The Erdős-Kac Central Limit Theorem

Brian Hong
brianhong73@gmail.com

Euler Circle

July 11, 2022

Table of Contents

Introduction

$\omega(n)$

This theorem is centered around the function $\omega(n)$, which counts all the distinct prime factors of positive integer n. For example:

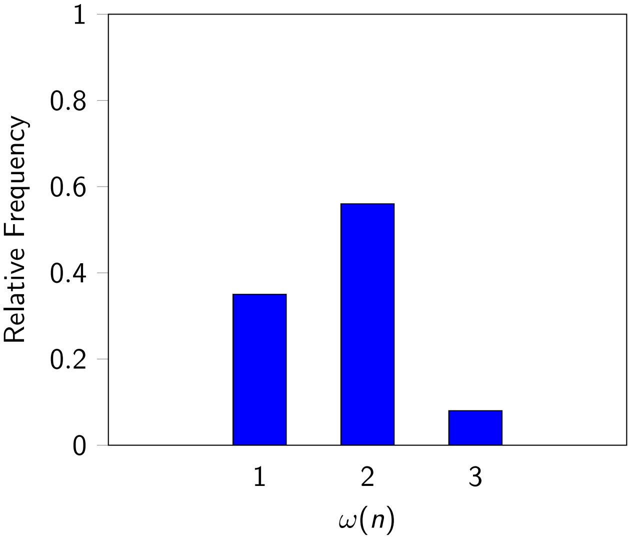
$$\omega(12) = \omega(2^2 \cdot 3) = 2$$

We will take a look at the values of $\omega(n)$ as n gets larger.

$n = 100$

```
1:  35
2:  56
3:  8
4:  0
5:  0
6:  0
7:  0
8:  0
```

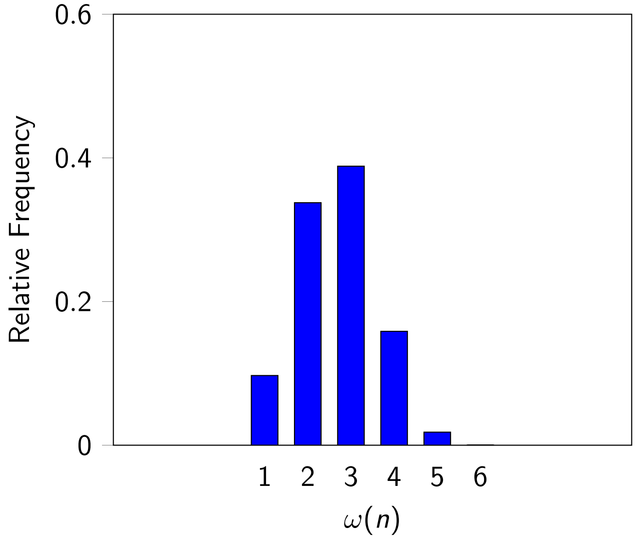
$n = 100$



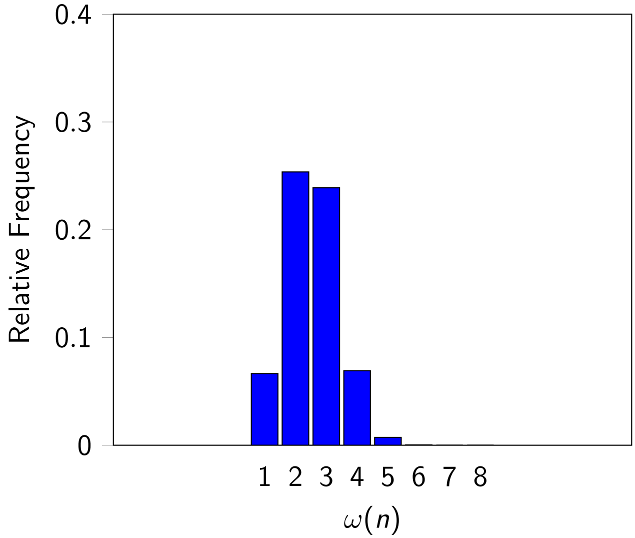
$n = 10^5$

```
1:   9700
2:   33759
3:   38844
4:   15855
5:   1816
6:   25
7:   0
8:   0
```

$n = 10^5$



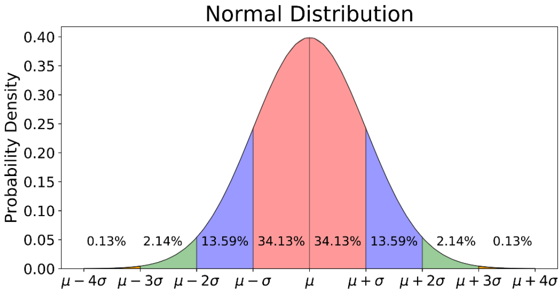
$n = 10^8$



The Normal Distribution

Definition 1.1

A common probability distribution seen in various situations is the normal distribution, which will form a bell-shaped curve.



The Normal Distribution Pt. 2

The normal distribution density function is expressed by the equation

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} exp\Bigg(- \frac{1}{2}\Bigg(\frac{z-\mu}{\sigma}\Bigg)^2\Bigg).$$

A key part of the proof of this theorem is proving that the behavior of the distribution function of Erdős-Kac matches the equation of the normal distribution. More on that later.

The Erdős-Kac Central Limit Theorem

To make the probability distribution of $\omega(n)$ even closer to the normal distribution, mathematicians Paul Erdős and Mark Kac in 1940 proved that the distribution of

Theorem 1.2

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

converges to a normal distribution as n approaches infinity.

Our probability model

We will use a probabilistic model for the occurrence of prime numbers by denoting $X(p)$ as

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

We can expect that the probability that an integer is divisible by a prime number is $1/p$, since the factor of p is every p number. The probability for the occurance of 0 can be found by taking the complementary, or $1 - 1/p$. The expected value of X(p) can be calculated as following:

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

Basics in Probability

Essential Terms

- σ^2 = Variance = The average of the differences from the mean within each value squared.
- σ = Standard Deviation = The measure of variance within a collection of data.
- $\mathbb{E}$ = Expected value.
- μ = Average of the data set.

The Central Limit Theorem

Theorem 2.1

The Central Limit Theorem states that when random, independent, and identically distributed variables are taken, the probability distribution of the averages will converge to a normal distribution.

- Rock Paper Scissors
- Rolling a Dice

Lindeberg-Feller Central Limit Theorem

Theorem 2.2

The Lindeberg-Feller Central Limit states that a sequence of random, independent variables will converge to a normal distribution.

This theorem differs from the original CLT by the fact that independent, and not necessarily identically distributed variables (along with a finite variance) meets conditions for the CLT.

Moments in Probability

The moments of a probability function allows us to see the asymptotic behavior of our probability function. In mathematical terms, the nth central moment of continuous random variable Z is

$$\mathbb{E}[(Z-\mu)^n] = \int_{-\infty}^{\infty} (z-\mu)^n f(z) dz.$$

- $n = 1$ represents mean
- $n = 2$ represents variance
- $n = 3$ represents skewness
- $n = 4$ represents kurtosis

Moments in Probability Pt. 2

The proof of this theorem will heavily depend on using the moments in the probability distribution based on the standard normal distribution that we are assuming matches the Erdos Kac. In this scenario, considering the even and odd moments and splitting them up works nicely.

$$\mathbb{E}(X^m) = \begin{cases} 2^{-m/2}\frac{m!}{(m/2)!} & \text{if m is even} \\ 0 & \text{if } m \text{ is odd} \end{cases}$$

A Small Problem...

It sounds correct to assume that if a sequence of random variables $X_1, X_2, \cdots$ converges, the moments should as well. Correct? Unfortunately, this cannot be completely certain due to the case of periodical outliers.

Proof.

Consider a sequence of functions $f_n(x) = 0$ with the case that $\int_{n=3}^{n=4} f_n(x)dx = 1$. While $\int_{n=0}^{n=\infty} f_n(x)dx = 1$, $\lim_{n\to\infty} f_n(x) = 0$. ∎

The Big 'O' Notation

Many times in mathematics in general, but particularly analytic number theory and probability, statements are phrased in terms of asymptotic notation. This theorem is largely based on observing the behavior of probability functions and their asymptotic natures, and that is where error terms come into play. We use Big O notation to indicate such error terms.

Definition 2.3

For example, if

$$f(x) = O(g(x)),$$

$|f(x)| \leq Cg(x)$ where C is a constant.

Some Proofs

Mobius Inversion

The Mobius inversion formula helps us relate summations that take into account positive integer divisors. This is possible due to the fact that the formulas are multiplicative, allowing us to play around with the numbers. This theorem shows that

$$\sum_{\substack{n \leq x \\ (n,a)=1}} 1 = x \sum_{d|a} \frac{\mu(d)}{d} + O(\tau(a)).$$

Now, we will look at the phi function, using a property discovered by Gauss.

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

Mobius Inversion Pt. 2

Plugging that in, we conclude with

$$\sum_{\substack{n \leq x \\ (n,a)=1}} 1 = x\frac{\phi(a)}{a} + O(\tau(a)),$$

where $\phi(a)$ counts the number of coprime integers up to a and $\tau(a)$ counts the number of factors of a. Recall our probability model:

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

We can achieve a counting function such that

$$\sum_{n \leq (x)} (f(p) - \frac{1}{p}) = \sum_{d|m} (f(p) - \frac{1}{p}) \sum_{\substack{n \leq x \\ (m,n)=d}} 1$$

$$= x\sum_{d|m}(f(p) - \frac{1}{p})\frac{\phi(m/d)}{m} + O(\tau(m)^2).$$

The origin of $\log\log(p)$

We will use the prime number theorem in this proof, which states that $\pi(x) \approx \frac{x}{\log(x)}$, where $\pi(x)$ represents the number of primes less than or equal to x. This approximation gives the asymptotic behavior of the value of $\pi(x)$.

This can be arranged to express P_n, or the nth prime number, as $P_n \approx n\log(n)$.

The origin of $\log\log(p)$ pt. 2

Next, we can express the original sum of counting primes by their reciprocals as:

$$\sum_{p \leq (x)} \frac{1}{p} \approx \int_2^x \frac{1}{n\log(n)}\, dn.$$

We set 2 as the lower boundary because it is lowest positive integer in which $log(n) > 0$. Now, we use $u-$substitution where $u = log(n)$ in order to achieve:

$$\log\log(x) - \log\log(2).$$

Since the value of $\log\log(2)$ is negligible, that can be dropped, thus giving our desired result. This is also known as Mertens' Second Theorem.

The origin of $\log\log(p)$ pt. 3

This was obviously a simplified proof. Mertens' Second Theorem can be more accurately described as

$$\sum_{p\leq(x)} \frac{1}{p} \approx \log\log(x) + c + O\left(\frac{1}{\log(x)}\right).$$

Concluding Remarks

Questions/Thoughts

Some parting questions:

- Would the distribution look different if all prime factors were considered instead of just once each?
- Can we make even closer approximations when we examine these prime numbers and their properties?
- And many more!

Thank you!

Thank you for your attention.