NUMBER THEORY IN STATISTICS: EXPLORING PRIME NUMBERS AND DIOPHANTINE EQUATIONS

Poonam Devi*

TGT Maths, Swami Ganeshananad Sanatan Dharam Middle School Uchana Mandi (Jind), Haryana, India

Email ID: poonamsharawat9@gmail.com

Accepted: 07.05.2023 **Published**: 01.06.2023

Keywords: Number Theory, Statistics, Prime Numbers, Diophantine Equations.

Abstract

The field of statistics has seen a number of exciting uses of number theory throughout the years. This survey paper's goal is to draw attention to certain particularly significant examples of uses of this kind. The study of prime numbers is a component of number theory that is both fascinating and difficult to investigate. Equations based on the diophantine approximation are at the heart of number theory. A Diophantine equation is a kind of problem that can only be solved by integrating the given numbers. In the first section of this article, we will talk about various issues that are associated with prime numbers as well as the function of Diophantine equations in Design Theory. A quasi-residual Metis design has its Fibonacci and Lucas numbers discussed, as have their contributions to the design. The Discrete Logarithm problem is a well-known issue that arises in the context of finite fields. The structure of the Discrete Logarithm is dissected in detail in the next section of this article.

Paper Identification

**Corresponding Author*

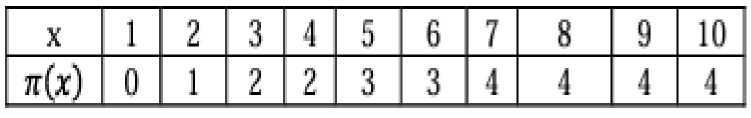
1. Introduction

Due to its ancient origins, number theory comprises a diverse range of subfields that are now being investigated. Strategies from other disciplines may be useful in addressing the issues posed by number theory, and conversely, number theory may offer insights for other fields of study. This article's objective is to draw attention to the necessity of adopting an interdisciplinary strategy to research, especially when investigating the relationships that exist between number theory and statistics. This article will go into concrete instances to illustrate how number theory may be utilized in the subject of statistics, highlighting the strong interaction that exists between the two academic fields of number theory and statistics.

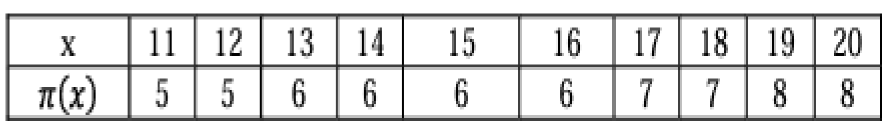
2. Prime Number

P. Ribenboim is the person to consult for further information on prime numbers. The behavior of prime numbers is the aspect of integers that causes the greatest confusion. In spite of the greatest efforts made by a variety of academics, the many features of prime numbers continue to present insurmountable challenges in terms of their ability to be understood. This is because prime numbers have a wide variety of features, not all of which are the same. A intriguing field of inquiry is the study of the distribution of prime numbers.

Let π(X) signify the number of primes that are less than or equal to x. The following is the table of values that we have:



x	1	2	3	4	5	6	7	8	9	10
$\pi(x)$	0	1	2	2	3	3	4	4	4	4



x	11	12	13	14	15	16	17	18	19	20
$\pi(x)$	5	5	6	6	6	6	7	7	8	8

Let's say that p_n stands for the nth prime. Taking into consideration this note, we have

$$\pi(p_n) = n \qquad 1$$

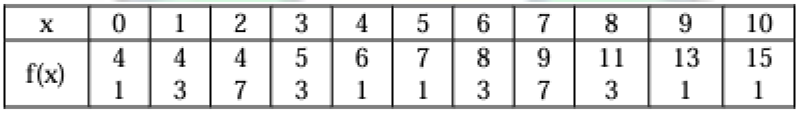
The following are some findings about primes that are widely known:

- An example of a prime number theorem states that as x rises, the number of primes not greater than x will tend to approach the value of x divided by the natural logarithm of x, written as $\frac{x}{\log x}$.
- The theorem of Tchebychef states that the order of magnitude of (x) is equal to $\frac{x}{\log x}$.

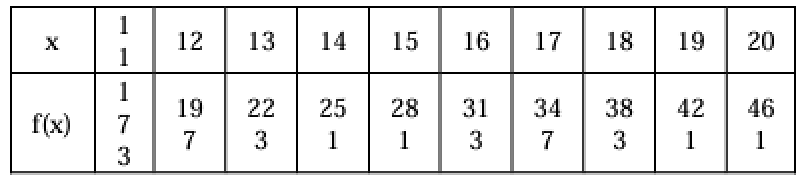
The distribution of the prime pairs p, p+2 is an intriguing subject.

3. The Polynomial of Euler

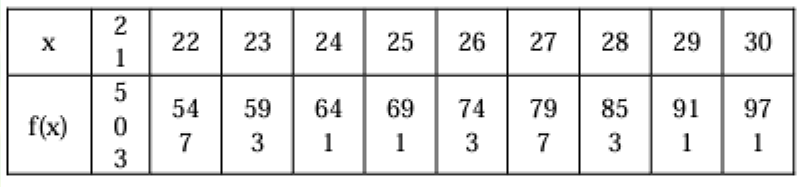
Researchers have undertaken many endeavors to ascertain polynomials that exclusively yield prime numbers as their output. “Leonhard Euler (1707–1783) examined the polynomial f(x) = x^2 + x + 41, assuming that x can only be whole numbers”. Remarkably, this polynomial exclusively accommodates whole numbers for a specific range of consecutive whole number values of x, commencing at 0, as evidenced by the subsequent tables:



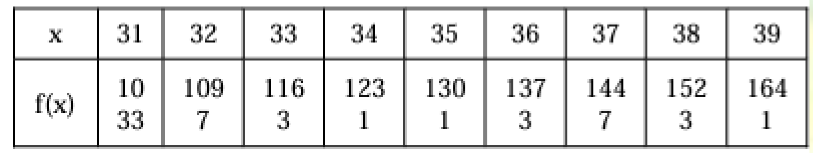
x	0	1	2	3	4	5	6	7	8	9	10
f(x)	41	43	47	53	61	71	83	97	113	131	151



x	11	12	13	14	15	16	17	18	19	20
f(x)	173	197	223	251	281	313	347	383	421	461



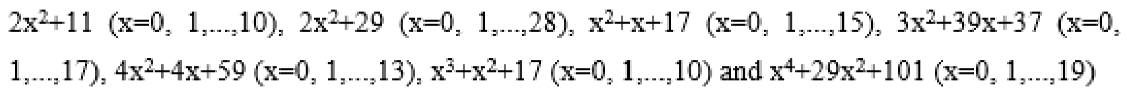
x	21	22	23	24	25	26	27	28	29	30
f(x)	503	547	593	641	691	743	797	853	911	971



x	31	32	33	34	35	36	37	38	39
f(x)	1033	1097	1163	1231	1301	1373	1447	1523	1641

However, when x equals 40, we find that f(x) is 40^2 + 40 + 41 = 40 (40 + 1) + 41 = 41^2, which is a composite value for f(x). It may be shown with the use of Euler's polynomial that there is no such thing as a polynomial that exclusively takes prime values.

Like Euler's polynomial, all subsequent polynomials take the special case when successive values of x within the parenthesis are prime numbers.



$2x^2+11$ (x=0, 1,...,10), $2x^2+29$ (x=0, 1,...,28), x^2+x+17 (x=0, 1,...,15), $3x^2+39x+37$ (x=0, 1,...,17), $4x^2+4x+59$ (x=0, 1,...,13), x^3+x^2+17 (x=0, 1,...,10) and x^4+29x^2+101 (x=0, 1,...,19)

The analysis of observational data has shown that, for every natural number x, there is no non-constant polynomial, denoted as f(x), with integral coefficients that stays prime for all values of x or for values of x that are sufficiently big. This conclusion was reached in light of the fact that there is no constant polynomial with integral coefficients that remains prime. You may find the writings of G.H. Hardy and E.M. Wright helpful if you're looking for additional in-depth information on this subject.

It is important to note that determining probability estimates for the successive prime (or composite) values taken on by Euler's polynomial when x is larger than 39 or for the other polynomials indicated when x reaches the required integral threshold would be an interesting task. This would be the case, as we have observed, since it would be a task that would need us to find out probabilistic estimates of these values.

- **Unresolved Prime Number Problems**

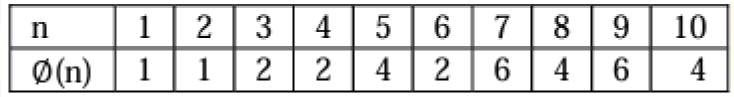
  Is there a limit to the number of primes that may be generated by the polynomial $f(x)=x^2+1$?

- Is there always a prime number that can be found between the x^2 and the $(x+1)^2$?

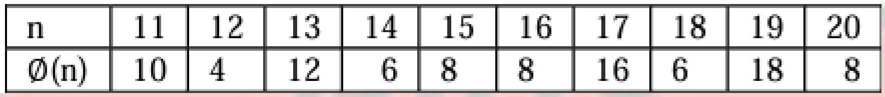
It might be useful to attempt solving the challenges described above using a probabilistic method.

4. An Issue Pertaining to EULER'S ARITHMETIC Function

Let n be a positive natural integer greater than of one. The number of positive integers that are less than and prime to n is what the -function of Euler is going to identify with n. By custom, the value of φ (n) is assumed to be 1. The following is the table of values that we have:



n	1	2	3	4	5	6	7	8	9	10
Ø(n)	1	1	2	2	4	2	6	4	6	4



n	11	12	13	14	15	16	17	18	19	20
Ø(n)	10	4	12	6	8	8	16	6	18	8

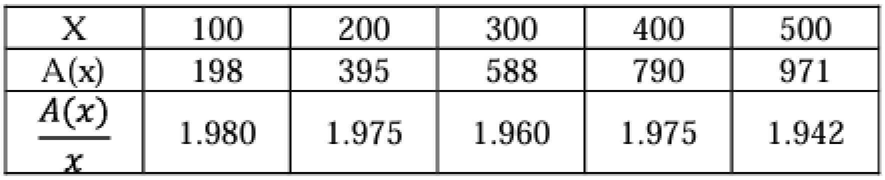
If we examine the prime factorization of the statement $n = p^{\alpha} q^{\beta}$, where p and q are both distinct primes, then we can see that the expression n=p q has the following prime factorization:

$$\emptyset(n) = n\left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right) \ldots. \qquad (2)$$

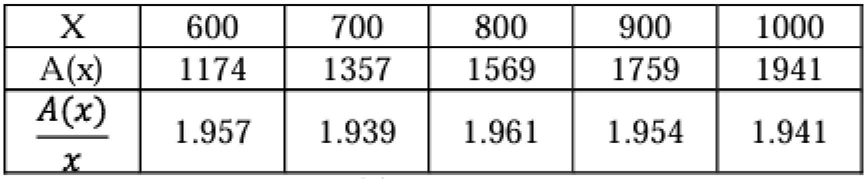
P.T. Bateman conducted an investigation into the study of the distribution of values connected to Euler's φ -function. He began his inquiry by defining the function by using "am" as the count of positive integers "n" for which φ (n) = "m." This was the first step in his study.

$$\mathrm{A(x)} = \sum_{m \leq x} a_m. \qquad (3)$$

That is to say, the value of A(x) is the number of positive integers n such that φ (n) ≤ x. He thought of the function written as (A(x))/x. He was successful in gathering the data listed below:



X	100	200	300	400	500
A(x)	198	395	588	790	971
$\frac{A(x)}{x}$	1.980	1.975	1.960	1.975	1.942



X	600	700	800	900	1000
A(x)	1174	1357	1569	1759	1941
$\frac{A(x)}{x}$	1.957	1.939	1.961	1.954	1.941

He hypothesized that there is a finite limit of 1.9435964 for (A(x))/x as the value of x increases. In addition, he proposed numerous methods to acquire estimates for the error component in (A(x))/x.

5. Diophantine Equations

A Diophantine equation is a type of problem that can only be resolved by the process of integrating the provided numerical values. Diophantus of Alexandria was inspired to develop the concept of Diophantine equations as a result of his fascination with integral solutions to algebraic equations. In the world of number theory, these equations are regarded as essential aspects of the field. L.J. Mordell is widely recognized as the definitive authority on diophantine equations.

A. Square- Free Natural Number

A natural number n is said to be "square-free" if it cannot be divided into squares by any integers bigger than 1. Because of this, n does not include any square roots if and only if it is the product of many different primes.

Determining the chance that a certain natural number, n, does not have a square root associated with it is an intriguing topic. Gauss made the observation that the likelihood that two integers should be relatively prime is $6/\pi^2$, and that this probability cannot be changed. The chance that a number does not include any squares is equal to $6/\pi^2$.

B. Pell's Equation

Let's say that D is a natural number that doesn't include any squares. This is the equation.

$$x^2 - Dy^2 = 1 \tag{4}$$

is often referred to as Pell's equation. This equation always has integer solutions for both "x" and "y," and the total number of possible solutions is unlimited. This is true for any natural number "d" that is not a perfect square. Other generalizations of Pell's equation include the following:

$$x^2 - Dy^2 = -1 \text{ and} \qquad (5)$$

$$x^2 - Dy^2 = N \qquad (6)$$

where N is a positive integer that is not zero. These generic formulas may or might not provide integral solutions for a given value of "N" or a square-free "D." Having said that, this cannot be considered a conclusive finding. It is interesting to note that Pell's equation for a certain "D" number is tied to a design, as touched on in the article that came after it. This makes for an unusual observation.

6. Design Theory

Design theory is an essential part of the discipline of statistics. One way to think about a design is as a point in the region R^5. The parameters that are connected with a design are expressed as a quintuple (v, b, r, k, and λ), which may be disassembled into the following basic parts:

This collection will be referred to as "V," and it will have precisely "v" different components. When we talk about blocks, we are really talking about subsets of V, and one kind of block is called a "b" block. It is assumed that each component of V is partitioned into "r" blocks, with "r" being smaller than "b." The letter "r" denotes the process of designing several identical copies of the same thing. There are "k" distinct kinds of veggies included inside each individual block. There is a presumption that each and every pair of components from V appears together in blocks of size " λ," where " λ " is less than "b." The word "co-valency" denotes the numerical value inside the design. The following are some valid relationships that may be drawn from the design's parameters:

$$vr = bk \qquad (7)$$

$$\lambda(v - 1) = r(k - 1) \qquad (8)$$

In the next section, we are going to talk about how design may benefit from number theory. In order to achieve this objective, we will look at a specific category of design.

A. Metis Design

By a Metis design we mean a block design with parameter set (v, b, r, k and λ) satisfying the additional relation

$$v = r + k + 1 \qquad (9)$$

B. Quasi- Residual Metis Design

The added attribute may be said to belong to a quasi-residual Metis design.

$$r = k + \lambda \tag{10}$$

Take a look at equations (7) through (10) here. We may determine that λ = r- k by using equation (10). When we apply this to equation (8), we get that rv - kv + K = Kr. When we use this substitution for v in equation (9), we get the connection

$$k^2 + kr = r^2 + r \tag{11}$$

When we consider equation (11) to be a quadratic in k, we arrive at the following relation:

$$k = \frac{-r \pm \sqrt{(5r^2 + 4r)}}{2}$$

Since k cannot accept values that are less than zero, we obtain

$$k = \frac{\sqrt{(5r^2 + 4r)} - r}{2} \tag{12}$$

In order for k to take on values that are integral, one of the conditions that must be met is for $5r^2+4r$ to be the square of a natural integer. Let's say that the greatest common divisor of $5r^2+4$ and r is denoted by the letter g. After that, g/ 4. This suggests that the ratio of $(5r^2+4r)/ g^2$ is a square. Because of this, both (5r+4)/g and r/g should be considered perfect squares. When modulo 4 is taken into consideration, it is clear that g cannot have the value of 2. So g must equal either 1 or 4. In all scenarios, 5r+4 and r are considered to be squares. As a result, there must be x and y natural integers in such a way that

$$5r + 4 = x^2 \text{ and} \tag{13}$$

$$r = y^2 \tag{14}$$

Since x and y are connected, we get the following formula:

$$x^2 - 5y^2 = 4 \tag{15}$$

Equation (15) is Pell's equation, written as $x^2 – Dy^2 = N$, with D equal to 5 and N equal to 4. Therefore, there is a connection between the Pell's equation and a quasi-residual Metis design.

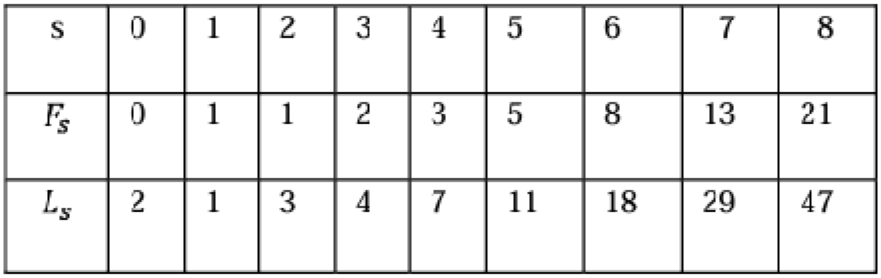
C. Relationship with Fibonacci and Lucas Numbers

Following is a recursive definition of the Fibonacci numbers, denoted by the letters (F_s), and the Lucas numbers, denoted by the letters (L_s) (see, for example, G.H.Hardy and E.M.Right).

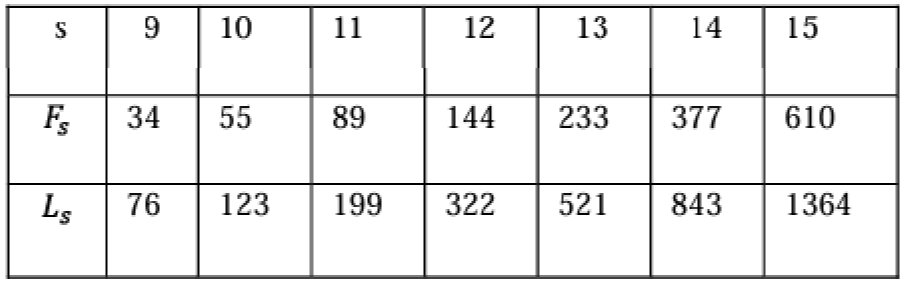
$$F_0 = 0, F_1 = 1 \; and \; F_{s+2} = F_{s+1} + F_s, \tag{16}$$

$$L = 0, L_1 = 1 \; and \; L_{s+2} = L_{s+2} + L_s \tag{17}$$

The following table contains the first few numbers based on the Fibonacci and Lucas sequences:



s	0	1	2	3	4	5	6	7	8
F_s	0	1	1	2	3	5	8	13	21
L_s	2	1	3	4	7	11	18	29	47



s	9	10	11	12	13	14	15
F_s	34	55	89	144	233	377	610
L_s	76	123	199	322	521	843	1364

One can see that the following trait is shared by the succeeding pairs of Fibonacci and Lucas numbers:

$$2^2 - 5.0^2 = 4, 1^2 - 5.1^2 = -4,$$
$$3^2 - 5.1^2 = 4, 4^2 - 5.2^2 = -4 \, etc.$$

These particular outcomes compel us to experiment with an induction strategy in order to get a general conclusion. We are able to deduce from this that

$$L_{2s}^2 - 5F_{2s}^2 \text{ and } L_{2s+1}^2 - 5F_{2s+1}^2 = -4$$

Therefore, the even subscripted terms that correspond to the Pell's equation (15) are satisfied by the Lucas and Fibonacci sequences, and as a consequence, these sequences result in a quasi-residual Metis design. Given these findings, the parameters of a quasi-residual Metis design may be expressed in terms of the Lucas numbers and the Fibonacci numbers as follows:

$$v = F_{2s}F_{2s+1} + 1, b = F_{2s}F_{2s+2}, r = F_{2s}^2,$$
$$\mathrm{k} = F_{2s-1}F_{2s} \text{ and } \lambda = F_{2s-2}F_{2s.}$$

7. The Problem of DISCRETE LOGARITHM

Allow p to represent an odd prime. In order to solve the discrete logarithm issue, you must discover x = $\log_b$ (y) in the finite field Z_p. Another way to put this is to locate the value or values of x in Z_p such that $b^x \equiv y \pmod p$. There is presently no algorithm that can solve this issue that has been developed. The practice of cryptography, which entails the skill of sending communications in a secret way to protect the secrecy of information, is one area in which this topic is put to use in a practical setting. D.Cloutier and J.Holden have given some thought to the issue of mapping the discrete logarithm. A. Hoffman has done research on the structure that may be found in the discrete logarithm.

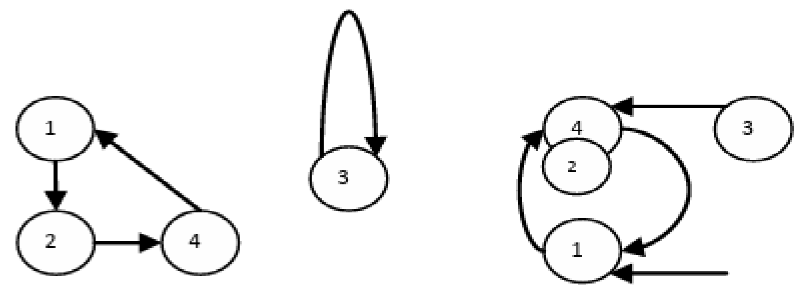
One way to look at the discrete logarithm is as a kind of function. The task at hand is to compute the inverse of the expression x → bx (mod p). It's possible that a tool like a functional graph may help with this issue. A graph's nodes may be used to represent the values of x, and arrows can be constructed to connect each mapping to its respective node. A functional graph is a directed graph in which every vertex must have precisely one edge leading out from it in order for the graph to be considered functional. A functional graph is said to be m-ary if each node has an in-degree that is either precisely zero or m.

Let us take a look at a few different scenarios to better understand the process that is involved. Consider the consecutive integral powers of 2 and reduce them modulo 5 to get the functional graph of 2 (mod 5).

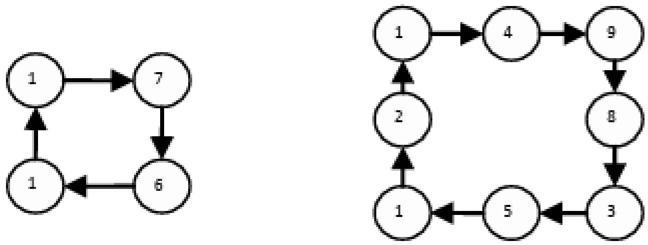
$2^1 \equiv 2 \pmod 5$, $2^2 \equiv 4 \pmod 5$, $2^3 \equiv 3 \pmod 5$, $2^4 \equiv 1 \pmod 5$. The forward correspondence is found by considering the exponent and the result of lowering modulo 5.

$$1 \to 2, 2 \to 4, 3 \to 3, 4 \to 1$$

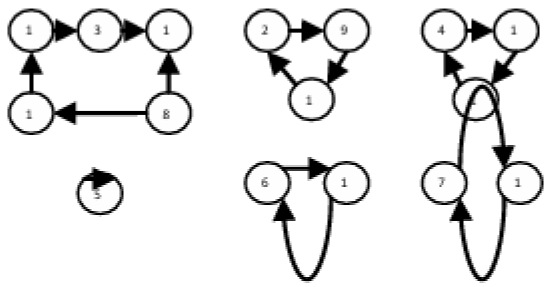
From this correspondence, we separate the cycles and get 1→2, 2→ 4, 4→1 and 3→ 3. A directed graph is used to illustrate each individual cycle. The functional graph for this instance, along with a few additional instances, can be seen further down the page.



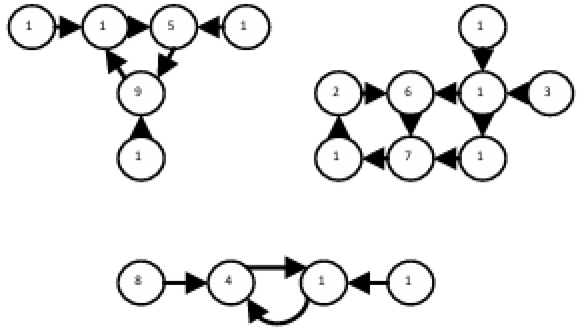
Functional graph for 2 (mod 5) Functional graph for
4(mod 5)



Functional graph for 7 (mod 13)



Functional graph for 3 (mod 17)



Functional graph for 5 (mod 19)

The nodes 2 and 3 in “the functional graph for 4 (mod 5) are not connected to any of the cycles in the graph. Taking into consideration a functional network in which every node represents a component of a cycle is an interesting exercise. Regarding this topic, the following are requirements:

Let r be an element of Z_p^*. Let e be “the smallest natural number such that r^e= 1 (mod p). We say that r is a primitive root modulo p if . e= φ(p). Let r be any primitive root modulo p and g $g \equiv r^{\alpha}$ (mod p). D. Cloutier and J. Holden have demonstrated that the values of "g" that yield an m-ary graph are exactly those for which the greatest common divisor of "α" and "p-1" is equal to "m."

A. Hoffman has selected b as a primitive root modulo p in the issue of discrete logarithm and has evaluated three parameters related with a functional graph. These parameters are the number of cycles, the maximum cycle length, and the weighted average cycle length. The number of cycles, the longest possible cycle, and the weighted average cycle length are the parameters that are being discussed here. He has shown that the structure of discrete logarithm may be broken down into its component parts by doing statistical study on the three components that were discussed before. He showed that it is possible to make comparisons between random permutations and those derived from the solution to the discrete logarithm problem by taking into consideration the expected values of the three parameters in both sets of conditions. He did this in order to illustrate that it is possible to make comparisons between the two types of permutations. This was done with the intention of demonstrating the plausibility of making comparisons.

As the distribution of cycle lengths follows the Poisson distribution, it has been shown how ANOVA tests can be carried out to determine the mean number of cycle components, the number of components variance, the mean maximum cycle length, the maximum cycle variance, the mean average cycle length, and the average cycle variance. All of these statistics can be used to determine the mean maximum cycle length, the maximum cycle variance, the mean average cycle length, and the average cycle variance. He was able to compute the statistical results for the three parameters of the functional graphs concerning the primes in order to emphasize the structure in the discrete logarithm by making use of the t-test and the Anderson-Darling test". In order to do this, he started by picking 30 prime numbers ranging from 99991 to 106921, and then he chose the prime numbers at random.

8. Conclusion

A few of the connections that exist between Number Theory and Statistics have been presented up to this point in the conversation. The applications of number theory in statistics and statistics in number theory both have a lot of room to be explored further. Research into the distribution of prime numbers is a difficult topic to investigate. When there are many factors in a design, the analysis of that design may get fairly complicated, which is why one has to have a higher level of computing expertise. High-end computer power is required to accomplish tasks such as comprehending the characteristics of primes and finding a solution to a discrete logarithm issue via the use of functional graphs. With the already available computing powers as a consequence of the progression of technology, the work that will be done in the future offers promise, and one may anticipate achieving concrete results in this fascinating area of study.

References

1. P.T. Bateman, "The distribution of values of Euler's ø-function", Acta Arith., Volume 21, Pp. 329 – 345, 1972
2. D. Cloutier and J. Holden, "Mapping the discrete logarithm", Involve, Volume 3, Issue 2, Pp. 197 – 213, 2010
3. G.H. Hardy and E.M. Wright, "An introduction to the theory of numbers", Oxford University Press, London, 1975.
4. Hoffman, "Statistical investigation of structure in the discrete logarithm", Rose-Hulman Undergraduate Mathematics Journal, Volume10, Issue 2, Pp. 1 – 20, 2009

5. L.J. Mordell, “Diophantine equations”, Academic Press, London, 1969 [6] A.M. Odlyzko, “Discrete logarithms in finite fields and their cryptographic significance” in “Advances in Cryptology: Proceedings of a EUROCRYPT 84”, Lecture Notes in Computer Science, Springer-Verlag, Volume 209, Pp. 242 – 314, 1985
6. P. Ribenboim, “The new book of prime number records”, Springer Verlag, New York, 1996.

.