Mausam, (1994), **45**, 1,75-78

551.577: 551.507.777

On the use of entropy in Markovian model of rainfall

S. K. MUKHOPADHYAY

Bidhan Chandra Krishi Viswavidyalaya,
Pundibari, Cooch Behar (W.B.)

(Received 27 August 1991, Modified 7 January 1993)

सार — कूच बिहार जिले की 1901 से 1988 की अवधि की वार्षिक वर्षा का सांख्यिकीय अध्ययन मार्कोव चेन मॉडल के आधार पर किया गया है। इस अध्ययन में एक पद 3×3 मार्कोव चेन मॉडल का प्रयोग किया गया है । इस मॉडल के उपयोग से इस जिले के दो केन्द्रों पर सामान्य, खराब और अच्छी वर्षा के वर्षों का पता चला है । एन्ट्रोपी के प्रयोग द्वारा स्वतंत्र परिकल्पना की जांच की गई और संभाव्यता अनुपात सिद्धान्त का प्रयोग करते हुए इसकी पुष्टि की गई। दोनों पद्धतियों से यह समान निष्कर्ष निकलता है कि जिले के दो केन्द्रों की वार्षिक वर्षा को एक दूसरे से स्वतंत्र माना जाए ।

ABSTRACT. A statistical study of the annual rainfall data at Cooch Behar district during the period 1901 to 1988 has been undertaken by using Markov chain model. One step 3×3 Marcov chain model has been used in this study. The outcomes of the model reveal normal, bad and good year of rainfall at the two stations of this district. The hypothesis of independence has been tested on the use of entropy and it has been verified using likelihood ratio criterion. The results of the two methods are the same that the yearly rainfall occurrence may be regarded as independent at the two places of the district.

Key words — Transition probability, Markov chain, Ergodic, Stationary, Stochastic matrix, Entropy, Markovian dependence.

1. Introduction

In the past decades, numerous authors utilizing stochastic model have presented the analysis of precipitation data for different places. With respect to this precipitation phenomenon, a major trust of this stochastic approach has been used to predict the behaviour of spell distribution and subsequently weather cycles under the validity of Markovian assumption. Mostly the work in this direction is confined to the two-state Markov chain model, considering only two phases of precipitation phenomenon, *i.e.*, wetting and drying. But, it is also natural to consider that some annual rainfall occurs around the normal or average rainfall over a long period of rainfall time series. Considering in this context, three outcomes, *i.e.*, normal, bad and good year of rainfall have been utilized here to develop the stochastic model. The Markov chain model, in terms of rainfall occurrence, has been considered to test the hypothesis of independence by likelihood ratio criterion or multinominal analogy. But the concept of entropy may also be applied to test the hypothesis of independence against Markovian dependence.

The present study is intended to test the hypothesis of independence of Markov chain model which is based upon the concept of entropy as well as the likelihood approach.

2. Data used

In this study, the 88 years (1901-1988) of annual rainfall data have been used for the two raingauge stations, Cooch Behar and Dinhata. Eleven years moving average method has been applied to obtain the missing monthly data.

3. Methodology

For the purpose of the present study the annual rainfall data are standardised by subtracting the mean and dividing by the standard deviation of the climatic time series. The standardised data would have high positive and negative values for the good and bad rainfall years respectively. The moderate positive and negative values would indicate the normal years. So, the standardised data have been used to identify the normal, bad and good years of rainfall of the two raingauge stations, Cooch Behar and Dinhata. The standardised values are used to classify the intensity of rainfall years as follows :

−0.49 to +0.49	Normal year (N)
−0.50 to −4.0	Bad year (B)
+0.50 to +4.0	Good year (G)

Similar method has been used by Ogallo (1986) to identify the climatic fluctuations.

The three states of rainfall (normal, bad and good) are the exhaustive and mutually exclusive states of Markov chain model at any time. We also assume that the chain is ergodic which is necessary to provide the asymptotic normality of the frequency counts which yields the asymptotic distribution of the likelihood ratio statistic. The mode of classification of annual rainfall yield the nine cell frequencies which are regarded as the three-state Markov chain. So, each year may be classified as one of the following nine possibilities :

Classification of yearly rainfall	*Probability* (p_{ij})
(*i*) A normal year preceded by a normal year	P_{11}
(*ii*) A normal year preceded by a bad year	P_{12}
(*iii*) A normal year preceded by a good year	P_{13}
(*iv*) A bad year preceded by a normal year	P_{21}
(*v*) A bad year preceded by a bad year	P_{22}
(*vi*) A bad year preceded by a good year	P_{23}
(*vii*) A good year preceded by a normal year	P_{31}
(*viii*) A good year preceded by a bad year	P_{32}
(*ix*) A good year preceded by a good year	P_{33}

Thus the nature of annual rainfall is classified as one of the nine possibilities depending upon the nature of previous year's annual rainfall (ignoring the initial year's annual rainfall). Repeating this process, each cell frequency is obtained. But the cell frequencies of the nine possibilities are denoted by n_{ij}

where, i & $j \in S$, $S = (1, 2, 3)$

S is the state-space. And the transition is,

$i \leftarrow j$.

These cell frequencies are arranged in a form of matrix for each annual rainfall series separately.

The transition probabilities (p_{ij}) are obtained by dividing the frequencies of one state of occurrence to the total frequency of that state of occurrence. The maximum likelihood method has been applied to obtain the estimation of each transition probability.

So, one step 3×3 transition probability matrix has been formed for each of the rainfall time series. This is 3×3 square matrix with non-negative element and each row sum is equal to unity.

Same method has been applied to estimate the probability of individual state of occurrence (p_i) and is given by

$$p_i = \frac{n_i}{n}$$

$$\text{where,} \quad n = \sum_i n_i$$

The mathematical model of Shannon has been used to measure the entropy of the stationary or initial probability as well as the transition probabilities of the respective informatrices as suggested by Theil (1973) and Bhattacharya and Waymire (1990).

The entropy of the stationary distribution is, H_1

$$H_1 = -\sum_i p_i \log p_i$$

The entropy of the pair observations is,

$$H_2 = -\sum_{i,j} p_i\, p_{ij} \log (p_i\, p_{ij})$$

The values of p_i and p_{ij} are estimated by maximum likelihood estimators as mentioned earlier.

If the sequence of the random variables are independent against Markovian dependence, *i.e.*,

H_0 : $p_{ij} = p_j$ then we have,

$H_2 = 2H_1$

The procedure have been suggested by Basawa and Rao (1980).

The accuracy and validity of this technique have been verified using the likelihood ratio criterion. The hypothesis of independence of annual rainfall, of the

TABLE 1

Measure of entropy and T_1 values

Variable	Cooch Behar	Dinhata
H_1	0.40	0.407
H_2	0.79	0.80
$H_2 - 2H_1$	$-.01$	$+.01$
T_1	2.61 (NS)	2.45 (NS)

NS — Not significant

two stations within the stationary Markovian dependence, has been tested by the following likelihood ratio criterion:

$$T_1 = 2 \sum_{i,j} n_{ij} \log \frac{n \, . \, n_{ij}}{n_i \, . \, n_j}$$

which has the limiting Chi-square distribution with $(m-1)^2$ degree of freedom. And the test statistic T_1 has been regarded as the appropriate test of independence against Markovian dependence (Medhi 1983).

4. Results and discussion

The transition matrices are so obtained from the two climatic time series. The matrices, p_{ij} (c) and p_{ij} (d) for Cooch Behar and Dinhata respectively are given below :

$$p_{ij}(c) = \begin{vmatrix} N & B & G \\ .13 & .62 & .25 \\ .06 & .48 & .46 \\ .12 & .61 & .27 \end{vmatrix} \qquad p_{ij}(d) = \begin{vmatrix} N & B & G \\ .13 & .75 & .12 \\ .10 & .52 & .38 \\ .08 & .41 & .51 \end{vmatrix}$$

The probabilities of individual state of occurrence for the two stations are presented by $p_i(c)$ and $p_i(d)$ respectively. These are given below :

	N	B	G
$p_i(c) =$	.09,	.53,	.38
$p_i(d) =$	.09,	.46,	.45

Among the probability of individual state of occurrence the bad rainfall year is larger than the other two outcomes of the rainfall for both the places.

The entropy of the probability of individual state of occurrence is computed and the values of H_1 for the two places are shown in Table 1. The entropy of the pair observations for the transition probabilities and the probability of individual state of occurrence are obtained and the values of H_2 are also shown in Table 1.

The values of $H_2 = 2H_1$ being very-very small, the hypothesis that the variables are independent, could be considered as insignificant for the two places.

The values of T_1 statistic are also shown in Table 1 for the two transition matrices. The values are found to be insignificant at the 5% level with four degrees of freedom for both the places. So, in this case the variables satisfy the hypothesis of independence.

The validity of the technique of entropy is verified using the likelihood ratio test. The result is the same. It may be concluded that the annual rainfall occurrence could be treated as independent at the district of Cooch Behar.

5. Conclusion

The two methods may be applied for the test of independence against the Markovian dependence. But more study will be required to come to this conclusion for different places.

Acknowledgement

The author would like to thank India Meteorological Department, Pune and C.T.R.I., Dinhata, Cooch Behar for supplying the meteorological data.

References

Bhattacharya, R. N. and Waymire, E.C., 1990, Stochastic processes with applications. John Wiley & Sons, New York.

Basawa, I.V. and Rao, B.L.S.P., 1980, Statistical inference for stochastic process, Academic Press, London.

Medhi, J. 1983, Stochastic processes., Wiley Eastern Ltd., Calcutta.

Ogallo, L.A.J., 1986, "Stochastic modelling of regional annual rainfall anomalies in East Africa, "*J. Appl. Statist.* , **13**, 1, pp. 49-56.

Theil, H., 1973, Principles of Econometrics, North Holland Publishing Company, London.