ERDŐS-KAC THEOREM

JOSH ZEITLIN

1. Abstract

In this paper, we will discuss the distribution of the number of prime factors a given integer has. We will discuss some initial work done by Hardy and Ramanujan which culminated in the Hardy Ramanujan Theorem. Finally, we'll discuss the Erdős-Kac theorem, which is also described as the fundamental theorem of probabilistic number theory. The Erdős-Kac theorem describes the distribution for the number of prime factors of the integers n up to x where $\omega(n)$ is the number of prime factors as the probability distribution of $\frac{\omega(n)-\log\log(n)}{\sqrt{\log\log(n)}}$ converges to the normal distribution. It extends with work of Hardy and Ramanujan in the Hardy-Ramanujan theorem which says the normal order of $\omega(n)$ is $\log\log(n)$.

2. Important Definitions

Definition 2.1. $\omega(n)$
The function $\omega(n)$ yields the number of prime factors dividing n, more formally:

$$\omega(n) = \#\{p_i \mid n \text{ s.t. } p_i \neq p_j \forall i, j\}$$

Definition 2.2. The Error Function (erf(x))
The error function yields the error encountered in integrating the normal distribution, explicitly written as,

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{\frac{-t^2}{2}} \text{dt}$$

Definition 2.3. Big O notation
$f(x) = O(g(x))$ means that f and g have the same asymptotic behavior as $x \to \infty$.

In more formal terms $f(x) = O(g(x))$ means that $\forall c \geq 0\ \exists k$ such that $|f(x)| \leq c{\cdot}g(x)\ \forall x \geq k$

Definition 2.4. Little o notation
The o notation is a stronger version of the O notation in the sense that $f(x) = o(g(x))$ means $\lim_{n\to\infty} \frac{f(x)}{g(x)} = 0$.

Definition 2.5. Normal Order
A function $f(x)$ has the normal order $g(x)$ if $f(x) \approx g(x)$ for almost all values of x. More formally:

$$(1-\epsilon)g(x) \leq f(x) \leq (1+\epsilon)g(x)\ \forall \epsilon > 0$$

Definition 2.6. Gaussian (Normal) Distribution
We defined the error function for this earlier to be our erf(x). However, the Gaussian distribution is defined here:

$$\Phi(x, y) = \frac{1}{\sqrt{2\pi}} \int_{x}^{y} e^{\frac{-t^2}{2}}$$

Date: May 2019.

3. The Hardy-Ramanujan Theorem

Theorem 3.1. ***The Hardy-Ramanujan Theorem***
The Hardy-Ramanujan Theorem says that if you define a function $f(n) = o(n^k)$, then

$$|\omega(n) - \log(\log(n))| < f(n)\sqrt{\log(\log(n))}$$

In simpler terms, this theorem essentially states that the normal order of the number of distinct prime factors of a number is approximately $\log(\log(n))$. We can thus write this theorem more compactly as

$$|\omega(n) - \log(\log(\text{n}))| < \log(\log(\text{n}))^{\frac{1}{2}+\epsilon}\text{for almost all } n \in \mathbb{Z}.$$

The almost always essentially means let $\rho(x)$ be the number of positive integers $n \leq x$ for which the inequality fails, then $\frac{\rho(x)}{x} \to 0$ as $x \to \infty$, so for larger and larger numbers, the ratio of failed integers up to x compared to total integers approaches zero.

Now, let's go into the history of this development. Hardy and Ramanujan proved this together in 1917; however, 17 years later this was actually proved in 1934 by Paul Turán using the Turán sieve, a much more innovative proof technique.

The Turán sieve is a technique used to estimate the size of sifted sets of positive integers which satisfy certain conditions expressed by congruences. This sieve gives the upper bound of the size of a sifted set and it is derived from an elementary form of inclusion/exclusion principle. Now, before we state the Turán sieve we will use a couple of definitions used in the proof, which should hopefully make the Sieve seem less daunting and abstract.

Definition 3.2. Pre-proof Definitions

(1) Let $\mathcal{S}$ be a finite set and $\mathcal{I}$ be an index set. Then, $\forall i \in \mathcal{I}$, let $\Omega(i)$ denote some arbitrary conditions to be satisfied, then, we define
$$\mathcal{S}_i = \{s \in \mathcal{S} \text{ s.t. } s \text{ satisfies } \Omega(i)\}$$
(2) Using the same $\Omega(i)$ and as in the last definition, $\forall s \in \mathcal{S}$, we'll define
$$\pi_s(\mathcal{I}) = \#\{i \in \mathcal{I} \text{ s.t. } s \text{ satisfies } \Omega(i)\}$$
(3) Now, let's define two constants, δ_i and ρ_i such that δ_i is significantly larger than ρ_i we get that
$$\frac{|\mathcal{S}_i|}{|\mathcal{S}|} = \delta_i + \rho_i.$$
Think of these as a quotient and remainder.
(4) If we choose some $i, j \in \mathcal{I}$ s.t. $i \neq j$, then
$$\frac{|\mathcal{S}_i \cup \mathcal{S}_j|}{|\mathcal{S}|} = \delta_i\delta_j + \rho_i\rho_j,$$
where $\delta_i\delta_j$ is significantly larger than $\rho_i\rho_j$. Again, δ_i is an approximation whereas ρ_i is a remainder/error term.

Now, we write out the Turán Sieve.

Theorem 3.3. ***The Turán Sieve***
Let $\nu = \sum_{i\in\mathcal{I}} \delta_i$. Then,

$$\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}(\pi_s(\mathcal{I}) - \nu^2) = \sum_{i\in\mathcal{I}}\delta_i(1-\delta_i) + \sum_{i,j\in\mathcal{I}} r_{i,j} - 2\nu\sum_{i\in\mathcal{I}} r_i$$

Before the proof a corollary:

Corollary 3.4.

$$\#\{s \in \mathcal{S} : \pi_s(\mathcal{I}) = 0\} \leq \frac{|\mathcal{S}|}{\nu} + \frac{|\mathcal{S}|}{\nu^2} \sum_{i,j \in \mathcal{I}} |\rho_{i,j}| + 2|\mathcal{S}| \sum_{i \in \mathcal{I}} |\rho_i|$$

Proof. **Proof of the Turán Sieve**

$$\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} (\pi_s(\mathcal{I}) - \nu)^2$$

$$= \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} (\pi_s(\mathcal{I})^2 - 2\nu\pi_s(\mathcal{I}) + \nu^2),$$

which we will then give the result as

$$\mathcal{S}_1 - \mathcal{S}_2 + \nu^2$$

From our definition of $\pi_s(\mathcal{I})$,

$$\mathcal{S}_1 = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} (\pi_s(\mathcal{I}))^2$$

$$= \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \Big(\sum_{\substack{i \in \mathcal{I} \\ s \in \mathcal{S}_i}} 1 \Big)^2$$

Now we switch the double sum and get

$$\mathcal{S}_1 = \frac{1}{|\mathcal{S}|} \sum_{i,j \in \mathcal{I}} \sum_{s \in \mathcal{S}_i \cap \mathcal{S}_j} 1$$

$$= \sum_{\substack{i,j \in \mathcal{I} \\ i \neq j}} \frac{|\mathcal{S}_i \cap \mathcal{S}_j|}{|\mathcal{S}|} \sum_{i \in \mathcal{I}} \frac{|\mathcal{S}_i|}{|\mathcal{S}|}$$

Now we will use $\delta_i, \rho_i, \rho_{i,j}$ and there definitions to get that

$$\mathcal{S}_1 = \sum_{\substack{i,j \in \mathcal{I} \\ i \neq j}} \delta_i \delta_j + \rho_{i,j} \sum_{i \in \mathcal{I}} \delta_i + \rho_i$$

$$= \Big(\sum_{i \in \mathcal{I}} \delta_i \Big)^2 - \sum_{i \in \mathcal{I}} \delta_i^2 + \sum_{i,j \in \mathcal{I}} \rho_{i,j} + \sum_{i \in \mathcal{I}} \delta_i$$

Now we show $\mathcal{S}_2$ as:

$$\mathcal{S}_2 = 2\nu \sum_{i \in \mathcal{I}} \delta_i + 2\nu \sum_{i \in \mathcal{I}} \rho_i$$

Now we combine our final equations for $\mathcal{S}_1$ and $\mathcal{S}_2$ we get that

$$\mathcal{S}_1 - \mathcal{S}_2 + \nu^2 = \Big(\sum_{i \in \mathcal{I}} \delta_i - \nu \Big)^2 + \sum_{i \in \mathcal{I}} \delta_i (1 - \delta_i) + \sum_{i,j \in \mathcal{I}} \rho_{i,j} - 2\nu \sum_{i \in \mathcal{I}} \rho_i$$

$$= \sum_{i \in \mathcal{I}} \delta_i (1 - \delta_i) + \sum_{i,j \in \mathcal{I}} \rho_{i,j} - 2\nu \sum_{i \in \mathcal{I}} \rho_i.$$

This essentially proves the Turán Sieve. ■

Remark 3.5. Before we go into Turán's proof of the Hardy-Ramanujan theorem we'll show some connections between the Turán Sieve and statistics which are some key insights that Turán used in his proof. The key insight is that our function $\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}(\pi_f(\mathcal{I})-\nu)^2$ can be viewed as variance.

For every $i \in \mathcal{I}$ we can create a random variable $X_i : \mathcal{S} \to \{0,1\}$ that has uniform distribution. Now, we can define

$$X_i = \begin{cases} 1 \text{ if } s \in \mathcal{S}_i \\ 0 \text{ if } s \notin \mathcal{S}_i \end{cases}$$

Using that definition we get $E(X_i) = \frac{|\mathcal{S}_i|}{|\mathcal{S}|} \approx \delta_i$. Now, let $X = \sum_{i\in\mathcal{I}} X_i$ be another discrete random variable with uniform distribution we can get that

$$X(s) = \#\{i \in \mathcal{I} : s \in \Omega(i)\} = \pi_s(\mathcal{I}),$$

and

$$E(X) = \sum_{i\in\mathcal{I}} E(X_i) \approx \sum_{i\in\mathcal{I}} \delta(i) = \nu.$$

Thus,

$$\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}(\pi_s(\mathcal{I})-\nu)^2 = \mathrm{Var}(X)$$

Again, Turán used the idea of variance and connections with statistics to prove the Hardy-Ramanujan theorem.

Proof. **Turán's proof of the Hardy-Ramanujan Theorem**
Turán created a function $R(N)$ such that showing $R(N) = O(n\log\log n)$ or in other words, $R(N) \ll n\log\log n$ implies that the Hardy-Ramanujan Theorem is true for $\omega(n)$. His $R(N)$ definition was like the variance idea which we just discussed:

$$R(N) = \sum_{n=1}^{N}(\omega(n) - \log\log N)^2$$

We will use a couple of lemmas to prove this theorem. I will omit their proofs and leave them as an exercise for the reader.

Lemma 3.6.

$$\sum_{n=1}^{N}(\omega(n))^2 = \sum_{i\neq j}\left[\frac{N}{p_ip_j}\right] + \sum_i\left[\frac{N}{p_i}\right]$$

Lemma 3.7.

$$\sum_{n=1}^{N}\omega(n) = \sum_i\left[\frac{N}{p_i}\right]$$

Lemma 3.8.

$$\sum_{p_ip_k\leq N}\frac{1}{p_ip_j}\log\log(N)^2 + O(\log\log(N))$$

Now we will use a consequence of the first and third lemmas we used.

$$\sum_{n=1}^{N}(\omega(n))^2 = N\sum_{p_ip_j\le N}\frac{1}{p_ip_j} + O(N) + N\sum_{p_i\le N}\frac{1}{p_i} + o(N)$$
$$= N(\log\,\log(N))^2 + O(N\log\,\log N)$$

Using our second lemma we get that

$$\sum_{n=1}^{N}\omega(n) = N\sum_{p_i\le N}\frac{1}{p_i} + o(N)$$
$$= N\log\,\log(N) + o(N),$$

so we get

$$R(N) = \sum_{n=1}^{N}(\omega(N) - \log\,\log\,(N))^2$$
$$= \sum_{n=1}^{N}(\omega(n))^2 - 2\log\,\log(N)\sum_{n=1}^{N}\omega(N) + (N\log\,\log(N))^2$$
$$= O(N\log\,\log(N))$$

Now it is simple to deduce the Hardy-Ramanujan theorem. ■

As I said Turán gave a much simpler proof 17 years after Hardy and Ramanujan's first proof. It was really long and extensive and frankly not as cohesive as Turán's proof. For those reasons I will leave you with only Turán's proof.

4. **Erdős-Kac Theorem**

Now we are at the home stretch, finally at the theorem we intended to prove all along. Now just to recall, the Erdős-Kac Theorem says that the probability distribution of

$$\frac{\omega(n) - \log\,\log N}{\sqrt{\log\,\log N}} = \Phi(x,y).$$

In other words the probability distribution of that function is the normal distribution.

Definition 4.1. Probability Distribution
The probability distribution is a mathematical function that provides the likelihood of possible results from an experiment. In other words, it is a description of a random phenomenon in terms of the probabilities of events.

Definition 4.2. Strongly Additive
A function f is said to be strongly additive if we have some $n = p_1^{e_1}\cdots p_k^{e_k}$ and $f(n) = f(p_1) + ... + f(p_k)$ where $|f(p)| \le 1$ for every prime number p.

First, lets define two functions and a couple of useful theorems.

Definition 4.3. Define two functions $A(n)$ and $B(n)$ such that

- $A(n) = \sum_{p\le n}\frac{f(p)}{p}$
- $B(n) = \sqrt{\sum_{p\le n}\frac{(f(p))^2}{p}}$

Definition 4.4. The Brun Sieve
Let A be a set of positive integers less than or equal to x and let P be a set of primes. For each p in P, let A_p denote the set of elements of A divisible by p and extend this to let A_d, the intersection of the A_p for p dividing d, when d is a product of distinct primes from P. Further let A_1 denote A itself. Let z be a positive real number and $P(z)$ denote the primes in $P \leq z$. The object of the sieve is to estimate

$$S(A, P, z) = \big|A \bigcup_{p \in P(z)} A_p\big|.$$

We let $|A_d|$ be written as

$$|A_d| = \frac{w(d)}{d} X + R_d$$

where w is multiplicative and $X = |A|$.

Definition 4.5. Lindeberg Condition
Let $(\Psi, \mathcal{Z}, \mathbb{M})$ be some probability space and $X_k : \Psi \to \mathbb{R}, k \in \mathbb{N}$ be independent random variables which are defined on that set. Let the expect value of X_k and variance of X_k be $\mathbb{E}[X_k] = \mu_k$ and $\mathbb{V}[X_k] = \sigma_k^2$ exist and be finite. The sequence X_k satisfies Lindeberg's condition if

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\big[\big(X_k - \mu_k\big)^2 \cdot \mathbb{1}_{\{X_k - \mu_k \text{ s.t. } \geq \epsilon \cdot s_n\}}\big] = 0$$

Here, we let $\epsilon > 0$ and $\mathbb{1}$ be the indicator function.

Definition 4.6. Central Limit Theorem
The central limit theorem means the random variables

$$Z_n = \frac{\sum_{k=1}^{n} (X_k - \mu_k)}{s_n}$$

converge to the standard normal, Gaussian, distribution. We get that if Lindeberg's condition holds, then so does the central limit theorem.

Theorem 4.7. ***Erdős-Kac Theorem***
For any fixed $a \leq b$

$$\lim_{x \to \infty} \left(\frac{1}{x} \# \Big\{ n \leq x \ s.t \ a \leq \frac{\omega(n) - log\ log(n)}{\sqrt{log\ log(n)}} \leq b \Big\} \right) = \Phi(a, b).$$

Moreover, if $f(n)$ is a strongly additive function then

$$\lim_{x \to \infty} \left(\frac{1}{x} \# \Big\{ n \leq x \frac{f(n) - A(n))}{B(n)} \Big\} \right) = \Phi(a, b)$$

For our proof, we'll use Erős' proof of the theorem. There are several other proofs that I'll encourage you to look at, done by both Halbertstam and Kac.

Proof. **Erdős's proof of the Erdős-Kac Theorem** We will first write out the definitions and theorems used in this proof:

- **Weak Convergence:** A sequence $\{F_n\}$ converges weakly to a function F if

$$\lim_{n \to \infty} F_n(x) = F(x) \text{ for all points where } F \text{ is continuous.}$$

- **Limiting Distribution Function:**

  Let f be an arithmetic function. Let N be a natural number.

  Now, we define $F_N(Z) = \nu_N\{n : f(n) \leq z\} = \dfrac{1}{N}\#\{n \leq N : f(n) \leq z\}$.

  We say that f posses a limiting distribution function F if the sequence $\{F_N\}$ converges weakly to a limit F that is a distribution function.
- **Characteristic Functions:**
  Let F be a distribution function. Then, its characteristic function is
  $$\varphi_F(\tau) = \int_{-\infty}^{\infty} \exp(i\tau z)\mathrm{d}F(z)$$
  A distribution function is completely characterized by its characteristic function and the characteristic function of Φ is $\varphi_\Phi(\tau) = \exp\left(\frac{-\tau^2}{2}\right)$
- **Levy's Convergence Theorem:**
  Let $\{F_n\}$ be a sequence of distribution functions and $\{\varphi_{F_n}\}$ be the corresponding sequence of their characteristic functions. Then $\{F_n\}$ converges weakly to a distribution function F if and only if φ_{F_n} converges pointwise on $\mathbb{R}$ to a function φ that is continuous at 0.

The atomic distribution function for some natural number N is
$$F_N(x) = \frac{1}{N}\#\Big\{n \leq N : \frac{\omega(n) - \log\log(N)}{\sqrt{\log\log(N)}} \leq x\Big\}$$
We will now write the characteristic function of F_N. We get
$$\varphi_{F_N(\tau)} = \int_{-\infty}^{\infty} e^{i\tau z}\mathrm{d}F_N(z).$$
If we take $P = \{\cdots \leq x_{-1} \leq x_0 \leq x_1 \leq \cdots \leq x_i \cdots\}$ be a partition of the real numbers. Then we get can simplify $\varphi_{F_N(\tau)}$
$$= \int_{-\infty}^{\infty} e^{i\tau z}\mathrm{d}F_N(z)$$
$$= \lim_{\mathrm{mesh}(P)\to 0} \sum_k e^{zi\tau}\big(F_N(x_k) - F_N(x_{k-1})\big)$$
$$= \lim_{\mathrm{mesh}(P)\to 0} \sum_k e^{zi\tau}\Big(\frac{1}{N}\#\Big\{n \leq N : \frac{\omega(n) - \log\log(N)}{\sqrt{\log\log(N)}} \leq x_k\Big\}$$
$$= \frac{1}{N}\#\Big\{n \leq N : \frac{\omega(n) - \log\log(N)}{\sqrt{\log\log(N)}} \leq x_{k-1}\Big\}\Big)$$
$$= \frac{1}{N}\Big[\lim_{\mathrm{mesh}(P)\to 0} \sum_k e^{zi\tau}\Big(\#\Big\{n \leq N : \frac{\omega(n) - \log\log(N)}{\sqrt{\log\log(N)}} \leq x_k\Big\}$$
$$= \#\Big\{n \leq N : \frac{\omega(n) - \log\log(N)}{\sqrt{\log\log(N)}} \leq x_{k-1}\Big\}\Big)\Big]$$

$$= \frac{1}{N} \sum_{k=0}^{\max\{\omega(n):n\leq N\}} e^{i\tau f(n)}$$

$$= \frac{1}{N} \sum_{n\leq N} e^{i\tau f(n)}$$

Now, to find the bounds for $\varphi_{F_n}(\tau)$ we get that

$$\varphi_{F_n}(\tau) = \exp\big(\frac{-\tau^2}{2}\big)\big(1 + O\big(\frac{|\tau| + |\tau|^3}{\sqrt{\log\,\log(N)}}\big)\big) + O\big(\frac{1}{\log\,\log(N)}\big)$$

Use the definition of $\varphi_{F_n}(\tau)$ we can let $n \to \infty$ and then we will get $\exp(\frac{-\tau^2}{2}) = \varphi_\Phi(\tau)$.
To put more simply, the characteristic function sequence converges pointwise to the characteristic function of the Gaussian distribution.

Now, after applying Levi's continuity theorem we get that

$$\frac{1}{N}\#\{n \leq N : \frac{\omega(n) - \log\,\log(N)}{\sqrt{\log\,\log(N)}} \leq x\} = \Phi(x, y).$$

Now we have completed the proof and we get that limit distribution the prime divisor counting function is in fact the Gaussian (or normal) distribution with both a mean and a variance of log $\log(N)$.

■

5. **References**

(1) Weisstein, Eric W. "Erdős-Kac Theorem." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Erdos-KacTheorem.html
(2) Riesel, H. "The Erdős-Kac Theorem." Prime Numbers and Computer Methods for Factorization, 2nd ed. Boston, MA: Birkhäuser, pp. 158-159, 1994.
(3) Kac, M. Statistical Independence in Probability, Analysis and Number Theory. New York: Wiley, 1959.
(4) Ross G. Pinsky, A Natural Probabilistic Model on the Integers and Its Relation to Dickman-Type Distributions and Buchstab's Function, Probability and Analysis in Interacting Physical Systems, $10.1007/978-3-030-15338-0_10, (267-294), (2019)$.
(5) Weisstein, Eric W. "Hardy-Ramanujan Theorem." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Hardy-RamanujanTheorem.html
(6) Yu-Ru Liu and M. Ram Murty, Sieve methods in combinatorics, Journal of Combinatorial Theory, Series A, 10.1016/j.jcta.2004.11.004, 111, 1, (1-23), (2005).
(7) A. E. Ingham, " The distribution of prime numbers " (Cambridge Tracts inMathematics, No. 30, 1932), 8.
(8) Hoheisel, Sitzungsberichte d. Preuss. A/fad, d. Wise., phys.-math. Kl. (1930),580.
(9) Turán, P. (2019). On a theorem of Hardy and Ramanujan. 11. 125-133.
(10) Paul Turán, On a Theorem of Hardy and Ramanujan, Journal of the London Mathematical Society, Volume s1-9,
Issue 4, October 1934, Pages 274–276, https://doi.org/10.1112/jlms/s1-9.4.274