{p})^n + \frac{n}{p}(1-\frac{1}{p})^{n-1}]$$

Now in our case, we calculate $P(X_1, ..., X_n$ pairwise coprime $)$ as follows.

$$P(\cap_{i<j}\{gcd(X_i, X_j) = 1\}) = \prod_p P(\cap_{i<j}\{p \not| X_i \text{ or } p \not| X_j\})$$

where the intersections is over all pairs.

But this is the same as "at most one of the X_I can be divisible by p". For if *two* of them *are* divisible by p, then $\{p \not| X_i \cup p \not| X_j\}$ does not occur, and so the intersection doesn't either. Therefore,

$$\cap_{i<j}\{p \not| X_i \cup p \not| X_j\} = \{p \not| X_j,\ j = 1, .., n\} \sqcup (\sqcup_{i=1}^n \{p|X_i, p \not| X_j\ \forall j \neq i\})$$

so by independence and mutual-exclusivity,

$$\prod_p P(\cap_{i<j}\{p \not| X_i \cup p \not| X_j\}) = \prod_p [(1-\frac{1}{p^s})^n + \frac{n}{p^s}(1-\frac{1}{p^s})^{n-1}].$$

We see in the limit $s \to 1$ that the same formula as the one found by Cai and Bach follows.

3.2.2 The distribution of the gcd of k *iid* zeta random variables

In this section, we show that the distribution of the gcd of k independent zeta random variables with common parameter s is the same as the distribution of one zeta random variable with parameter ks.

Theorem 3.2. *Let $Z_1, ..., Z_k$ be iid zeta random variables with common parameter s. Let n be any positive integer, and write $n = \prod p^{a_p}$ where all but finitely many of the exponents a_p are nonzero. Let $c_p(Z_i)$ denote the p-th exponent in the prime factorization of Z_i as in (1.13). Then $gcd(Z_1, ...Z_k) \stackrel{d}{=} X_{ks} \sim Zeta(ks)$.*

Proof. By the definition of *gcd*, the independence of the Z_i's, and the principle of inclusion-exclusion, we have

$$\begin{aligned} P\{gcd(Z_1, ...Z_k) = n\} &= \prod_p P(\min\{c_p(Z_1), ..., c_p(Z_k)\} = a_p) \\ &= \prod_p P(\cup_{i=1}^k \{c_p(Z_i) = a_p, c_p(Z_j) \geq a_p \ \forall j \neq i\}) \\ &= \prod_p \textstyle\sum_{l=1}^k (-1)^{l+1} \binom{k}{l} P(c_p(Z_1) = a_p)^l P(c_p(Z_1) \geq a_p)^{k-l} \\ &= \prod_p \textstyle\sum_{l=1}^k (-1)^{l+1} \binom{k}{l} [\frac{1}{p^{a_p s}}(1 - \frac{1}{p^s})]^l [\frac{1}{p^{a_p s}}]^{k-l} \\ &= \prod_p \textstyle\frac{1}{p^{k a_p s}} \sum_{l=1}^k (-1)^{l+1} \binom{k}{l} (1 - \frac{1}{p^s})^l \end{aligned}$$

Using the binomial theorem,

$$\sum_{l=1}^{k}(-1)^{l+1}\binom{k}{l}(1-\frac{1}{p^s})^l = -\textstyle\sum_{l=1}^{k}\binom{k}{l}(-1+\frac{1}{p^s})^l$$
$$= -[(1-1+\tfrac{1}{p^s})^k - 1] = (1-\tfrac{1}{p^{ks}}).$$

Plugging this in to the above, we see that

$$P\{gcd(Z_1,...Z_k)=n\} = \prod_p \frac{1}{p^{ka_p s}}(1-\frac{1}{p^{ks}}) = \frac{1}{\zeta(ks)}\frac{1}{n^{ks}}.$$

We notice this is exactly the probability $P(X_{ks}=n)$, where X_{ks} is zeta with parameter ks. Since this is true for any n, the random variable $gcd(Z_1,...Z_k)$ must be zeta distributed with parameter ks. □

Several distributions (e.g. Poisson, gamma, etc) have the property that the sum of two independent versions is again in that same family of distributions, where the location parameter is shifted. Here, the location parameter may be regarded as s, and instead of adding two random variables X and Y, we take the greatest common divisor of X and Y. The location parameter is *shifted to the left*, in the sense that $gcd\{X,Y\}$ is always smaller or equal to X or Y. This begs the question, does the *lcm* operator *shift to the right*? One can show that for any positive integer $n=\prod p^{a_p}$,

$$P(lcm\{Z_1,...,Z_k\}=n) = \prod_p [(1-\frac{1}{p^{(a_p+1)s}})^k - (1-\frac{1}{p^{a_p s}})^k].$$

Unfortunately, this does not seem to factor nicely in terms of n, k and s, as in the *gcd* case.

3.2.3 Working with Gut's remarks

As noted in the outline, Gut ([10]) shows that $\log X_s$ is compound poisson where the number of terms N is poisson distributed with parameter $\lambda = \log \zeta(s)$, and the terms are independent and distributed according to $\log V$ as in (1.14). How can we use this decomposition of the log of the zeta random variable in a useful way? One approach to this is the following theorem.

Theorem 3.3. *Define V as in (1.14), and let $m(n) = P(\Omega(V) = m)$, where $\Omega(V)$ is the number of prime factors of V. Then*

$$P(\Omega(X_s) = n) = e^{-\log \zeta(s)} \sum_{k=1}^{n} \frac{(\log \zeta(s))^k}{k!} m^{*k}(n),$$

where m^{*k} is the k-fold convolution of m.

Proof. Using that $\log X_s$ is compound poisson, we can rewrite

$$X_s \stackrel{d}{=} \prod_{i=1}^{N} p_i^{m_i}$$

where $V_i = p_i^{m_i}$ is a random prime power. If we fix a positive integer k, we attain

$$P(\Omega(V) = k) = P(V = p^k \text{ for some prime } p) = \frac{1}{\log \zeta(s)} \sum_p \frac{1}{kp^{ks}} = \frac{1}{\log \zeta(s)} \frac{P(ks)}{k}.$$

Conditioning on the value of N gives

$$
\begin{aligned}
P(\Omega(X_s) = n) = & \sum_{k=0}^{n} P(\Omega(X_s) = n | N = k) P(N = k) \\
= & \sum_{k=0}^{n} P(\Omega(V_1) + ... + \Omega(V_k) = n) P(N = k) \\
= & e^{-\log \zeta(s)} \sum_{k=1}^{n} \frac{(\log \zeta(s))^k}{k!} m^{*k}(n)
\end{aligned}
$$

□

See ([8], VI.4) for a more general account of compound poisson processes.

3.3 Sampling non-negative integers using the Poisson distribution

The Poisson Distribution is another way we can sample a non-negative integer at random. Just as the harmonic Y_n, zeta X_s, and the uniform variable X_n behave similarly in their respective limits, probabilities involving a Poisson variable X with parameter λ also seem to converge to that of the other three cases, as its rate parameter $\lambda \to \infty$. In the following section, we study one of these probabilities. Namely, if we take m independent copies of the Poisson distribution with a common rate parameter, we show that as $\lambda \to \infty$, the probability these m variables are relatively prime converges to $\frac{1}{\zeta(m)}$.

We start with a lemma involving roots of unity.

Lemma 3.3.1. *Let $n \in \mathbb{N}$, and X be Poisson distributed with parameter λ. Then*

$$P\{n|X\} = \frac{e^{-\lambda}}{n}(e^{\lambda} + e^{\omega\lambda} + ... + e^{\omega^{n-1}\lambda}),$$

where ω is a primitive n-th root of unity.

Proof. We use the maclaurin series for the e^{λ}, and classify each term in the series into the n residue classes $\mod n$. This gives

$$\begin{aligned} e^{\lambda} + e^{\omega\lambda} + e^{\omega^2\lambda} + ... + e^{\omega^{n-1}\lambda} &= \sum_{k=0}^{\infty}\sum_{r=0}^{n-1} \frac{\lambda^{nk+r}}{(nk+r)!}(1 + \omega^{nk+r} + (\omega^2)^{nk+r} + ... + (\omega^{n-1})^{nk+r}) \\ &= \sum_{k=0}^{\infty}\sum_{r=0}^{n-1} \frac{\lambda^{nk+r}}{(nk+r)!}(1 + \omega^{r} + (\omega^2)^{r} + ... + (\omega^{n-1})^{r}). \end{aligned}$$

using that $\omega^{nk} = 1$. Now if $r \neq 0$,

$$1 + \omega^r + (\omega^2)^r + ... + (\omega^{n-1})^r = 1 + \omega^r + (\omega^r)^2 + ... + (\omega^r)^{n-1}$$

$$= \frac{(\omega^r)^n - 1}{\omega^r - 1} = \frac{0}{\omega^r - 1}$$

since ω^r is an nth root of unity. If $r = 0$, we have instead $1 + 1 + ... + 1$ n times. Therefore,

$$e^{\lambda} + e^{\omega\lambda} + e^{\omega^2\lambda} + ... + e^{\omega^{n-1}\lambda} = n\sum_{k=0}^{\infty} \frac{\lambda^{nk}}{(nk)!}.$$

The lemma follows from multiplying both sides by $\frac{e^{-\lambda}}{n}$.

□

From the above lemma, we can deduce a formula for $P(gcd(X,Y)=1)$. We note that

$$P(gcd(X,Y)=1)=1-P(gcd(X,Y)>1)=1-P(\cup_p\{p|X,p|Y\}).$$

By Inclusion-Exclusion and independence,

$$\begin{aligned} P(gcd(X,Y)=1) =& 1-\sum_p P(\{p|X,p|Y\})+\sum_{p\neq q} P(\{p|X,p|Y\}\cap\{q|X,q|Y\})-...\\ =& 1-\sum_p P(p|X)^2+\sum_{p\neq q} P(pq|X)^2-...+(-1)^n\sum_{p_1,...,p_n} P(p_1p_2...p_n|X)^2+... \end{aligned}$$

where the general term in the equation is a sum over all sets of n distinct primes.

Using the mobius function $\mu(n)$, where

$$\mu(n)=\begin{cases} 1 & n=1\\ (-1)^t & n \text{ is a product of } t \text{ distinct primes}\\ 0 & \text{otherwise} \end{cases}$$

and noting that $P(1|X)=1$, we write the above equation as

$$P(gcd(X,Y)=1)=\sum_{n=1}^{\infty}\mu(n)P(n|X)^2.$$

This immediately generalizes to a sum of m iid random variables.

Theorem 3.4. *Let n be any integer, and $X_1, ...X_m$ be a sequence of m iid Poisson random variables with common parameter λ. Then*

$$P\{gcd(X_1, ..., X_m) = 1\} = \sum_{n=1}^{\infty} \mu(n) P\{n|X\}^m.$$

Putting Lemma 3.3.1 and Theorem 3.4 together, we get

Corollary 3.3.1.

$$P\{gcd(X_1, ..., X_m) = 1\} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^m} (\sum_{k=0}^{n-1} \phi(\frac{2k\pi}{n}))^m.$$

where $\phi(t) = e^{\lambda(e^{it}-1)}$ is the characteristic function of X.

3.3.1 Taking the limit as $\lambda \to \infty$

If we analyze the probability in Theorem 3.4 as λ goes to infinity, we get the following theorem.

Theorem 3.5. *Let $X_1, ..., X_m$ be independent poisson random variables with common parameter λ. If we send $\lambda \to \infty$, then*

$$P\{gcd(X_1, ..., X_m) = 1\} \to \frac{1}{\zeta(m)}.$$

Proof. We use Lemma 3.3.1. Apart from ± 1, roots of unity have nonzero imaginary part. If we pair these roots of unity into complex conjugate pairs, we see that

$$
\begin{aligned}
|e^{\lambda}+e^{\omega\lambda}+...+e^{\omega^{n-1}\lambda}| &\leq |e^{\lambda}+e^{-\lambda}+\textstyle\sum_{k=1}^{\lfloor\frac{n}{2}\rfloor-1}(e^{\omega^k\lambda}+e^{\overline{\omega}^k\lambda})| \\
&\leq e^{\lambda}+e^{-\lambda}+\textstyle\sum_{k=1}^{\lfloor\frac{n}{2}\rfloor-1}2e^{\cos(\frac{2k\pi}{n})\lambda}\cos(\sin(\frac{2k\pi}{n})\lambda) \\
&\leq e^{\lambda}+e^{-\lambda}+2(\lfloor\tfrac{n}{2}\rfloor-1)\max_k\{e^{\cos(\frac{2k\pi}{n})\lambda}\cos(\sin(\tfrac{2k\pi}{n})\lambda)\} \\
&\leq e^{\lambda}+e^{-\lambda}+ne^{\cos(\frac{2\pi}{n})\lambda}.
\end{aligned}
$$

The first inequality is an *equality* if n is even. If n is odd, -1 is not an n-th root of unity, but adding $e^{-\lambda}$ to the sum only makes it bigger. Multiplying by $\frac{e^{-\lambda}}{n}$, we see that

$$P\{n|X\} \leq \frac{1}{n}+\frac{e^{-2\lambda}}{n}+e^{-\lambda(1-\cos(\frac{2\pi}{n}))}.$$

For the lower bound, by a similar argument, but with the *reverse* triangle inequality, we can show that

$$P\{n|X\} \geq \frac{1}{n}-\frac{e^{-2\lambda}}{n}-e^{-\lambda(1-\cos(\frac{2\pi}{n}))}.$$

These two together imply the probability n divides X converges exponentially in λ to that of the uniform distribution. That is,

$$P\{n|X\} = \frac{1}{n}+o(e^{-\lambda})$$

as $\lambda \to \infty$. Plugging into Theorem 3.4, we get that

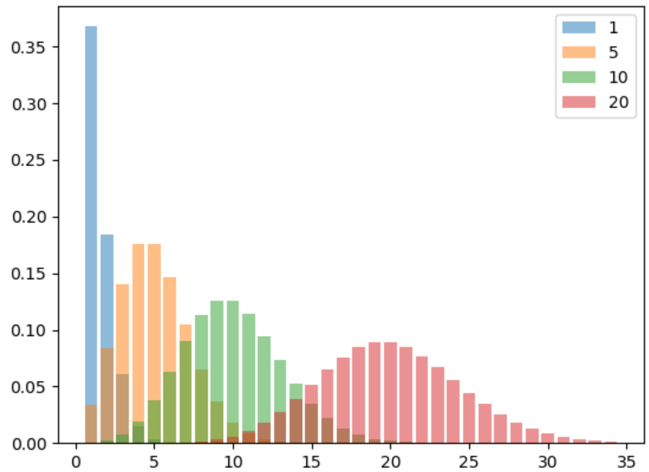

Figure 3.1: Different pmfs of the Poisson Distribution

These are four different probability mass functions of poisson variables with expected value $\lambda = 1, 5, 10, 20$. As the rate parameter λ increases, the mass becomes more evenly distributed among all non-negative integers.

$$P\{gcd(X_1, ..., X_m) = 1\} = \sum_{n\geq 1} \mu(n)(\frac{1}{n} + o(e^{-\lambda}))^m,$$

so that as $\lambda \to \infty$,

$$P\{gcd(X_1, ..., X_m) = 1\} \to \sum_{n\geq 1} \frac{\mu(n)}{n^m} = \frac{1}{\zeta(m)}.$$

□

We see that, as $\lambda \to \infty$, the probability that m Poisson random variables are relatively prime converges to the same limit as when the m random variables are uniformly distributed from $\{1, 2, ..., n\}$ and $n \to \infty$. This is intuitively clear; the probability mass function of the Poisson "flattens out" as $\lambda \to \infty$, and becomes more and more uniform across the set of non-negative integers.

Bibliography

[1] T. M. Apostol. *Introduction to analytic number theory.* Springer-Verlag, New York-Heidelberg, 1976. Undergraduate Texts in Mathematics.

[2] R. Arratia. *On the amount of dependence in the prime factorization of a uniform random integer*, volume 10 of *Bolyai Soc. Math. Stud.* János Bolyai Math. Soc., Budapest, 2002.

[3] J.-Y. Cai and E. Bach. On testing for zero polynomials by a set of points with bounded precision. pages 15–25. 2003. Computing and combinatorics (Guilin, 2001).

[4] P. Diaconis. Examples for the theory of infinite iteration of summability methods. *Canadian J. Math.*, 29(3):489–497, 1977.

[5] P. Donnelly and G. Grimmett. On the asymptotic distribution of large prime factors. *J. London Math. Soc. (2)*, 47(3):395–404, 1993.

[6] R. Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 2019. Fifth edition of [MR1068527].

[7] P. Erdös and M. Kac. The Gaussian law of errors in the theory of additive number theoretic functions. *Amer. J. Math.*, 62:738–742, 1940.

[8] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[9] P. Flajolet and M. Soria. Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combin. Theory Ser. A*, 53(2):165–182, 1990.

[10] A. Gut. Some remarks on the Riemann zeta distribution. *Rev. Roumaine Math. Pures Appl.*, 51(2):205–217, 2006.

[11] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers.* Oxford University Press, Oxford, sixth edition, 2008. Revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles.

[12] M. Kac. *Statistical independence in probability, analysis and number theory.* The Carus Mathematical Monographs, No. 12. Published by the Mathematical Association of America. Distributed by John Wiley and Sons, Inc., New York, 1959.

[13] J. F. C. Kingman, S. J. Taylor, A. G. Hawkes, A. M. Walker, D. R. Cox, A. F. M. Smith, B. M. Hill, P. J. Burville, and T. Leonard. Random discrete distribution. *J. Roy. Statist. Soc. Ser. B*, 37:1–22, 1975.

[14] D. E. Knuth and L. Trabb Pardo. Analysis of a simple factorization algorithm. *Theoret. Comput. Sci.*, 3(3):321–348, 1976/77.

[15] G. D. Lin and C.-Y. Hu. The Riemann zeta distribution. *Bernoulli*, 7(5):817–828, 2001.

[16] S. P. Lloyd. Ordered prime divisors of a random integer. *Ann. Probab.*, 12(4):1205–1212, 1984.

[17] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.

[18] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, third edition, 2015. Translated from the 2008 French edition by Patrick D. F. Ion.

Appendix A

Appendix

A.1 The same decompositon as Lloyd's

Our factorization of the moment generating function for $\Omega(X)$ is exactly the same result as Lloyd's. As noted earlier, the moment generating function of $\Omega(X)$ can be factored as a product of moment generating functions

$$Ee^{t\Omega(X)} = e^{P(s)(e^t-1)} \exp\{\sum_{m=2}^{\infty} \frac{P(ms)}{m}(e^{tm}-1)\}.$$

This rightmost exponent corresponds to the moment generating function of $\Omega(N")$, where $N"$ is as in Lloyd's paper. Let us verify this claim.

$$\exp\{\sum_{m=2}^{\infty}\frac{P(ms)}{m}(e^{tm}-1)\} = \exp\{\sum_p \sum_{m=2}^{\infty} \tfrac{1}{mp^{ms}}(e^{tm}-1)\}$$
$$= \prod_p \exp\{\sum_{m\geq 2} \tfrac{1}{mp^{ms}}(e^{tm}-1)\}$$
$$= \prod_p \exp\{\sum_{m=1}^{\infty} \tfrac{(e^t/p^s)^m}{m} - \sum_{m=1}^{\infty} \tfrac{1}{mp^{ms}} + \tfrac{1}{p^s} - \tfrac{e^t}{p^s}\}$$
$$= \prod_p \exp\{-\log(1-e^t/p^s)\}\exp\{\log(1-1/p^s)\}e^{1/p^s}\exp\{-e^t/p^s\}$$
$$= \prod_p (1-\tfrac{1}{p^s})e^{1/p^s}\tfrac{\exp\{-e^t/p^s\}}{1-e^t/p^s}.$$

Let us simplify the notation and write ρ for $1/p^s$ (note that ρ still depends on p). Using the power series for e^a and $1/(1-a)$ and convolution, we see that

$$\prod_p (1-\rho)e^{\rho}\frac{\exp\{-\rho e^t\}}{1-\rho e^t} = \prod_p (1-\rho)e^{\rho}\sum_{m=0}^{\infty}\rho^m e^{tm}\sum_{m=0}^{\infty}(-1)^m\tfrac{\rho^m e^{tm}}{m!}$$
$$= \prod_p (1-\rho)e^{\rho}\sum_{m=0}^{\infty}(\sum_{j=0}^{m}\tfrac{(-1)^j}{j!}\rho^{js}\rho^{(m-j)s})e^{tm}$$
$$= \prod_p (1-\rho)e^{\rho}\sum_{m=0}^{\infty}\rho^m\sigma(m)e^{tm},$$

where $\sigma(m) = \sum_{j=0}^{m}\frac{(-1)^j}{j!}$ is as before.

If we substitute $z = e^t$ to turn the moment generating function into an ordinary generating function, we recover Lloyd's result.

Thus we have seen that the argument behind Lloyd's factorization $N = N'N$" of the zeta random variable N into a product of random variables N' and N" corresponds to the same decomposition we have made of $\Omega(X_s) = \Omega(X_s') + \Omega(X_s")$, where $\Omega(X_s')$ is a poisson random variable with parameter $P(s)$, and $\Omega(X_s")$ is a random variable with moment generating function given by the residue above.

A.2 The statistics of Euler's ϕ function

The average value of euler's totient function converges to $6/\pi^2$ for a randomly chosen integer. We prove here the case where the random integer is chosen uniformly from $\{1,..,N\}$, following Mark Kac's argument in ([12], Section 4.2).

Theorem A.1. *Suppose X_N is chosen uniformly from $\{1,...,N\}$. Then*

$$\lim_{N\to\infty} E\frac{\varphi(X_N)}{X_N} = \frac{6}{\pi^2}$$

Proof. For any n, we have $\frac{\varphi(n)}{n} = \prod_{p|n}(1-\frac{1}{p})$. From this, one easiliy deduces that

$$\frac{\varphi(n)}{n} = \sum_{d|n}\frac{\mu(d)}{d},$$

where $\mu(n)$ is the mobius function.

It follows that

$$\begin{aligned} E\frac{\varphi(X_N)}{X_N} =& \frac{1}{N}\sum_{n=1}^{N}\frac{\varphi(n)}{n} \\ =& \frac{1}{N}\sum_{d=1}^{N}\frac{\mu(d)}{d}\lfloor\frac{N}{d}\rfloor, \end{aligned}$$

where $\lfloor x\rfloor$ denotes the greatest integer less than or equal to x.

Writing $\lfloor\frac{N}{d}\rfloor = \frac{N}{d} - r_{N,d}$, where $0 \le r_{N,d} < 1$, we see that

$$\begin{aligned}\lim_{N\to\infty} E\frac{\varphi(X_N)}{X_N} &= \lim_{N\to\infty}\sum_{d=1}^{N}\frac{\mu(d)}{d}\{\frac{1}{d}-\frac{r_{N,d}}{N}\}\\ &= \lim_{N\to\infty}\sum_{d=1}^{N}\frac{\mu(d)}{d^2}-\lim_{N\to\infty}\frac{1}{N}\sum_{d=1}^{N}\frac{\mu(d)}{d}r_{N,d}\\ &=\frac{1}{\zeta(2)}.\end{aligned}$$

□