ON THE DISTRIBUTIONS OF THE NUMBER OF DISTINCT PRIME FACTORS

HARSHIL NUKALA

Abstract. The function $\omega(n)$ is a wonderful function. In our paper, we will be exploring $\omega(n)$ in a probabilistic number theory setting. It is known that the normal order of $\omega(n)$ is $\log\log n$ and the variance of $\omega(n)$ is also $\log\log n$. In this paper, we will be studying how the distribution $\omega(n)$ relates to the normal distribution, which is the result of the Erdös-Kac Theorem–our key result in this paper. We will be assuming very little knowledge to be able to understand such an unintuitive result.

1. Introduction

The Erdös-Kac Theorem is a well celebrated theorem in probabilistic number theory that is quite unintuitive at the start but becomes more and more engaging and beautiful as you study further. Define $\omega(n)$ to be the number of distinct prime factors of n. For example $\omega(2) = 1$, since 2 is prime and $\omega(2024) = \omega(2^3 \times 11 \times 53) = 3$. Before we go any further, we would like to clarify some notation that we will be using throughout the paper.

Definition 1.1. (Big O Notation) Denote f and g to be two functions such that $g(x) > 0$ for all $x \geq a$. Then, $f(x) = \mathcal{O}(g(x))$, if there exists a constant $M > 0$ such that

$$|f(x)| \leq Mg(x)$$

for all $x \geq a$.

Definition 1.2. (Little O Notation) Denote f and g to be two functions. If

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0,$$

then $f(x) = o(g(x))$.

When the index of the sum is $n \leq x$ we are summing over all positive integers n less than x, but when the index is $p \leq x$ we will be summing over all primes p less than x. In this paper, the natural logarithm of x is denoted as $\log x$. Finally, we denote $[r]$ to be the set $\{1, 2, \cdots, r\}$.

The function $\omega(n)$ has been a prime focus for many well known mathematicians. Turán [Tur34] showed that

$$\sum_{n\leq x} (\omega(n) - \log\log x)^2 = \mathcal{O}(x \log\log x).$$

Then, Hardy and Ramanujan [HR17] showed that for some real number δ the number of positive integers $n \leq N$ that satisfy

$$|\omega(n) - \log\log n| \geq (\log\log N)^{\frac{1}{2}+\delta} \tag{1.1}$$

Date: July 10, 2024.

is $o(N)$. But now, we might ask if there is a stronger result that may give the distribution for $\omega(n)$. After Turán's simple, probabilistic proof of 1.1, the ideas further developed and led Erdös and Kac to discover that there exists a normal distribution for

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \tag{1.2}$$

with mean 0 and variance 1. In particular, Erdös and Kac asserted the Central Limit Theorem for $\omega(n)$. More precisely, we have that Theorem 1.3 is our center of focus.

Theorem 1.3 (Erdös-Kac Theorem)**.** *For* $\gamma \in \mathbb{R}$,

$$\lim_{x\to\infty} \frac{1}{x} \# \left\{ 3 \le n \le x : \frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \le \gamma \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{-\frac{t^2}{2}}\, dt.$$

The original proof of Erdös-Kac theorem uses Brun's method from Sieve Theory [EK40]. However, in 1953 Delange [Del62] and in 1955 Halberstam [Hal56] used the idea of the method of moments to proof Erdös-Kac Theorem, but were rather complex. In 2006, Granville and Soundarajan [GS06] proved Erdös-Kac Theorem in a more understandable way by comparing the moments of 1.2 to the moments of the normal distribution.

In this paper, we will accomplish two things: the first is proving the Erdös-Kac Theorem (by a method similar to Granville and Soundarajan but removing the necessity of introducing C_k which is in terms of the gamma function as done in [Les15]) and the second is demonstrating the beauty of Erdös-Kac Theorem through multiple different variants. In the second section, we will start of with some probability review that is necessary to understand the Central Limit Theorem. Then, in the third section, we will review some number theory results which will be essential for understading the behavior of $\omega(n)$. Then, in the fourth section, we will be exploring characteristic functions which is a vital part in understanding the Central Limit Theorem. Then, in the fifth section, we will talk about the method of moments and understand how they relate to what we wish to show. Then, in the sixth section, we will discuss the Central Limit Theorem and understand how 1.2 relates to what we discussed in Section 2. Then, in the seventh section , we define some final notation and models that will be used to prove the Erdös-Kac Theorem. From here, in the eighth section, we prove the Erdös-Kac Theorem. Finally, in the ninth section, we will define final notation to be able to discuss the variants of Erdös-Kac Theorem. For our ultimate section, we will be windowing watching some of the magnificent analogues for the Erdös-Kac Theorem.

2. Probability Theory Review

We begin by reviewing the necessary knowledge of probabilistic theory to understand the Erdös-Kac theorem and its proof. Denote $P(\bullet)$ to denote probability. The results from this section will be carrying over to other sections as well. Denote X to be a *discrete* random variable which is defined to be a process that is associate to a particular value which is a natural number. Here, the natural number statement is necessary since we are talking about a discrete random variable. Soon, we will talk about *continuous* random variables that will have a different definition.

We have that the expected value of X, denoted as $\mathbb{E}[X]$, can be represented as

$$\mathbb{E}[X] = \sum_{x\in\mathbb{N}} P(X = x) \cdot x$$

Hence, we can define μ, the *mean*, to equal to the expected value of X. That is $\mu = \mathbb{E}[X]$. Now define

$$\sigma^2 = \mathbb{E}[(X-\mu)^2] = \sum_{x\in\mathbb{N}} P(X=x)\cdot(X-\mu)^2.$$

We refer to σ^2 as the *variance*, which can also be written as $\mathrm{Var}[X]$. The variance tells us how distributed our data is.

Until now, we were talking about a single random sample X; however, most of the time we are processing X more than just a single time–suppose N times. Hence, suppose that $(X_1, X_2, X_3, \cdots, X_N)$ is a sequence of independent, identically distributed, real-valued random variables. To demonstrate this, suppose the process of a coin flip with a fair coin. This process is referred to as a *random experiment.* Denote X_i to be a random variable that takes values of 0 or 1, where 1 represents heads and 0 represents tails. We say $X_i = 1$ with probability p, and 0 with probability $1-p$. Since our coin is fair, $p = 1/2$. Throughout this paper, we will shorten the definition of the random variable, such as the one defined above, to the following:

$$X_i = \begin{cases} 1 & \text{with probability} \quad \frac{1}{2}, \\ 0 & \text{with probability} \quad \frac{1}{2}. \end{cases}$$

A coin flip is an example of a *Bernouli Trail*, which is defined when a random experiment has one random variable X_i which has two possible values 0 and 1 with not necessarily the same probability (a weighted coin flip is still considered a Bernouli Trail). Now we are equipped to talk about binomial distributions and their wonderful mathematical properties.

Definition 2.1 (Binomial Distribution)**.** Assume $X_1, X_2, \cdots, X_N$ are independent and identically distributed (idd) Bernoulli random variables, where $P(X_i = 1) = p$. Then let

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

We say that $\mathcal{Y}_N$ is a *binomial distribution.*

Binomial distributions are one type of *partial sum process* (others include negative binomial distribution and gamma distribution). We can now explore mean and variance in our binomial distribution setting.

Proposition 2.2 (Lineraity of Expectation)**.**

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbb{E}[X_i]$$

Proof. It suffices to show that

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}\left[\sum_{i=1}^{2} X_i\right] = \sum_{i=1}^{2} \mathbb{E}[X_i] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

We have that by definition of expected value,

$$\begin{aligned}\mathbb{E}[X_1+X_2] &= \sum_{n\geq 0} nP\left(X_1+X_2=n\right)\\ &= \sum_{n\geq 0}\sum_{x_1+x_2=n}(x_1+x_2)P(X_1=x_1)P(X_2=x_2)\\ &= \sum_{x_1+x_2\geq 0}(x_1+x_2)P(X_1=x_1)P(X_2=x_2)\\ &= \sum_{x_1\geq 0}\sum_{x_2\geq 0}x_1P(X_1=x_1)P(X_2=x_2)+\sum_{x_2\geq 0}\sum_{x_1\geq 0}x_2P(X_1=x_1)P(X_2=x_2)\\ &= \sum_{x_1\geq 0}x_1P(X_1=x_1)+\sum_{x_2\geq 0}x_2P(X_2=x_2)\\ &= \mathbb{E}[X_1]+\mathbb{E}[X_2].\end{aligned}$$

This completes the proof. ■

Notice that in the above proof we did not use the fact that $(X_1, X_2, \cdots, X_N)$ are independent random variables. Hence, Proposition 2.2 is true even when the random variables are non-independent. In fact linearity extends to variance as well (but with a small caveat). For, linearity of variance to be satisfied the random variables must be independent (since then we can undermine the *covariance*). That is,

$$\operatorname{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N}\operatorname{Var}[X_i]$$

only when $(X_1, X_2. \cdots, X_N)$ are independent random variables. Linearity of Variance will not be used in this paper and an interested reader can read the proof here [Cha21]. One nice property, as a consequence from the definition of variance and Proposition 2.2 is that

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Until now, we have been referring to events happening in a discrete manner (that is, all random variables are associated with a value in the natural numbers); however, this may not necessarily be the case. We will now study when our random variable is *continuous*. A *continuous* random variable X is defined to be taking on values between a fixed range $[a, b]$. We wish to move away from our binary binomial distribution and define a *normal distribution*, a valuable player in Erdös-Kac Theorem. Before defining a normal distribution, we will first define a *probability density function*.

Definition 2.3 (Probability Density Function)**.** A function $f(x)$ is a probability density function if $f(x) \geq 0$ for all $-\infty \leq x \leq \infty$, and

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

I would like to highlight the integral is analogous to saying that the sum of all discrete events probabilities for a given experiment is equal to 1. For example, in our coin flip example, we have that if the probability of flipping heads is p the probability of flipping tails is $1-p$, so that the sum of the two events probabilities will be 1. We can now define probability, expected value, and variance for continuous random variables.

Definition 2.4 (Probability of a Continuous Random Variable)**.** Denote X to be a continuous random variable such that $a \leq X \leq b$ and has a probability density function $f(x)$. Then,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

Note that $P(X = x) = 0$. That is the probability of a continuous random variable taking on a single value is 0 since its is assumed that X has infinitely mainly values it could take.

Definition 2.5 (Expected Value of a Continuous Random Variable)**.** Denote X to be a continuous random variable and $f(x)$ to be the probability density function of X. Then

$$\mathbb{E}[X] = \mu = \int_{-\infty}^{\infty} x f(x)\,dx.$$

Definition 2.6 (Variance of a Continuous Random Variable)**.** Denote X to be a continuous random variable and $f(x)$ to be the probability density function of X. Then

$$\text{Var}[X] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

We will now define the normal distribution:

Definition 2.7 (Normal Distribution)**.** Denote X to be a continuous random variable. Then X has a *normal distribution* if it has a probability density function $\Phi(x)$ give by

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \text{Var}[X]$. We call $\Phi(x)$ the *normal density function.*

As an example, Figure 1 is a normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

The normal density function has multiple wonderful properties of which we will be highlighting four of them in Lemma 2.8.

Lemma 2.8 (Normal Density Function Properties)**.** *Denote $f(x)$ to be a normal density function. Then $f(x)$ satisfies the following properties:*

- *$f(x)$ is symmetric about $x = \mu$.*
- $\lim_{x\to\pm\infty} f(x) = 0$.
- *$f'(x) > 0$ when $x < \mu$ and $f'(x) < 0$ when $x > \mu$. Moreover, $f(x)$ attains its maximum value at $x = \mu$.*
- *$f(x)$ has two inflections points (change in concavity) at $x = \mu \pm \sigma$.*

Proof. To show the first property, it suffices to show that $f(\mu - x) = f(\mu + x)$. Indeed,

$$f(\mu - x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{-x}{\sigma}\right)^2\right] = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right] = f(\mu + x).$$

The next three properties are all simple calculus exercises. The second property is a direct consequence of

$$\lim_{x\to\pm\infty} -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 = -\infty.$$

Hence,

$$\lim_{x\to\pm\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma\sqrt{2\pi}} \lim_{x\to\pm\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = 0,$$

as needed.

To show the third property, note that

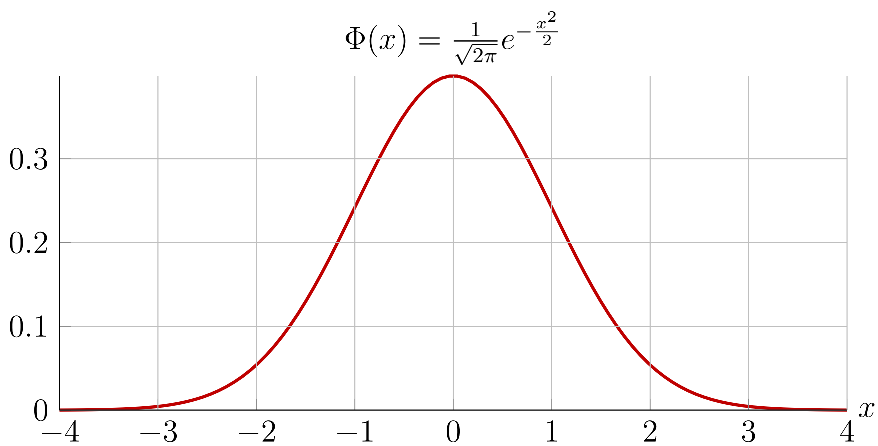
$$f'(x) = -\left(\frac{x-\mu}{\sigma^2}\right) f(x).$$

Since $f(x)$ is always positive by definition, we have that when $x < \mu$, $f'(x) > 0$ and when $x > \mu$, $f'(x) < 0$. Hence, the third property follows by the first derivative test and $f'(x) = 0$ when $x = \mu$. To show the fourth property, note that

$$f''(x) = f(x)\left(-\frac{1}{\sigma^2} + \frac{(x-\mu)^2}{\sigma^4}\right).$$

Note that $f''(x) = 0$ when $x = \mu \pm \sigma$. Since both $\frac{1}{\sigma^2}$ and $\frac{(x-\mu)^2}{\sigma^4}$ are both positive, $f''(x) > 0$ when $\frac{(x-\mu)^2}{\sigma^4} > \frac{1}{\sigma^2}$, which occurs when $x \in (-\infty, \mu-\sigma) \cap (\mu+\sigma, \infty)$. Similarly $f''(x) < 0$ when $\frac{(x-\mu)^2}{\sigma^4} < \frac{1}{\sigma^2}$, which occurs when $x \in (\mu-\sigma, \mu+\sigma)$. Hence, we have that $x = \mu \pm \sigma$ are inflections points, proving the fourth property. ∎

Figure 1. Example of a normal distribution with $\mu = 0$ and $\sigma^2 = 1$.



We will end our discussion here to review some necessary number theory. However we will return to some of the ideas when talking about *characteristic functions* and *moments*.

3. Number Theory Review

We will start with a result that we will not proof but will use to proof Theorem 3.2.

Theorem 3.1 (Abel Summation). *Denote* $\{a_n\}_{n=1}^{\infty}$ *to be a sequence of complex numbers. For* $t > 0$, *denote* $A(t) = \sum_{n\le t} a_n$. *Denote* $b(t)$ *to be a continuously differentiable function on the interval* $[1, x]$, *where* $x > 1$ *is a real number. Then,*

$$\sum_{1\le n\le x} a_n b(n) = A(x)b(x) - \int_1^x A(t)b'(t)\,dt.$$

We can now discuss a beautiful theorem that will be continuously used throughout this paper. The proof has been adapted from [Vil05].

Theorem 3.2 (Mertens' Second Theorem). *There exists a constant C such that for $x \geq 2$,*

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + C + \mathcal{O}\left(\frac{1}{\log x}\right) = \log\log x + \mathcal{O}(1).$$

Proof. Our proof is via Abel Summation across continuous functions. We define our a_n such that

$$a_n = \begin{cases} \frac{\log p}{p} & \text{if} \quad n = p \\ 0 & \text{if} \quad n \neq p, \end{cases}$$

and $b(x) = \frac{1}{\log x}$. Then, we have that $A(x) = \sum_{p \leq x} \frac{\log p}{p}$ (since only primes give a non-zero value) and then Theorem 3.1 gives us that

$$\sum_{p \leq x} \frac{1}{p} = \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2}\, dt.$$

Writing $A(x) = \log x + R(x)$, where $R(t) = \sum_{x \leq p} \frac{\log p}{p} - \log t$. Note that this means that $|R(t)| = 2$. Hence, we have that

$$\begin{aligned}
\sum_{p \leq x} \frac{1}{p} &= \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2}\, dt \\
&= \frac{\log x + R(x)}{\log x} + \int_2^x \frac{\log t + R(t)}{t(\log t)^2}\, dt \\
&= 1 + \frac{R(x)}{\log x} + \int_2^x \frac{1}{t \log t}\, dt + \int_2^x \frac{R(t)}{t(\log t)^2}\, dt \\
&= \log\log x - (\log\log 2 - 1) + \left(\frac{R(x)}{\log x} + \int_2^\infty \frac{R(t)}{t(\log t)^2}\, dt - \int_x^\infty \frac{R(t)}{t(\log t)^2}\right) dt \\
&= \log\log x + C + \mathcal{O}\left(\frac{1}{\log x}\right) \\
&= \log\log x + \mathcal{O}(1).
\end{aligned}$$

The reason why we go from the fifth equality to the sixth equality is due to the fact that C can be treated as an error term of $\mathcal{O}(1)$ and since $x \geq 2$, we have that $\mathcal{O}\left(\frac{1}{\log x}\right)$ gets absorbed into $\mathcal{O}(1)$. This completes the proof. ■

We will now begin our discussion on *arithmetic functions* [Apo76] which will allow us to proof Lemma 3.11. This lemma is in turn used in the proof of Erdös-Kac Theorem. We will start with the definition:

Definition 3.3. (Arithmetic Functions) Denote $f : \mathbb{N} \to \mathbb{C}$ as a function. We refer to f as an *arithmetic function.*

There are many basic arithmetic functions. A few examples are $\tau(n)$ which counts the number of (positive) integer divisors of n and $\sigma(n)$ which is the sum of the divisors of n. For example, $\sigma(12) = 1+2+3+4+6+12 = 28$ and $\tau(12) = 6$. Also, another arithmetic function is $N(n) = n$. However, in the context of the Erdös-Kac theorem we will be highlighting $\mu(n)$ and $\phi(n)$, the Möbius and Totient arithmetic functions.

Definition 3.4. (Möbius Function) Denote $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ as the prime factorization of n for primes $p_1, p_2, \cdots, p_k$. Then the *Möbius function*, $\mu(n)$, is defined as follows:

$$\mu(n) = \begin{cases} (-1)^k & \text{if} \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

and $\mu(1) = 1$.

Hence, we have that $\mu(n) = 0$ if n has any prime factor divides it more than once, otherwise $\mu(n) = -1$ if n has an odd number of primes factors and $\mu(n) = 1$ if n has an even number of prime factors. For example $\mu(19) = -1, \mu(20) = 0$, and $\mu(21) = 1$. We can jump straight into a very helpful theorem which be used to prove Theorem 3.7.

Theorem 3.5. *If $n \geq 1$, then*

$$\sum_{d|n} \mu(n) = \left\lfloor \frac{1}{x} \right\rfloor = \begin{cases} 1 & \text{if} \quad n = 1, \\ 0 & \text{if} \quad n > 1. \end{cases}$$

Proof. First when $n = 1$, we have that

$$\sum_{d|1} \mu(n) = \mu(1) = 1 = \left\lfloor \frac{1}{1} \right\rfloor,$$

as claimed. Now suppose that $n \geq 2$. Denote $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ as the prime factorization of n. Note that

$$\begin{aligned} \sum_{d|n} \mu(n) &= \mu(1) + \sum_{1 \leq i \leq k} \left(\sum_{1 \leq a_1 < a_2 < \cdots < a_i \leq k} \mu(p_{a_1} p_{a_2} \cdots p_{a_i}) \right) \\ &= \mu(1) + \mu(p_1) + \mu(p_2) + \mu(p_3) + \cdots + \mu(p_n) + \mu(p_1 p_2) + \cdots + \mu(p_{k-1} p_k) + \cdots + \mu(p_1 p_2 \cdots p_k), \end{aligned}$$

since the only divisors of n that have a non-zero $\mu(\bullet)$ value are 1 and product of distinct primes. However, note that there are $\binom{k}{1}$ primes that have $\mu(\bullet) = (-1)^1 = -1$, $\binom{k}{2}$ pairs of pair of primes such that the $\mu(\bullet)$ of their product is $(-1)^2 = 1$, and so on. Hence,

$$\begin{aligned} \sum_{d|n} \mu(n) &= \mu(1) + \sum_{1 \leq i \leq k} \left(\sum_{1 \leq a_1 < a_2 < \cdots < a_i \leq k} \mu(p_{a_1} p_{a_2} \cdots p_{a_i}) \right) \\ &= \mu(1) + (-1)^1 \binom{k}{1} + (-1)^2 \binom{k}{2} + (-1)^3 \binom{k}{3} + \cdots + (-1)^k \binom{k}{k} \\ &= (1-1)^k \\ &= 0, \end{aligned}$$

as claimed for $n \geq 2$. This completes the proof. ∎

We will now introduce the Euler Totient Function.

Definition 3.6. For $n \geq 1$, the *Euler Totient Function*, $\phi(n)$, counts the number of values of $1 \leq a \leq n$ that are relatively prime to n.

For example, $\phi(12) = 4$ since $\gcd(1, 12) = \gcd(5, 12) = \gcd(7, 12) = \gcd(11, 12) = 1$. We now can explore a relationship between $\mu(n)$ and $\phi(n)$:

Theorem 3.7. *For $n \geq 1$, we have*

$$\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}.$$

Proof. By the definition of $\phi(n)$, we can write

$$\phi(n) = \sum_{k=1}^{n} \left\lfloor \frac{1}{\gcd(n, k)} \right\rfloor.$$

Using Theorem 3.5 with $n \equiv \gcd(n, k)$ we have,

$$\phi(n) = \sum_{k=1}^{n} \sum_{d|\gcd(n,k)} \mu(d) = \sum_{k=1}^{n} \sum_{d|n, d|k} \mu(d).$$

We now bring our attention to the innermost sum. We have that for a fixed value d, we need to sum over all the values of d that divides n. Moreover, for $1 \leq k \leq n$, we would also need to sum the values of k that are multiples of d. Hence, if $k = qd$ then $1 \leq k \leq n$ if and only if $1 \leq q \leq \frac{n}{d}$. Hence,

$$\phi(n) = \sum_{k=1}^{n} \sum_{q=1}^{n/d} \mu(d) = \sum_{d|n} \mu(d) \sum_{q=1}^{n/d} 1 = \sum_{d|n} \mu(d) \frac{n}{d}.$$

This completes the proof. ∎

What we saw in Theorem 2.5 is a frequent sum that we see in number theory, that is sums of the form

$$\sum_{d|n} f(d) g\left(\frac{n}{d}\right),$$

where f and g are two arithmetic functions. In fact, this sum has a special name for it:

Definition 3.8 (Dirichlet Convolution)**.** If f and g are two arithmetic functions, we define h, another arithmetic function, to be the Dirichlet convolution of f and g by

$$h(n) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right).$$

As a shorthand, we can write $h(n) = (f * g)(n)$ or $h = f * g$. Notice, by Theorem 3.7, $\phi = \mu * N$. To see how Dirichlet Convolution can act like normal operations which we are familiar with, we will define one last arithmetic function:

Definition 3.9. (Identity Function) For $n \geq 1$, we have

$$I(n) = \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1 & \text{if} \quad n = 1, \\ 0 & \text{if} \quad n > 1. \end{cases}$$

A simple theorem with $I(n)$ makes it clear why the function is named "identity":

Theorem 3.10. *Denote f as an arithmetic function and I as the identity function. Then,*

$$f * I = I * f = f.$$

Proof. We have that

$$(f * I)(n) = \sum_{d|n} f(d) I\left(\frac{n}{d}\right) = \sum_{d|n} f(d) \left\lfloor \frac{d}{n} \right\rfloor = f(n) \quad \text{and}$$
$$(I * f)(n) = \sum_{d|n} I f\left(\frac{n}{d}\right) = \sum_{d|n} \left\lfloor \frac{1}{n} \right\rfloor f\left(\frac{n}{d}\right) = f(n),$$

since $\lfloor n/k \rfloor = 0$ when $k > n$. ■

With our knowledge on arithmetic functions, we can study the following lemma which is pivotal in the proof of Erdös-Kac Theorem. The proof has been adapted from [GS06]

Lemma 3.11. *Suppose a positive integer $r > 2$ and denote R as the product of the distinct primes powers of r. Then,*

$$\sum_{\substack{n \le x \\ \gcd(n,R)=d}} 1 = \frac{x}{R}\phi\left(\frac{R}{d}\right) + \mathcal{O}\left(\tau\left(\frac{R}{d}\right)\right).$$

Proof. Write $n = md$. Note that $R = \left(\frac{R}{d}\right) d$ since $d = \gcd(n, R) \implies d \mid R$. From here we can do some manipulation in the indices:

$$\begin{aligned} \sum_{\substack{n \le x \\ \gcd(n,R)=d}} 1 &= \sum_{\substack{md \le x \\ \gcd(md,R)=d}} 1 \\ &= \sum_{\substack{md \le x \\ \gcd(md,(R/d)d)=d}} 1 \\ &= \sum_{\substack{m \le x/d \\ \gcd(m,R/d)=1}} 1 \\ &= \sum_{m \le x/d} \sum_{k | \gcd(m,R/d)} \mu(k). \end{aligned}$$

Since R is square-free by definition, we have that this implies that m and k are square-free as well. Since $k \mid \gcd(m, R/d) \implies k \mid m$. Hence,

$$\sum_{m \le x/d} \sum_{k|\gcd(m,R/d)} \mu(k) = \sum_{k|R/d} \mu(k) \sum_{\substack{m \le x/d \\ k|m}} 1.$$

We have that the inner sum is $\lfloor \frac{x}{dk} \rfloor = \frac{x}{dk} - \{ \frac{x}{dk} \} = \frac{x}{dk} + \mathcal{O}(1)$. Hence, substituting and using Theorem 3.7,

$$\begin{aligned}
\sum_{k|R/d} \mu(k) \sum_{\substack{m \leq x/d \\ k|m}} 1 &= \sum_{k|R/d} \mu(k) \left(\frac{x}{dk} + \mathcal{O}(1) \right) \\
&= \frac{x}{d} \sum_{k|R/d} \frac{\mu(k)}{k} + \sum_{k|R/d} \mu(k)\mathcal{O}(1) \\
&= \frac{x}{d}\frac{d}{R} \sum_{k|R/d} \frac{\mu(k)}{k}\frac{R}{d} + \sum_{k|R/d} \mu(k)\mathcal{O}(1) \\
&= \frac{x}{R}\phi\left(\frac{R}{d}\right) + \mathcal{O}\left(\sum_{k|R/d} 1 \right) \\
&= \frac{x}{R}\phi\left(\frac{R}{d}\right) + \mathcal{O}\left(\tau\left(\frac{R}{d}\right) \right).
\end{aligned}$$

This completes the proof. ∎

4. Characteristic Functions

The proof of Erdös-Kac Theorem uses central limit theorems which we will cover in Section 6; however, before then, we need to understand *characteristic functions*. This section is an adaption of Chapter 5.1 in [PB95]. We will start with the formal definition:

Definition 4.1 (Characteristic Function)**.** The *characteristic function* of a probability measure μ on the line is define for $t \in \mathbb{R}$ by

$$\begin{aligned}
\varphi(t) &= \int_{-\infty}^{\infty} \mu e^{itx}\,dx \\
&= \int_{-\infty}^{\infty} \mu \cos(tx)\,dx + i\int_{-\infty}^{\infty} \mu \sin(tx)\,dx,
\end{aligned}$$

where $i^2 = -1$.

However, for our purposes we naturally have the following to be true:

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} \mu e^{itx}\,dx.$$

From here, we will consider the multiplicative property of characteristic functions. However, here we need to consider the independence of our two events.

Proposition 4.2 (Characteristic Functions Multiplication Property)**.** *Suppose that X_1 and X_2 are independent random variables with characteristic functions φ_1 and φ_2. Then,*

$$\varphi_1(t)\varphi_2(t) = \mathbb{E}[e^{it(X_1+X_2)}].$$

Proof. Denote $Y_i = \cos(X_i)$ and $Z_i = \sin(tX_i)$, then note that (Y_1, Z_1) and (Y_2, Z_2) are independent since X_1 and X_2 are independent by assumption. Then from the expected value definition of the characteristic function and Linearity of Expectation,

$$\begin{aligned}\varphi_1(t)\varphi_2(t) &= \mathbb{E}[Y_1 + iZ_1]\mathbb{E}[Y_2 + iZ_2] \\ &= (\mathbb{E}[Y_1] + i\mathbb{E}[Z_1])(\mathbb{E}[Y_2] + i\mathbb{E}[Z_2]) \\ &= \mathbb{E}[Y_1]\mathbb{E}[Y_2] - \mathbb{E}[Z_1]\mathbb{E}[Z_2] + i(\mathbb{E}[Y_1]\mathbb{E}[Z_2] + \mathbb{E}[Y_2]\mathbb{E}[Z_1]) \\ &= \mathbb{E}[Y_1Y_2 - Z_1Z_2 + i(Y_1Z_2 + Z_1Y_2)] \\ &= \mathbb{E}[e^{it(X_1+X_2)}],\end{aligned}$$

as needed. ■

In fact, this property can be generalized. If $X_1, X_2, \cdots, X_n$ are independent, then

$$\mathbb{E}[e^{it\sum_{k=1}^{n} X_k}] = \prod_{k=1}^{n} \mathbb{E}[e^{itX_k}].$$

Finally, to conclude our discussion on characteristic functions, we have the follow property: if X has the characteristic function $\varphi(t)$, then $aX + b$ has the characteristic function

$$\mathbb{E}[e^{it(aX+b)}] = e^{itb}\varphi(at).$$

In particular, when $(a, b) = (-1, 0)$, the characteristic function of $-X$ is $\varphi(-t)$; the complex conjugate of $\varphi(t)$.

5. The Method of Moments

The wonderful proof of Erdös-Kac theorem by Granville and Soundarajan utilizes the idea of showing that the moments of the assumption of Erdös-Kac theorem converges to the moments of the normal distribution (with $\mu = 0$ and $\sigma^2 = 1$). Hence, we will need to study some theory on moments; we will start with the definition, courtesy of [Tsu].

Definition 5.1. (Moments) Denote X to be a random variable and a scalar $c \in \mathbb{R}$. Then, the k^{th} *moment* of X is

$$\mathbb{E}[X^k],$$

and the k^{th} *moment* of X (about c) is

$$\mathbb{E}[(X - c)^k].$$

For the purposes of probabilistic number theory and our main focus on Erdös-Kac theorem, we are interested when $c = \mu$, refereed to as the central moment.

Definition 5.2. (Central Moments) Denote X to be a continuous random variable. Then, the k^{th} *central moment* is defined as

$$\mathbb{E}[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k f(x)\, dx.$$

In this paper we have already saw the first moment, being μ, and the second moment, being σ^2. We have that the third moment represents *skewness* of the distribution measuring the two tails of the distribution and the fourth moment represents the *kurtosis* of the distribution measuring the total size of the tails relative to the entire distribution. Interestingly, higher moments behave in a similar way as skewness and kurtosis: with the k^{th} moment ($k \geq 3$)

when $k \equiv 1 \pmod 2$, measuring the relative nature of the two tails with respect to each other, while when $k \equiv 0 \pmod 2$ measures the total tail-ends together as a whole relative to the entire distribution. With definitions in place we can explore a theorem which is vital for the proof of Erdös-Kac Theorem:

Theorem 5.3. *Denote X as a continuous random variable that exhibits the normal probability distribution. Denote $\mathbb{E}[(X-\mu)^k]$ as the k^{th} central moment of X. Then,*

$$\mathbb{E}[(X-\mu)^{2k+1}] = 0 \quad \textit{and} \quad \mathbb{E}[(X-\mu)^{2k}] = \frac{(2k)!\sigma^{2k}}{k!2^k}.$$

Proof. We will first start with evaluating $\mathbb{E}[(X-\mu)^{2k+1}]$. Note that, by the first property of Lemma 2.8, we have that,

$$\int_{-\infty}^{\mu} (x-\mu)^{2k+1} f(x)\, dx = -\int_{\mu}^{\infty} (x-\mu)^{2k+1} f(x)\, dx.$$

Hence,

$$\begin{aligned}
\mathbb{E}[(X-\mu)^{2k+1}] &= \int_{-\infty}^{\infty} (x-\mu)^{2k+1} f(x)\, dx \\
&= \int_{-\infty}^{\mu} (x-\mu)^{2k+1} f(x)\, dx + \int_{\mu}^{\infty} (x-\mu)^{2k+1} f(x)\, dx \\
&= -\int_{\mu}^{\infty} (x-\mu)^{2k+1} f(x)\, dx + \int_{\mu}^{\infty} (x-\mu)^{2k+1} f(x)\, dx \\
&= 0,
\end{aligned}$$

as needed. We will now show the even moment case by induction. Before we start the induction, we will clean up $\mathbb{E}[(X-\mu)^{2k}]$. Denote $u = \frac{z-\mu}{\sigma}$, so $du = \frac{dz}{\sigma}$. Also, denote $\Phi(u)$ as the normal density function in terms of u. Hence, by Definition 5.2,

$$\begin{aligned}
\mathbb{E}[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \int_{-\infty}^{\infty} u^2\sigma^3 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du \\
&= \sigma^2 \int_{-\infty}^{\infty} u^2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\right) du \\
&= \sigma^2 \int_{-\infty}^{\infty} u^2 \Phi(u)\, du. \\
&= \sigma^2 \mathbb{E}[U^2] \\
&= \sigma^2(\sigma^2 - (\mathbb{E}[U])^2) \\
&= \sigma^2(1-0^2) \\
&= \sigma^2.
\end{aligned}$$

Hence, we have established that $\mathbb{E}[(X-\mu)^2] = \sigma^2$. Note that when $k=1$, $\sigma^2 = \frac{(2k)!\sigma^{2k}}{k!(2^k)}$, which is our base case. Now suppose that $\mathbb{E}[(Z-\mu)^{2n}] = \frac{(2n)!\sigma^{2n}}{(n!)2^n}$. Before proceeding to our inductive step, we must proof the following proposition:

Proposition 5.4. *For $n \in \mathbb{N}$,*

$$\mathbb{E}[(X-\mu)^{n+1}] = n\sigma^2\mathbb{E}[(X-\mu)^{n-1}].$$

Proof. Recall from the proof of Lemma 2.8 property 3, $-\sigma^2 f'(x) = (x-\mu)f(x)$. Hence,

$$\begin{aligned}\mathbb{E}[(Z-\mu)^{n+1}] &= \int_{-\infty}^{\infty}(x-\mu)^{n+1}f(x)\,dx\\ &= \int_{-\infty}^{\infty}(x-\mu)^{n}(x-\mu)f(x)\,dx\\ &= -\sigma^2\int_{-\infty}^{\infty}(x-\mu)^{n}f'(x)\,dx.\end{aligned}$$

From here we can proceed with the classic integration by parts technique. With $u=(x-\mu)^n$, $du = k(x-\mu)^{n-1}$, $dv = f'(x)$, and $v = f(x)$,

$$\begin{aligned}-\sigma^2\int_{-\infty}^{\infty}(x-\mu)^{n}f'(x)\,dx &= \left[-\sigma^2(x-\mu)^n f(x)\right]_{-\infty}^{\infty} + \sigma^2\int_{-\infty}^{\infty} n(x-\mu)^{n-1}f(x)\,dx.\\ &= n\sigma^2\int_{-\infty}^{\infty}(x-\mu)^{n-1}f(x)\,dx\\ &= n\sigma^2\mathbb{E}[(X-\mu)^{n-1}],\end{aligned}$$

as needed. Note that reason why the first summand is zero is due to L'Hôpital's rule applied $n+1$ times. ■

Now we can come back and finish the induction. We have that

$$\begin{aligned}\mathbb{E}[(X-\mu)^{2n+2}] &= (2n+1)\sigma^2\mathbb{E}[(Z-\mu)^{2n}]\\ &= (2n+1)\sigma^2\mathbb{E}[(X-\mu)^{2n}]\\ &= \frac{(2n+1)!\sigma^{2n+2}}{(n!)2^n}\frac{2n+2}{2n+2}\\ &= \frac{(2n+2)!\sigma^{2n+2}}{(n+1)!2^{n+1}},\end{aligned}$$

which is just what we wanted. Hence, we conclude by induction. ■

We finish this section by defining *moment-generating function*.

Definition 5.5. (Moment-Generating Function) Denote X be a continuous random variable. The *moment-generating function* $m(t)$ of X is

$$m(t) = \mathbb{E}[e^{tX}].$$

6. The Central Limit Theorem

As stated in Section 1, the Erdös-Kac Theorem asserts the Central Limit Theorem on $\omega(n)$. We have that the expression 1.2 does not come out of nowhere and is actually a result of the following Central Limit Theorem:

Theorem 6.1. *Denote $\{X_n\}$ to be an independent sequence of random variables having the same distribution with mean μ and finite positive variance σ^2, then*

$$\lim_{x\to\infty} \frac{S_n - \mu n}{\sigma\sqrt{n}} \to \chi,$$

where $S_n = X_1 + X_2 + \cdots + X_n$ and χ has the normal distribution with mean 0 *and variance* 1.

This theorem is essentially modeling the idea that the sum of many independent random variables will be approximately normally distributed given that each summand has a high probability of being small. As mentioned in Section 2, the normally distribution can be graphically represented by a bell curve as seen in Figure 1.

We have that Theorem 5.2 is a subcase of the bigger central limit theorem, the Lindeberg-Feller Central Limit Theorem, which is pivotal for the proof of Erdös-Kac Theorem. We will start our discussion with a definition:

Definition 6.2. Denote $\{a_n\}_n$ to be a sequence of natural numbers such that if for each $n \in \mathbb{N}$, we have that there exists random variables $X_{n,1}, \cdots, X_{n,r_n}$, the collection $\{X_{n,k}\}_{n,k}$ is called a *triangular array.*

Visually, we can represent a triangular array as we did in Figure 2.

Figure 2. Triangular Array

$$\begin{array}{llllll}
X_{1,1} & & & & & \\
X_{2,1} & X_{2,2} & & & & \\
X_{3,1} & X_{3,2} & X_{3,3} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
X_{i,1} & X_{i,2} & X_{i,3} & \cdots & X_{i,i} & \\
\vdots & \vdots & \vdots & & & \ddots
\end{array}$$

From here we can define the Lindeberg-Levy-Feller CLT.

Theorem 6.3. *For each n and $1 \le m \le n$, let $X_{n,m}$ be independent random variables with* $\mathbb{E}[X_{n,m}] = 0$. *If*

(1) $\sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2] \to \sigma^2 > 0$, *and*
(2) *For all $\epsilon > 0$,* $\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \epsilon] = 0$

Then $S_n = X_{n,1} + \cdots X_{n,n} \Rightarrow \sigma_\chi$ as $n \to \infty$.

We have that (2) is called the *Lindeberg condition.* The theorem is basically saying that the sum of a large number of small independent effects (not necessarily identically distribution) has a normal distribution. This idea is the reason why we will be using the Lindeberg-Feller CLT rather than the the first CLT mentioned. This will be talked about more in Section 7. We will first show how Theorem 6.3 implies Theorem 6.1.

Proposition 6.4. *Theorem 6.3 implies Theorem 6.1*

Proof. Suppose that $(Y_1, Y_2, \cdots, Y_n)$ are independent and identically distributed random variables under the assumption of the Central Limit Theorem. WLOG, assume that $\mathbb{E}[Y_n] = 0$ and $\mathbb{E}[Y_n^2] = \sigma^2$. As per the condition of Lindeberg-Feller Central Limit Theorem, $X_{n,m} = \frac{Y_m}{\sqrt{n}}$. By the first condition of Lindeberg-Feller Central Limit Theorem,

$$\lim_{x\to\infty} \sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2) = \sigma^2.$$

Also, we have that

$$\sum_{m=1}^{n} \mathbb{E}(X_{n,m}^2) = n\mathbb{E}(X_{n,m}^2) = \mathbb{E}(Y_1^2) = 0,$$

as n approaches infinity. Hence, we have showed what we asserted. This completes the proof. ■

It can be shown that the characteristic function of the normal distribution with mean μ and variance σ^2 is

$$\varphi(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right).$$

A proof of this can be found in [Dur19]. The below proof of Theorem 6.3 is adapted from [Dur19] as well, with many parts omitted for the interested reader to check thoroughly in the book for further clarity.

Proof of Theorem 6.3. Denote $\varphi_{n,m}(t) = \mathbb{E}[\exp(itX_{n,m})]$ and $\sigma_{n,m}^2 = \mathbb{E}[X_{n,m}^2]$. It suffices to show that

$$\prod_{m=1}^{n} \varphi_{n,m}(t) \to \exp(-t^2\sigma^2/2).$$

It can be show that

$$|z_{n,m} - w_{n,m}| \leq \epsilon t^3 \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| \leq \epsilon] + 2t^2 \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \epsilon],$$

where $z_{n,m} = \varphi_{n,m}(t)$ and $w_n = 1 - \frac{t^2\sigma_{n,m}^2}{2}$. Now, we have that the conditions implies that

$$\lim_{n\to\infty} \sum_{m=1}^{n} |z_{m,n} - y_{m,n}| \leq \epsilon t^3 \sigma^2 \quad \text{and} \quad \left|\prod_{m=1}^{n} \varphi_{n,m}(t) - \prod_{m=1}^{n}\left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| \to 0.$$

We have that,

$$\left|\prod_{m=1}^{n}\left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| \to \exp\left(-\frac{t^2\sigma_{n,m}}{2}\right).$$

Hence, we see that

$$\left|\prod_{m=1}^{n} \varphi_{n,m}(t) - \exp\left(-\frac{t^2\sigma_{n,m}}{2}\right)\right| \to 0.$$

■

Hence we get that,

$$\prod_{m=1}^{n} \varphi_{n,m}(t) \to \exp(-t^2\sigma^2/2),$$

as needed. This completes the proof.

7. Heuristics

We are almost to the goal of proving Erdös-Kac theorem but we would need to define a few more terminologies to get ready for the proof. What we have shown till now is that if m_k are the moments of a random variable under the normal distribution (when $\mu = 0$ and $\sigma^2 = 1$), then

$$m_k = \begin{cases} \frac{(2\ell)!}{\ell! 2^\ell} & \text{if} \quad k = 2\ell, \\ 0 & \quad \text{otherwise.} \end{cases}$$

We would like to show that the moments of

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}}$$

are *asymptotically* equal to the moments of the normal distribution. We would like to stress the word asymptotically–we are not showing that the moments of the two actually match, but rather match to a small degree of error. Hence, to prove the Erdös-Kac Theorem, it suffices to show that

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \le x} (\omega(n) - \log\log x)^k = m_k (\log\log x)^{k/2} + o(\log\log x)^{k/2}. \tag{7.1}$$

Note that this argument can be made due to the fact that the normal distribution is determined by its moments. This is not always the case. As an example, the *lognormal function* is not defined by its moments. We will now define some probabilistic heuristics which will be a common sight in our next section. Denote

$$g_p(n) = \begin{cases} 1 & \text{if} \quad p \mid n, \\ 0 & \text{otherwise.} \end{cases}$$

From this definition, we can construct a probabilistic model for $g_p(n)$. As defined in Section 2, we denote $X(p)$ to be a Bernoulli trail such that

$$X(p) = \begin{cases} 1 & \text{with probability} \quad \frac{1}{p}, \\ 0 & \text{with probability} \quad 1 - \frac{1}{p}. \end{cases}$$

From here, note that we can find the mean and variance very easily from the model defined:

$$\begin{aligned}
\mathbb{E}[X(p)] &= 1 \cdot \frac{1}{p} + 0 \cdot \left(1 - \frac{1}{p}\right) = \frac{1}{p}, \\
\sigma_p^2 &= \mathbb{E}\left[\left(X(p) - \frac{1}{p}\right)^2\right] \\
&= \mathbb{E}[X(p)^2] - \frac{2}{p}\mathbb{E}[X(p)] + \frac{1}{p^2} \\
&= \mathbb{E}[X(p)] - \frac{2}{p}\mathbb{E}[X(p)] + \frac{1}{p^2} \\
&= \frac{1}{p} - \frac{2}{p^2} + \frac{1}{p^2} \\
&= \frac{1}{p}\left(1 - \frac{1}{p}\right).
\end{aligned}$$

Now, notice by Theorem 3.2,

$$\sum_{p \le n} \sigma_p^2 = \sum_{p \le n} \frac{1}{p}\left(1 - \frac{1}{p}\right) = \sum_{p \le n} \frac{1}{p} - \sum_{p \le n} \frac{1}{p^2} = \log\log n + \mathcal{O}(1).$$

As a result from Theorem 6.3, as $y \to \infty$,

$$\mathbb{E}\left[\left(\sum_{p \le y}\left(X_p - \frac{1}{p}\right)\right)^k\right] = m_k(\log\log y)^{k/2} + o(\log\log y)^{k/2} \tag{7.2}$$

Note that we must use the Lindeberg-Feller Central Limit Theorem opposed to the Central Limit Theorem because we are not guaranteed that X_p has the same distribution for all p, which is an assumption of the Central Limit Theorem and not the Lindeberg-Feller Central Limit Theorem.

Hence, we see that the probabilistic model which was defined is very similar to what we are dealing with in Erdös-Kac Theorem, in particular the sum of the $X(p)$ values over primes at most n. Before we proceed, denote the following function $f_p(n)$:

$$f_p(n) = \begin{cases} 1 - \frac{1}{p} & \text{if} \quad p \mid n, \\ -\frac{1}{p} & \text{otherwise.} \end{cases}$$

Note that $f_p(n) = g_p(n) - 1/p$. We can now proof Erdös-Kac Theorem: by finding a connection between equation 7.2 and $f_p(n)$ and finding a relation between $f_p(n)$ and $\omega(n)$. Hence, these two connections can establish a relation between 7.2 and $\omega(n)$, which we wish to obtain due to the right-hand side of 7.2.

8. Proof of Erdös-Kac Theorem

We have that Theorem 8.1 prompts us in the right direction of what we would like to show and will give a proof of it towards the end of this section. For now, assume it to be true.

In this section we have another variable y which is defined to be a function of x, $y(x)$, such that $y(x) \to \infty$ as $x \to \infty$. This particular value of y will be analyzed just before Lemma 8.2 and shown to be suitable in our setting of Erdös-Kac Theorem in that lemma itself.

Theorem 8.1. *For $k \geq 1$,*

$$\frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^k = \mathbb{E}\left[\left(\sum_{p\leq y}\left(X(p)-\frac{1}{p}\right)\right)^k\right] + \mathcal{O}\left(\frac{y^k}{x(\log y)^k}\right).$$

With Theorem 8.1 we will first proof that equation 7.1 is indeed true. From there, we will proof the above assumption made.

We will begin by finding the moments of $\omega(x) - \log\log x$. For $1 \leq n \leq x$, denote

$$\omega(n) - \log\log x = \sum_{p\leq y} f_p(n) + R(x),$$

where $R(x)$ is an error term such that it will get replaced as $\mathcal{O}(\bullet)$. Note that $R(x)$ does not depend on n. From here, we can specifically look at the quantity we want in equation 7.1.We have,

$$\begin{aligned}
\frac{1}{x}\sum_{n\leq x}(\omega(n)-\log\log x)^k &= \frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n) + R(x)\right)^k \\
&= \frac{1}{x}\sum_{n\leq x}\sum_{j=0}^{k}\binom{k}{j}\left(\sum_{p\leq y} f_p(n)\right)^j R(x)^{k-j} \\
&= \frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^k + \frac{1}{x}\sum_{n\leq x}\sum_{j=0}^{k-1}\binom{k}{j}\left(\sum_{p\leq y} f_p(n)\right)^j R(x)^{k-j} \\
&\leq \frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^k + R(x)^k\sum_{j=0}^{k-1}\binom{k}{j}\frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^j \\
&\leq \frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^k + 2^k R(x)^k \max_{j\in\{0,1,\cdots,k-1\}}\frac{1}{x}\sum_{n\leq x}\left|\sum_{p\leq y} f_p(n)\right|^j
\end{aligned}$$

For brevity, denote

$$\mathcal{E} = \frac{1}{x}\sum_{n\leq x}(\omega(x)-\log\log x)^k.$$

Hence by Cauchy-Schwarz,

$$\mathcal{E} \leq \frac{1}{x}\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^k + 2^k R(x)^k \max_{j\in\{0,1,\cdots,k-1\}}\frac{1}{x}\left(\sum_{n\leq x} 1\right)^{\frac{1}{2}}\left(\sum_{n\leq x}\left(\sum_{p\leq y} f_p(n)\right)^{2j}\right)^{\frac{1}{2}}.$$

Using Theorem 8.1 twice and equation 7.2, we have that

$$\mathcal{E} = m_k(\log\log y)^{k/2} + o(\log\log y)^{k/2} + \mathcal{O}\left(\frac{y^k}{x(\log y)^k}\right) + \mathcal{O}\left(\frac{R(x)^k}{\sqrt{x}}(\log\log y)^{\frac{k-1}{4}}\right).$$

Looking at what we would like to proof and the expression above, we are very close. Since y is a function of x, we want to find a value of y such that $\log\log y$ is asymptotic to $\log\log x$ and such that the the two rightmost $\mathcal{O}(\bullet)$ error terms can get absorbed into $o(\log\log x)^{k/2}$.

Such a y with this property can be found in Lemma 8.2:

Lemma 8.2. *Denote $y = x^{1/\log\log\log x}$. Then for $n \leq x$,*

$$\omega(n) - \log\log x = \sum_{p \leq y} f_p(n) + \mathcal{O}(\log\log\log x).$$

Proof. By definition of $\omega(n)$ and the asymptotic equivalence established by Theorem 3.2, we can represent the left hand side in terms of a sum:

$$\begin{aligned}
\omega(n) - \log\log x &= \sum_{p|n} 1 - \sum_{p \leq x} \frac{1}{p} + \mathcal{O}(1) \\
&= \sum_{\substack{p|n \\ p \leq y}} 1 + \sum_{\substack{p|n \\ p > y}} 1 - \sum_{p \leq y} \frac{1}{p} - \sum_{y < p \leq x} \frac{1}{p} + \mathcal{O}(1) \\
&= \sum_{\substack{p|n \\ p \leq y}} 1 + \sum_{\substack{p|n \\ p > y}} 1 - \left(\sum_{\substack{p|n \\ p \leq y}} \frac{1}{p} + \sum_{\substack{p\nmid n \\ p \leq y}} \frac{1}{p} \right) - \sum_{y < p \leq x} \frac{1}{p} + \mathcal{O}(1) \\
&= \sum_{\substack{p|n \\ p \leq y}} \left(1 - \frac{1}{p}\right) + \sum_{\substack{p\nmid n \\ p \leq y}} \frac{1}{p} + \sum_{\substack{p|n \\ p > y}} 1 - \sum_{y < p \leq x} \frac{1}{p} + \mathcal{O}(1) \\
&= \sum_{p \leq y} f_p(n) + \sum_{\substack{p|n \\ p > y}} 1 - \sum_{y < p \leq x} \frac{1}{p} + \mathcal{O}(1) \\
&= \sum_{p \leq y} f_p(n) + \sum_{\substack{p|n \\ p > y}} 1 - (\log\log x - \log\log y) + \mathcal{O}(1) \\
&= \sum_{p \leq y} f_p(n) + \sum_{\substack{p|n \\ p > y}} 1 - (\log\log\log\log x) + \mathcal{O}(1)
\end{aligned}$$

where we go from the fourth to fifth equality from the definition of $f_p(n)$ and from the fifth to sixth equality from Theorem 3.2 again. Now we will study the second summand. If there are ℓ distinct primes, $p_1, p_2, \cdots, p_\ell$ that satisfy the second summand, then we have that following inequality:

$$y^\ell \leq \prod_{i=1}^{\ell} p_i \leq n \leq x \Longleftrightarrow y^\ell \leq x.$$

Hence, we have that that the second sum is bounded above by $\frac{\log x}{\log y} = \log\log\log x$. Hence, we have that,

$$\begin{aligned}\omega(n) - \log\log x &= \sum_{p \le y} f_p(n) + \sum_{\substack{p \mid n \\ p > y}} 1 - (\log\log\log\log x) + \mathcal{O}(1) \\ &= \sum_{p \le y} f_p(n) + \mathcal{O}(\log\log\log x),\end{aligned}$$

from our bounding argument. This completes the proof. ■

We have that Lemma 8.2 can meet all the conditions stated before proving the lemma and as a result shows that equation 7.1 is true. First, it can be shown that $\log\log y$ is asymptotic to $\log\log x$ when $y = x^{1/\log\log\log x}$. Moreover, recall that $R(x)$ is an error term that can be expressed in terms of $\mathcal{O}(\bullet)$. We take the error from Lemma 8.2 to denote $R(x)$ (that is, $R(x) = \log\log\log x$). Note that we can also replace y with x everywhere. Hence, we have that

$$\begin{aligned}\mathcal{E} &= m_k(\log\log y)^{k/2} + o(\log\log y)^{k/2} + \mathcal{O}\left(\frac{y^k}{x(\log y)^k}\right) + \mathcal{O}\left(\frac{R(x)^k}{\sqrt{x}}(\log\log y)^{\frac{k-1}{4}}\right) \\ &= m_k(\log\log x)^{k/2} + o(\log\log x)^{k/2} + \mathcal{O}\left(\frac{x^k}{x(\log x)^k}\right) + \mathcal{O}\left(\frac{(\log\log\log x)^k}{\sqrt{x}}(\log\log x)^{\frac{k-1}{4}}\right) \\ &= m_k(\log\log x)^{k/2} + o(\log\log x)^{k/2}.\end{aligned}$$

Hence, we have shown that

$$\mathcal{E} = \frac{1}{x}\sum_{n \le x}(\omega(n) - \log\log x)^k = m_k(\log\log x)^{k/2} + o(\log\log x)^{k/2},$$

which is exactly what we wanted. Hence, we have proved Erdös-Kac Theorem! Now, we must proof Theorem 8.1, that we assumed. We will now define our last set of notations. Denote $m = \prod_{i=1}^r p_i^{\alpha_i}$, $M = \prod_{i=1}^r p_i$, and

$$f_m(n) = \prod_{j=1}^r \left(f_{p_j}(n)\right)^{\alpha_j},$$

as an altered version to our previously define $f_p(n)$. Note that as a consequence of Proposition 2.2,

$$\mathbb{E}\left[\prod_{j=1}^r \left(X(p_j) - \frac{1}{p_j}\right)^{a_j}\right] = \prod_{j=1}^r \mathbb{E}\left[\left(X(p_j) - \frac{1}{p_j}\right)^{\alpha_i}\right].$$

Hence, by a standard calculation of expectation from the model of $X(p)$ defined in Section 7, we have

$$\prod_{j=1}^r \mathbb{E}\left[\left(X(p_j) - \frac{1}{p_j}\right)^{\alpha_i}\right] = \prod_{j=1}^r \left(\frac{1}{p_j}\left(1 - \frac{1}{p_j}\right)^{\alpha_i} + \left(-\frac{1}{p_j}\right)^{\alpha_i}\left(1 - \frac{1}{p_j}\right)\right). \tag{8.1}$$

Note from Lemma 3.11, when $(R,d)=(a,1)$ we have that

$$\sum_{\substack{n\le x\\ \gcd(n,a)=1}} 1 = x\frac{\phi(a)}{a}+\mathcal{O}(\tau(a)).$$

This representation of the lemma is crucial in Proposition 8.3:

Proposition 8.3.

$$\frac{1}{x}\sum_{n\le x} f_m(n)=\prod_{j=1}^{r}\left(\frac{1}{p_j}\left(1-\frac{1}{p_j}\right)^{\alpha_j}+\left(-\frac{1}{p_j}\right)^{\alpha_i}\left(1-\frac{1}{p_j}\right)\right)+\mathcal{O}\left(\frac{2^{2r}}{x}\right)$$

Proof. We have that

$$\begin{aligned}\sum_{n\le x} f_m(n) &= \sum_{d\mid M}\sum_{\substack{n\le x\\ \gcd(M,n)=d}} f_m(n)\\ &= \sum_{d\mid M} f_m(d)\sum_{\substack{n/d\le x/d\\ (M/d,n/d)=1}} 1\\ &= x\sum_{d\mid M} f_m(d)\frac{\phi(M/d)}{M}+\mathcal{O}\left(\tau(M)^2\right).\end{aligned}$$

Since $|f_m(d)|\le 1$, and $\tau(M)=2^r$ since M has r distinct primes that are square-free, by construction. Hence, we can simply look at the sum without the coefficient of x and have an error term of $\mathcal{O}\left(\frac{2^{2r}}{x}\right)$ as we have divided by x to get what we wanted on the left-hand side. From the definition of $f_m(n)$,

$$\begin{aligned}\sum_{d\mid M} f_m(d)\frac{\phi(M/d)}{M} &= \sum_{d\mid M}\prod_{p_j\mid d}\left(1-\frac{1}{p_j}\right)^{\alpha_j}\prod_{p_j\mid M/d}\left(-\frac{1}{p_j}\right)^{\alpha_j}\frac{\phi(M/d)}{M}\\ &= \sum_{d\mid M}\prod_{p_j\mid d}\frac{1}{p_j}\left(1-\frac{1}{p_j}\right)^{\alpha_j}\prod_{p_j\mid M/d}\left(-\frac{1}{p_j}\right)^{\alpha_j}\frac{\phi(M/d)}{M}\\ &= \sum_{d\mid M}\prod_{p_j\mid d}\frac{1}{p_j}\left(1-\frac{1}{p_j}\right)^{\alpha_j}\prod_{p_j\mid M/d}\left(-\frac{1}{p_j}\right)^{\alpha_j}\left(1-\frac{1}{p_j}\right).\end{aligned}$$

We can now exploit the fact that since $d\mid M$, then $d=\sum_{j\in S}p_j$ for $S\subset[r]$. Denote $S^c=[r]\setminus S$, the complement of S. Hence, summing over a set,

$$\begin{aligned}\frac{1}{x}\sum_{n\le x} f_m(n) &= \sum_{S\subset[r]}\prod_{j\in S}\frac{1}{p_j}\left(1-\frac{1}{p_j}\right)^{\alpha_j}\prod_{j\in S^c}\left(-\frac{1}{p_j}\right)^{\alpha_j}\left(1-\frac{1}{p_j}\right)\\ &= \prod_{j=1}^{r}\left(\frac{1}{p_j}\left(1-\frac{1}{p_j}\right)^{\alpha_j}+\left(-\frac{1}{p_j}\right)^{\alpha_j}\left(1-\frac{1}{p_j}\right)\right),\end{aligned}$$

as needed. ■

Hence, from equation 8.1,

$$\frac{1}{x}\sum_{n\le x} f_m(n) = \mathbb{E}\left[\prod_{j=1}^{r}\left(X(p_j)-\frac{1}{p_j}\right)^{a_j}\right] + \mathcal{O}\left(\frac{2^{2r}}{x}\right).$$

With this, we can proof Theorem 8.1.

Proof of Theorem 8.1. We have,

$$\begin{aligned}\frac{1}{x}\sum_{n\le x}\left(\sum_{p\le y} f_p(n)\right)^k &= \sum_{p_1,p_2,\cdots,p_k\le y}\frac{1}{x}\sum_{n\le x} f_{p_1\cdots p_k}(n)\\ &= \sum_{p_1,p_2,\cdots,p_k\le y}\left(\mathbb{E}\left[\prod_{j=1}^{r}\left(X(p_j)-\frac{1}{p_j}\right)^{a_j}\right]+\mathcal{O}\left(\frac{2^{2k}}{x}\right)\right)\\ &= \mathbb{E}\left[\sum_{p\le y}\left(X_p-\frac{1}{p}\right)\right]^k + \mathcal{O}\left(\frac{y^k}{x(\log y)^k}\right),\end{aligned}$$

since $a_i = 1$ for all p_i where $1 \le i \le k$. This completes the proof. ■

Hence, we have proved Erdös-Kac Theorem!

9. Further Exploration

In Section 10, we will be discussing some further analogues of the Erdös-Kac theorem to see how beautiful it is. To do so, we would like to share some definitions and notations.

Definition 9.1. Denote $n \ge 0$ to be a natural number. Then $\pi(x)$ counts the number of primes $p \le x$.

Actually, $\pi(x)$ comes up a lot in analytical number theory and most prominently in the *Prime Number Theorem*, which states that

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

We have that Definition 9.1 establishes the prime analogue of the Erdös-Kac Theorem. The following definition is used to generalize the Erdös-Kac Theorem. Note that $\pi(n)$ appeared quite subtle in our paper already–the error term in Theorem 8.1 can be written as

$$\mathcal{O}\left(\frac{\pi(y)^k}{x}\right).$$

Definition 9.2 (Strongly Additive)**.** Denote m and n to be positive integers such that $\gcd(m,n) = 1$. Denote p to be a prime and $e > 2$ to be an integer. We say f is *strongly additive* if

(1) $f(mn) = f(m) + f(n)$, and
(2) $f(p^e) = f(p)$

As an example, $\omega(n)$, is a strongly additive function.

We will end of this section, quite surprisingly, with some theory on elliptic curves that will give us one of the most fascinating variants of Erdös-Kac Theorem.

Definition 9.3 (Elliptic Curves)**.** For constants $A, B \in \mathbb{R}$, an *elliptic curve* is defined as

$$E := y^2 = x^3 + Ax + B.$$

The discriminant of of the elliptic curve is defined as $\Delta(E) = -16(4A^3 + 27B^2)$. Will will denote an elliptic curve E to be over $\mathbb{Q}$ as $E/\mathbb{Q}$. Now, denote a finite field $\mathbb{F}_p$ and $E(\mathbb{F}_p)$ as the set of rational points defined over $\mathbb{F}_p$. We denote $\#E(\mathbb{F}_p)$ to be the number of such points in $E(\mathbb{F}_p)$.

Definition 9.4. Denote p as a prime. Then p is said to be *of good reduction* of E if

$$p \nmid \Delta(E).$$

With all these definitions, we will be able to see three different variants of the Erdös-Kac Theorem.

10. Variants of Erdös-Kac Theorem

In 1955, Halberstam [Hal55] proved that for a prime p,

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \# \left\{ p \leq x : \frac{\omega(p-1) - \log\log p}{\sqrt{\log\log p}} \leq \gamma \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{\frac{-t^2}{2}}\, dt.$$

This is often called the "prime analogue" of Erdös-Kac Theorem.

We can also define a generalized version of the Erdös-Kac Theorem that Halberstam [Hal55] also showed. If f is a strongly additive function, then

$$\lim_{x\to\infty} \frac{1}{x} \left\{ n \leq x : a \leq \frac{f(n) - A(n)}{\sqrt{B(n)}} \leq b \right\} = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{\frac{-t^2}{2}}\, dt,$$

where

$$A(n) = \sum_{p\leq n} \frac{f(p)}{p} \quad \text{and} \quad B(n) = \sqrt{\sum_{p\leq n} \frac{f(p)^2}{p}}.$$

Notice this generalization applies to Erdös-Kac Theorem: $f(n) \equiv \omega(n)$, and by Theorem 3.2 ,setting $f(p) = 1$ for all p, we get the mean and variance of from the Erdös-Kac Theorem.

Finally, we will finish the paper by stating the prime analogue of the Erdös-Kac theorem for elliptic curves [Liu06].

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \# \left\{ p \leq x : p \textit{ is of good reduction and } \frac{\omega(\#(E(\mathbb{F}_p) - \log\log p}{\sqrt{\log\log p}} \leq \gamma \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma} e^{-\frac{t^2}{2}}\, dt.$$

Acknowledgements. The author would like to thank Dr.Simon-Rubinstein Salzedo for the suggestion of the topic and Ryan Catullo for the incredible feedback making this paper possible.

References

[Apo76] Tom M. Apostol. *Introduction to Analytic Number Theory*. Springer-Verlag New York Inc, 1976.

[Cha21] Deeparnab Chakrabarty. Probability: Variance. 2021.

[Del62] Hubert Delange. Sur le nombre des diviseurs premiers de n. *Acta Arithmetica*, 7:191–215, 1962.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. 2019.

[EK40] Paul Erdös and Mark Kac. The gaussian law of errors in the theory of additive number theoretic functions. *American Journal of Mathematics*, 62:738, 1940.

[GS06] Andrew Granville and K. Soundararajan. Sieving and the erdős-kac theorem, 2006.

[Hal55] H. Halberstam. On the Distribution of Additive Number-Theoretic Functions. *Journal of the London Mathematical Society*, s1-30(1):43–53, 01 1955.

[Hal56] Heini Halberstam. On the distribution of additive number-theoretic functions (ii). *Journal of The London Mathematical Society-second Series*, pages 43–53, 1956.

[HR17] G.H Hardy and S. Ramanujan. The normal number of prime factors of a number n. *Quart. J. Pure. Appl. Math., Vol. 48*, 1917.

[Les15] Steve Lester. Normal order of $\omega(n)$ and the erdös-kac theorem". 2015.

[Liu06] Yu-Ru Liu. Prime analogues of the erdős–kac theorem for elliptic curves. *Journal of Number Theory*, 119(2):155–170, 2006.

[PB95] 3rd ed. Patrick Billingsley. *Probability and Measure*. A Wiley-Interscience, 1995.

[Tsu] Alex Tsun. Method of moments estimation.

[Tur34] Paul Turán. On a theorem of hardy and ramanujan. *Journal of The London Mathematical Society-second Series*, pages 274–276, 1934.

[Vil05] Mark B. Villarino. Mertens' proof of mertens' theorem, 2005.

Email address: harshiln14@gmail.com