Journal of Applied Mathematics and Computation, 2021, 5(2), 124-133
https://www.hillpublisher.com/journals/JAMC/
ISSN Online: 2576-0653
ISSN Print: 2576-0645

Blake's "A Poisson Tree" Statistically Climbed

Jharna Pradhan, Soubhik Chakraborty*

Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India.

How to cite this paper: Jharna Pradhan, Soubhik Chakraborty. (2021) Blake's "A Poisson Tree" Statistically Climbed. *Journal of Applied Mathematics and Computation*, 5(2), 124-133.
DOI: 10.26855/jamc.2021.06.007

Received: April 15, 2021
Accepted: May 12, 2021
Published: June 4, 2021

***Corresponding author:** Soubhik Chakraborty, Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India.
Email: soubhikc@yahoo.co.in

Abstract

"A Poisson Tree" is an exemplary poem composed by William Blake in 1794. The present work encompasses a statistical study intensifying several features such as finding the appropriate probability distribution that fits the data corresponding to the number of words per line, word length, the number of vowels per stanza and the number of letters per line. Additionally, the various parts of speech used in the poem are studied and the pronoun is found to be used most. It is found that the number of words per line is distributed uniformly in the poem which is tested by Chi Square goodness of fit test. A Zero truncated Poisson distribution is found to be a good fit for the frequency distribution of word length. Binomial distribution turns out to be a good fit for the number of vowels per stanza in the poem. The number of letters per line is also distributed uniformly in the poem which is again tested by Chi Square goodness of fit test. The poem is following the rhythmic scheme AABB CCDD EEFF GGHH. The number of words between two successive rhythmic words can be dichotomized.

Keywords

Zero truncated Poisson distribution, Binomial distribution, Chi Square goodness of fit test, Word length, Rhythm

The paper is organized as follows. Section 1 is the introduction. Section 2 gives the methodology. Section 3 gives the experimental results and discussion. Finally, Section 4 gives the concluding remarks.

1. Introduction

1.1 Mathematics and poetry

Mathematics and poetry have a deep connection in that both mathematicians and poets look for patterns. It is another matter that in mathematics, it is the idea behind the patterns that is more important while in poetry, it is not the idea but how this idea is expressed through words that is crucial. Mathematicians and poets however repeat the same dictom as correctly assessed by Ron Aharoni: "The beauty of mathematics resemble that of poetry" [1]. The other works connecting mathematics and poetry include those of Shmukler and Ziskin [2] which, apart from providing a substantial list of mathematicians who wrote poetry, also highlight the similarities and differences between creativity in mathematics and poetry. Growney [3] has explained how mathematics influences poetry. Growney's poem "Geometry" is the best example to illustrate the use of mathematical images and themes in the poems. The other famous poem "My Dance is Mathematics" by her connects poetry and mathematical forms. Growney's contribution in the interesting field of mathematics and poetry is remarkable. Glaz [4] has discussed mathematical ideas (importantly decimal and place value system) in ancient Indian poetry. She has used poetic forms in the mathematics classroom to enhance the pedagogy and course content. All these remarkable contributions focus on the interlink between mathematics and arts. Encouraged by these works, we take a statistical look on one of the famous Blake's poems "A Poisson Tree".

1.2 Central Idea of the poem

This poem is an example of Blake's acute psychological insight. How true it is, as in the first two lines, that if those who quarrel 'have it out', their wrath subsides! But if wrath is suppressed, a poisonous tree will grow. The use of the word 'friend' in line 1 and 'foe' in line 3 bring out the sound relationship one can have with a friend but show that wrath with a foe is difficult to suppress because of the wrong relationship. "Till it bore an apple bright" is a reference, of course, to the Fall of Man in the Garden of Eden; "the pole" refers to the North pole. (see Green [5], p. 238). The poem in question is given below.

A Poisson Tree by William Blake

I was angry with my friend:
I told my wrath, my wrath did end,
I was angry with my foe:
I told it not, my wrath did grow.

And I watered it in fears,
Night and morning with my tears;
And I sunned it with smiles,
And with soft deceitful wiles.

And it grew both day and night,
Till it bore an apple bright;
And my foe beheld it shine,
And he knew that it was mine,

And into my garden stole
When the night had veiled the pole;
In the morning glad I see
My foe outstretched beneath the tree.

1.3 Statistical Analysis

The statistical study on the aforesaid poem studies several interesting features such as the number of words per line, the number of words per stanza, the number of letters per line and those per stanza all of which seem to follow uniform distribution. A zero truncated distribution is a very good fit for the frequency distribution of word length. Binomial distribution fits the number of vowels per stanza, as tested by Chi-square goodness of fit test. Other studies are on rhythm and parts of speech.

2. Methodology (Gupta and Kapoor [6])

2.1 Chi Square Goodness of Fit Test:

Chi Square Goodness of fit test is the non parametric test which is used to find the suitability of a probability model to be a good fit or otherwise for a given data by finding out how the observed frequencies differ from the expected ones.

If $O_{i,}(i=1,2,\ldots.n)$ is a set of experimental frequencies and $E_i(i=1,2,\ldots n)$ is the corresponding set of theoretical (expected) frequencies then Chi-Square is given by

Chi-Square$=\Sigma(O_i-E_i)^2/E_{i,}$ with the linear restriction $\Sigma O_i=\Sigma E_{i,}$

Which follows Chi Square distribution with (n-1) degree of freedom where n is the number of classes. The summation extends over all the n classes. One degree of freedom is lost due to the linear restriction $\Sigma O_i=\Sigma E_{i,}$

The calculated Chi-Square is compared with the corresponding tabulated Chi Square value at (n-1) degrees of freedom and α% level of significance. If calculated Chi Square exceeds the tabulated value then we reject the null hypothesis and conclude that there is significant difference between the experimental and theoretical values else we may accept the null hypothesis that the difference between the theoretical and the experimental frequencies is insignificant statistically at α% level of significance. Here we shall take $\alpha = 5\%$.

Remark: If any class frequency is less than 5, adjacent classes have to be combined because Chi Square is actually a continuous probability distribution (square of a standard normal variate is a chi square variate with one degree of freedom) and it loses its character of continuity if any class frequency is less than 5.

2.2 Binomial Distribution:

Binomial distribution is a discrete distribution. A random variable X is said to follow binomial distribution if it assumes only non negative values and its probability mass function is given by:

$P(X=x)={}^{n}C_{x}\, p^{x}\, (1-p)^{n-x}$; x=0,1,2…n; 0<p<1

Where n is the number of independent Bernoullian trials, p is the fixed probability of success in a trial.

2.3 Poisson Distribution

Poisson distribution is a limiting case of binomial distribution, the following being the limiting conditions:-

i. Number of trials, n is large. That is, n tends to infinity.

ii. The probability p of success is small. That is, p tends to zero.

iii.np = μ (say) is constant and finite.

The probability mass function of Poisson distribution is given by:

$$P(X=x)= (e^{-\mu}\mu^{x})/x!;\ x=0,1,2,3\ldots..$$

Where μ is the parameter of the distribution.

For this distribution, mean = variance = μ.

2.4 Zero Truncated Poisson distribution

Zero Truncated Poisson distribution is a Poisson distribution conditioned on being non zero.

since, $\sum_{x=0}^{\infty}(e^{-\mu}\mu^{x}/x!)=1$

$$P(0)+P(1)+\ldots\ldots=1$$

$$P(1)+P(2)\ldots\ldots=1-P(0)$$

$$\sum_{x=1}^{\infty}(e^{-\mu}\mu^{x}/x!)=1-P(0)$$

$$\sum_{x=1}^{\infty}1/((1-P(0))(e^{-\mu}\mu^{x}/x!)=1$$

Where X is now a Zero truncated Poisson variate with parameter μ. Thus the probability mass function of a zero truncated poisson variate is similar to that of a Poisson variate except for the constant coefficient 1/(1-P(0)).

And $P(0)=e^{-\mu}$

3. Results and Discussions

3.1 Study 1: words per line: Uniform or non Uniform?

Table 1 depicts the number of words per line in the poem.

Table 1. Number of words per line in the poem

Lines	No. of words per line
1	6
2	8
3	6
4	8
5	6
6	6
7	6
8	5
9	7
10	6
11	6
12	7
13	5
14	7
15	6
16	6

Here we set up null hypothesis H_0: The no. of words per line of the poem is distributed uniformly. H_0 would be tested against H_1: The no. of words per line of the poem is not distributed uniformly at 5% level of significance.

Chi Square calculations are shown in Table 1a.

Table 1a. Calculations for Chi-Square

Line	Observed frequencies(O_i)	Expected Frequencies(E_i)	$(O_i-E_i)^2/E_i$
1	6	101/16	729/1616
2	8	101/16	25/1616
3	6	101/16	729/1616
4	8	101/16	25/1616
5	6	101/16	729/1616
6	6	101/16	729/1616
7	6	101/16	729/1616
8	5	101/16	441/1616
9	7	101/16	121/1616
10	6	101/16	729/1616
11	6	101/16	729/1616
12	7	101/16	121/1616
13	5	101/16	441/1616
14	7	101/16	121/1616
15	6	101/16	729/1616
16	6	101/16	729/1616

$\Sigma O_i= \Sigma E_i= 101 \Sigma(O_i-E_i)^2/E_i =1.81$

Calculated Chi-Square from table-1a $= \Sigma (O_i-E_i)^2/E_i = 1.81$

Degrees of freedom= 16-1=15 (since we are given 16 frequencies and $\Sigma O_i= \Sigma E_i$linear constraint).

Tabulated Chi-Square value at 5% level of significance for 15 degree of freedom is 24.996. Since calculated chi square value is less than the tabulated value, and hence the null hypothesis may be accepted at 5% level of significance.

Therefore, we can conclude that number of words per line may be assumed to be distributed uniformly in the poem.

3.2 Study 2: Frequency distribution of word length

Table 2. Frequency distribution of word length

Word length r	f= Number of words (frequency) having word length r
1	7
2	20
3	26
4	21
5	14
6	7
7	4
8	0
9	1
10	0
11	0
12	1

Table 2 gives the frequency distribution of word length ie number of letters in the word. Total number of words= 7+20+26+21+14+7+4+0+1+0+0+1=101 (From Table 2). Total number of letters = Sum of (word length X frequency) =1X7+2X20+3X26+4X21+5X14+6X7+7X4+8X0+9X1+10X0+11X0+12X1=370

Mean and variance of r Table 2 is 3.66 and 3.213 respectively (Mean of r = $\sum fr/\sum f$ and variance of r = $\sum fr^2/\sum f$ - $\{\sum fr/\sum f\}^2$) which are close. Hence, Poisson distribution may be a good fit for Table 2, but a poisson variate can take a zero value whereas the word length can never be zero. Therefore, Zero truncated Probabilty Distribution model is opted to be fitted to the word length data. In order to fit the Zero truncated Poisson distribution, we have to find the expected frequencies. The frequencies for word length r are computed using the formula:

$$f(r)= \text{Total no. of words in the poem X } P(X=r)$$

where X has zero truncated Poisson distribution. In our case,

$$f(r)= \text{Total no. of words in the poem X}(e^{-3.66}(3.66)^r)/r! \text{X } 1/(1-P(0)$$

where r= 1, 2, 3……12
and $P(0)=e^{-3.66}= 0.0257$

$$f(1)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66))/1! \ = 9.7632$$

$$f(2)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^2)/2!= 17.4075$$

$$f(3)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^3)/3! \ = 21.7973$$

$$f(4)=101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^4)/4! \ = 19.9445$$

$$f(5)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^5)/5! = 14.5994$$

$$f(6)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^6)/6! \ = 8.9056$$

$$f(7)=101\frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^7)/7! = 4.6564$$

$$f(8)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^8)/8! \ = 2.1302$$

$$f(9)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^9)/9! \ = 0.8663$$

$$f(10)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^{10})/10! \ = 0.3170$$

$$f(11)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^{11})/11! \ = 1.0549$$

$$f(12)= 101X\ \frac{1}{(1-0.0257)} X(e^{-3.66}(3.66)^{12})/12! \ = 0.0321$$

Accordingly, the theoretical zero truncated Poisson frequencies correct to three decimal places are tabulated in Table 2a:

Table 2a. Theoretical Poisson frequencies

r=word length	1	2	3	4	5	6	7	8	9	10	11	12
Expected frequencies	9.7632	17.4075	21.7973	19.9445	14.5994	8.9056	4.6564	2.1302	0.8663	0.3170	1.0549	0.0321

Chi Square calculations based on the data in Table 2a are shown in Table 2b. We pool the last seven classes to ensure sufficiently large class frequencies.

Table 2b. Calculations of Chi-Square (After pooling the last seven expected frequencies)

Word length	Observed Frequencies (O_i)	Expected Frequencies (E_i)	$(O_i-E_i)^2/E_i$
1	7	9.7632	0.7820
2	20	17.4075	0.3861
3	26	21.7973	0.8103
4	21	19.9445	0.0559
5	14	14.5994	0.0241
6	13	17.9625	6.6904

Calculated Chi square, $\Sigma(O_i-E_i)^2/E_i$= 8.749 (From Table 2b)

Degrees of freedom =6-1= 5

Tabulated chi square value at 5% level of significance and 5degrees of freedom equals 11.07. Calculated chi square is less than the corresponding tabulated value. Hence, Zero truncated Poisson distribution is a good fit to the data on word length.

3.3 Study 3: Number of vowels per stanza in the poem

Table 3. Number of vowels per stanza

Stanza	Number of vowels
1	25
2	33
3	31
4	37

Chi square goodness of fit test is used to check whether the overall proportion of vowels in the entire poem is maintained in the respective stanzas or not. Hence, the number of vowels in a word is taken as binomial distribution with parameters n and p where n is the word length and p is the probability of a vowel which is estimated by the proportion of vowels in the entire poem. Therefore we set up null hypothesis H_0 that overall proportion of vowels in the entire poem is maintained in the respective stanzas.

H_0: The overall proportion of vowels in the entire poem is maintained in the respective stanzas

H_1: The overall proportion of vowels in the entire poem is not maintained in the respective stanzas.

H_0 would be tested against H_1 at 5% level of significance.

Table 3 gives the number of vowels per stanza in the poem.

Total number of vowels=25+33+31+37=126.

Total number of letters= 370.

The poem has four stanzas.

The proportion of vowels in the poem=126/370.

Expected frequency of vowels in stanza i= No. of letters in stanza i X$\frac{126}{370}$

where $\frac{126}{370}$ is the empirical probability of a vowel (considering the entire poem).

Table 3a gives the Chi square calculations.

Table 3a. Calculation of Chi-Square

Stanza	Observed Frequencies(O_i)	Expected Frequencies(E_i)	$(O_i-E_i)^2/E_i$
1	25	88x$\frac{126}{370}$= 29.968	0.8236
2	33	93x$\frac{126}{370}$=31.670	0.0559
3	31	90x$\frac{126}{370}$=30.649	0.1803
4	37	99x$\frac{126}{370}$=33.714	0.3203

$\Sigma O_i= \Sigma E_i$= 126$\Sigma(O_i-E_i)^2/E_i$ =1.3801 (From table-3a)

Number of degrees of freedom = 4-1 = 3

Calculated Chi-Square, $\Sigma(O_i-E_i)^2/E_i$ =1.3801 (From table-3a)

Tabulated Chi square value at 5% level of significance for 3 degree of freedom = 9.34888

Since calculated Chi-square is less than the tabulated Chi-square and hence we accept the null hypothesis.

Hence we conclude that binomial distribution is a good fit to the no. of vowels per stanza in the poem.

3.4 Study 4: Rhythm of the poem

Mathematically, rhythm can be evaluated by counting the number of words between two successive words in rhyme which reflects the pause between two successive rhythmic words.

The poem follows AABB CCDD EEFF GGHH in the four stanzas, similar letter indicating a rhyme. Table 4 illustrates the data of number of words in between the pairs of rhyming words in the poem.

Table 4. Number of words between the rhyming pairs

Sl.no	Rhyming Words	Number of words in between the rhyming pairs
1	(Friend, End)	7
2	(Foe, Grow)	7
3	(Fears, Tears)	5
4	(Smiles, Wiles)	4
5	(Night, Bright)	5
6	(Shine, Mine)	6
7	(Stole, Pole)	6
8	(See, Tree)	5

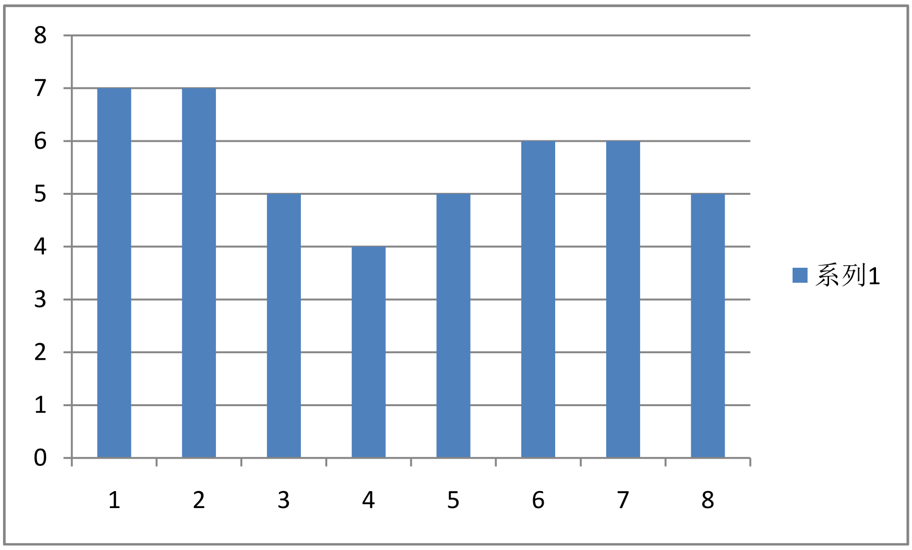

Figure 1. Bar graph of Table 4.

Figure 1 represents a bar graph of Table 4. From the bar graph, it is clear that although all the bars are not of uniform height yet there are clearly two types of bars: higher ones (numbered 1, 2, 6, 7) and lower ones (numbered 3, 4, 5, 8).

3.5 Study 5: Number of letters per line

Table 5. Number of letters per line

Line	Number of letters
1	21
2	25
3	18
4	24
5	20
6	26
7	22
8	25
9	24
10	23
11	21
12	22
13	20
14	28
15	20
16	31

Table 5 depicts the number of letters per line in the poem.

Here we set up null hypothesis H_0: The no. of letters per line of the poem is distributed uniformly.

To be tested against H_1: The no. of letters per line of the poem is not distributed uniformly.

H_0 would be tested against H_1at 5% level of significance.

Table 5a gives the Chi Square calculations.

Table 5a. Calculations of Chi-Square

Line	Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2/E_i$
1	21	370/16	0.1952
2	25	370/16	0.1520
3	18	370/16	1.1358
4	24	370/16	0.0331
5	20	370/16	0.4222
6	26	370/16	0.3574
7	22	370/16	0.0547
8	25	370/16	0.1520
9	24	370/16	0.0331
10	23	370/16	0.0067
11	21	370/16	0.1952
12	22	370/16	0.0547
13	20	370/16	0.4222
14	28	370/16	1.0277
15	20	370/16	0.4222
16	31	370/16	2.6818

$\Sigma O_i= \Sigma E_i = 370$ $\Sigma (O_i-E_i)^2/E_i=7.346$ (From table-5a)

Calculated Chi-Square, $\Sigma (O_i-E_i)^2/E_i=7.346$

Degree of freedom= 16-1=15 (since we are given 16 frequencies and $\Sigma O_i= \Sigma E_i$ is a linear constraint).

Tabulated Chi-Square value at 5% level of significance for 15 degree of freedom is 24.996. Since calculated Chi Square value is less than the corresponding tabulated value, hence the null hypothesis may be accepted at 5% level of significance. Therefore, we can conclude that the number of letters per line may be assumed to be distributed uniformly in the poem.

3.6 Study 6: Number of words per stanza

Table 6 depicts the number of words per stanza.

Table 6. Number of words per stanza

Stanza	Number of words
1	28
2	23
3	26
4	24

Here we set up null hypothesis H_0: The no. of words per stanza of the poem is distributed uniformly.

To be tested against H_1: The no. of words per stanza of the poem is not distributed uniformly.

H_0 would be tested against H_1 at 5% level of significance.

Table 6a. Calculations of Chi-Square

Stanza	Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2/E_i$
1	28	101/4	121/404
2	23	101/4	81/404
3	26	101/4	9/404
4	24	101/4	25/404

$\Sigma O_i = \Sigma E_i = 101$ $\Sigma (O_i-E_i)^2/E_i = 238/404 = 0.584$

Calculated Chi-Square, $\Sigma (O_i-E_i)^2/E_i = 0.584$(From Table 6a)

Degree of freedom= 4-1=3 (since we are given four frequencies and $\Sigma O_i = \Sigma E_i$is a linear constraint).

Tabulated Chi-Square value at 5% level of significance for 3 degree of freedom is 7.815. Since calculated Chi Square value is less than the corresponding tabulated value, hence the null hypothesis may be accepted at 5% level of significance. Therefore, we can conclude that number of words per stanza may be assumed to be distributed uniformly in the poem.

3.7 Study 7: Analysis of parts of speech used in the poem

Table 7 gives the distinct parts of speech used in the poem.

Table 7. Number of parts of speech used per stanza

Parts of speech used in the poem	Stanza 1	Stanza 2	Stanza 3	Stanza 4	Total number of parts of speech
Noun	6	5	4	6	21
Pronoun	10	5	7	3	25
Adjective	2	2	2	5	11
Verb	7	3	6	5	21
Adverb	1	0	0	0	1
Preposition	2	4	1	3	10
Conjunction	0	4	6	2	12
Interjection	0	0	0	0	0

An interesting observation is that in each stanza Pronoun is the maximum used part of speech while adverb is used least. Also, interjection is not used in any of the stanza. Moreover, the pronoun, noun and verb are mostly used in that order followed by conjunction, adjective and preposition.

4. Conclusion

The statistical study is performed on the exemplary poem "A Poisson Tree" by William Blake which reveals several statistical features which depict Blake's style as embedded in this particular poem. The number of words per line, number of letters per line and the number of words per stanza are all distributed uniformly in the poem. A zero truncated Poisson distribution is a good fit for the frequency distribution of word length while Binomial distribution is found to be a good fit for number of vowels per stanza. The poem follows the rhythmic sequence AABB CCDD EEFF GGHH. The number of words between two successive rhythmic words can be dichotomized. It is further observed that the pronoun is the part of speech that is used most in the poem while interjection is not used at all.

A similar study on Blake's other poems would enlighten on the commonality and diversity of such statistical features. This is reserved as a rewarding future work.

Ethical declaration

The authors hereby declare that this research is not funded and that they have no conflict of interest.

References

[1] Ron Aharoni. (2014). Mathematics, Poetry and Beauty. Journal of Mathematics and the Arts, Vol. 8, 2014, issue-1-2,

pp. 5-12.
[2] Alla Shmukler and Clara Ziskin. (2014). Through the looking glass of history: mathematicians in the land of poetry. Journal of Mathematics and the Arts, 2014, Vol. 8, issue 1-2, pp. 78-86.
[3] Jo Anne Growney. (2008). Mathematics influences poetry. Journal of Mathematicians and the Arts, 2008, Vol. 2, issue 1, pp. 1-7
[4] Sarah Glaz. Mathematical Ideas in Ancient Indian poetry, www.2.math.uconn.edu/glaz/My_Articles/Mathematical IdeasInAncientIndianPoetry.Bridges13.pdf accessed on 13th Feb., 2020.
[5] David Green. (2018). The Winged Word. TRINITY Press, 2018.
[6] S. C. Gupta and V. K. Kapoor. (2014). Fundamentals of Mathematical Statistics. Sultan Chand and Sons.