Relations Between The Erdos-Kác Theorem and Ergodic theory (Self Contained Version)

Syed Ismail

June 2026

Abstract

One of the most popular results of probabilistic number theory, the Erdős–Kac theorem, states that the function $\omega(n)$, which denotes the number of distinct prime divisors of n, satisfies that, as $n \to \infty$, the normalized arithmetic function

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

converges in distribution to the standard normal distribution. Ergodic theory is the study of dynamical systems. Although they appear as very separate concepts, in reality they share the same underlying philosophy: deterministic systems can exhibit statistical behavior. Through this expository, we first explore ergodic theory, including theorems like Birkhoff's ergodic theorem, followed by the heuristic derivation of the Erdős–Kac theorem. After which, we explore topics where both of these themes seem to have a correlation, such as additive functions, Fourier analysis, Möbius randomness, logarithmic averages, and more. Furthermore, we examine how techniques from ergodic theory provide new perspectives on classical problems in number theory, highlighting the deep connections between deterministic dynamics and probabilistic models. The goal of this exposition is to demonstrate how these 2 different fields of mathematics are unified through the study of statistical behavior arising from deterministic structures.

1 Introduction

People underestimate the uncertainty and relations of statistical behavior with everyday randomness around us. Many objects in mathematics are deterministic, yet when we upscale these objects,they exhibit behavior that resembles randomness. Take for example the digits of irrational numbers, the distribution of prime numbers, and the trajectories of dynamical systems which all reveal statistical patterns despite having no innate randomness.

Each integer has a unique prime factorization, hence we would assume that there is a irregular behavior regarding the collection of prime numbers for every integer and hence no interesting patters. However this would prove to be untrue, when Paul erdös and mark kac found a relation by discovering that the number of distinct prime divisors of a typical integer follows a remarkably precise statistical law which states the distribution for the number of prime divisors over the set of integers model as the normal distribution as $n \to \infty$.

A similar relation occurs with ergodic theory which studies how deterministic dynamical systems change over time and how their long-term averages can be statistically described. One of the most prominent examples being Birkhoff's ergodic theorem which demonstrates that the average behavior observed along a single trajectory is related to the average behavior of the entire space-providing a framework for understanding how these processes which are deterministic in nature derive further statistical laws.

Although ergodic theory and erdos-kac theorem come from very distinct math fields, they share the same underlying principle : deterministic objects can exhibit statistical regularity when viewed from an appropriate perspective. Erdos-kac theorem studies this underlying principle through arithmetic functions while ergodic theory focus on this principle through the evolution of points under dynamical systems.

This exposition explores the relationship between these two perspectives. We first develop the necessary foundations from real analysis, measure theory, probability theory, and analytic number theory after which we describe a heuristic derivation of the Erdös–Kac theorem by modeling prime divisibility through Bernoulli random variables and connecting the resulting distribution to the central limit theorem. After which, we examine deeper connections between arithmetic functions and dynamical systems, including additive functions, Fourier analytic methods, Möbius randomness, logarithmic averages, and modern problems in arithmetic dynamics. The aim of this expository is to describe the relation between these two distinct fields of mathematics in order to demonstrate how these seemingly different fields can be unified unified by a shared study of statistical behavior.

2 Ergodic Theory

2.1 Preliminaries for Ergodic Theory

Before jumping into ergodic theory, it is of best interest to discuss some basic preliminaries. The main preliminaries are real analysis and measure theory which are used in subsequent chapters to define probability theory, measure theoretic definitions for ergodic theory, and other such applications.

2.1.1 Real Analysis

We shall first discuss some ideas on the topology of real numbers, followed by sequences and series, and finally the Riemann integral which is required when defining its limitations and Lebesgue's measure theoretic integral. [1]

Assume the set A such that $A \subseteq \mathbb{R}$. Set A is defined bounded above if there exists a number $b \in \mathbb{R}$ such that $a \leq b, \forall a \in A$. By contrast, set A is defined bounded below if there exists a number $b \in \mathbb{R}$ such that $a \geq b, \forall a \in A$. The supremum of A(notation : $sup(A)$)is defined as the least upper bound of A if it is an upper bound for A such that for any upper bound b, $b \in \mathbb{R}$, the $sup(A) \leq b$. likewise, we define the infimum of A (notation : $inf(A)$) as the greatest lower bound of A such that for a lower bound $b \in \mathbb{R}$, $inf(A) \leq b$.

Given $a \in \mathbb{R} \wedge \epsilon > 0$,The epsilon-neighborhood for the point a (notation : $V_\epsilon(a)$) is the open interval $(a - \epsilon, a + \epsilon)$, centered at a with radius ϵ. $V_\epsilon =$

[1] During the preliminary, we do not go into the proofs of most real analysis theorems and definitions. If you are interested in reading further then refer to [Abb15]

$\{x \in \mathbb{R} || x - a| < \epsilon\}$. Using the epsilon neighborhood, we now define open and closed sets : The set A, A $\subseteq \mathbb{R}$ is defined as an open set if $\forall x \in A$, there exists a $E_\epsilon(x) \in A$. Both the union of an arbitrary collection of open sets and the intersection of a finite collection of open sets is open.

We define the point x as the limit point of the set A if for every $V_\epsilon(x)$ of x intersects the set A at some point other than x and $\lim a_n \to x$ for some sequence $a_n \in A$ such that $a_n \neq x, \forall n \in \mathbb{N}$. If x doesn't meet the requirement to be considered a limit point, then the point x is defined as the isolated point of the set A. We define set A as a closed set if it contains its limit points. The closure of the set A is defined as the union between set A and the set of all its limit points (L) (notation - $\overline{A} = A \cup L$). Note that $\overline{A}$ is a closed set and the smallest closed set containing A.

Take set A and its compliment A^c, set A is a closed set if and only if (A^c) is an open set and vice-versa. Furthermore, the union of a finite collection of closed sets is closed and the intersection of an arbitrary collection of closed sets remain closed (proved via de-morgans's laws).

We shall now shift our attention to defining open covers and finite subcovers - Let us assume the set $J \subseteq \mathbb{R}$, An open cover for J is the collection of open sets $(\{O_i | i \in I\})$whose union contains the set J $[J \subseteq \bigcup_{i \in I} O_i]$. A finite subcover is a finite collection of $\{O_i | i \in F\}$, where $F \subseteq I$ is finite, such that $J \subseteq \bigcup_{i \in F} O_i$. Using the idea of open covers and finite subcovers, we now define the Heine-Borel Theorem :

Theorem (Heine-Borel Theorem) - *Let $A \subseteq R$. All of the following statements are equivalent as they imply each other :*

- *A is compact*
- *A is closed and bounded*
- *Every open cover for A has a finite subcover*

This concludes the prerequisites required for the topology of $\mathbb{R}$, we shall now move on to discussing sequences and series.

Let us assume the sequence b_n ,n $\in \mathbb{N}$, be convergent if there exists some b $\in \mathbb{R}$ such that $\lim_{n->\infty} b_n = b$ for the $\epsilon-$neighborhood $V_\epsilon(b)$, $\epsilon > 0$, where there exists a point N $\in \mathbb{N}$ in the sequence after which all the terms of the sequence b_n are within V_ϵ(b). Else, the sequence is divergent. b_n can be defined either increasing if $b_n \leq b_{n+1}$ $\forall n \in \mathbb{N}$ or decreasing if $b_n \geq b_{n+1}$ $\forall n \in \mathbb{N}$. If the sequence is either increasing or decreasing, we state the sequence is **monotone**. In accordance with the **Monotone Convergence Theorem**, if a sequence is bounded and monotone, then it must converge.

The subsequence of b_n is of the form b_{n_k} such that $n_1 < n_2 < ....$ Note that the order of terms in a subsequence is the same as that of the main sequence, and there is no repetition of terms. Furthermore, if b_n converges to some L $\in \mathbb{R}$, then its subsequence must also converge at L. We shall now define the

Bolzano-Weierstrass Theorem which is required to grasp [reason for defining the Bolzano-Weierstrass theorem] :

Bolzano-Weierstrass Theorem -*Every bounded sequence must contain a convergent subsequence.*

We shall also define the algebraic and order limit theorems of sequences to act as a supplemental aid to the subsequent theorems and proofs :

Theorem (Algebraic Limit Theorem) - *Assume we have sequences a_n and b_n such that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$ such that $a, b \in \mathbb{R}$, then we state the following :*

- $\lim ca_n -> c \lim a_n = ca, \forall c \in \mathbb{R}$
- $\lim(a_n + b_n) = \lim a_n + \lim b_n = a + b$
- $lim(a_n \cdot b_n) = a \cdot b$
- $\lim(\frac{a_n}{b_n}) = \frac{a}{b}$

Theorem (Order Limit Theorem) - *Using the same assumptions to that of the algebraic limit theorem, we state the following :*

- If $a_n \geq 0, \forall n \in \mathbb{N}$, then $a \geq 0$
- If $a_n \leq b_n$, $\forall n \in \mathbb{N}$, then $a \leq b$
- If there exists c$\in \mathbb{R}$ for which $c \leq b_n$, $\forall n \in \mathbb{N}$, then c $\leq b$. If $a_n \leq c, \forall n \in \mathbb{N}$, then $a \leq c$.

Moving away from sequences, we shall now discuss its counterpart, the Infinite Series. For the infinite series $\sum_{k=1}^{\infty} a_k$, its convergence is always equal to its expression in partial sums. Due to this property, we can render most of the axioms of order and algebraic limit theorem of sequences into series. However, it does not translate the product of 2 convergent sequences. Instead, to find the product for any two given convergent series, we use cauchy's product which states :

Theorem (Cauchy Product) - *Assume that $\sum_{i=1}^{\infty} a_i$ converges absolutely to A and $\sum_{j=1}^{\infty} b_j$ converges absolutely to B. Then :*

$$\left(\sum_{i=1}^{\infty} a_i\right) \cdot \left(\sum_{j=1}^{\infty} b_j\right) = \sum_{k=2}^{\infty} d_k = AB$$

where $d_k = a_1 b_{k-1} + a_2 b_{k-2} + ... + a_{k-1} b_1$

Finally, we define pointwise and uniform convergence. Let (f_n) be a sequence of functions defined on a set $A \subseteq \mathbb{R}$, where each $f_n : A \to \mathbb{R}$. The sequence (f_n) converges pointwise on A to a function $f : A \to \mathbb{R}$ if, for every $x \in A$, the sequence of real numbers $(f_n(x))$ converges to $f(x)$. By contrast, (f_n)

converges uniformly to f on A if, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for every $n \geq N \wedge x \in A$.

Lastly, we shall discuss the derivation of the Riemann integral to end this section of preliminaries.

Let us start of by defining partitions which are essential to understanding how Bernhard Riemann defined the integral. A partition of [a,b] is defined as a finite set of points from [a,b] (including a and b). The notation for a partition is $P = \{x_0, x_1, ..., x_n\}$ such that $a = x_0 < x_1 < .. < x_n = b$.

For any subinterval $[x_{k-1}, x_k]$ of P, let :

- $m_k = \inf\{f(x) \mid x \in [x_{k-1}, x_k]\}$
- $M_k = \sup\{f(x) \mid x \in [x_{k-1}, x_k]\}$

Then, the lower darboux sum of f with respect to the partition P is given by :

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1})$$

Likewise, the upper darboux sum of f with respect to the partition P is given by :

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1})$$

In high school, we were taught the approach of taking the limit of these sums, hence as the partitions get smaller, the upper sum's overestimate and lower sum's underestimate decreases. However we shall use an alternate idea by using the axiom of completeness to define the riemann integral. Let V be the collection of all possible partitions of the interval [a,b]. Then the upper integral of f is defined as

$$U(f) = \inf\{U(f, P) \mid P \in V\}$$

And let the lower integral of f be defined as :

$$L(f) = \sup\{L(f, P) \mid P \in V\}$$

Riemann Integrability - a bounded function f on the interval [a,b] is defined Riemann-integrable if $U(f) = L(f)$

$$\int_a^b f = U(f) = L(f)$$

The Darboux characterization above is equivalent to Riemann's original definition, which is stated in terms of tagged partitions and Riemann sums. Given a partition P, a tagged partition (P,c_k)is obtained by choosing a point $c_k \in [x_{k-1}, x_k]$ in each subinterval. The corresponding Riemann sum is

$$R(f, P) = \sum_{k=1}^{n} f(c_k)\Delta x_k$$

where $\Delta x_k = x_k - x_{k-1}$

Riemann's original definition of the integral states a bounded function $f : [a, b] \to \mathbb{R}$ is integrable on [a,b] with $\int_a^b f(x)dx = A, \forall \epsilon > 0, \ \exists \delta > 0$ such that for any tagged partition (P,$\{c_k\}$) satisfying $\Delta x_k < \delta, \forall k$, it follows that:

$$|R(f, P) - A| < \epsilon$$

Finally we shall define Lebesgue's criterion for the Riemann integral-Let f be a bounded function defined on the interval [a,b]. Then, f is Riemann-integrable if and only if the set of points where f is not continuous has measure zero.

Measure zero is defined for a set X, $X \subseteq \mathbb{R}$ where for all $\epsilon > 0$, there exists a set of open intervals $\{I_1, , I_k\}$, $1 \leq k \leq \infty$, such that $X \subseteq \bigcup_{k=1}^{\infty} I_k \wedge \sum_{k=1}^{\infty} |I_k| \leq \epsilon$. We will go more into detail about measure theory in the subsequent section.

2.1.2 Measure Theory

The idea of measure theory arises due to some consequences of the Riemann integral. The Riemann integral gives us a powerful framework for continuous and sufficiently well defined functions. However, this framework doesn't hold for functions which are unbounded, have many discontinuities , or the limits of sequences of functions. In order to fix these issues, there comes a general idea to interchange the Riemann integral and limit, however this approach requires strict conditions on the functions involved, making many important results difficult to establish within the Riemann framework. Hence, this raised the requirement to create another framework more powerful than the Riemann Integral - Lebesgue's integral. However, before we go into what a Lebesgue's integral is we must define measurable spaces.

The length of the open interval I (notation : $\ell(I)$) is defined as

$$\ell(I) = \begin{cases} b - a & \text{if } I = [a, b] \text{ for some } a, b \in \mathbb{R} \ \wedge a < b \\ o & \text{if } I = \emptyset \\ \infty & \text{if } I = (-\infty, a) \vee (a, \infty) \text{ for some } a \in \mathbb{R} \\ \infty & \text{if } I = (-\infty, \infty) \end{cases}$$

Suppose the set A $\subseteq \mathbb{R}$, the outer measure of set A (notation : $|A|$ is defined as the greatest lower bound of the sum of the open intervals whose union contains A. In formal notation - $\{|A| = \{\sum_{k=1}^{\infty} \ell(I_k) \mid I$ are open intervals such that $A \subseteq \bigcup_{k=1}^{\infty} I_k\}$. For any countable set, the outer measure of the set is always equal to 0. Furthermore, outer measure shall always preserve order. For example, assume set A and B such that $A \subseteq b \subseteq \mathbb{R}$, then $|A| < |B|$. Also, outer measure is translation [2] invariant, this means $|A| = |k + A|$ for any $k \in \mathbb{R}$. For a closed interval [a,b], its outer measure $|[a, b]| = b - a$.[3] In addition, for two disjoint sets A and B, $|A \cup B| \neq |A| + |B|$. Hence outer measure is not additive.

Theorem : *There does not exist a function μ which has all the following properties:*

- *Axiom 1 - μ is a function from the set of subsets of $\mathbb{R}$ to $[0, \infty]$.*
- *Axiom 2 - $\mu(I) = \ell(I)$ for every open interval I of $\mathbb{R}$.*
- *Axiom 3 - $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$ for every disjoint sequence $A_1, A_2, .. \subseteq \mathbb{R}$. (countable additivity).*
- *Axiom 4 - $\mu(t + A) = \mu(A)$ for every $A \subseteq \mathbb{R}$ and every $t \in \mathbb{R}$.*

We must keep axiom 2 as we need the size of an interval to equate to its length. Furthermore, We must keep countable additivity from axiom 3 for subsequent proofs which involve limits. We require translation invariance to ensure that measure serves as a generalized notion of length. Since we require these 3 axioms while also ensuring μ does not possess all the properties as mentioned in the theorem above, we must restate axiom 1 to slacken its constraints. Hence, we define σ-algebra.

Theorem (σ-algebra) - *suppose X is a set and S is a set of subsets of X. Then S is called a σ-algebra on X if the following conditions are satisfied :*

- $\emptyset \in S$
- if $E \in S$, then $X \setminus E \in S$
- if $E_1, E_2, ...$ is a sequence of elements of S, then $\bigcup_{k=1}^{\infty} E_k \in S$

Now that we have defined measure and σ-algebra. We should define measurable space and measurable set. A measurable space is an ordered pair (X,S), where X is a set and S is a σ-algebra on X. Any element of S is called measurable set for S (S-measurable set). Another thing to note : The smallest σ-algebra on $\mathbb{R}$ containing all open subsets of R is called the collection of Borel subsets of $\mathbb{R}$. An element of this σ-algebra is called the Borel Set.

We shall now shift our attention to measurable functions. Suppose (X, S) is a measurable space. A function $f : X \to \mathbb{R}$ is called S-measurable if $f^{-1}(B) \in S$

[2] A translation is where every element of the set moves by k units to the right (if $k > 0$) or left (if $k < 0$). let $k \in \mathbb{R}$ and $A \subseteq \mathbb{R}$, then the translation $k + A = \{k + a \mid a \in A\}$

[3] This is proved by the Heine-Borel Theorem. (refer to chapter 2 of[Axl20])

[4]for every Borel set $B \subseteq \mathbb{R}$. We should also define a characteristic function (χ_E) - Suppose E is a subset of a set X. The characteristic function of E is the function $\chi_E : X \to \mathbb{R}$ such that:

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \neq E \end{cases}$$

A measuring function f is considered a Borel measurable function if $f^{-1}(B)$ is a Borel set for every Borel set $B \subseteq \mathbb{R}$.

Suppose (X, S) is a measurable space and $f_1, f_2, ...$ is a sequence of S-measurable functions from X to $\mathbb{R}$. Then $\lim_{k\to\infty} f_k(x)$ exists for each $x \in X$ is an S-measurable function aswell.

We defined the infimum and supremum of a function in the previous prerequisite section. But it is only natural for us to redefine it now that we have a new type of function : S-measurable function.

Suppose (X, S) is a measurable space and $f_1, f_2, ...$ is a sequence of S-measurable functions from X to $[-\infty, \infty]$. let $a, b : X \to [-\infty, \infty]$ where

$$a(x) = \inf\{f_k(x) \mid k \in \mathbb{Z}^+\} \wedge b(x) = \sup\{f_k(x) \mid k \in \mathbb{Z}^+\}$$

Then a and b are S-measurable functions.

Moving onwards we shall now define measure : Suppose X is a set and S is a σ-algebra on X. A measure on (X, S) is a function $\mu : S \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

for every disjoint sequence $E_1, E_2, ...$ of sets in S.

Furthermore, we define a measure space is an ordered triple (X, S, μ) where X is a set, S is a σ-algebra on X, and μ is a measure on (X,S). Note that measure preserves order and $\mu(E \setminus D) = \mu(E) - \mu(D),\ \mu(D) < \infty$. Furthermore, the measure space allows countable additivity.

Lebesgue's measure - The measure on ($\mathbb{R}$,B), where B is the σ-algebra of Borel subsets of $\mathbb{R}$, that assigns to each Borel set its outer measure. [5] Furthermore, A set $A \subseteq \mathbb{R}$ is Lebesgue measurable if there exists a Borel set $B \subseteq A$ such that $|A \setminus B| = 0$.

[4]If $f : X \to Y$ is a function and $G \subseteq Y$, then the set $f^{-1}(G)$ is defined : $f^{-1}(G) = \{x \in X \mid f(x) \in G\}$.

[5]Note that Lebesgue's sets does not apply to arbitrary sets but only to Borel sets.

The set $\mathcal{L}$ of Lebesgue measurable sets of $\mathbb{R}$ is a σ-algebra on $\mathbb{R}$. Also, the outer measure is a measure on $(\mathbb{R}, \mathcal{L})$

We can redefine Lebesgue measure as the measure on $(\mathbb{R}, \mathcal{L})$, where $\mathcal{L}$ is the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$, which assigns each Lebesgue measurable set its outer measure.

We shall finally now look at the Lebesgue Integral. We shall start of by defining the S-partition. Suppose S is a σ-algebra on a set X. A S-partition of X is a finite collection $A_1, ..., A_m$ of disjoint sets in S such that $\bigcup_{k=1}^{m} A_k = X$. Using this new defined partitions we now define the lower Lebesgue sum which states for the measure space $(X, S, \mu), f : X \to [0, \infty]$ is an S-measurable function, with P being a S-partition of X. The lower Lebesgue sum $\mathcal{L}(f, P)$ is defined

$$\mathcal{L}(f, P) = \sum_{j=1}^{m} \mu(A_j) inf(f)$$

Suppose we have a measurable space (X, S, μ) and $f : X \to [0, \infty]$ is a S-measurable function. Then the integral with respect to μ is defined as

$$\int f \, d\mu = \text{int}\{\mathcal{L}(f, P) \mid \text{P is a S-partition of X}\}$$

Note that the integral with respect to μ for χ_E equates to $\mu(E)$.

Theorem (Monotone Convergence Theorem) - Suppose (X,S,μ) is a measurable space where $0 \leq f_1 \leq ...$ is an increasing sequence of $S-$measurable functions. Furthermore we define the function $f : X \to [0, \infty]$ sucht that $f(x) = \lim_{k\to\infty} f_k(x)$. Then :

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu$$

If we have real valued function f^+ and f^- such that $\int f^+ \, d\mu$ [6] and $\int f^- \, d\mu$ [7] are finite. The the integral of f with respect to μ is defined :

$$\int f \, d\mu = \int f^+ \, d\mu + \int f^- \, d\mu$$

Lastly we shall briefly discuss the relation between the Riemann integral and Lebesgue's Integral. Suppose $a < b$ and $f : [a, b] \to \mathbb{R}$ is a bounded function.

[6] $f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}$

[7] $f^-(x) = \begin{cases} 0) & \text{if } f(x) \geq 0 \\ -f(x) & \text{if } f(x) < 0 \end{cases}$

Then f is Riemann integrable if at any discontinuity in the closed interval [a,b] the function f equates to 0. IF f is Riemann Integrable and λ denotes Lebesgue measure on $\mathbb{R}$. Then f is Lebesgue measurable and

$$\int_a^b f = \int_{[a,b]} f \, d\lambda$$

This marks the end of the prerequisites required for the ensuing section of Ergodic theory. We shall now discuss dynamical systems, how ergodic theory relates to dynamical systems, and discuss some notable theorems.

2.2 Ergodic Theory

A dynamical system in mathematics consists of functions defined $f : X \to X$ such that its domain and codomain are the same. Ergodic theory is a subfield, studying the statistical behavior of these dynamical systems. We shall start defining some key ideas of ergodic theory.

We shall start with defining a A Measure Preserving Dynamical System - A measure preserving system is the tuple $(X, \mathcal{A}, \mu, T)$ where $(X, \mathcal{A}, \mu)$ and $T : X \to X$ is a measure preserving transformation.

We shall now define the orbit of a point. Suppose $(X, \mathcal{A}, \mu, T)$ is a measure-preserving dynamical system and let $x \in X$. The orbit of x under the transformation T is defined as the set

$$\mathcal{O}(x) = \{x, T(x), T^2(x), T^3(x), \ldots\}.$$

The orbit represents the collection of all future positions of the point under repeated applications of the transformation T.

Suppose $A \in \mathcal{A}$. We define the inverse image of A under T by

$$T^{-1}(A) = \{x \in X : T(x) \in A\}.$$

A transformation T is said to preserve measure if

$$\mu(T^{-1}(A)) = \mu(A)$$

for every measurable set $A \in \mathcal{A}$.

We now define an invariant function. Let $(X, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system. A measurable function $f : X \to \mathbb{R}$ is said to be T-invariant if

$$f(T(x)) = f(x)$$

for μ-almost every $x \in X$.

Recall from the previous section that a property is said to hold almost everywhere if the set of points for which the property fails has measure zero. Hence we write

$$\mu\left(\{x \in X : P(x) \text{ fails}\}\right) = 0.$$

Throughout ergodic theory, almost every convergence theorem is interpreted in this sense.

Suppose $f : X \to \mathbb{R}$ is an integrable function. The time average of f along the orbit of x is defined by

$$A_n(f, x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)).$$

This quantity measures the average value of f observed along a single trajectory.

By contrast, the space average of f is defined as

$$\int_X f \, d\mu.$$

Unlike the time average, the space average averages over the entire measure space rather than a single orbit.

One of the principal questions of ergodic theory is determining under what conditions the time average

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

coincides with the space average

$$\int_X f \, d\mu.$$

This question is answered by one of the most fundamental results in ergodic theory, namely Birkhoff's Ergodic Theorem.

We shall now define the concept fo ergodicity - A measure-preserving transformation T on a measure space $(X, \mathcal{A}, \mu)$ is said to be ergodic if whenever $A \in \mathcal{A}$ is a T-invariant set. Then either, $\mu(A) = 0$ or $\mu(A^c) = 0$. If T is ergodic, then $(X, \mathcal{A}, \mu, T)$ is an ergodic system. We say A set A is T-invariant if whenever $x \in A$, then $T(x) \in A$ aswell.

Theorem [Birkhoff's Ergodic Theorem] *-Let $(X, \mathcal{A}, \mu, T)$ be an ergodic-measure preserving system . Then for any $L^1(\mu)$wehave*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^n(x)) = \int_X f d\mu$$

as $n \to \infty$

In simpler terms, It means for an ergodic measure-preserving dynamical system, the long-term average of a function along a trajectory is equal to its average over the entire space.

Example of Birkhoff's Ergodic Theorem

- Assume an interval X = [0,1) and let $T(x) = 2x \pmod 1$ with usual measure μ.
- We shall take the observable $f(x) = x$. Then according to Birkhoff's Ergodic Theory :

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^n(x)) = \int_X f d\mu$$

$$\frac{x + T(x) + T^2(x) + ... + T^{N-1}(x)}{N} = \int_0^1 dx$$

$$\frac{x + T(x) + T^2(x) + ... + T^{N-1}(x)}{N} = 0.5$$

We come to an observation that the time average equates to the space average.

Another important theorem we shall state is the Poincaré recurrence theorem.

Theorem (Poincaré Recurrence Theorem) - *Let $(X, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system with $\mu(X) < \infty$. Then for every measurable set $A \in \mathcal{A}$, almost every point $x \in A$ returns to A infinitely many times.*

Birkhoff's Ergodic Theorem and the Poincaré Recurrence Theorem demonstrate that deterministic dynamical systems can exhibit predictable statistical behaviour over time. This underlying principle is one of the central ideas of ergodic theory and serves as the foundation for many modern connections with probability theory and analytic number theory, including the Erdős–Kac theorem discussed in the following section.

3 The Erdos-Kac Theorem

3.1 Preliminaries for the Probabilistic methods in number theory

3.1.1 Probability Theory

Before defining the erdös-kac theorem, we should define some basic probability definitions. To begin with we shall first define the probability measure. We define the probability measure by the ordered triple (Ω,$\mathcal{F}$, P) where Ω represents the collection of all possible outcomes, $\mathcal{F}$ is the set of events and P is a function defined $P : \mathcal{F} \to [0, 1]$. $\mathcal{F}$ is a σ-algebra of Ω.

Next we define random variables. A random variable X is defined as a measurable function that maps from a probability space to the set of all real numbers. So, if for every Borel set B $\in \mathbb{R}$ we have

$$X^{-1}(B) = \{w \mid X(w) \in B\} \in \mathcal{F}$$

Bernoulli Random Variable - It is a discrete random variable which can take up two values- 1 and 0. We assign the probability p to 1 when we get the favorable outcome, and we assign 1-p to 0 if we do not get the favorable outcome we are looking for.

$$B_p = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

For a bernoulli random variable with parameter p, we say the expected value of the bernoulli variable is p and the variance is p(1-p).

For a random variable X on the probability space (Ω,$\mathcal{F}$, P). We define its expected value as :

$$E(X) = \int_\Omega X \, dP$$

it is the weighted average of all possible outcomes. After defining expected value, we can also define variance. The variance of a random variable measures how far its values tend to spread out from their expected value. Var(X) $= E(X^2) - [E(X)]^2$.

Now we move onto the normal distribution. The normal (or Gaussian) distribution is one of the most important probability distributions in statistics and probability theory. A random variable X is said to follow a normal distribution with mean μ and variance σ^2, denoted by

$$X \sim N(\mu, \sigma^2),$$

if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty.$$

The parameter μ determines the centre of the distribution, while σ^2 determines its spread.

The normal distribution is a continuous probability distribution that is symmetric about its mean and has the characteristic bell-shaped curve. Furthermore, the mean, median, and mode all coincide at μ, and the total area under the density curve is equal to 1.

If

$$X \sim N(\mu, \sigma^2),$$

then the expectation and variance of X are given by

$$E(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

Thus, the mean specifies the location of the distribution, while the variance measures how dispersed the values are about the mean.

A particularly important special case is the standard normal distribution, defined by

$$Z \sim N(0, 1).$$

Any normally distributed random variable can be transformed into a standard normal random variable by the standardization

$$Z = \frac{X - \mu}{\sigma}.$$

This transformation allows probabilities associated with any normal distribution to be computed using the standard normal distribution.

Finally, one of the most essential theorem's we define throughout all prerequisite sections : The Central Limit Theorem

Theorem (Central Limit Theorem) = *Let $Y_1, Y_2, ..., Y_n$ be independent and identically distributed variables such that $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2 < \infty$. Then we define*

$$U_n = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$$

The distributive function U_n converges to the standard normal distribution as $n \to \infty$.

We now move on to the final prerequisite section throughout this expository - Number Theory. [8]

[8] The preliminary for probability has been kept to the absolute bare minimum. For a more rigorous understanding to probability theory. I recommend going through [Dem23]

3.1.2 Number Theory

We shall start of defining the fundamental theorem of arithmetic.

Theorem (Fundamental Theorem of arithmetic) - *Every integer greater than 1 can be written in the form* $p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$ *where* $n_i > 0$ *and* p_i *are distinct primes. Each integer has a unique factorization ignoring the order of the actors.*

An arithmetic function is defined as any function whose domain is the set of natural numbers $f : \mathbb{N} \to \mathbb{R}$. Following this definition, we define the following arithmetic functions :

- $\tau(n)$ = The number of positive divisors that n contains.

$$\tau(n) = \sum_{d|n} 1$$

[9]

- $\sigma(n)$ = The sum of all the positive divisors of n.

$$\sigma(n) = \sum_{d|n} d$$

- $\pi(n)$= The number of primes $\leq n$
- $\omega(n)$ = the number of prime divisors than n contains
- $\Omega(n)$ = the number of prime divisors that n contains counted with multiplicity

We define an additive function as multiplicative if f is an arithmetic function such that for $m, n \in \mathbb{N} \wedge \gcd(m, n) = 1$, then $f(m)f(n) = f(mn)$. if $f(m)f(n) = f(mn)$ for all m and n regardless of them being co-prime, then we call the function f as completely multiplicative.

Theorem : *if* f *is a multiplicative function and* F *is defined* $F(n) = \sum_{d|n} f(d)$ *then* F *is also multiplicative.* Proof : Let $m, n \in \mathbb{N} \wedge \gcd(m, n) = 1$. since m and n are relatively prime numbers, we observe that d is product of all the prime divisors of m and n. Hence d can be rewritten as d = $l_1 l_2$ such that $l_1 | m$ and $l_2 | n$. This gives

$$F(m, n) = \sum_{d|mn} f(d) = \sum_{l_1|ml_2|n} f(l_1)f(l_2) = \sum_{l_1|m} \sum_{l_2|n} f(l_1) \cdot f(l_2)$$

[9] $\sum_{d|n}$ means to take the sum of all values of $f(d)$ where d runs over all the positive divisors of n.

$$= \left(\sum_{l_1|m} l_1|n \right) \cdot \left(\sum_{l_2|n} f(l_2) \right) = F(m) \cdot F(n)$$

Hence F is multiplicative.

We shall now shift our attention to the convolution between 2 functions [10]. The convolution of two arithmetic functions f and g is defined by

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

Dirichet's product is commutative and associative. If both functions f and g are multiplicative, then $f * g$ is commutative.

We shall now define the möbius-μ function in order to derive the möbius inversion formula. for a $n \in \mathbb{Z}^+$, we define $\mu(n)$ as the following :

$$\mu(n) \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } p^2|n \text{ for some prime } p \\ (-1)^r & \text{if } n = p_1 p_2 ... p_r, \text{where } p_i \text{ are distinct primes} \end{cases}$$

$\mu(n)$ is a multiplicative function. We can prove this by assuming m,n are relatively prime such that their prime decomposition contains no repeated factors. let $m = p_1 p_2 ... p_r$ and $n = q_! q_2 ... q_S$, where no prime $p_i = q_j$. Then

$$\mu(mn) = \mu((p_1 p_2 ... p_r) \cdot (q_1 q_2 ... q_s)) = (-1)^{r+s} = (-1)^r \cdot (-1)^s = \mu(n) \cdot \mu(m)$$

Hence showing that the möbius-μ function is multiplicative. Furthermore, if we were to take the sum of all $\mu(d)$ (where d consists of all the prime divisors of n) then we get the following result :

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}$$

[11] We shall now define Dirichlet's identity function (I_n) which states :

$$I_n = \sum_{d|n} \mu(d) = \left[\frac{1}{n}\right] = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}]$$

[10] convolution is also known as Dirichlet's product

[11] Assume n = p^k, then $\sum_{d|p^k} \mu(d) = \mu(1) + ... + \mu(p^k) = \mu(1) + \mu(p) = 1 + (-1)^k = 0$.This is the reason for $n > 1$, $\sum_{d|p^k} \mu(d) = 0$

Theorem (Dirichlet's Inverse) - *If f is an arithmetical function with $f(1) \neq 0$, there is a unique arithmetical function f^{-1} such that $f * f^{-1} = f^{-1} * f = I$*

Now that we have defined the möbius-μ function and Dirichlet's identity function, we can now state möbius's inversion formula.

Theorem (Möbius Inversion Formula) - *Let f and g be two number theoretic functions such that :*

$$f(n) = \sum_{d|n} g(n)$$

Then it implies

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

Euler's totient Function - the number of positive integers not exceeding n which are relatively prime to n.

$$\varphi(n) = \sum_{k=1 \ \gcd(k,n)=1} 1$$

This can be rewritten in the product form :

$$\varphi(n) = n \prod_{p|n} (1 - \frac{1}{p})$$

A key theorem which labels the relationship between euler's totient function and möbius-μ function is shown below [12] :

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

We shall now shift our attention to some non-exhaustive elements of analytic number theory required for comprehending the heuristic derivation of the erdos-kac theorem.

We shall first define big O notation - Let us assume b, $b \in \mathbb{R}$, and let $g(x)$ be a real-valued function such that $g(x) > 0$ when $x \geq b$. Then we write $f(x) = O(g(x))$ which means the quotient $\frac{f(x)}{g(x)}$ is bounded by $x \geq b$, such that there exist some $M > 0$ such that $|f(x)| \leq Mg(x)$, $\forall x \geq A$. Note that sometimes we << to represent the big O notation.

[12] [Apo76] - To refer to the proof of this theorem

Another thing to note is if $\lim_{n\to\infty}\frac{f(x)}{g(x)} = 1$ then we state $f(x) \sim g(x)$ as $n \to \infty$. However if $\lim_{n\to\infty}\frac{f(x)}{g(x)} = 0$ then we say $f(x) = o(g(x))$ as $n \to \infty$.

Finally we shall look at two last things to end this section of prerequisites: The prime number theorem and Merten's estimates.

Theorem (Prime Number Theorem) - *Let x be a real positive number and* $\pi(x)$ *be the number of primes less than x. Then*

$$\pi(x) \sim \frac{x}{\ln(x)}$$

Merten's theorems are a set of estimates which help quantify the distribution and density of prime numbers up to a given limit x. Hence we define it as the following :

Theorem (Merten's Estimates) - *Let x be a positive real number such that* $x > 1$. *We get the following :*

- $\sum_{p\leq x}\frac{\ln(p)}{p} = \ln(x) + O(1)$
- $\sum_{p\leq x}\frac{1}{p} = \ln(\ln(x)) + A + O\left(\frac{1}{\ln(x)}\right)$ where A is some constant. [13]

This marks the end of all the prerequisites required for the expository. We shall now move on to the derivation of the erdos-kac theorem.

3.2 The Erdos-Kac theorem and its Derivation

The Erdös-kac theorem is one of the most popular results in the field of probabilistic number theory. The theorem was proven by Paul Erdös and Mark kac in 1940 by combining the ideas of the central limit theorem with Brun's sieve. (However for the heuristic derivation, we shall not use Brun's sieve approximation as that would be a separate expository in itself).

Theorem (Erdos Kac Theorem) - *Let* $\omega(n)$ *denote the number of distinct prime divisors of n. Then, as* $n \to \infty$, *the normalized arithmetic function*

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

converges in distribution to the standard normal distribution.

Heuristic Proof: let the arithmetic function $\omega(n)$ denote the number of prime divisors that n contains.

[13] Note that there are 2 more mertan estimates which i have omitted from this paper due to it being out of the scope of this expository. To refer to the complete set of merten's estimates, refer to Chapter 4 [Apo76]

We shall model the bernoulli random variable such that it creates an equivalence to $\omega(n)$. To do this, we can model the conditions such that for the prime $p \in \mathbb{Z}^+$; we assign 1 to $p|n$ and 0 to $p \nmid n$. Therefore we redefine $\omega(n)$ such that :

$$B_p(n) = \begin{cases} 1 & \text{if } p|n \\ 0 & \text{if } p \nmid n \end{cases}$$

To explain this model further, what has been done is that we are defining the bernoulli random variable as an indicator function for all prime numbers ,$p \in \mathbb{Z}^+$, for which divide by n. By summing up all the indicators which which equate to 1, we get the number of prime divisors that n has. This may seem very unintutive at first, so to best understand this we shall use a simple example.

$$\omega(n) = \sum_{p \leq n} B_p(n)$$

For example, let n = 15. The number of primes which are less than 15 = $\{2, 3, 5, 7, 11, 13\}$ out of which $\{3, 5\}$ are divisible by n.

$p \leq n$	0 or 1
2	0
3	1
5	1
7	0
11	0
13	0

$$\omega(n) = \sum_{p \leq n} B_p(n) = \sum_{p \leq n} B_2(15) + B_3(15) + B_5(15) + B_7(15) + B_{11}(15) + B_{13}(15)$$

$$0 + 1 + 1 + 0 + 0 + 0 = 2$$

$\omega(n) = 2$ [which is true since the prime factors of 15 is 3 and 5]

Now that we have created an equivalence for $\omega(n)$ using random variables, we can find its expectation and variance. Recall that the expectation of a binomial variable is the probability of success, which in this case is $\frac{1}{p}$.Therefore the expectation for $\mathbb{E}(\omega(n))$ equals to

$$\mathbb{E}(\omega(n)) = \sum_{p \leq n} \mathbb{E}[B_p(n)] = \sum_{p \leq n} \frac{1}{p}$$

Using Mertens estimate which states $\sum_{p \leq x} \frac{1}{p} = \log(\log(\text{x})) + A + o\left(\frac{1}{\log(\text{x})}\right)$:

$$\sum_{p\leq n}\frac{1}{p}=\log(\log(\mathrm{n}))+A+o\left(\frac{1}{\log(\mathrm{n})}\right)$$

$$\sum_{p\leq n}\frac{1}{p}=\log(\log(\mathrm{n}))+A+o\left(1\right)$$

as $n\to\infty$ there is asymptotic equivalence with $\mathbb{E}[\omega(n)]$ and loglogn.

$$\sum_{p\leq n}\frac{1}{p}=\log(\log(\mathrm{n}))+A+o\left(1\right)$$

$$\frac{\sum_{p\leq n}\frac{1}{p}}{\log(\log\,n)}=\frac{\log(\log(\mathrm{n}))}{\log(\log\,n)}+\frac{A}{\log(\log\,n)}+\frac{o\left(1\right)}{\log(\log\,n)}$$

$$\frac{\sum_{p\leq n}\frac{1}{p}}{\log(\log\,n)}=1+0+0$$

which therefore shows :

$$\sum_{p\leq n}\frac{1}{p}\sim\log(\log\,n)$$

We shift our focus to the variance of the Bernoulli random variable. the variance for a Bernoulli random variable is equal to $\frac{1}{p}\left(1-\frac{1}{p}\right)$. Therefore the variance of $\omega(n)$ equals to

$$Var(\omega(n))=\sum_{p\leq n}Var(B_p(n))=\sum_{p\leq n}\frac{1}{p}\left(1-\frac{1}{p}\right)=\sum_{p\leq n}\frac{p-1}{p^2}$$

This algebraically simplifies to

$$\sum_{p\leq n}\frac{p-1}{p^2}=\sum_{p\leq n}\frac{1}{p}-\sum_{p\leq n}\frac{1}{p^2}$$

as $n\to\infty$, $\sum_{p\leq n}\frac{1}{p^2}$ gets smaller and smaller eventually converging to zero. While $\sum_{p\leq n}\frac{1}{p}$ is a classic p series which diverges. Because of this we state the following approximation

$$\sum_{p\leq n}Var(B_p(n))=\sum_{p\leq n}\frac{1}{p}$$

We can now apply Merten's estimate :

$$\sum_{p \leq n} \frac{1}{p} = \text{loglog(n)} + A + o\left(\frac{1}{\log(\text{x})}\right)$$

as $n \to \infty$ there is asymptotic equivalence between $\sum_{p \leq n} \frac{1}{p}$ and loglog(n)

$$\frac{\sum_{p \leq n} \frac{1}{p}}{\text{loglog}n} = \frac{\text{loglog}n}{\text{loglog}n} + \frac{A}{\text{loglog}n} + \frac{o(1)}{\text{loglog}n}$$

which simplifies to

$$\frac{\sum_{p \leq n} \frac{1}{p}}{\text{loglog}n} = 1$$

hence

$$\sum_{p \leq n} \frac{1}{p} = Var(\omega(n)) = \text{loglog}n$$

$$\sum_{p \leq n} \frac{1}{p} \sim \text{loglog}n$$

$$Var(\omega(n)) \sim \text{loglog}n$$

we have now shown that expectation and variance of $\omega(n)$ approximates to loglog(n). Recall the central limit theorem when many independent random variables are added together, the normalized sum tends toward a normal distribution. Hence we can model the central limit theorems the following :

$$\frac{\omega(n) - \mathbb{E}(\omega(n))}{\sqrt{Var(\omega(n)}}$$

$$\frac{\omega(n) - \text{loglog(n)}}{\sqrt{loglog(n)}} \to \mathcal{N}(0,1)$$

Hence, heuristically proving the theorem.

However there is a minor consequence of this heuristic proof of the erdos-kác theorem. Throughout this entire proof we have implicitly assumed that each bernoulli random variable is independent to each other. However in reality this is not the case, assume two distinct primes a and b :

$$P(a|n) = \frac{\lfloor A \rfloor}{a}, \qquad P(b|n) = \frac{\lfloor B \rfloor}{b}.$$

If they were independent, we would expect:

$$P(a|n, b|n) = P(a|n)P(b|n) = \frac{\lfloor N/ab \rfloor}{N}.$$

However our assumption is

$$P(a|n, b|n) = P(a|n)P(b|n) = \frac{1}{ab}$$

This mall error compared to the actual value due to probability being calculated over a finite interval from $\{1, 2, .., N\}$ decreases as $N \to \infty$ such that

$$\frac{\lfloor N/ab \rfloor}{N} \sim \frac{1}{ab}, [\text{as } n \to \infty]$$

Because of this we assume that

$$P(a|n, b|n) = P(a|n)P(b|n)$$

and hence can ignore the small error and still be able to gain relatively accurate data. Though to fix this consequence, we must use sophisticated methods which are not included in this expository as it is out of scope.

4 Relations between the Erdos-Kac theorem and Ergodic Theory

4.1 Deterministic Systems with statistical behavior

A central idea connecting ergodic theory and number theory is that deterministic objects can exhibit statistical patterns.

In ergodic theory, Birkhoff's theorem studies the long-term average of an observable along a trajectory:

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k(x)).$$

This shows how repeated observations of a deterministic system can produce predictable statistical behavior.

Similarly, the Erdös–Kac theorem studies the statistical distribution of

$$\omega(n) = \sum_{p \leq n} 1_{\{p|n\}},$$

as n ranges over the integers.

Although the integers themselves are deterministic, their prime divisors follow a probabilistic pattern, with $\omega(n)$ approaching a normal distribution after normalization.

A central idea connecting ergodic theory and number theory is that deterministic objects can exhibit statistical patterns.

In ergodic theory, Birkhoff's theorem studies the long-term average of an observable along a trajectory:

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k(x)).$$

This shows how repeated observations of a deterministic system can produce predictable statistical behavior.

Similarly, the Erdös–Kac theorem studies the statistical distribution of

$$\omega(n) = \sum_{p \leq n} 1_{\{p|n\}},$$

as n ranges over the integers.

Although the integers themselves are deterministic, their prime divisors follow a probabilistic pattern, with $\omega(n)$ approaching a normal distribution after normalization.

This illustrates a common principle shared by both fields. Rather than studying individual points or integers, we instead examine their collective behaviour over sufficiently large scales. By considering averages, limiting distributions, and asymptotic behaviour, deterministic mathematical objects begin to exhibit statistical laws which resemble those of genuinely random systems.

As a simple example, Birkhoff's Ergodic Theorem states that for an ergodic measure-preserving dynamical system,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(T^k(x)) = \int_X f \, d\mu$$

for almost every $x \in X$. Likewise, the Erdős–Kac theorem states that

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \Longrightarrow N(0,1),$$

demonstrating that the arithmetic function $\omega(n)$ satisfies a universal statistical law despite arising from deterministic prime factorization.

One of the major open problems motivated by this philosophy is **Sarnak's Möbius Disjointness Conjecture**. The conjecture predicts that deterministic dynamical systems of zero entropy are asymptotically uncorrelated with the Möbius function, further strengthening the connection between ergodic theory and analytic number theory.

4.2 Additive Functions as Dynamical Observables

Let $f : \mathbb{N} \to \mathbb{R}$ be an additive arithmetic function satisfying

$$f(mn) = f(m) + f(n), \qquad \gcd(m,n) = 1.$$

The classical example is the prime divisor counting function

$$\omega(n) = \sum_{p|n} 1,$$

which can be interpreted as an observable on the probability space

$$(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathbb{P}_N),$$

where

$$\mathbb{P}_N(A) = \frac{1}{N}\#\{n \le N : n \in A\}.$$

For each $n \in \mathbb{N}$ define the orbit

$$n \mapsto (T^k n)_{k \ge 0},$$

where the transformation T acts on the factorization structure of n. The arithmetic function f can therefore be viewed as an observable:

$$X_n = f(n).$$

The Erdős–Kac theorem states that for many additive functions,

$$\frac{f(n) - \mu_N}{\sigma_N} \Rightarrow N(0,1),$$

where

$$\mu_N = \sum_{p \leq N} \frac{f(p)}{p},$$

and

$$\sigma_N^2 = \sum_{p \leq N} \frac{f(p)^2}{p} - \left(\sum_{p \leq N} \frac{f(p)}{p} \right)^2.$$

For the function $\omega(n)$,

$$\mu_N \sim \log\log N,$$

and

$$\sigma_N^2 \sim \log\log N,$$

giving

$$\frac{\omega(n) - \log\log n}{\sqrt{\log\log n}} \Rightarrow N(0,1).$$

Let's take an example for N = 10^6, the empirical distribution of

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

can be approximated by

$$\frac{1}{N} \sum_{n=1}^{N} \delta_{Z_n}.$$

The first moments satisfy approximately

$$\frac{1}{N} \sum_{n \leq N} Z_n \approx 0,$$

and

$$\frac{1}{N}\sum_{n\leq N} Z_n^2 \approx 1,$$

matching the limiting Gaussian measure.

Similarly, for the additive function

$$\Omega(n) = \sum_{p^k||n} 1,$$

we obtain

$$\frac{\Omega(n) - \log\log n}{\sqrt{\log\log n}} \Rightarrow N(0,1).$$

Hence different arithmetic observables produce the same limiting statistical behavior, analogous to ergodic systems where different observables may share identical limiting distributions. We can connect this to ergodic avergaes as in ergodic theory, an observable $g : X \to \mathbb{R}$ satisfies

$$\frac{1}{N}\sum_{k=0}^{N-1} g(T^k x) \to \int_X g\, d\mu.$$

For additive functions, the analogous statistical average is

$$\frac{1}{N}\sum_{n\leq N} f(n) \to \mathbb{E}(f),$$

with fluctuations governed by

$$\frac{f(n) - \mathbb{E}(f)}{\sqrt{\mathrm{Var}(f)}}.$$

The Erdős-Kac phenomenon therefore suggests that prime divisibility behaves like a weakly dependent dynamical system whose observables satisfy a central limit theorem.

An open problem is to Determine the largest class of additive functions f for which the normalized observable

$$Z_n = \frac{f(n) - \mu_N(f)}{\sigma_N(f)}$$

satisfies

$$\lim_{N\to\infty} \frac{1}{N} \#\left\{n \leq N : Z_n \leq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

for every

$$x \in \mathbb{R}.$$

Equivalently, characterize all additive observables for which the arithmetic dynamical system exhibits Gaussian statistical behavior.

5 Acknowledgements

I would like to express my sincere gratitude to my TA Ophir Horowitz and Simon Rubinstein-Salzedo, for their guidance and support throughout this project. I am especially grateful for the opportunity they provided me to undertake my first research paper and express this idea.

6 Bibliography

References

[Abb15] Stephen Abbott. *Understanding analysis*. Undergraduate Texts Math. New York, NY: Springer, 2nd ed. edition, 2015.

[Apo76] Tom M. Apostol. *Introduction to analytic number theory*. Undergraduate Texts Math. Springer, Cham, 1976.

[Axl20] Sheldon Axler. *Measure, integration & real analysis*, volume 282 of *Grad. Texts Math.* Cham: Springer, 2020.

[Dem23] Amir Dembo. Probability theory course notes by amir dembo, Mar 2023.