# The Erdös-Kac Theorem

Harshil Nukala

Euler Circle

July 9, 2024

# Preliminary Notation

## Definition

(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 $x \geq a$.

## Preliminary Notation

### Definition

(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

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

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

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

# The Star of the Show: $\omega(n)$

**Definition**

Denote $\omega(n)$ to count the number of unique prime factors of $n$.

# The Star of the Show: $\omega(n)$

**Definition**

Denote $\omega(n)$ to count the number of unique prime factors of $n$.

**Example**

$\omega(2) = 1$, since 2 is prime and $\omega(2024) = 3$, since $2024 = 2^3 \times 11 \times 23$.

# Mertens' Second Theorem

Theorem (Mertens' Second Theorem)

*There exists a constant $C$ such that for $x \geq 2$,*

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

# Mertens' Second Theorem

**Theorem (Mertens' Second Theorem)**

*There exists a constant $C$ such that for $x \geq 2$,*

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

Mertens' Second Theorem can be used to find the expected value of $\omega(n)$.

# Mean and Variance

There are two definitions that we are particularly interested in when it comes to understand the Erdös-Kac Theorem: **mean** and **variance**.

# Mean and Variance

There are two definitions that we are particularly interested in when it comes to understand the Erdös-Kac Theorem: **mean** and **variance**.

### Definition (Mean)

Denote $X$ as a discrete random variable. Then, the mean of $X$, $\mu$, is defined as

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

# Mean and Variance

There are two definitions that we are particularly interested in when it comes to understand the Erdös-Kac Theorem: **mean** and **variance**.

## Definition (Mean)

Denote $X$ as a discrete random variable. Then, the mean of $X$, $\mu$, is defined as

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

## Definition (Variance)

Denote $X$ as a discrete random variable. Then, the variance of $X$, $\sigma^2$, is define as

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

# Mean of $\omega(n)$

We can calculate the mean of $\omega(n)$, $\mathbb{E}[\omega(n)]$, to be $\log \log n$ by the definition of $\omega(n)$ and Mertens' Second Theorem:

# Mean of $\omega(n)$

We can calculate the mean of $\omega(n)$, $\mathbb{E}[\omega(n)]$, to be $\log\log n$ by the definition of $\omega(n)$ and Mertens' Second Theorem:

$$\begin{aligned}
\frac{1}{x}\sum_{n\leq x}\omega(n) = \frac{1}{x}\sum_{n\leq x}\sum_{p|n} 1 &= \frac{1}{x}\sum_{p\leq x}\sum_{\substack{n\leq x\\ p|n}} 1 \\
&= \frac{1}{x}\sum_{p\leq x}\left\lfloor \frac{x}{p}\right\rfloor \\
&= \frac{1}{x}\sum_{p\leq x}\left(\frac{x}{p}+\mathcal{O}(1)\right) \\
&= \frac{1}{x}\left(x\log\log x + O(x)\right).
\end{aligned}$$

# Variance of $\omega(n)$

In 1934, Turán computed the variance of $\omega(n)$:

**Theorem (Turán 1934)**

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

# Variance of $\omega(n)$

In 1934, Turán computed the variance of $\omega(n)$:

**Theorem (Turán 1934)**

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

Hence, we have that the $\mathrm{Var}[\omega(n)] = \log\log n$.

# Motivation for Erdös-Kac Theorem

We have that the motivation for Erdös-Kac Theorem is due to the the following theorem from Hardy and Ramanujan.

**Theorem (Hardy-Ramanujan 1917)**

*For some real number $\delta$,*

$$\lim_{N\to\infty} \# \left\{ n \leq N : |\omega(n) - \log\log N| > (\log\log N)^{\frac{1}{2}+\delta} \right\} = o(N).$$

# Motivation for Erdös-Kac Theorem

We have that the motivation for Erdös-Kac Theorem is due to the the following theorem from Hardy and Ramanujan.

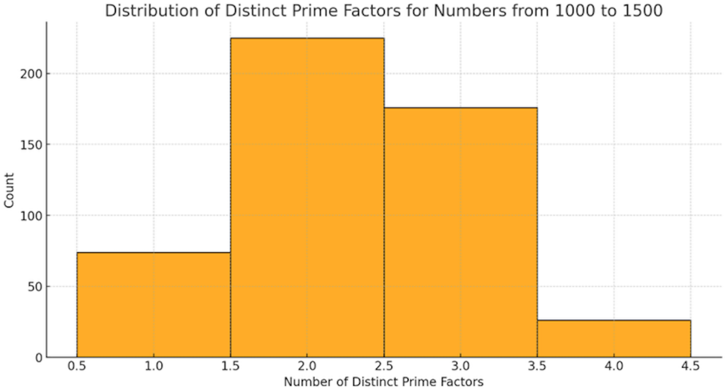
Theorem (Hardy-Ramanujan 1917)

*For some real number $\delta$,*

$$\lim_{N\to\infty} \# \left\{ n \leq N : |\omega(n) - \log\log N| > (\log\log N)^{\frac{1}{2}+\delta} \right\} = o(N).$$

Notice how Turán's Theorem can deduces the above theorem. This motivates us to ask the question if we can find a distribution for $\omega(n)$ to find a stronger result than the above Theorem. We will do some exploration.

# Distribution of $\omega(n)$ $(1000 \leq n \leq 1500)$

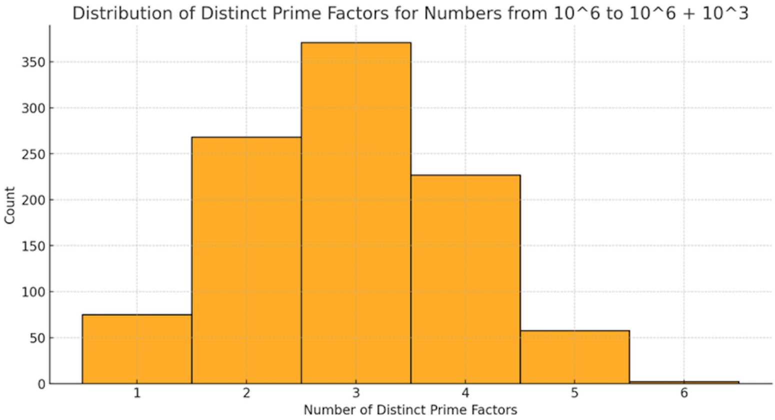


# Distribution of $\omega(n)$ Data ($1000 \leq n \leq 1500$)

| | |
|---|---|
| $\mu$ | 2.308 |
| $\sigma^2$ | 0.613 |
| $\log\log 1500$ | 1.990 |

| $\omega(n)$ | $\#(\omega(n))$ |
|---|---|
| 1 | 74 |
| 2 | 224 |
| 3 | 176 |
| 4 | 26 |

# Distribution of $\omega(n)$ $(10^6 \leq n \leq 10^6 + 10^3)$



# Distribution of $\omega(n)$ Data ($10^6 \leq n \leq 10^6 + 10^3$)

| | |
|---|---|
| $\mu$ | 2.931 |
| $\sigma^2$ | 1.039 |
| $\log\log(10^6 + 10^3)$ | 2.626 |

| $\omega(n)$ | $\#(\omega(n))$ |
|---|---|
| 1 | 75 |
| 2 | 268 |
| 3 | 371 |
| 4 | 227 |
| 5 | 58 |
| 6 | 2 |

# Normal Distribution

As you may have seen from the images of the proceeding slides, the graphs look pretty similar in nature. Indeed, the graphs look like the normal distribution!

# Normal Distribution

As you may have seen from the images of the proceeding slides, the graphs look pretty similar in nature. Indeed, the graphs look like the normal distribution!

**Definition (Normal Distribution)**

Denote $X$ to be a continuous random variable. Then $X$ has a *normal distribution* if it has a probability density function $f(x)$ give by

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

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

# The Erdös-Kac Theorem

In 1940, M. Kac and P. Erdös showed that

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

follows the normal distribution with mean 0 and variance 1.

# The Erdös-Kac Theorem

In 1940, M. Kac and P. Erdös showed that

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

follows the normal distribution with mean 0 and variance 1.

Which can be written more concisely as.

**Theorem (Erdös-Kac Theorem)**

*For* $\gamma \in \mathbb{R}$,

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

# Moments

**Definition (Moments)**

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

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

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

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

# Moments

## Definition (Moments)

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

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

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

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

## Definition (Central Moments)

Denote $X$ to be a continuous random variable and $f(x)$ be its normal density function. Then, the $k^{\text{th}}$ *central moment* is defined as

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

# Why Moments?

The proof by Granville and Soundarajan of the Erdös-Kac theorem is to show that the moments of

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

are asymptotic to the moments of the normal distribution.

# Why Moments?

The proof by Granville and Soundarajan of the Erdös-Kac theorem is to show that the moments of

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

are asymptotic to the moments of the normal distribution.

This can be done because the normal distribution are defined by its moments. Note that not all distributions exhibit this property (for example, the *lognormal* distribution cannot).

# Moments of the Normal Distribution

**Theorem**

*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,*

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

# Sufficiency of Erdös-Kac Theorem

Hence, by the previous theorem we can develop the following shorthand that represents the moments of the normal distribution for Erdös-Kac Theorem. Since $\sigma^2 = 1$,

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

# Sufficiency of Erdös-Kac Theorem

Hence, by the previous theorem we can develop the following shorthand that represents the moments of the normal distribution for Erdös-Kac Theorem. Since $\sigma^2 = 1$,

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

Hence, it suffices to show that

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

# Probabilistic Model

Denote

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

and consider the following probabilistic model:

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

# Probabilistic Model

Denote

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

and consider the following probabilistic model:

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

# Mean and Variance of $X(p)$

The mean can be calculated as

$$\mathbb{E}[X(p)] = 1 \cdot \frac{1}{p} + 0 \cdot \left(1 - \frac{1}{p}\right) = \frac{1}{p}.$$

# Mean and Variance of $X(p)$

The mean can be calculated as

$$\mathbb{E}[X(p)] = 1 \cdot \frac{1}{p} + 0 \cdot \left(1 - \frac{1}{p}\right) = \frac{1}{p}.$$

The variance can be calculated as

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

# Final Remarks on X(p)

The reason why we have defined this model is because of the following (which comes from Lindeberg-Feller Central Limit Theorem) is the following:

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

# Thank you

Thank you for your attention!