IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, March 2024

Statistical Study of Euler's Zeta Function by using Riemann Hypothesis

Vijay P Sangale[1] and Govardhan K Sanap[2]
Professor, Department of Mathematics, R. B. Attal College, Georai, Beed, Maharashtra, India[1]
Professor, Department of Mathematics, S. S. Mahavidyalaya, Majalgaon, Beed, Maharashtra, India[2]

Abstract: *The aim of this paper to study the statistical measures for the zeta function and the relative comparison of the function. Here we give an examples of some of the suggested models and compare these models. The development of these models were done in a collection of machine learning algorithms for data mining tasks. As a measure of success of any model, we study the statistical measures CC, MAE, RMSE, RAE and RRSE. We would like to mention that there is no specific rational behind the choice of the statistical measures studied; except that they give good point-wise and overall statistics on the validity and success of the models. The statistical measures Root Mean Square Error and Mean Absolute Error are concerned with the average difference errors.*

Keywords: Zeta function, Statistical measures, Euler's function, Riemann Hypothesis

I. INTRODUCTION

The Euler's zeta function is defined as the infinite sum as the following:

$$\zeta(s) = 1+ \frac{1}{n^s} \qquad \ldots\ldots\ldots(1)$$

Where n is natural number and s is any number greater than 1 [8].Firstwe are going to look the derived results from theEuler's function. Then we will constitute the model which is useful result in meta-querying to access the discovered information from giant database.

Euler first defined the zeta function and then he showed that it has a deep and profound connection with the pattern of the prime numbers. We are giving here the definitions of Zeta function given by mathematician Euler

$$\zeta(s) = л\left(\frac{1}{1-n^s}\right) \qquad \ldots\ldots\ldots (2)$$

Where n are the prime numbers and s is any real number greater than 1.

Then the above stated definition indicates that the Euler's function can be observed as multiple of prime terms.Now we are focusing the result derived from the above study like Euler's discovery to write the following formula

$$(1-x)^{-1} = 1/(1-x)$$
$$= 1+x +x^2 + x^3 +x^4 + \qquad \ldots\ldots\ldots\ldots(3)$$

So, we have,

$$\frac{1}{1-p^s} = 1+ 1/p^s +1/p^{2s} + 1/p^{3s}\ldots\ldots\ldots\ldots (4)$$

In the above equation on the left is a typical term in Euler's infinite product of course, so the above equation provides an infinite sum expression for each term in the infinite product. Then Euler did next was multiply together all of these infinite sums to give an alternative expression for his infinite product. Using the ordinary algebraic rules for multiplying a finite number of finite sums, but applying them this time to an infinite number of infinite sums, one can see that when you write out the right-hand side as a single infinite sum, its terms are all the expressions of the product form[1]. From the point of view of the subsequent development of mathematical models it was not so much the fact that the prime harmonic series has an infinite sum that is important, even though it did provide a completely new proof of Euclid's result that there are infinitely many prime numbers [1, 4]. The above stated Euler's infinite product formula for $\zeta(s)$ marked the beginning of modern number theory.

Lejeune Dirichlet generalized Euler's method to show that in any arithmetic progression $a, a+k, a+2k, a+3k, a+4k, \ldots,$ where a and $k$numbers such that they have no common factors, there are infinitely many prime numbers. Euclid's theorem can be regarded as the special case of this for the arithmetic progression 1, 3, 5, 7, ... of all odd numbers[7].The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, March 2024

principal modification to Euler's method that Dirichlet made was to modify the zeta function so that the primes were separated into separate categories, depending on the remainder they left when divided by *k*.

Statistical Measures:

The Dirichlet L_{+1} series became an essential element in mathematics, and is known as the zeta function. It is formulated like the following:

$$\zeta(s) = \sum \frac{1}{n^s} \quad \ldots\ldots\ldots\ldots (5)$$

Where n is the natural number.

In 1859, in his paper "On the Number of Primes Less Than a Given Quantity", Bernard Riemann relied on analytic continuation to extend Euler's zeta function to the entire complex plane, with a single pole at s = 1 [5].

Riemann did not talk about analytically continuing Euler's zeta function beyond the half-plane ®(s) > 1, but rather talked about finding a valid formula that defines the zeta function for all complex numbers say "s". This differs from the current view of analytic continuation. From here on, by zeta we will mean Riemann's zeta function given by the following equation 6, where $\Pi(s-1)$ is the Gaussian notation for the extended factorial function for all complex numbers with ® (s) > −1.

$$2\,\mathrm{SIN}(\pi s)\,\pi(s\text{-}1)\,\zeta(s) = \int \frac{(-x)^{s-1}}{e^x-1}\,dx \quad \ldots\ldots\ldots. (6)$$

The relation between the zeta function and the primes extends beyond equation 6. The hypothesis due to Riemann regarding the zeros of zeta, and hence known as the Riemann hypothesis. It has many implications about the distribution of the primes, as well as better approximations to various arithmetic functions such as the difference between the prime counting function π(x) and the logarithmic integral Li(x) [8-11].

The zeta function also appears in physics; especially in areas relevant to chaos in classical and quantum mechanics. For instance, one study uses prime numbers to define an abstract numerical gas, and thus uses the zeta function as a thermodynamical partition function [7]. Here in our proposed research we are using to retrieve data and information from complex pattern like special databases and Data ware House.

The reason is that, in contrast to mining in relational databases, spatial data mining algorithms have to consider the neighbors of objects in order to extract useful knowledge. This is necessary because the attributes of the neighbours of some objects of interest may have the significant influence on that objects itself [3].

In this study, we will introducethe general framework for spatial data mining which takes into account that mentioned characteristics of spatial data. We also show that the approach allows a tight and efficient integration of spatial data mining algorithms with spatial database systems.We will further present algorithms for this tasks of spatial clustering, spatial characterization, spatial trend detection and spatial classification utilizing the proposed framework. Furthermore, an example applications are discussed for these algorithms[6]. Here it gives a short summary and shows some directions for future research.

We have listed the following Riemann Hypothesisformulae generated as data mining tools.As a measure of success of any model, we define the statistical measures CC, MAE, RMSE, RAE and RRSE as follows:

$$\text{Correlation Coefficient (CC)} = \frac{(n-1)\sum_{i=1}^{n}(p_i-\bar{p})(a_i-\bar{a})}{[\sum_{i=1}^{n}(p_i-\bar{p})^2][\sum_{i=1}^{n}(a_i-\bar{a})^2]} \quad \text{(R 1)}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n}|p_i-a_i|}{n} \quad \text{(R 2)}$$

$$\text{Root Mean –Squared Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n}(p_i-a_i)^2}{n}} \quad \text{(R 3)}$$

$$\text{Relative Absolute Error (RAE)} = \frac{\sum_{i=1}^{n}|p_i-a_i|}{\sum_{i=1}^{n}|a_i-\bar{a}|} \quad \text{(R 4)}$$

$$\text{Root Relative Squared Error (RRSE)} = \sqrt{\frac{\sum_{i=1}^{n}(p_i-a_i)^2}{\sum_{i=1}^{n}(a_i-\bar{a})^2}} \quad \text{(R 5)}$$

Mathematicians have since been struggling to prove the hypothesis[2-4]. This includes Riemann himself, who admits to having given up, at least temporarily, on a proof of Riemann hypothesis after several unsuccessful attempts.Riemann conjectured that all the non-trivial zeros of zeta lie on this critical line. This is known as the Riemann Hypothesis , considered to be one of the most important open problems in mathematics.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, March 2024

We study the statistical measures for zeta function on the critical line by studying the sequence given by the above definitions[7-11]. All the non-trivial zeros of the zeta function calculated thus far have real part equal to ½, it suffices to only study their imaginary parts.

Statistical Analysis of Zeta Function:

The difference errors relative to the actual values taken for Relative Absolute Error and Root Relative Squared Error give more intuition about statistical measure. Take for example an error of 1% in a calculation where the actual outcome is known to be 100. The approximate answer given is 100 ± 1. On the other hand, say we approximate a function with smaller relative error 0.5%, but are approximating a value that is actually 10^6. The estimated answer would be off by $\pm 10^6 \times 0.5\% = \pm 5 \times 10^3$.

Although the relative error is smaller in the second example, the difference error which is the mean absolute error for data of size one is much smaller in the first case. Therefore, to obtain strong local approximations, we attempt to minimize the mean absolute and the root mean square errors[2,9].

For practicality in the speed of the study, we calculated all our models and their statistical measures of accuracy using the first 10^4 elements of $\{t_n\}_n$, split into two subsets: The first 5×10^3 elements used as training data, while the rest being test data.

Sample of input data for the model of our study say, $n\frac{t_{n+1}}{t_n}$.

n	$n\frac{t_{n+1}}{t_n}$
5	5.706104125321350
50	51.009398160818996
500	500.978044956544980
2000	2001.021407482380000
8000	8000.783730290239000
10000	10000.882909059201000

Table 1: Sample of input data for the model

For each study, we examined the relationship between the proposed model $\{f(t_n)\}_n$ and the number of the instance n. Tables 1 show samples of the training and test instances for two of the proposed models [8-10].Table-1 lists the models that were proposed, their approximate linear models, and their statistical measures of validity[4].

It is clear from Table-2 that the mapping model $n\frac{\ln(t_{n+1})}{\ln(t_n)}$ has the best overall statistical measures, and thus has the best linear approximation. This shows a sample of terms of the sequence $\{t_n\}_n$ along with their corresponding approximations, an + b, calculated using this model, with the sample inputs.

Sample of input data for the model $n\frac{\ln(t_{n+1})}{\ln(t_n)}$.

n	$n\frac{\ln(t_{n+1})}{\ln(t_n)}$
5	5.189007239929700
50	50.201333522284500
500	500.145867031175040
2000	2000.130412290700200
8000	8000.087023205120200
10000	10000.095984553700000

Table 2: Sample of input data for the model

We search for possible models for t_n that are linear in all or some of t_1, t_2, t_3, ... , t_{n-1}. Unfortunately, no good model could be found. This is expected; for it is generally believed that the zeros of the zeta function in particular, and any L-function in general, are linearly independent.

In order to get past this problem, we chose a different approach for a solution. Instead of studying the sequence $\{t_n\}_n$, we map the sequence to another sequence $\{f(t_n)\}_n$, and then study thisnew sequence for possible linear patterns.

The Model	Generated Model (an + b)	CC	MAE	RMSE	RAE(%)	RRSE(%)
$n\frac{t_{n+1}}{t_n}$	n + 0.79	1	0.279	0.344	0.006	0.007
$ne\left(\frac{t_{n+1}}{t_n}\right)$	2.72n + 2.15	1	0.757	0.935	0.006	0.007
$n\pi\left(\frac{t_{n+1}}{t_n}\right)$	3.14n + 2.84	1	0.001	1.237	0.006	0.008
$n\frac{\ln(t_{n+1})}{\ln(t_n)}$	n + 0.12	1	0.032	0.041	0.001	0.001
$n\ln\left(\frac{\ln(t_{n+1})}{\ln(t_n)}\right)$	0.12	0.04	0.032	0.041	98.492	103.552
$n\ln\left(\frac{t_{n+1}}{t_n}\right)$	0.79	0.01	0.28	0.344	102.886	100.918

Table 3: various models and statistical measures of their accuracy

Using data mining techniques, we studied the first 10^5 elements of $\{t_n\}_n$, and have determined that the function $n\frac{\ln(t_{n+1})}{\ln(t_n)}$ is "almost" linear in n. That is, for certain real constants a and b [2, 6]. The concern that immediately arises is the best value(s) for the constants a and b. Random actual values from {tn} n and their approximations given by the model $n\frac{\ln(t_{n+1})}{\ln(t_n)}$ choice of b = .12244.

The remainder of the study is devoted to this model since it has the best performance.

Actual Value	**Approximation**	**Difference Error**
21.02204	20.75236	+0.26968
30.42488	29.22134	+1.20354
37.58618	36.44748	+1.13870
4819.028	4820.397	−1.36900
4824.678	4826.047	−1.36900
45186.26	45187.50	−1.24000
51317.06	51318.29	−1.23000
51321.28	51322.51	−1.23000

Table 4: Difference error for best performance model

Here identified the satisfactory model for the sequence $\{t_n\}_n$, we turn to this mathematical analysis process of our study. We can validates these models mathematically also.

II. CONCLUSION

The Relative Absolute Error and Root Relative Squared Error statistics give more intuition about the difference errors relative to the actual values studied. The latter statistics might be misleading. Statistical measures applies to both the efficiency of algorithms as well to the complexity of possible patterns that can be found in a spatial database.This gives

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, March 2024

us the interesting idea to utilize the defined measures and Riemann hypothesis based Zeta function in doing data mining.

REFERENCES

[1] Edwards H. M. Riemann's Zeta Function. Academic Press. ISBN 0-486-41740-9.
[2] V P Sangale, G K Sanap, A study of properties of Euler's Zeta function by using Riemann Hypothesis, IJARSET, Vol. 10, Issue 5, May 2023
[3]. Hutchinson, J. I. On the Roots of the Riemann Zeta Function. Trans-actions of the American Mathematical Society. Vol. 27. No. 1. Jan.1925, 49-60.
[4]. Wang F. T. ,A Formula on Riemann Zeta Function. The Annals of Mathematics, 2nd Series. Vol. 46. No. 1. Jan. 1945, 88-92.
[5].Apostol, T. M. Some Series Involving the Riemann Zeta Function. Proceedings of the American Mathematical Society. Vol. 5. No. 2. Apr.1954, 239-243.
[6]. Hochstadt, H. Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable. SIAM Review. Vol. 22.No. 3. Jul. 1980, 378.
[7]. Walter Rudin, Functional Analysis, second edition, McGraw-Hill (1991).
[8]. V P Sangale, A Critical Study of Zeta Function and Riemann Hypothesis from Various Fields of Mathematics, Multidisciplinary Research, Dec.2020, 127-135
[9]. Titchmarsh E. C. (1986) Heath-Brown (ed). The Theory of the Riemann Zeta Function (2nd rev. ed.) Oxford University Press.
[10]. Choudhury, B. K. The Riemann Zeta-Function and Its Derivatives. Proceedings: Mathematical and Physical Sciences. Vol. 450. No.1940. Sep. 1995, 477-499.
[11]. Titchmarsh, E. C. The Theory of the Riemann Zeta-Function. Second Edition. Oxford Science Publications. 1986.