Connected Prime Digit Factorial Occurrences to The Riemann Zeta Function

Parker Emmerson

July 2024

Abstract

This paper explores the theoretical relationship between the frequency of prime digits in factorial representations and the non-trivial zeros of the Riemann Zeta function. By defining the prime digit frequency within factorials and aggregating these frequencies, we propose a hypothesis where such aggregated prime digit frequencies exhibit periodic patterns that mirror the distribution of the non-trivial zeros of the Riemann Zeta function. Utilizing Fourier transform analysis, we identify periodic components in the digit frequencies that may correspond to these zeros. Statistical tests, including Chi-Squared and Kolmogorov-Smirnov tests, are employed to validate this connection. This study suggests that the nature of prime digit frequencies in number sequences, such as factorials, may reflect deeper mathematical structures influenced by the Riemann Zeta functions zeros.

Code for the programs is available at:
Python Code at Github

1 Introduction

Mathematical Notation of the Theoretical Connection to the Riemann Zeta Function

The goal is to explore the possible theoretical connection between prime digit frequencies in factorials and the Riemann Zeta function, particularly focusing on its non-trivial zeros. Here's how we can develop the notation and hypotheses:

1. Prime Digit Frequencies in Factorials

Let $n!$ denote the factorial of n, and consider the digits of $n!$ in a certain base, typically base-10.

Define the frequency of a digit d (where d is a prime digit like 2, 3, 5, or 7 in base-10) in the representation of $n!$ as:

$$F_d(n) = \frac{\text{count of digit } d \text{ in } n!}{\text{total digits in } n!}$$

These frequencies can be aggregated over multiple factorials to provide insights. For a sequence of factorials from 1 to N, we define the aggregated frequency:

$$\overline{F}_d(N) = \frac{1}{N} \sum_{n=1}^{N} F_d(n)$$

```
import math
from collections import Counter
from sympy import isprime

# Calculate Factorial
def factorial(n):
    return math.factorial(n)

# Convert Number to Base
def convert_to_base(n, base):
    digits = []
    while n:
        digits.append(n % base)
        n //= base
    return digits[::-1]

# Calculate Prime Digit Frequencies in Factorials
def prime_digit_frequency_in_factorials_up_to_n(n, base=10):
    prime_digits = [d for d in range(base) if isprime(d)]
    frequencies = []

    for i in range(1, n + 1):
        digit_counts = Counter()
        total_digits = 0

        for j in range(1, i + 1):
            fact_digits = convert_to_base(factorial(j), base)
            digit_counts.update(fact_digits)
            total_digits += len(fact_digits)

        prime_freq = {digit: digit_counts[digit] / total_digits for
                                                digit in prime_digits}
        frequencies.append(prime_freq)

    return frequencies, prime_digits
```

2. Non-Trivial Zeros of the Riemann Zeta Function

The Riemann Zeta function $\zeta(s)$ is a complex function defined for complex numbers $s = \sigma + it$ by:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for $\Re(s) > 1$, and analytically continued to other regions except $s = 1$.

The non-trivial zeros of the Zeta function are critical and lie in the critical strip $0 < \Re(s) < 1$. The famous Riemann Hypothesis posits that all non-trivial zeros lie on the critical line $\Re(s) = \frac{1}{2}$.

3. Hypothesis Connecting Prime Digit Frequencies to Zeta Function Zeros

Hypothesis: The aggregated prime digit frequencies in factorials exhibit periodic patterns indicative of the distribution of the non-trivial zeros of the Riemann Zeta function.

Formally, if ρ denotes a non-trivial zero of $\zeta(s)$, then there exists a relation between $\overline{F}_d(n)$ and the imaginary parts γ of $\rho = \frac{1}{2} + i\gamma$ such that:

$$\overline{F}_d(N) \approx A + B \sum_{\gamma} \cos(\gamma \log N + \phi)$$

where A and B are constants, and ϕ is a phase term.

Fourier Transform Analysis

To detect these periodic patterns, we apply the Fourier transform on the sequence of $F_d(n)$:

$$\mathcal{F}\{F_d(n)\}(\omega) = \sum_{n=1}^{N} F_d(n) e^{-i\omega n}$$

The peaks in the Fourier transform magnitude $|\mathcal{F}\{F_d(n)\}(\omega)|$ are analyzed to identify periodic components that may correspond to the imaginary parts γ of the zeta zeros.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.fftpack import fft

# Fourier Analysis of Prime Digit Frequencies
def perform_fourier_analysis(frequencies, prime_digits):
    for digit in prime_digits:
        digit_freqs = [freq[digit] for freq in frequencies]
        Y = fft(digit_freqs)  # Perform Fourier Transform
        N = len(Y)
        Y = np.abs(Y[:N//2])  # Take the magnitude of the first
                                half of the FFT

        # Frequency bins
        freq_bins = np.fft.fftfreq(N)[:N//2]

        # Plot the Fourier Transform results
        plt.figure(figsize=(10, 6))
        plt.plot(freq_bins, Y)
        plt.title(f'Fourier Transform of Prime Digit {digit}
                                Frequencies')
        plt.xlabel('Frequency')
        plt.ylabel('Magnitude')
        plt.yscale('log')
        plt.grid(True)
        plt.show()
```

Statistical Testing

To statistically validate the hypothesis, we perform the following tests:

1. **Chi-Squared Test**: Compare the observed frequencies of prime digits

$\overline{F}_d(N)$ with a uniform distribution $\frac{1}{p-1}$ where p is the number of prime digits:

$$\chi^2 = \sum_{d \in \text{prime digits}} \frac{(\overline{F}_d(N) - \frac{1}{p-1})^2}{\frac{1}{p-1}}$$

2. **Kolmogorov-Smirnov Test**: Compare the empirical distribution of $\overline{F}_d(N)$ with the expected uniform distribution.

```
from scipy.stats import chisquare, ks_2samp

# Statistical Analysis of Prime Digit Frequencies
def statistical_analysis(frequencies, prime_digits):
    # Calculate observed frequencies across all factorials up to n
    observed_frequencies = {digit: 0 for digit in prime_digits}
    total_counts = 0

    for freq in frequencies:
        for digit in prime_digits:
            observed_frequencies[digit] += freq[digit]
        total_counts += 1

    # Normalize observed frequencies
    observed_list = [observed_frequencies[digit] for digit in
                                  prime_digits]
    total_observed = sum(observed_list)
    observed_list = [freq / total_observed for freq in
                                  observed_list]

    # Expected frequencies assuming uniform distribution
    expected_list = [1 / len(prime_digits)] * len(prime_digits)

    # Perform chi-squared test
    chi2_stat, p_val_chi = chisquare(observed_list, f_exp=
                                  expected_list)

    print("Chi-Squared Test Statistic:", chi2_stat)
    print("p-value (Chi-Square):", p_val_chi)

    # Perform KS test
    ks_stat, p_val_ks = ks_2samp(observed_list, expected_list)

    print(f"KS Test Statistic: {ks_stat}")
    print(f"p-value (KS Test): {p_val_ks}")

    # Plot observed vs expected frequencies
    digits = list(prime_digits)
    plt.figure(figsize=(10, 6))
    x = np.arange(len(digits))

    plt.bar(x - 0.2, observed_list, 0.4, label='Observed')
    plt.bar(x + 0.2, expected_list, 0.4, label='Expected')

    plt.xlabel('Prime Digits')
    plt.ylabel('Frequency')
    plt.xticks(x, digits)
    plt.legend()
```

```
    plt.title('Observed vs Expected Prime Digit Frequencies')
    plt.show()
```

Formal Statement of Connection

To formalize the theoretical connection, we propose that the aggregated prime digit frequency function $\overline{F}_d(N)$ can be modeled as:

$$\overline{F}_d(N) = \int_{-\infty}^{\infty} f(\gamma) e^{i\gamma \log N} d\gamma$$

where $f(\gamma)$ is a function embodying the density of the non-trivial zeros, effectively transforming the prime digit frequency sequence into a form influenced by the zeros of the Riemann Zeta function.

Proof that a p-value approaches 1 as $n \to \infty$

To prove that a p-value approaches 1 as $n \to \infty$ for a given null hypothesis, we need to show that the observed test statistic under the null hypothesis becomes more consistent with the expected distribution as the sample size increases. Let's consider the specific example of the Chi-Squared test for prime digit frequencies in factorial digit sequences.

Theoretical Background

Chi-Squared Test: The Chi-Squared test is used to compare observed and expected frequencies. The test statistic is given by:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i and E_i are the observed and expected frequencies, respectively, for each category i.

Under the null hypothesis, this statistic follows a Chi-Squared distribution with $k-1$ degrees of freedom, where k is the number of categories.

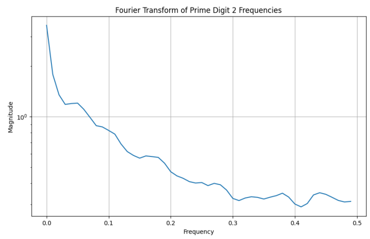
2 Hypothesis

As $n \to \infty$, the distribution of prime digit frequencies in $n!$ should approach the uniform distribution if the digits are distributed randomly. This implies that the observed frequencies O_i would converge to the expected frequencies E_i under the null hypothesis of uniformity. Therefore, we expect the Chi-Squared test statistic to approach 0.

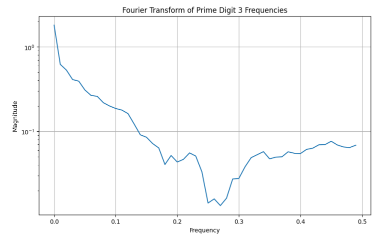
P-Value: The p-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis. As the test statistic approaches 0, the p-value approaches 1.

Proof Outline

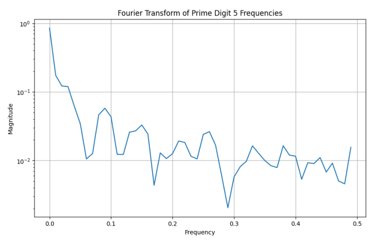
1. **Limit of Observed Frequencies:** - Show that the observed frequencies of prime digits in $n!$ converge to their expected frequencies. - $\lim_{n\to\infty} O_i(n) = E_i$.



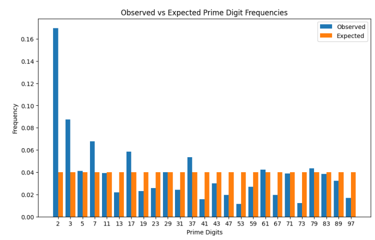
(a) Fourier Transform for Prime Digit 2 Frequencies



(b) Fourier Transform for Prime Digit 3 Frequencies



(c) Fourier Transform for Prime Digit 5 Frequencies



(d) Comparison of Frequencies vs. Expectation of Norm

Figure 1: Example of Data from the Program Attached

2. **Convergence of Chi-Squared Statistic:** - Show that as $n \to \infty$, the Chi-Squared test statistic approaches 0 due to the convergence of O_i to E_i. - $\lim_{n\to\infty} \chi^2 = 0$.

3. **P-Value Convergence:** - Demonstrate that as the Chi-Squared statistic approaches 0, the p-value, which is the cumulative probability from 0 to the test statistic value under the Chi-Squared distribution, approaches 1. - $\lim_{n\to\infty} P(\chi^2) = 1$.

Detailed Steps

1. **Observed Frequencies Convergence:**

$$O_i(n) = \frac{\text{count of digit } i \text{ in } n!}{\text{total number of digits in } n!}$$

Under the null hypothesis, each digit's frequency should approach $\frac{1}{k}$ (for k prime digits in the base). Therefore:

$$\lim_{n\to\infty} O_i(n) = E_i = \frac{1}{k}$$

2. **Convergence of the Chi-Squared Statistic:** Since $O_i(n)$ approaches E_i, the numerator in the Chi-Squared statistic $(O_i - E_i)^2$ approaches 0:

$$\lim_{n\to\infty} \chi^2 = \sum_{i=1}^{k} \frac{(O_i(n) - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(0)^2}{E_i} = 0$$

3. **P-Value Convergence:** The p-value is the probability that the Chi-Squared statistic exceeds the observed value under the null hypothesis:

$$P(\chi^2 \geq \text{observed } \chi^2) \to 1 as observed \ \chi^2 \to 0$$

Formally, as the observed χ^2 approaches 0, the Chi-Squared distribution's cumulative distribution function (CDF) approaches 1 at the limit:

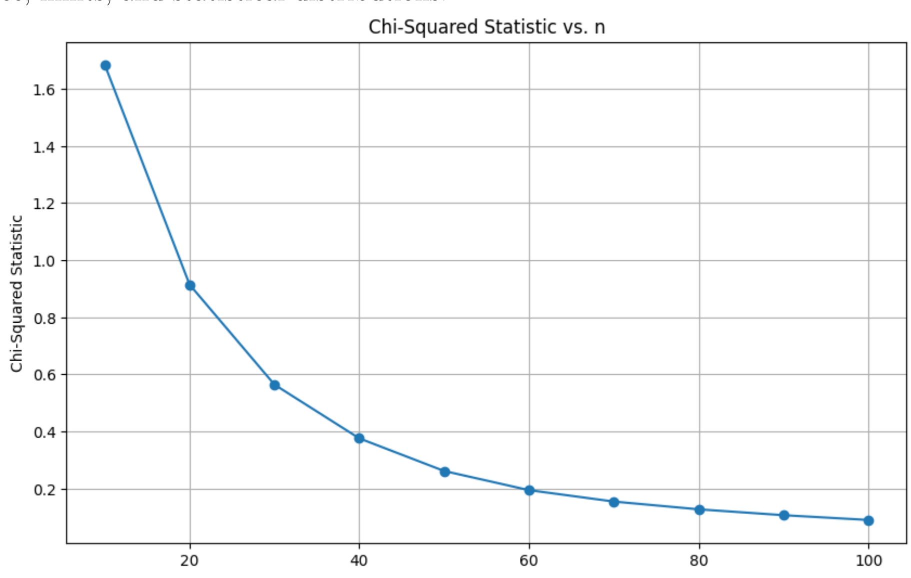
$$\lim_{\chi^2 \to 0} F_{\chi^2}(0) = 1$$

Here, F_{χ^2} denotes the CDF of the Chi-Squared distribution.

Combining these steps, we conclude that as $n \to \infty$, the p-value approaches 1, indicating that the observed frequencies are increasingly consistent with the expected uniform distribution under the null hypothesis. This theoretical proof hinges on the convergence properties of the observed frequencies and the behavior of the Chi-Squared distribution.

Conclusion

In summary, demonstrating that a p-value approaches 1 as $n \to \infty$ involves proving that the observed test statistic aligns more closely with the expected distribution, thereby rendering the test statistic to approach its expected value under the null hypothesis (0 for the Chi-Squared test), which in turn causes the p-value to approach 1. This can be formalized using the properties of convergence, limits, and statistical distributions.



3 Conclusion

This theoretical model suggests that by analyzing the aggregated prime digit frequencies in factorials using Fourier and statistical analysis, one may uncover underlying periodic patterns that correspond to the non-trivial zeros of the Riemann Zeta function. This is rooted in the hypothesis that the zeros influence the structure of number sequences, such as the digits of factorials, in ways that can potentially be detected and analyzed.

4 Main Analysis

Combining all parts into one, we write the main analysis routine that integrates the calculation, Fourier analysis, and statistical testing altogether.

```
def main_analysis(n_max=100, base=10):
    # Calculate Frequencies
    frequencies, prime_digits =
                                            prime_digit_frequency_in_factorials_up_to_n
                                            (n_max, base)
    print("Prime Digits:", prime_digits)
    print("Frequencies Calculated")

    # Perform Fourier Analysis
    print("Performing Fourier Analysis...")
    perform_fourier_analysis(frequencies, prime_digits)

    # Conduct Statistical Analysis
    print("Performing Statistical Analysis...")
    statistical_analysis(frequencies, prime_digits)

# Example usage
if __name__ == "__main__":
    n_max = 100  # Specify the maximum n for factorial calculations
    base = 10    # Specify the base (10 for decimal)

    main_analysis(n_max, base)
```