CHAPTER - VI

MEASURE OF UNCERTAINTY ON DAILY RAINFALL

CHAPTER - VI

MEASURE OF UNCERTAINTY ON DAILY RAINFALL

6.1. INTRODUCTION :

Many scientists and meteorologists have studied the probabilities of the occurrence of dry and wet days only by fitting a Markov chain model and geometric and Markovian-geometric spell distribution. But the studies of uncertainty about the nature of a day's rainfall are rarely available. Every probability distribution has some uncertainty associated with it. The concept of entropy is introduced here to provide a quantitative measure of this uncertainty and to test the stochastic dependence.

The main objectives of this chapter are :

i) To examine the quantitative values of degree of uncertainty.

ii) To study the nature of rainfall event by Markov chain method.

6.2. STOCHASTIC MATRICES FOR EACH MONTH

A one step 5 × 5 transition probability matrix from one transitional state to another has been formed for each of the rainy months, May to September from the frequency of the occurrence of daily rainfall amount at Cooch Behar.

The nature of the daily rainfall has been classified according to the intensity of daily rainfall. The categories are given below :

Category No.	Rainy days	Intensity of daily rain-fall (in m.m.)
0	Non-rainy day	0 - 2.4
1	Light rainy day	2.5 - 10.0
2	Moderate rainy day	10.1 - 50.00
3	Heavy rainy day	50.1 - 124.9
4	Very heavy rainy day	125 and above

The mode of classification yields a sequence of non-rainy, light rainy, moderate rainy, heavy rainy and very heavy rainy days which can be regarded as a five-state Markov chain. Thus, for each year, transition to 1st May is classified as one of the twenty five possibilities depending on the weather of 30th April which is a non-rainy day or other type of the nature of rainy days. Repeating the process each year, the cell frequencies for the above twenty five possibilities are obtained. Here we use the daily rainfall data from 1971 to 1988. In this study the daily rainfalls are utilized from 30th April to 30th September for each year. So, each year consists of 153 observations and the total number of observations is 2754.

Let the cell frequencies of these twentyfive possibilities be denoted by n_{ij} where

i & $j \in S$; $S = (0,1,2,3,4)$ and $i \rightarrow j$

where 'S' is the state-space.

These frequencies are arranged in the form of a matrix for each monsoon month i.e., June, July, August, September and one pre-monsoon month i.e. May. The cell frequencies are arranged in matrix form for the five months which are given in Table 6.1. The conditional probabilities which can be estimated by maximum likelihood method are defined as

$$P_{ij} = P\left\{ X_k = j \mid X_{k-1} = i \right\}$$

$$i \,\&\, j \in S$$

$$S = (0,1,2,3,4)$$

The conditional probabilities p_{ij} are calculated by dividing the cell frequencies of one state by the total frequencies of that state. This yields the estimation of each cell probability and the cell probabilities are given by

$$P_{ij} = \frac{n_{ij}}{n_i}$$

where $$n_i = \sum_{j=0}^{4} n_{ij}$$

subject to the restrictions

$$\sum_j P_{ij} = 1 , \quad \text{for each } i$$

and $$P_{ij} \geq 0 \quad \text{for all } i \,\&\, j.$$

Each row constitutes the probability vector and it is convenient to give the set of all such vectors as a matrix. This matrix is known as the transition probability matrix or the matrix of transition probabilities of the Markov chain or simply the Markov matrix. This is a square matrix (5×5) with non-negative elements and unit row sum.

The one-step transition probability matrix for each monsoon month and one pre-monsoon month are given in Table 6.2(a).

The estimate of the probability of individual state of occurrence (π_i) is given by

$$\pi_i = \frac{n_i}{n} \qquad \text{where} \quad n = \Sigma\, n_i,\ \Sigma\pi_i = 1,\quad i \in S.$$

The values of n_i and n are given in Table 6.2(b) and the values of π_i are given in Table 6.2(b).

6.3. REDUNDANCY TEST

The concept of entropy given by Shannon,in his mathematical model serves as a measure of uncertainty of the transition probabilities (P_{ij}) and this is defined by

$$H_i = - \sum_{j=0}^{4} P_{ij} \log_{10} P_{ij}$$

for each i.

Where H_i denotes the entropy of the ith state of occurrence.

The values of H_i for different states and for different months are given in Table 6.3.

The weighted entropy value for each of the months is calculated as a sum of the entropy values of all the different categories with the probabilities of the corresponding states.

The weighted entropy H can be calculated as

$$H = - \Sigma\, \pi_i \quad H_i$$

Where π_i is the probability of individual state of occurrence i.e.in the ith state.

And H_i is the entropy of the i th state. The values of the weighted entropy are given in Table 6.3.

The weighted values over the transition probabilities during the monsoon months and the pre-monsoon month have showed a typical nature. The weighted entropy value for the month of July is the highest and it is followed by June. It is observed that

June and July are considered to be the most active monsoon months in Cooch Behar district. The most interesting matter is that the weighted entropy for the month of May has a significant role for May to be considered as a monsoon month. The weighted entropy of this month is likely to be equivalent to that of the month of August. Now we may consider May as the beginning month of monsoon at Cooch Behar.

The entropy of the stationary distribution i.e. the entropy of the individual state of occurrence is given by

$$H_{ii} = -\Sigma\ \pi_i \log\pi_i$$

Where π_i is the probability of the ith state of occurrence.

The values of the entropy of the individual state of occurrence have been given in Table **6.3**. The entropy has an important use in measuring the uncertainty as well as in testing the hypothesis of Markov dependence.

A measure of uncertainty,M,of the stationary model is obtained from the individual states of occurrence over the Markovian model in the system and is given by

$$M = H_{ii} - H = \left(-\Sigma\ \pi_i \log\pi_i \right) - \Sigma\ \pi_i\ H_i$$

The values of M for all the months are given in Table **6.3.**

The measure of uncertainty of stationary model and that of the Markovian model have some difference which is very small,almost negligible. The difference in uncertainty between the Markovian model and the stationary model at Cooch Behar lies between + 8% to −12 percent.

The redundancy of the state of occurrence, R, is obtained as the difference from one of the ratio of the weighted entropy value H to the maximum possible entropy (H_{max})

Here, H_{max} = Log5

as we have here only 5 states of occurrence as Thail (1973).

So,

$$R = 1 - \frac{H}{H_{max}}$$

This redundancy value is used to determine the favourableness or unfavourableness of the Markovian system. As the redundancy value, R, tends to 1, the Markovian system tends to maximum favourable condition i.e. almost certain. Now on the light of this argument, we may examine the Markovian dependency on the monsoon months or rainy months. From Table 6.3 it is observed that the redundancy value of the month of May is the maximum but it is very low in comparison to the value one. The redundancy values of the monsoon months lie between 0.14 and 0.29. Considering these values, we may conclude that the one-day dependence cannot be considered as explaining the rainfall pattern at Cooch Behar.

In the next sub-section, we would verify this conclusion by adopting another method to test the Markovian dependence through informatrix of the rainy months.

6.4. LIKELIHOOD-RATIO TEST BY ENTROPY

In this section we use the informational measure to test the hypothesis of Markovian dependence.

The mathematical model of Shannon is used to obtain the measure of entropy in individual state of occurrence which we have calculated in the previous section, denoted by H_{11} and also the transitional probabilities of informatrices. We have H_{11} as given in Table 6.3.

The average conditional uncertainty can be measured by

$$H_{21} = - \Sigma \pi_i p_{ij} \log p_{ij}$$

$$= \frac{1}{n} \left\{ \Sigma n_i \log n_i - \Sigma n_{ij} \log n_{ij} \right\}$$

This is the same as the weighted entropy which is stated in the previous section. These values are also given in Table 6.4 for each of the months.

The hypothesis testing, involving Markov chains has been considered by several ways. Mainly the Chi-square and the likelihood ratio-criterion have been used for testing the hypothesis of independence of the random variable. Here we introduce a test criterion which involves the entropy but this is equivalent test of the likelihood-ratio criterion. The test statistic is $T_1 = 2n\ (H_1 - H_{21}) = 2\ \Sigma\ n_{ij} \log \frac{n_{ij}}{n_i n_j / n}$

$$= 2\ \Sigma\ n_{ij} \log \frac{n . n_{ij}}{n_i n_j}$$

That is T_1 is the same as the likelihood-ratio criterion. The test statistic has a limiting Chi-square distribution with $(m - 1)^2$ degrees of freedom (here m = 5) and the large values of the statistic correspond to rejection of the hypothesis.

Before applying this test, we now set up the null hypothesis: the occurrence of one day rainfall is independent against Markov dependence. The test statistic has been applied to each of the months separately. The test procedure followed here has been described in Basawara & Rao (1980). Here we have to indicate that the total numbers of observations (n) are different for different months. The total number of observations for each ofthe months of May, July and August is (31×18) = 558 , ignoring the initial observation,the last day of April. And the total number of observations for each of the months of June and September is (30 × 18) =540 each. These total numbers of observations of each month are given in Table **6.4.**

The tabulated value of Chi-square with $(m-1)^2 = 16$ degrees of freedom,at 5% level of significance is 26.30 and that at 1% level is 32.00. It is also shown in Table **6.4.** The calculated value of T_1 for each month is also shown in the same table. The computed value of T_1 for the month of May is 19.64 which is less than the theoretical value of Chi-square with 16 degrees of freedom at 5% level of significance. This is non-significant. Therefore,we cannot reject the null hypothesis . Hence,null hypothesis is accepted. Then we may come to the conclusion that the occurrences of daily rainfall in the month of May are really independent. In the month of May, the day's precipitation does not depend on the precipitation of the previous day. Hence,we may say that the daily weather occurs randomly in the month of May at Cooch Behar.

Now we consider the case of active monsoon months which indicates an interesting result. The computed values of thetest statistic for these four months are greater than the tabulated value of Chi-square with 16 degrees of freedom at 5 percent level of significance. So,these are all significant. Thus,we can reject the null hypothesis for each of the four months. The active monsoon months except September are also significant at one percent level of significance.

Above test suggests that the weather of a day is influenced by the immediately preceding day's weather only during the monsoon months at Cooch Behar.

The analysis of the behaviour of the daily weather has established that our result is in good agreement with that of Medhi (1976). But he considered only the two state Markov chain model of the daily rainfall in Guwahati, Assam.

6.5 CONCLUSION.

The daily weather pattern of Cooch Behar during the monsoon months has been established to follow the Markov chain model.The weighted entropy for the month of July is the highest during the monsoon months. Among the probabilities of individual states of occurrence the probabilities of non-rainy days are highest, followed by moderate rainy days during the monsoon season except the month of July where this feature is reversed.

The validity of the redundancy test is verified by using the test equivalent to the likelihood-ratio test. So,it is observed that the redundancy test is not so powerful test of Markov dependence. But the test equivalent to the likelihood-ratio test has shown better result against the Markov dependence on the same observations. But considering the empirical result of Medhi (1976), it would have been suggested that the latter test is more powerful test as well as the most appropriate test against the Markov dependence of daily rainfall.

TABLE – 6.1.

Transitional cell frequencies in matrix form for rainy months at Coochbehar

May

0	1	2	3	4
203	46	51	8	1
36	11	31	3	0
56	23	55	10	0
14	3	7	0	0
0	0	0	0	0

June

0	1	2	3	4
128	39	43	9	3
26	9	38	4	1
55	19	61	22	11
10	8	20	12	4
3	3	9	3	0

July

0	1	2	3	4
80	33	36	9	5
37	45	30	10	3
33	30	68	33	6
14	10	25	15	9
6	2	8	9	2

August

0	1	2	3	4
172	44	46	13	2
46	21	15	6	5
40	20	23	19	2
15	7	20	14	7
3	1	6	8	3

September

0	1	2	3	4
153	39	44	13	1
37	28	22	11	3
46	28	42	12	2
10	7	12	9	6
1	2	7	2	2

'0' – Non-rainy day, '1'-Light rainy day,
'2' – Moderate rainy day,
'3'– Heavy rainy day,
'4' – Very heavy rainy day.

TABLE : 6.2(a).

One-step(5×5) stochastic matrix for rainy months at Cooch Behar.

May

P_{ij} =	0	1	2	3	4
0	.66	.15	.16	.03	00
1	.44	.14	.38	.04	00
2	.39	.16	.38	.07	00
3	.58	.13	.29	00	00
4	00	00	00	00	00

June

P_{ij} =	0	1	2	3	4
0	.58	.18	.19	.04	.01
1	.33	.12	.49	.05	.01
2	.33	.11	.36	.13	.07
3	.19	.15	.37	.22	.07
4	.17	.17	.50	.16	.00

July

P_{ij} =	0	1	2	3	4
0	.49	.20	.20	.06	.03
1	.30	.36	.24	.08	.02
2	.19	.18	.40	.19	.04
3	.19	.14	.34	.21	.12
4	.23	.07	.30	.33	.07

August

P_{ij} =	0	1	2	3	4
0	.62	.16	.17	.04	.01
1	.50	.23	.16	.06	.05
2	.39	.19	.22	.18	.02
3	.24	.11	.32	.22	.11
4	.14	.05	.29	.38	.14

September

P_{ij} =	0	1	2	3	4
0	.61	.16	.17	.05	.01
1	.36	.28	.22	.11	.03
2	.35	.22	.32	.09	.02
3	.23	.16	.27	.20	.14
4	.07	.20	.47	.13	.13

(ij = 0,1,2,3,4) where,

'0' – Non-rainy day, '1'–Light rainy day,

'2' – Moderate rainy day, '3'–Heavy rainy day,

'4' – Very heavy rainy day.

T A B L E - 6.2(b)

Stationary vectors and probabilities of rainy months at Cooch Behar.

STATIONARY VECTOR

	0	1	2	3	4	Total
May	309	81	144	24	0	558
June	222	78	168	54	18	540
July	163	125	170	73	27	558
August	277	93	104	63	21	558
September	250	101	130	44	15	540
				GRAND TOTAL	=	2754

STATIONARY PROBABILITY

	0	1	2	3	4
May	.55	.15	.26	.04	0
June	.41	.15	.31	.10	.03
July	.30	.22	.30	.13	.05
August	.50	.17	.19	.11	.03
September	.46	.19	.24	.08	.03

- '0 ' – Non-rainy day, '1' – Light rainy day,

'2 ' – Moderate rainy day,

'3 ' – Heavy rainy day,

'4 ' –Very heavy rainy day.

TABLE : 6.3.

MEASURE OF ENTROPY AND REDUNDANCY TEST

(For the experimental months)

	May	June	July	August	September
Non-Rainy Day(H_0)	.415	.484	.555	.341	.474
Light Rain (H_1)	.492	.506	.587	.563	.610
Moderate Rain(H_2)	.527	.620	.623	.609	.591
Heavy Rain (H_3)	.422	.646	.669	.663	.697
Very Heavy Rain(H_4)	.0	.540	.624	.620	.605
Entropy (H_i)	.47	.58	.48	.57	.57
Weighted Entropy	.45	.54	.60	.49	.55
Redundancy	.348	.217	.140	.288	.215
M	.02	.04	– .12	.08	.02

TABLE 6.4.

Entropy of individual state, weighted entropy and the value of test statistic for the experimental months.

	May	June	July	August	September
H_{1i}	.4744	.5847	.6398	.5767	.574
H_{2i}	.4568	.5477	.601	.5344	.549
n	558	540	558	588	540
X^2 (cal)	19.64	41.04**	43.30**	47.21**	27.0*
X^2 5%	26.30				
DF	16				

* Significant at 5% level.

** Significant at 1% level