

Analysing the Poetic Structure of *Jana-Gana-Mana* in Entirety: A Statistical Approach

Niraj Kumar Singh¹, Soubhik Chakraborty^{2,*}, Mihul Roy³

¹Department of Computer Science and Engineering, Birla Institute of Technology, Mesra-835215, Jharkhand, India.

²Department of Mathematics, Birla Institute of Technology, Mesra-835215, Jharkhand, India.

³IT Services, MECON Limited, Ranchi-834002, Jharkhand, India.

How to cite this paper: Niraj Kumar Singh, Soubhik Chakraborty, Mihul Roy. (2021) Analysing the Poetic Structure of *Jana-Gana-Mana* in Entirety: A Statistical Approach. *Journal of Applied Mathematics and Computation*, 5(4), 264-272. DOI: 10.26855/jamc.2021.12.004

Received: September 3, 2021

Accepted: September 28, 2021

Published: October 18, 2021

*Corresponding author: Soubhik Chakraborty, Department of Mathematics, Birla Institute of Technology, Mesra-835215, Jharkhand, India.

Email: soubhikc@yahoo.co.in

Abstract

Measurable investigation of abstract content so as to bring bits of knowledge into its expressive highlights has been a shared zone of enthusiasm among the aficionados of writing and measurements. The thought of applying the equivalent for mining critical perceptions with respect to our huge abstract legacy has likewise intrigued the antiquarians. There have been several works on automated analysis of poetry. Notwithstanding, adequate consideration has not been laid on such works in even Hindi-prose, leaving Hindi-poetry with an almost total nonattendance of any examination in such manner. The curiosity of the work proposed in this paper is emphasized by the way that it includes the utilization of binomial model to literary compositions which means to take a new way towards statistical interpretation of literary pieces. Any poetic composition in Hindi may be viewed as a sequence of numeral symbols 1 and 2 only which correspond to a *laghu* (short syllable) and a *guru* (long syllable) respectively. Determining short and long syllables in Hindi-poetry is not a trivial activity as this determination is not independent rather is a function of a number of factors. The candidate poem is statistically analysed using three case studies. In the first study, the *laghu* and *guru* data is successfully fitted to a binomial model. Short syllables themselves can be classified into two types. The second case study is an attempt towards fitting a binomial model to the type_1 and type_2 *laghu* data. In Hindi, a vowel can be written either independently or as a *mātrā* (diacritic mark) along with a consonant letter. As our final case study, we have successfully fitted a binomial model over diacritic marks data in *Jana-Gana-Mana*. Importance of this task lies in the observations leading to fit the various phonological features of the candidate song through a single model type, that is, binomial model.

Keywords

Hindi, Language, Statistical analysis, Stylistic features, Stylometry, Binomial model, Poetry

1. Introduction

Composed by extraordinary writer and Nobel laureate poet Rabindranath Tagore, the melody *Jana-Gana-Mana* is the public hymn of India. This melody was initially created as *Bhāroto-Bhāgyo-Bidhātā* in Bengali and later was written in *tatsama* Bengali which is vigorously sanskritised. It was embraced in its Hindi variant by the Constituent Assembly as the National Anthem of India on 24 January 1950 [1]. The complete melody comprises of five stanzas. The current work

is a measurable examination of the full form of this melody, of which the principal verse, the first one, establishes the National Anthem.

Apart from enjoying the status of being National Anthem of a major and culturally rich nation like India, this song is well known for its significantly delicate, new and delightful stanzas. Statistical analysis of specific highlights of this song is certainly not a ground-breaking thought. India's national anthem has been modelled successfully using a simple exponential smoothing by Chakraborty et al. [2]. Their fit is found to be explaining the note progression well enough with smoothing factor Alpha 0.7164. Their paper is a factual investigation on melodic perspective, though, the current undertaking centres around investigating the content based measurable highlights of this song.

Most of the compositions of Rabindra Nath Tagore are written in Bengali. Nonetheless, his scholarly pieces have been converted into different languages including English and Hindi. Over a span, the literary work of Rabindra Nath have been dissected by several researchers. Few remarkable studies include those by Das [3], Gokak [4] and Hassan et al. [5]. One such undertaking has been carried out to automatically extract stylistic and phonetic data which are valuable for investigation and examination of poems (see Rakshit et al., [6]). This undertaking is an endeavour, utilizing a Multiclass SVM classifier, to categorize Tagore's assortment of poetry into four categories: devotional, love, nature and nationalism. Authors of this paper claim that this is a first such attempt on Bangla poetry. As referenced before, the song *Jana-Gana-Mana* was originally written in Bangla and later embraced in Hindi adaptation. Through this case study, the present task is an attempt towards exploring the content based stylistic features in Hindi composition of a non-Hindi poet.

There have been several works on automated analysis of poetry in Indian languages such as Bangla, Odia, Punjabi and Marathi (see Deshmukh et al., [7]; Kaur and Saini [8]; Nayak and Panigrahi [9]; Rakshit et al., [6]). Automated analysis of Hindi-prose, in recent past, has become a notable exploration territory (see Joshi and Kushwah [10]; Magdum, Patil and Suman [11]). Be that as it may, adequate consideration has not been laid on such works in Hindi-poetry. Majority of such tasks belong to classification problem solved through machine learning approaches (see Bafna and Saini [12]; Pal and Patel [13]; Pandian, Maurya, and Jaiswal [14]). Kushwah and Joshi [15] in their paper have worked on identifying Chanda based Hindi-poetry. However, their task is a basic one and is severely limited in its functionality.

In Hindi, a poetic composition (metrical or non-metrical) can always be viewed as a sequence of numerals 1s and 2s. The symbols 1 and 2 correspond to the terms *laghu* and *guru* discussed in the section 2 of this paper. Unlike classifying letters simply either as vowels or consonants, *laghu* and *guru* determination in a Hindi poetic composition is a non-trivial activity. The idea of designating the occurrence of a *varna* as one binomial trial is a novel approach with potential of lending a new angle to the contemporary research in the field of stylometry. Through this case study, the current errand is an investigation on how some perspectives related to Hindi-poetry are relatable through mathematical models, especially binomial.

The paper has been sorted out in the following order: Section 2 gives a prologue to the Hindi-alphabet set and discusses about the issues related to the concept of *varna* in Hindi poetry. Section 3 examines the results obtained and related discussion. Lastly, conclusion and future scope have been provided in Section 4.

2. An Introduction to Hindi alphabet and Varna

2.1 An Introduction to Hindi alphabet

In Hindi grammar, an *aksara* is defined as a root sound which cannot be decayed further (see Kumar [16]). For example, अ [a], उ [u], त [t] etc. are root sounds. A serialized set of *aksaras*, of Hindi language, is known as the Hindi-alphabet. There are two types of *aksaras* in Hindi-alphabet: *svara* (vowels) and *vyamjana* (consonants).

Svara (vowel)- is defined as an *aksara* which can be pronounced independently, and is helpful in pronouncing *vyamjana aksaras*. Hindi language consists of eleven *svara aksaras*. These are further divided into two classes namely, *hrsva* and *dirgha* depending upon the time they consume. The time elapsed in *dirgha* is nearly the twofold than that required in *hrsva* (see Kumar [16]).

Svara (vowel) in Hindi-alphabet are- अ [a], आ [ā], इ [i], ई [ī], उ [u], ऊ [ū], ऋ [r], ए [e], ऐ [ai], ओ [o] and औ [au].

vyamjana (consonant)- is defined as an *aksara* which can be pronounced with help of *svara aksaras*. Hindi-alphabet consists of thirty-three basic consonants. Apart of these basic consonants we have two additional and four conjunct consonants as well. Any consonant in its unadulterated form is written with a *halanta* mark underneath it. However, for simplicity, in common practice very often it is written without the *halanta* mark. See Table 1 for insight.

In Hindi texts, any vowel can appear either independently or along with a consonant as a diacritic mark. For instance, in the word आम [āma], the vowel आ [ā] appears independently whereas, in another word काम [kāma], it appears as a diacritic mark ा [ā] along with consonant क [ka]. The list of vowel diacritics in Hindi includes: - ा, ि, ी, ु, ू, े, ै, ो, ौ. Example of consonant letter क [k] along with vowel diacritics- क [ka], का [kā], कि [ki], की [kī],

कु [ku], कू [kū], कृ [kr], के [ke], कै [kai], को [ko] and कौ [kau].

Table 1. Vyamjana in Hindi-alphabet

Basic consonants									
क [ka]	ख [kha]	ग [ga]	घ [gha]	ङ [ṅa]	च [ca]	छ [cha]	ज [ja]	झ [jha]	ञ [ña]
ट [ṭa]	ठ [ṭha]	ड [ḍa]	ढ [ḍha]	ण [ṇa]	त [ta]	थ [tha]	द [da]	ध [dha]	न [na]
प [pa]	फ [pha]	ब [ba]	भ [bha]	म [ma]	य [ya]	र [ra]	ल [la]	व [va]	
श [śa]	ष [ṣa]	स [sa]	ह [ha]						
Additional consonants									
ड़ [ṛa]	ढ़ [ṛha]								
Conjunct consonants									
क्ष [kṣa]	त्र [tra]	ज्ञ [jña]	श्र [śra]						

2.2 An Introduction to varṇa and Laghu-Guru Determination

Varṇa or syllable- A vowel with or without one or more consonants is considered as one syllable. A syllable can be short (*laghu*) or long (*guru*), depending on whether its vowel is short or long (see Pandey [17]).

Laghu or short syllables- The vowels अ [a], इ [i], उ [u] and ऋ [r] are short. Whenever any of these is used in a verse separately or with one or more consonants, it will be considered as a short syllable. For example, क [ka], कि [ki], etc. are short syllables. The numeral symbol 1 is used to represent a short syllable in scansion and metrical analysis.

Guru or long syllables-The vowels आ [ā], ई [ī], ऊ [ū], ए [e], ऐ [ai], ओ [o], औ [au] are long. Whenever any of these is used in a verse separately or with one or more consonants, it will be considered as long. For example, का [kā], की [kī] etc. are long syllables. The numeral symbol 2 is used to represent a long syllable in scansion and metrical analysis.

A short vowel gets the practical status of long under the following three conditions (see Dvivedi and Singh [18]; Pandey [17]).

- (1) If a vowel is followed by an *anusvara* (◌ं). For example, तं [tam̐] and गं [gam̐] are long syllables.
- (2) If a vowel is followed by *visarga*. For example, तः [taḥ] and गः [gaḥ] are long syllables.
- (3) If a vowel is followed by a conjunct consonant. For example, बन्ध [bandha]. Here, the short syllable ष [ba] is to be counted long, since it is followed by the conjunct न्ध [ndha].

Varna determination in Hindi is a non-trivial activity. The reason is, this determination does not solely depend on the current character (*aksara*), rather is a function of preceding and/or succeeding characters (*aksaras*) as well.

3. Experimental results & Discussion

All the five stanzas of the song *Jana-Gaṇa-Mana* (in *devanāgarī* as well as in romanised IAST transliteration) are given in Table 2. A number of useful information can be drawn from this text. The information includes number and types of syllables, number of lines per stanza, number of words in a line, number of characters per line, word lengths along with their respective frequencies, average word length etc. A statistician would always love to look for numbers even in textual data. Following it, in this paper, we have attempted to look for significant mathematical patterns in corresponding data, if any. All the computational tasks have been performed using Python programs written by the first author of this paper.

3.1 Fitting Binomial Models to the Song *Jana-Gaṇa-Mana*

3.1.1 Verifying the Stability of Success Probability: *Laghu* versus *Guru*

The idea of fitting binomial model to Hindi poetic compositions is entirely a novel approach. A Bernoulli trial is defined as a random experiment with exactly two possible outcomes, success and failure, in which the probability of success is the same, every time the experiment is conducted. Let p and q denote the probabilities of success and failure respectively in a Bernoulli trial, where p and q are related by the fact: $q = 1 - p$. A poetic composition in Hindi may be viewed as a sequence of numeral symbols 1 and 2 only. The symbol 1 represents a *laghu* or short syllable and the symbol 2 represents a *guru* or long syllable. Restriction to only two values hints towards possible application of binomial model

to the corresponding data set. Each syllable in such a text may be termed as one Bernoulli trial provided the other conditions hold true. Going this way, occurrence of a *laghu* or *guru* is designated as a success or failure respectively. Next, the number of trials (n_i), success count (r) and corresponding probability ($p_i = r/n_i$) for each stanza in the candidate poetry is computed, where i denotes the stanza number. The population success probability p is computed as

$$p = \sum r_i / \sum n_i$$

which for the given data set is $368/551 = 0.667876$.

Table 2. Text of the song *Jana-Gaṇa-Mana*

Stanza#	Devanāgarī Script	Romanised IAST Transliteration
1	जन गण मन अधिनायक जय हे भारत भाग्य विधाता पंजाब, सिन्ध, गुजरात, मराठा द्राविड़, उत्कल, बंग विन्ध्य, हिमाचल, यमुना, गंगा उच्छल जलधि तरंग तव शुभ नामे जागे तव शुभ आशिष मांगे गाहे तव जय गाथा जन गण मंगल दायक जय हे भारत भाग्य विधाता जय हे, जय हे, जय हे जय जय जय जय हे	<i>jana gaṇa mana adhināyaka jaya he</i> <i>bhārata bhāgya vidhātā</i> <i>pañjāba, sindha, gujarāta, marāṭhā</i> <i>drāviḍa, utkala, baṅga</i> <i>vindhya, himācala, yamunā, gaṅgā</i> <i>ucchala jaladhi taraṅga</i> <i>tava śubha nāme jāge</i> <i>tava śubha āśiṣa māṅge</i> <i>gāhe tava jaya gāthā</i> <i>jana gaṇa maṅgala dāyaka jaya he</i> <i>bhārata bhāgya vidhātā</i> <i>jaya he, jaya he, jaya he</i> <i>jaya jaya jaya jaya he</i>
2	अहरह तव आह्वान प्रचारित, शुनि तव उदार वाणी हिन्दु बौद्ध शिख जैन पारसिक मुसलमान क्रिस्टानी पूरब पश्चिम आसे तव सिंहासन पाशे प्रेमहार हय गाथा। जनगण ऐक्य विधायक जय हे भारत भाग्य विधाता! जय हे, जय हे, जय हे जय जय जय जय हे	<i>aharaha tava āhvāna pracārita,</i> <i>śuni tava udāra vāṇī</i> <i>hindu bauddha śikha jaina pārasika</i> <i>musalamāna kristānī</i> <i>pūraba paścima āse</i> <i>tava siṃhāsana pāše</i> <i>premahāra haya gāthā </i> <i>janagaṇa aikya vidhāyaka jaya he</i> <i>bhārata bhāgya vidhātā!</i> <i>jaya he, jaya he, jaya he</i> <i>jaya jaya jaya jaya he</i>
3	पतन अभ्युदय वन्धुर पंथा, युग युग धावित यात्री, हे चिर सारथी, तव रथ चक्रेमुखरित पथ दिन रात्रि दारुण विप्लव माझे तव शंखध्वनि बाजे, संकट-दुख-श्राता, जन गण पथ परिचायक जय हे भारत-भाग्य-विधाता,	<i>patana abhyudaya vandhura paṁthā,</i> <i>yuga yuga dhāvita yātrī,</i> <i>he cira sārathī,</i> <i>tava ratha cakremukharita patha dina rātri</i> <i>dāruṇa viplava mājhe</i> <i>tava śaṅkhadhvani bāje,</i> <i>saṅkaṭa-dukha-śrātā,</i> <i>jana gaṇa patha paricāyaka jaya he</i> <i>bhārata-bhāgya-vidhātā,</i>

	जय हे, जय हे, जय हे, जय जय जय जय हे	<i>jaya he, jaya he, jaya he, jaya jaya jaya jaya he</i>
4	घोर तिमिर घन निविड निशीथ पीडित मुच्छित देशे जाग्रत दिल तव अविचल मंगल नत नत नयने अनिमेष दुस्वप्ने आतंके रक्षा करिजे अंके स्नेहमयी तुमि माता, जन गण दुखत्रायक जय हे भारत भाग्य विधाता, जय हे, जय हे, जय हे, जय जय जय जय हे	<i>ghora timira ghana niviṇa niśītha pīṇita murcchita deśe jāgrata dila tava avicala maṅgala nata nata nayane animeṣa dusvapne ātaṅke rakṣā karije aṅke snehamayī tumi mātā, jana gaṇa dukhatrāyaka jaya he bhārata bhāgya vidhātā, jaya he, jaya he, jaya he, jaya jaya jaya jaya he</i>
5	रात्रि प्रभातिल उदिल रविच्छवि पूरब उदय गिरि भाले, साहे विहङ्गम, पून्नो समीरण नव जीवन रस ढाले, तव करुणारुण रागे निद्रित भारत जागे तव चरणे नत माथा, जय जय जय हे, जय राजेश्वर, भारत भाग्य विधाता, जय हे, जय हे, जय हे, जय जय जय जय हे	<i>rātri prabhātīla udīla ravicchavi pūraba udaya giri bhāle, sāhe vihangama, pūnno samīraṇa nava jīvana rasa ḍhāle, tava karuṇāruṇa rāge nidrita bhārata jāge tava carane nata mātā, jaya jaya jaya he, jaya rājeśvara, bhārata bhāgya vidhātā, jaya he, jaya he, jaya he, jaya jaya jaya jaya he</i>

The claim of fitting binomial model can be established only if the computed population success probability is maintained even across the smaller units (stanza in our case) of the candidate poem. Further, the expected number of successes corresponding to each stanza is obtained as,

$$E_i(\text{laghu}) = n_i * p$$

where p is the population success probability.

All of the aforesaid computations have been summarized in Table 3.

Table 3. Laghu and gurudata in the poem *Jana-Gaṇa-Mana*

<i>Stanza</i>	Var na count	Observed laghu count	Observed laghu probability	Expected laghu count
<i>s(i)</i>	n_i	r_i	p_i	$E_i(\text{laghu})$
1	126	82	0.6507937	84.15245009
2	105	68	0.647619	70.12704174
3	106	73	0.6886792	70.79491833
4	105	72	0.6857143	70.12704174
5	109	73	0.6697248	72.79854809

The maximum and minimum of p_i are $\max(p) = 0.6886792$ and $\min(p) = 0.647619$ respectively. Mean p is given as

$mean(p) = \sum p_i$, over $i = 1, 2, \dots, 5 = 0.668506$.

A very small difference (0.0410602) between the minimum and maximum p values strongly indicates towards the stability of success probability p_i across the various stanzas of the candidate poem. This stability is verified through Chi-square goodness of fit test.

The probability of a *laghu* over the entire data set is

$$P(laghu) = Total\ laghu / (Total\ laghu + Total\ guru) = 0.667876$$

$$E(laghu)\ in\ a\ stanza = Total\ letters\ in\ stanza * P(laghu).$$

Hence, for each stanza S_i , the expected success count is,

$$E_i(laghu) = Total\ letters\ in\ stanza\ S_i * P(laghu).$$

The Chi-Square statistic (χ^2) is calculated as follows,

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

where $O_i = r_i$ and $E_i = E_i(laghu)$ are the observed and the expected values respectively for i^{th} stanza, i.e., S_i .

For our data set with 4 degree of freedom at 5% significance level, the calculated chi-square value is 0.23883 against the table chi-square value 9.488. Since the experiment passes the chi-square goodness of fit test, the distribution of *laghu* and *guru* in the given text in general (and across each stanza in particular) can be regarded as *binomial* as the probability of success remains constant.

3.1.2 Verifying the Stability of Success Probability: Laghu versus Laghu

As known, in context of *Devanāgarī*, a syllable is characterized either as an independent vowel or a vowel with one or more consonants. Further, whenever a short vowel is used in a verse separately or with one or more consonants, it is considered as a short syllable. For example, क [ka], कि [ki], कु [ku] etc. are short syllables. Here it must be noted that the constant क [ka] in its pure form is written as क् [k], that is, with a *halant* symbol beneath the letter क [ka]. Similarly, letter ख [kha] in its pure form is denoted as ख् [kh]. The same holds true for all the remaining basic consonants in Hindi alphabet. It must also be noted that, in contemporary context and practice, these consonants are very frequently written without the underneath *halant* symbol. In an unusual observation, short syllables may be dichotomized in two disjoint sets, the one consisting of syllables (labelled as type_1) क [ka], ख [kha], ग [ga], घ [gha] etc. such that, each of such is also an element in the Hindi-alphabet set. Another set consisting of syllables (labelled as type_2) like कि [ki] and कु [ku] which are not present in the Hindi-alphabet set. The idea of fitting a binomial model may again be exercised for the reasons mentioned in Section 3.1.1. Referring to the arguments made in that section, each short syllable in our candidate poem may be termed as one Bernoulli trial. Occurrence of a type_1 or type_2 *laghu* is designated as a success or a failure respectively. Next, the number of trials (n_i), success count (r) and corresponding probability ($p_i = r/n_i$) for each stanza is computed, where i denotes the stanza number. The population success probability p is computed as

$$p = \sum r_i / \sum n_i$$

which for the given data set is $305/368 = 0.828804$.

Further, the expected number of successes corresponding to each stanza is obtained as

$$E_i(laghu_type_1) = n_i * p$$

where p is the population success probability.

All the aforesaid computations are summarized in Table 4.

Table 4. Laghu versus laghudata in the poem Jana-Gaṇa-Mana

Stanza	Laghu	Observed Type-1 Laghu	Observed Type-1 Laghu probability	Expected Type-1 Laghu
$s(i)$	n_i	r_i	p_i	$E_i(laghu_type_1)$
1	82	71	0.865853	67.961928
2	68	58	0.852941	56.358672

3	73	58	0.794520	60.502692
4	72	56	0.777777	59.673888
5	73	62	0.849315	60.502692

Again, in order to fit a binomial model to *laghu* data, we need to verify the stability of above computed population success probability p .

The probability of a type_1 *laghu* over the entire data set is

$$P(\text{laghu_type_1}) = \text{Total laghu_type_1} / (\text{Total laghu_type_1} + \text{Total laghu_type_2}) = 0.828804.$$

$$E(\text{laghu_type_1}) \text{ in a stanza} = \text{Total laghuleters in stanza} * P(\text{laghu_type_1}).$$

Hence, for each stanza S_i , the expected success count is

$$E_i(\text{laghu_type_1}) = \text{Total laghuleters in stanza } S_i * P(\text{laghu_type_1}).$$

The Chi-Square statistic (χ^2) is calculated as follows,

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

where $O_i = r_i$ and $E_i = E_i(\text{laghu_type_1})$ are the observed and the expected values respectively for i^{th} stanza, i.e., S_i .

For our data set with 4 degree of freedom at 5% significance level, the calculated chi-square value is 0.550375 against the table chi-square value 9.488. The experiment passes the chi-square goodness of fit test as the probability of type_1 *laghu* for the entire data set is maintained in the individual stanzas as well. So, the distribution of type_1 and type_2 *laghu* in the candidate poem can be regarded as *binomial*.

3.1.3 Fitting a Binomial Model to Diacritic Marks Data in *Jana-Gaṇa-Mana*

In this section, we attempt to fit a binomial model over diacritic marks data in *Jana-Gaṇa-Mana*. In Hindi, a vowel can be written either independently or as a *mātrā* (diacritic mark) along with a consonant letter. A *mātrā* is essentially a form of vowel which is used to represent it, as per the requirement, in any text. For example, the word आईए [*āīe*] is made up of three independent vowels आ [*ā*], ई [*ī*], and ए [*e*]. The word नीला [*nīlā*], consists of two syllables नी [*nī*] and ला [*lā*]. In this word, the vowel ई [*ī*] is used as diacritic mark ि [*ī*] with the consonant न [n] to form the first syllable नी [*nī*]. Similarly, the vowel आ [*ā*] is used as diacritic mark ः [*ā*] with the consonant ल [la] to yield the latter syllable ला [*lā*].

Given below is Table 5 of vowels in Hindi-alphabet along with their corresponding diacritical forms.

Table 5. Vowels with their diacritical forms

Vowel	अ[a]	आ[ā]	इ[i]	ई[ī]	उ[u]	ऊ[ū]	ए[e]	ऐ[ai]	ओ[o]	औ[au]	अं[am]
Diacritic Mark	NA	ः	ि	ी	ु	ू	े	ै	ो	ौ	ं

As the vowel अ [a] is implicit to all the consonants it does not have an equivalent diacritic mark. That is the reason, it is frequently referred to as an inert vowel.

Table 6. Diacritic marks data in *Jana-Gaṇa-Mana*

Vowel as Diacritic Mark (→)	ः	ि	ी	ु	ू	े	ै	ो	ौ	ं	ं
Stanza (↓)											
1	25	09	×	04	×	10	×	×	×	08	×
2	19	10	02	03	01	08	02	×	01	02	×
3	15	09	02	07	×	09	×	×	×	04	×
4	10	13	03	04	×	14	×	01	×	03	×
5	16	11	02	01	02	12	×	01	×	01	×

The Diacritic marks data in the poem *Jana-Gaṇa-Mana* is provided in Table 6. In order to fit a binomial model, occur-

rence of a vowel as a diacritic mark (VADM) is designated as a success, whereas, occurrence of an independent vowel or consonant is designated as a failure. The number of trials (n_i), success count (r) and corresponding probability ($p_i = r/n_i$) for each stanza in the poetry is computed, where i denotes the stanza number. The population success probability p is computed as

$$p = \sum r_i / \sum n_i$$

which for the given data set is $244/551 = 0.442831$.

The expected number of successes corresponding to each stanza is obtained as

$$E_i(VADM) = n_i * p$$

where p is the population success probability.

All the aforesaid computations are summarized in Table 7.

Table 7. Diacritic marks data summary

Stanza	trials	success	Observed success probability	Expected success count
$s(i)$	n_i	r_i	p_i	$E_i(VADM)$
1	126	56	0.444	55.7967
2	105	48	0.457	46.4972
3	106	46	0.433	46.9401
4	105	48	0.457	46.4972
5	109	46	0.422	48.2686

A very small difference of 0.035 between the minimum and maximum p values strongly indicates towards the stability of success probability p_i across the various stanzas of the candidate poem and the same is verified through Chi-square goodness of fit test.

The Chi-Square statistic (χ^2) is calculated as follows,

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

where $O_i = r_i$ and $E_i = E_i(laghu)$ are the observed and the expected values respectively for i^{th} stanza, i.e., S_i .

For our data set with 4 degree of freedom at 5% significance level, the calculated chi-square value is 0.2233 against the table chi-square value 9.488. The experimental result suggests binomial model for the diacritic marks data set corresponding to the poem *Jana-Gana-Mana*.

The various statistical definitions have been referred from (see Gupta and Kapoor [19]; Pal and Sarkar [20]).

4. Concluding Remarks

The statistical analysis performed over numerical data corresponding to the candidate poem provides valuable insights pertaining to its stylistic features. Such a feat might find its audience equally in the fields of arts as well as mathematics. The endeavour to seek for significant mathematical patterns in the corresponding data may be deemed as satiated as the experiments aimed at fitting binomial model to the various attributes of the candidate poem have been successful.

The data, in terms of short and long syllables, corresponding to the candidate poem is successfully fitted to a binomial model. Similarly, short syllables when dichotomized in two disjoint sets based on the properties mentioned in the corresponding section, have been found to follow a binomial model. Finally, diacritic marks data in *Jana-Gana-Mana* is successfully as well as convincingly fitted to a binomial model.

The up-and-comer poem may be investigated further to get additional statistical patterns so as to writing style of the poet may be established more prominently. Such findings may be useful in resolving contentious issues in authorship attribution problem.

The poem *Jana-Gana-Mana* has its own importance and significance for a number of reasons cited in the introduction section of this paper. After it was composed, over a span of more than hundred years, several other Hindi poets wrote poems, influenced by it. The likeness degree of such poems with the current one is liable to be explored. Along these lines, this study may give a base to any such future examinations.

References

- [1] National Portal of India. (2020). National anthem. <https://www.india.gov.in/india-glance/national-symbols>.
- [2] Chakraborty, S., Sarkar, S., Tewari, S., and Pal, M. (2012). Modeling India's national anthem: a statistical approach. *International Research Journal of Social Sciences*, 1(2), 17-24.
- [3] Das, S. K. (2002). *The English Writings of Rabindranath Tagore, Vol III, A Miscellany*. Sahitya Akademi.
- [4] Gokak, V. K. (1961). *Tagore's influence on modern poetry, Indian Literature*. Sahitya Akademi.
- [5] Hassan, S., Mohammed, N., and Momen, S. (2016). Learning from tagore. In: *International Workshop on Computational Intelligence*, Dhaka, Bangladesh, 12-13 December 2016, pp. 79-84. IEEE.
- [6] Rakshit, G., Ghosh, A., Bhattacharyya, P., and Haffari, G. (2015). Automated analysis of Bangla poetry for classification and poet identification. In: *12th International Conference on Natural Language Processing* (eds D. S. Sharma et al.), Trivandrum, India, December 2015, pp. 247-253. India: NLP Association of India.
- [7] Deshmukh, R., et al. (2019). Marathi poem classification using machine learning. *International Journal of Recent Technology and Engineering*, 8(2), 2723-2727.
- [8] Kaur, J. and Saini, J. R. (2020). Designing Punjabi poetry classifiers using machine learning and different textual features. *International Arab Journal of Information Technology*, 17(1), 38-44.
- [9] Nayak, S. K. and Panigrahi, A. K. (2018). Modern poetic trends of Odisha versus the neoteric strives of Guruprasad. *International Journal of Trends in Scientific Research and Development*, 11(2), 1140-1160.
- [10] Joshi, B. K. and Kushwah, K. K. (2016). Sandhi: the rule-based word formation in Hindi. *International Journal of Computer Science and Information Security*, 14(12), 781-785.
- [11] Magdum, D., Patil, T., and Suman, M. (2019). Schwa deletion in Hindi language speech synthesis. *International Journal of Engineering and Advanced Technology*, 8(6S), 211-214.
- [12] Bafna, P. and Saini, J. R. (2020). On exhaustive evaluation of eager machine learning algorithms for classification of Hindi verses. *International Journal of Advanced Computer Science and Applications*, 11(2), 181-185.
- [13] Pal, K. and Patel, B. V. (2020). Model for classification of poems in Hindi language based on ras. In: Somani, A., Shekhawat, R., Mundra, A., Srivastava, S., Verma, V. (eds.), *Smart Systems and IoT: Innovations in Computing*. Singapore: Springer, pp. 655-661.
- [14] Pandian, A., Maurya, P., and Jaiswal, N. (2020). Author identification of Hindi poetry. *International Journal of Scientific and Technology Research*, 9(3), 3791-3795.
- [15] Kushwah, K. K. and Joshi, B. K. (2017). Rola: an equi-matrikchhand of Hindi poems. *International Journal of Computer Science and Information Security*, 15(3), 362-364.
- [16] Kumar, V. (2018). *Vrihat Vyakarana Bhaskar*. Bharati Bhawan.
- [17] Pandey, S. (2015). *Chhand Manjari (1st ed.)*. Anjuman Prakashan.
- [18] Dvivedi, K. D. and Singh, S. L. (2013). *The Prosody of Pingala*. Varanasi: Vishwavidyalaya Prakashan.
- [19] Gupta, S. C. and Kapoor, V. K. (2002). *Fundamentals of Mathematical Statistics (11th ed.)*. Sultan Chand & Sons.
- [20] Pal, N. and Sarkar, S. (2009). *Statistics: Concepts and Applications (2nd ed.)*. PHI.