Applications of Number Theory in Statistics

Mrs. Gurinder Kaur
PG Department of Mathematics
RR Bawa DAV College for Girls Batala,
Gurdaspur, Punjab – 143 505, INDIA

Abstract--- *There have been several fascinating applications of Number Theory in Statistics. The purpose of this survey paper is to highlight certain important such applications. Prime numbers constitute an interesting and challenging area of research in number theory. Diophantine equations form the central part of number theory. An equation requiring integral solutions is called a Diophantine equation. In the first part of this paper, some problems related to prime numbers and the role of Diophantine equations in Design Theory is discussed. The contribution of Fibonacci and Lucas numbers to a quasi-residual Metis design is explained. A famous problem related to finite fields is the Discrete Logarithm problem. In the second part of this paper, the structure of Discrete Logarithm is discussed.*

Keywords--- *Distribution of Primes, Diophantine Equations, Design, Fibonacci and Lucas Numbers, Discrete logarithm Problem*

INTRODUCTION

NUMBER theory is perhaps the oldest branch of Mathematics and consequently there are several research areas within the realm of number theory. Techniques from other branches of knowledge may prove handy in solving some of the problems in number theory and vice versa. The aim of this paper is to stress the importance of inter-disciplinary approach in research, especially the linkages between number theory and Statistics. Certain specific problems are discussed to illustrate the applications of number theory in Statistics and to bring out the scope of inter-dependence of the two subjects.

PRIME NUMBERS

For detailed account of prime numbers, one may refer P.Ribenboim. The most perplexing behaviour of integers is that of prime numbers. In spite of the best efforts put in by different researchers, understanding the several properties of prime numbers continues to pose insurmountable difficulties. This is due to the variations in the properties possessed by prime numbers. The distribution of primes is a fascinating area of research.

Let $\pi(X)$ denotes the number of prime's not exceeding x. We have the following table of values:

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 p_n denote the n[th] prime. With this notation, we have

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

The following are well known results on primes:

- Prime number theorem: The number of prime's not exceeding x is asymptotic to $\frac{x}{\log x}$.
- Tchebychef's theorem: The order of magnitude of $\pi(x)$ is $\frac{x}{\log x}$.

An interesting question is to find how the prime pair's p, p+2 are distributed.

THE POLYNOMIAL OF EULER

There have been several attempts by researchers to find out polynomials which would yield prime numbers only. Leonhard Euler (1707-1783) considered the polynomial $f(x) = x^2 + x + 41$ where x is presumed to take integral values only. Surprisingly, this polynomial takes integral values only for several consecutive integral values of x, starting from 0, as shown in the following 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	1601

However, when x = 40, we have f(x) = 40^2 +40+41= 40(40+1)+41 = 41^2, associating a composite value for f(x). Euler's polynomial is an example to show that there cannot be a polynomial taking prime values only.

Similar to Euler's polynomial, the following polynomials also assume prime values only for the consecutive values of x provided within parentheses.

$2x^2$+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).

When x is a natural number, it has been proved that no polynomial f(x) with integral coefficients, not a constant, can be prime for all x, or for sufficiently large x (see for e.g. G.H. Hardy and E.M. Right).

We observe that it would be an interesting problem to find out probabilistic estimates of consecutive prime (or composite) values assumed by Euler's polynomial for x > 39 or the other polynomials specified above for x exceeding the specified integral value.

➢ ***Some Unsolved Problems Pertaining To Primes***

- Are there infinitely many primes given by the polynomial $f(x) = x^2 + 1$?
- Is there always a prime between x^2 and$(x + 1)^2$?

It is worthwhile to try the above problems with probabilistic approach.

A PROBLEM RELATED TO EULER'S ARITHMETIC FUNCTION

Let n be a given natural number> 1. Euler's Ø-function associates with n the number of positive integers less than and prime to n. By convention, $\emptyset(n)$ is taken as 1. We have the following table of values:

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

Consider the prime factorization of n. If $n = p^{\alpha}q^{\beta}$ ... where p, q ... are distinct primes, then

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

P.T. Bateman [1] considered the distribution of values of Euler's **Ø**-function. He took a_m as the number of positive integers n with $\emptyset(n) = m$ and defined the function

$$A(x) = \sum_{m \le x} a_m. \quad (3)$$

i.e., $A(x)$ is the number of positive integer's n with $\emptyset(n) \le x$. He considered the function $\frac{A(x)}{x}$. The following values were obtained by him:

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 conjectured that $\frac{A(x)}{x}$ has a finite limit of 1.9435964... as$x \to \infty$. In [1], he has presented several techniques to obtain the estimates for the error term in $\frac{A(x)}{x}$

DIOPHANTINE EQUATIONS

An equation requiring integral solutions is called a Diophantine equation. Diophantus of Alexandria was interested in the integral solutions of algebraic equations and hence the nomenclature of Diophantine equations. These equations form the central part of number theory. A standard reference for Diophantine equations is L.J. Mordell.

➢ ***Square-Free Natural Number***

A natural number n is said to be square-free if it is not divisible by the square of a number > 1. Therefore n is square-free if and only if it is the product of distinct primes.

An interesting problem is to determine the probability that a given natural number n is square-free. Gauss observed that the probability that two integers should be relatively prime is $\frac{6}{\pi^2}$The probability that a number should be square-free is $\frac{6}{\pi^2}$ (see for e.g. G.H. Hardy and E.M. Right).

➢ ***Pell's Equation***

Let D be a given square-free natural number. The equation

$$x^2 - Dy^2 = 1 \quad (4)$$

is known as Pell's equation. For a given square-free natural number d, this equation always has integer solutions in x and y and the number of solutions is infinite. Other general forms of Pell's equation are

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

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

where N is a non-zero integer. These general forms may not possess integral solutions for a given N or a square-free D. It is of interest to note that Pell's equation for a special value of D is related to a Design as brought out in the sequel.

DESIGN THEORY

An important branch of Statistics is Design Theory. A design can be thought of as a point in R^5. The parameters associated with a design form a quintuple (v, b, r, k, λ) as descrfibed below:

Let V denote a finite set consisting of v elements. By a block we mean a subset of V. We consider b blocks. It is assumed that each element of V is in r blocks where r ≤ b. We refer to r as the replication number of the design. Let k denote the number of varieties in each block. It is assumed that every pair elements of V appears together in λ blocks where λ ≤ b. The number λ is called the co-valency for the design. The following relations hold for the parameters of the design:

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

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

The contribution of number theory to designs will be considered in the sequel. Towards this purpose, we consider a special type of a design.

A. Metis Design

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

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

B. Quasi-Residual Metis Design

A quasi-residual Metis design has the additional property

$$r = k + \lambda \quad (10)$$

Consider equations (7) through (10). From (10) we have $\lambda = r - k$. Using this in (8), we get $rv - kv + k = kr$. Substituting for v from (9), we obtain the relation

$$k^2 + kr = r^2 + r \quad (11)$$

Treating (11) as a quadratic equation in k, we are led to the relation

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

Since k cannot take negative values, we get

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

In order that k assumes integral values, a necessary condition is that $5r^2 + 4r$ is the square of a natural number. Let g denote the greatest common divisor of 5r+4 and r. Then g/4. This implies that $\frac{5r^2+4r}{g^2}$ is a square. Hence each one of $\frac{5r+4}{g}$ and $\frac{r}{g}$ shall be perfect squares. Considering modulo 4, it is seen that g cannot take the value of 2. Hence g = 1 or 4. In either case 5r+4 and r are both squares. Therefore there exist natural numbers x and y such that

$$5r + 4 = x^2 \text{ and} \quad (13)$$

$$r = y^2 \quad (14)$$

Thus we see that x and y are related by the following equation

$$x^2 - 5y^2 = 4 \quad (15)$$

Equation (15) is the Pell's equation $x^2 - Dy^2 = N$ with D = 5 and N = 4.Thus a quasi-residual Metis design is related to the Pell's equation.

C. Relationship with Fibonacci and Lucas Numbers

Fibonacci numbers $\{F_s\}$ and Lucas numbers $\{L_s\}$ are recursively defined as follows (see for e.g. G.H.Hardy and E.M.Right [3]).

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

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

The first few Fibonacci and Lucas numbers are furnished in the following table:

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 observe that the successive pairs of Fibonacci and Lucas numbers have the following property:

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

These specific results prompt us to try an induction approach to have a general result. By induction we see that

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

Thus the corresponding even subscripted terms in the Lucas and Fibonacci sequences satisfy the Pell's equation (15) and consequently they lead to a quasi-residual Metis design. In view of this result, the parameters of a quasi-residual Metis design are obtained in terms of the Lucas and Fibonacci numbers as:

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

THE PROBLEM OF DISCRETE LOGARITHM

Let p be an odd prime. The discrete logarithm problem is to find $x = \log_b(y)$ in the finite field Z_p i.e., to find the value(s) x in Z_p such that $b^x \equiv y \pmod p$. No algorithm is currently available for this problem. There are applications of this problem in cryptography which is the subject of sending messages in a secret way, ensuring the security of the information (see for e.g., A.M.Odlyzko). The problem of mapping the discrete logarithm has been considered by D.Cloutier and J.Holden. The structure in the discrete logarithm has been studied by A.Hoffman.

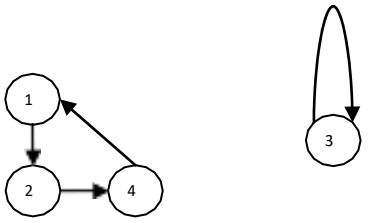
The discrete logarithm can be viewed as a function. The problem is to determine the inverse of $x \to b^x \pmod p$. A functional graph may be used a tool for this problem. The values of x can be represented by nodes of a graph and arrows may be drawn for each one of the mappings. A functional graph is a directed graph such that each vertex must have exactly one edge directed out from it. An m-ary functional graph is a functional graph where each node has in-degree of exactly zero or m.

Let us consider a few specific cases to illustrate the procedure involved. For the functional graph of 2 (mod 5), consider the successive integral powers of 2 and reduce them modulo 5. We have

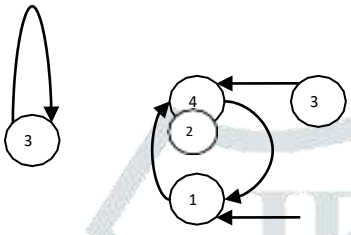
$2^1 \equiv 2 \pmod 5$, $2^2 \equiv 4 \pmod 5$, $2^3 \equiv 3 \pmod 5$, $2^4 \equiv 1 \pmod 5$.Taking into account the exponent and the result after reducing modulo 5, we obtain the forward correspondence

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

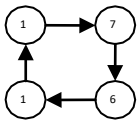
From this correspondence, we separate the cycles and get $1 \to 2, 2 \to 4, 4 \to 1 \; and \; 3 \to 3$. Each cycle is represented by means of a directed graph. The functional graph for this case and a few other examples are shown below.

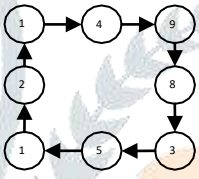


Functional graph for 2 (mod 5)

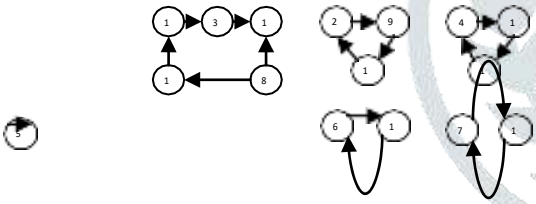


Functional graph for 4(mod 5)

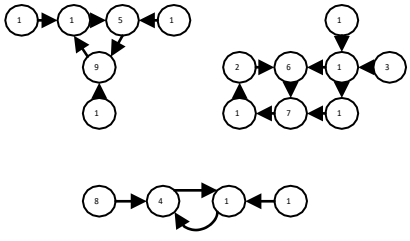




Functional graph for 7 (mod 13)



Functional graph for 3 (mod 17)



Functional graph for 5 (mod 19)

In the functional graph for 4 (mod 5), the nodes 2 and 3 are not parts of any cycles. It is of interest to consider a functional graph wherein each node is part of a cycle. In this regard, we need the following:

(16) Let r be an element of Z_p^*. Let e be the smallest natural number such that = 1 (mod p). We say that r is a primitive root modulo p if . Let r be any primitive root modulo p and g D. Cloutier

r^e

$e = \emptyset(p)$

$g \equiv r^\alpha \pmod p$. It has been shown by and J. Holden [2] that the values of g that produce an m-ary graph are precisely those for which gcd (α, p-1) = m.

(17) In the problem of discrete logarithm, A. Hoffman [4] has taken b as a primitive root modulo p and considered three parameters associated with a functional graph, viz. the number of cycles, the maximum cycle length and the weighted average cycle length. He has shown that the structure of discrete logarithm can be analysed by statistical investigation of these three parameters. He has illustrated how comparisons are possible between random permutations and those constructed from the solution to the discrete logarithm problem by considering the expected values of the three parameters in both cases.

With the distribution of cycle lengths following Poisson distribution, has shown how ANOVA tests can be carried out for mean number of cycle components, number of components variance, mean maximum cycle length, maximum cycle variance, mean average cycle length and average cycle variance. Selecting 30 primes in the range 99991 – 106921 and employing t-test and Anderson-Darling test, he has derived the statistical results for the three parameters of the functional graphs concerning the primes to illustrate the structure in the discrete logarithm.

CONCLUSION

In the foregoing discussion, some of the linkages between Number Theory and Statistics have been furnished. There is much scope for probing into the applications of Number Theory in Statistics and vice versa. Distribution of prime numbers is a challenging area of research. When the parameters in a design become large, analysis of the design becomes quite complex and so one requires more computational skill. Understanding of the properties of primes and solving a discrete logarithm problem by means of functional graphs require high-end computing power. With the presently available computational capabilities due to technological development, the future research work holds promise and one may expect tangible results in this interesting field of research.

ACKNOWLEDGEMENT

The author is thankful to the referee for the suggestions towards the improvement of the paper.

REFERENCES

- P.T. Bateman, “The distribution of values of Euler’s ∅-function”, Acta Arith., Volume 21, Pp. 329 – 345, 1972
- D. Cloutier and J. Holden, “Mapping the discrete logarithm”, Involve, Volume 3, Issue 2, Pp. 197 – 213, 2010
- G.H. Hardy and E.M. Wright, “An introduction to the theory of numbers”, Oxford University Press, London, 1975.
- A. Hoffman, “Statistical investigation of structure in the discrete logarithm”, Rose-Hulman Undergraduate Mathematics Journal, Volume10, Issue 2, Pp. 1 – 20, 2009
- L.J. Mordell, “Diophantine equations”, Academic Press, London, 1969
- A.M. Odlyzko, “Discrete logarithms in finite fields and their cryptographic significance” in “Advances in Cryptology: Proceedings of EUROCRYPT 84”, Lecture Notes in Computer Science, Springer-Verlag, Volume 209, Pp. 242 – 314, 1985
- P. Ribenboim, “The new book of prime number records”, Springer Verlag, New York, 1996.