CHAPTER IV

TEST OF RANDOMNESS OF WORD-LENGTH AND SENTENCE-LENGTH
4.1 Introduction:

The length of words of a given text recorded in the normal reading order, gives rise to a series generally known as a word-length series. Word-length may be measured in terms of syllables or phonemes or letters. The randomness of such series which forms the basis of a number of investigations about authors characteristics was discussed by Fucks (1954). He measured word-length in syllables and carried out many important calculations such as (i) the correlation between lengths of consecutive words and (ii) correlation between lengths of words which are not consecutive. He found the auto-correlation coefficient of the first order to be insignificant for German and English literature by considering randomly selected pieces of work. Herdan (1962) generalised the above result to the extent that randomness will not be disturbed even if sampling is of chapters or pages or of some other suitable units. Bhattacharya (1965) studied
the randomness of the series of word-length in a number of works in Bengali prose and estimated the auto-correlation coefficients of various orders. He also examined the randomness of sentence-length in words by employing certain statistical tests.

Mazumder (1985) tested the randomness of word-length in syllables and sentence-length in words occurring in the prose works of Assamese literature, and reached the conclusion that autocorrelation coefficients of order up to seven are all near zero for word-length and are quite small, for sentence-length series.

In this chapter, we test the randomness of word-length in syllables as well as that of sentence-length in words for the prose work of Sanskrit literature. The sample data giving the word-lengths and sentence-lengths series are based on simple random sampling. In fact, twenty randomly selected pages from each of four selected books have been used for the test of randomness. The autocorrelations of various orders between successive values of word-length as well as sentence-length have been obtained by the well known formula:

\[
\xi_L = \frac{\sum_{t=1}^{N-L} Z_t Z_{t+L} - \left( \sum_{t=1}^{N-L} Z_t \right) \left( \sum_{t=L+1}^{N} Z_t \right)}{\left( \sum_{t=1}^{N-L} Z_t^2 - \left( \sum_{t=1}^{N-L} Z_t \right)^2 / N-L \right) \left( \sum_{t=L+1}^{N} Z_t^2 - \left( \sum_{t=L+1}^{N} Z_t \right)^2 / N-L \right)^{1/2}}
\] (4.1.1)
where \( r_L \) represents the autocorrelation coefficient of order \( L \) and \( \{ z_t \} \) is the series of the values of the characteristic under consideration, \( N \) is the total number of values in the series.

Since in our investigation neither word-length nor sentence-length distribution of any text under consideration follows normal law, a test of randomness based on serial correlation coefficient for non-normal parent population given by Von Neumann (1941) has been applied.

Von Neumann used the statistic

\[
\frac{\delta^2}{s^2} = \frac{\sum_{t=1}^{N-1} (z_{t+1} - z_t)^2}{(N-1)} \]  

(4.1.2)

which follows asymptotic normal distribution with

Mean = \( 2N / (N-1) \) and

Variance = \( \frac{4n^2(N-2)}{(n-1)^3(n+1)} \)  

(4.1.3)

The statistic

\[
U = \frac{\delta^2/s^2 - 2N/(N-1)}{[4n^2(N-2)/(N-1)^3(n+1)]} 
\]  

(4.1.4)
which then follows standard normal distribution could be conveniently held for the test.

4.2 Test of Randomness of Word-Length Series

In this section, we would examine the randomness of word-lengths. We assume that all the words occurring in the twenty randomly selected pages of each book are taken in the order of their occurrence in the sample and their length in syllables are recorded in the same order. This gives a series of word-length within sample. The null hypothesis is that the series of word-lengths from a particular book is in random order. This hypothesis of randomness is equivalent to that of no dependence, that is zero autocorrelation coefficient.

In order to see, the dependence of the words in a given series, we calculate the autocorrelation coefficients of different orders by measuring lengths of words in syllables of the sampled data consisting of 1188 words from Book 1 (Kādambarī), 2112 words from Book 2 (Tilakmaṇjarī), 1793 words from Book 3 (Mandārmaṇjarī) and 1749 words from Book 4 (Sivarājvijyām). The autocorrelation coefficients of order 1 to 7 are presented in the following Table 4.1.
Note: These computations were performed on an HCL Computer available in the Department. The program can be made available on request.

<table>
<thead>
<tr>
<th>Number</th>
<th>Autocorrelation Coefficients ($r_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.992 0.9944 0.9968 0.9990 1.0000</td>
</tr>
<tr>
<td>2</td>
<td>2.221 0.9650 0.9837 0.9914 0.9974</td>
</tr>
<tr>
<td>3</td>
<td>3.179 0.8340 0.9686 0.9912 0.9995</td>
</tr>
<tr>
<td>4</td>
<td>6.194 0.7360 0.9320 0.9962 0.9999</td>
</tr>
</tbody>
</table>

*Table 4.1: Autocorrelation Coefficients of Word-Length Series.*
The values of $r_L$, given in the Table 4.1, reveal that in most of the cases the autocorrelation coefficient of different orders for individual book varies between $0.979$ to $0.1714$ excepting an extreme values viz. $0.2934$. For authentic conclusions, regarding its significance, we apply the non-parametric test of randomness based on statistic $U$ given in (4.1.4). The values of the test statistic $U$ have been estimated for each book and these are found to be $1.66284$, $0.924063$, $0.00223$ and $3.2897$, for Book 1, Book 2, Book 3 and Book 4 respectively. Comparing all the values with $1.96$ and $2.58$, it is found that at $5\%$ and $1\%$ probability level, the null hypothesis may be accepted for first three books and rejected for Book 4.

4.3 TEST OF RANDOMNESS OF SENTENCE - LENGTH:

All the sentences occurring in the twenty randomly selected pages of each book are calculated in their order of occurrence in the sample and their length in words are recorded in the same order. This gives a series of sentence-lengths. Since the sampled data are from randomly selected pages rather than from continuous pages, the observations for autocorrelation coefficient have been constructed in such a way that the number of words in the first sentence of any page is not matched with that of the last sentence of previous page. The same procedure for estimating the
<table>
<thead>
<tr>
<th>Sample size</th>
<th>$r_1$</th>
<th>$r_2$</th>
<th>$r_3$</th>
<th>$r_4$</th>
<th>$r_5$</th>
<th>$r_6$</th>
<th>$r_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>0.432051</td>
<td>0.195429</td>
<td>0.005041</td>
<td>-0.03835</td>
<td>-0.03687</td>
<td>-0.13899</td>
<td>-0.148566</td>
</tr>
<tr>
<td>51</td>
<td>0.033479</td>
<td>-0.03372</td>
<td>-0.19845</td>
<td>-0.100012</td>
<td>-0.001544</td>
<td>0.22111</td>
<td>0.03634</td>
</tr>
<tr>
<td>42</td>
<td>0.030096</td>
<td>0.106421</td>
<td>-0.019315</td>
<td>-0.074413</td>
<td>0.076419</td>
<td>0.24166</td>
<td>-0.214348</td>
</tr>
<tr>
<td>126</td>
<td>-0.041528</td>
<td>-0.084929</td>
<td>0.10424</td>
<td>-0.012061</td>
<td>-0.022535</td>
<td>-0.03665</td>
<td>0.132791</td>
</tr>
</tbody>
</table>

Table 4.2: Autocorrelation coefficients ($r_i$) of sentence-length.
autocorrelation coefficient has been repeated in case of Book 2, Book 3 and Book 4. The autocorrelation coefficient of order one to seven of the sentence-length in words have been estimated by applying the equation (4.1.1) and presented in Table 4.2.

After examination of the values in Table 4.2, it is found that in most of the cases the autocorrelation coefficient of different orders are below (0.24) in magnitude except $r_1$ for books. The same test of randomness used for word length series has been applied here for sentence length series.

The statistic $U$ has been estimated for different books and found to be 0.40193, 0.2579175, 0.0656202 and 0.314699 respectively. All these values are insignificant at both 1% and 5% probability levels. It seems that our hypothesis may be accepted. Hence we arrive at the conclusion that the nature of the sentence length series is random.

4.4 AUTO-CORRELATION COEFFICIENTS OF WORD-LENGTH WITHIN SENTENCE:

In this last section, we have examined the nature of word-length series within sample. Here we shall study word-length series within sentences. For this purpose, 70, 54, 52 and 126 sentences from twenty randomly pages of four
experimental texts were subjected to test. From each sentence, the length of words in syllables were recorded and the autocorrelation coefficient of first order i.e. $r_1$ for each sentence were computed for all the four books. The range of the variation of $r_1$ for different texts are presented in the following Table.

**TABLE 4.3 : AUTOCORRELATION COEFFICIENTS OF WORD-LENGTH WITHIN SENTENCE.**

<table>
<thead>
<tr>
<th>Book</th>
<th>Range of $r_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.99024 to 0.559511</td>
</tr>
<tr>
<td>2</td>
<td>-0.711237 to 0.533129</td>
</tr>
<tr>
<td>3</td>
<td>-0.70993 to 0.523507</td>
</tr>
<tr>
<td>4</td>
<td>-0.693375 to 0.644602</td>
</tr>
</tbody>
</table>

From the above Table, it is apparent that the autocorrelation coefficients between word-length within sentences are quite high in magnitude for all the books. Hence, it may be concluded that there is a dependence between word-lengths within a sentence.