Dependence of Production of Paddy on the Total Annual Rainfall: A Different Approach*

S. Chattopadhyay

Calcutta, India

E-mail: binu_2008@yahoo.com

Abstract

Some statistical techniques have been employed to discern the dependence of paddy production on total annual rainfall. The study area is West Bengal, a state of India. The study is based upon the computation of Pearson Correlation Coefficient, Entropy, and testing for Poisson distribution.

Key words: West Bengal, Paddy production, Pearson Correlation Coefficient, Entropy, Poisson distribution

Introduction

This brief article aims at finding out a statistical relationship between the paddy production and total rainfall in West Bengal, a state in India. A positive correlation between these two features is well established. Most studies, however, are based on a traditional statistical approach. The present study deviates a bit from the earlier ones. The newness of this study is the application of the concept of entropy as explained by Chaudhuri and Chattopadhyay in *Solstice* (2003). One limitation of the traditional approaches is that they are based on the assumption that the yearly values of the aforesaid features in successive decision periods are serially independent.

This paper develops an approach to incorporating serial correlation (Wilks, 1991) into the decision making process. The underlying idea is to show that in the future both the production of paddy in the aforesaid state as well as probable maximum rainfall can be investigated through the theory of Markov Chain (Wilks, 1995) and that uncertainty in the production of paddy in the coming years can be discerned through the predicted value of maximum probable rainfall.

Experimentation setup

The experimentation set up consists of the following steps:

- Testing for the Markov status through lagged autocorrelations (Wilks, 1995)
- Finding out the interdependence between total yearly rainfall (R) and the production of paddy (P) through the Pearson Correlation Coefficient (Chattopadhyay, 2002)
- Checking for the Poisson distribution in the data series of ‘P’ considering it as a variable dependent on ‘R’ (Box and Jenkins, 1976)
- Calculating the entropies in the probability distribution of ‘P’ with different changes (%) in the value of ‘R’.

The study is based on data for the period 1995-2000 made available from *The Statesman*, a leading newspaper of India.

Testing for Markov Status

We have two time series, one for the values of ‘R’ and the other for the values of ‘P’.

For each variable, we consider the null hypothesis:

H0: The data are serially independent

This is to be tested against the alternative hypothesis

H1: The data are serially dependent.

Under the null hypothesis a Chi-square statistic is calculated for each parameter using the formula:

$$X^2 = [(\text{Observed value} - \text{Expected value})^2 / \text{Expected value}]$$

If the observed value of the statistic is found to exceed the tabular value the null hypothesis is rejected, otherwise accepted .

In our study we have found that

For 'R' Chi-square= 10. 319

For 'P' Chi-square= 14.319

Both of the values are found to exceed the tabular value (Wilks, 1995) of Chi-square at 1% level of significance, leading us to reject the null hypothesis H0. It can therefore be concluded that on the basis of the body of evidence, we have nothing to believe that either 'R' or 'P' are serially independent. As the decision is true at 1% level of significance, we have enough reason to infer that in the long run, in 99% cases the data will remain to be serially dependent.

Next to see their Markov status:

Lag-k autocorrelation coefficient (ACC) is computed as

ACC=

(Covariance between k-lagged data pair)/{(sd for first (k-1)data values)(sd for last (k-1)data values)}

……………………………………………………….(1)

where, sd= Standard Deviation.

From the Markovian point of view, Lag-1 ACC, denoted as r1 is the measure of persistence. So if both of the series are found to have significant r1, we can go ahead to test the Markov status defined as

$$rk = (r1)^k \qquad \text{……………………………………………………………..(2)}$$

The Lagged ACCs in our study are presented in table-1.

Table-1

Parameter	Lag-1ACC	Lag-2ACC	Lag-3ACC	Lag-4ACC
R	0.4126	0.1703	0.0706	0.0291
P	0.5311	0.2310	0.1501	0.0811

The Lag-1 ACC being of significant value (compared to 1), and the four lagged ACCs being found to obey equation (2), it can be concluded that the series are generated by first-order-two-state Markov Chain (Wilks, 1995). Thus, serial dependence with a specific pattern is established.

Pearson Correlation Coefficient

The values of 'P' and 'R' have been standardized by using the formula:

Standardized X=[(Actual X-Average of X)/sd of X]……………………………………….(3)

Their scatterplot with a trend line is shown in Fig.01.

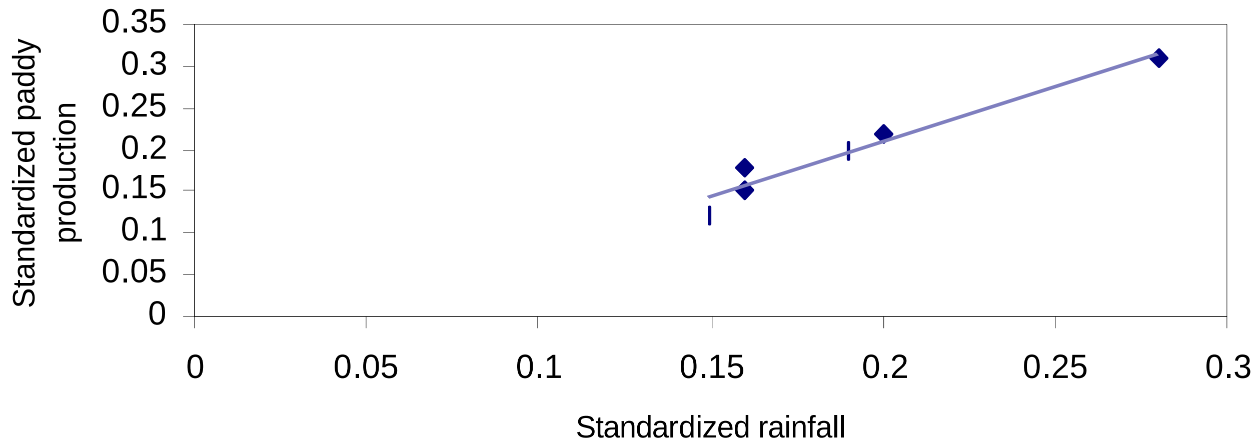
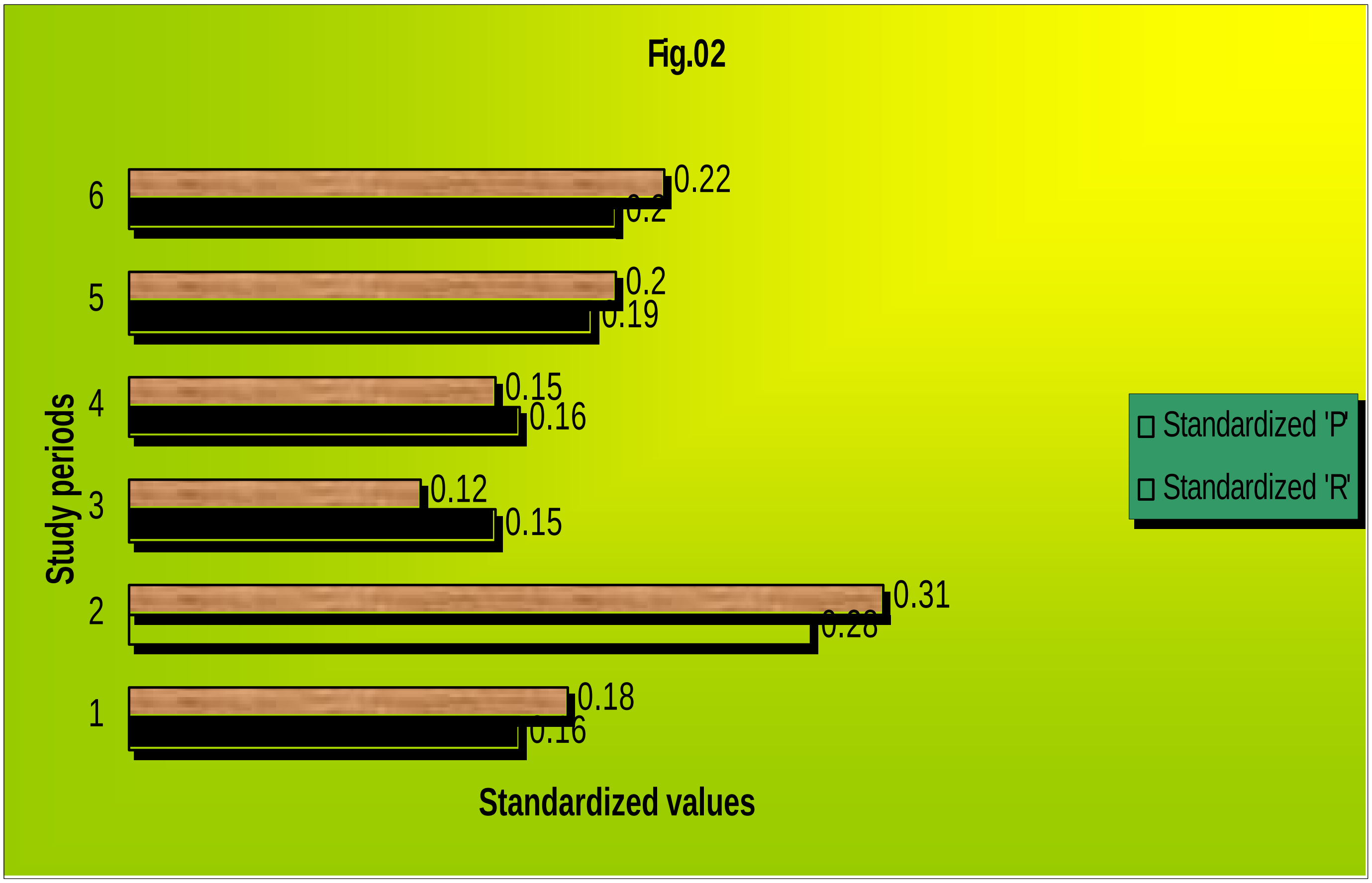


Fig.01-Scatter showing a linear relationship between rainfall and paddy production

The linear trend leads us to calculate a Pearson Correlation Coefficient between 'R' and 'P' which in this case is found to be 0.97, supporting quantitatively the linear relationship. The interrelationship has also been presented in figure-02.



Check for Poisson distribution

We now consider 'R' as an independent variable and 'P' as the variable dependent on it. Next, consider the null hypothesis:

H0: 'P' is not distributed as Poisson.

This is to be tested against,

H1: 'P' is distributed as Poisson.

Poisson distribution is presented as

$$f(x)=\exp(-\mu)(\mu^{x})/x! \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Using H0 and (4), a Chi-square statistic is formed and the value at 1% level of significance and with 5 degrees of freedom. The value is found to be, 19.286. Comparing this value with the tabular value, it is found that 'P' is Poisson distributed. Thus 'P' has randomness with respect to 'R'.

Entropy calculation

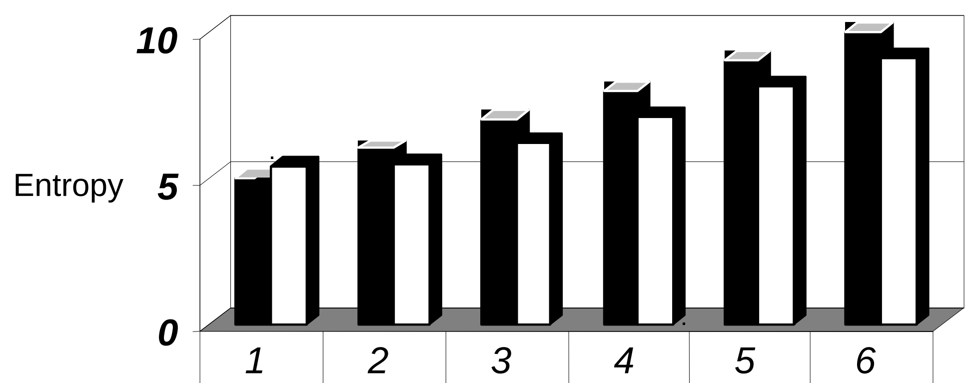
Maximum entropy probability distribution is calculated for 'P' with change (%) in 'R'. Results are presented in table-2.

Table-2

Change (%) in R	Entropy for P
5%	5.412
6%	5.469
7%	6.181
8%	7.121
9%	8.118
10%	9.111

The figure below (Fig. 03) shows that Paddy production is very much vulnerable to change in the value of total rainfall.

Fig.03-Relationship between change in R(%) and entropy for 'P'



	1	2	3	4	5	6
Change in R (%)	5	6	7	8	9	10
Entropy for P	5.412	5.469	6.181	7.121	8.118	9.111

Conclusion

It is no surprise that paddy production is dependent on total annual rainfall. The use of an entropy calculation shows the extent to which paddy production is vulnerable to change in the value of total rainfall. The degree to which randomness and uncertainty in production depend on rainfall is characterized by a Markov pattern.

*Acknowledgement:

The author wishes to thank Professor Sandra Arlinghaus for helping in various ways while preparing the manuscript

References

Wilks, D.S., 1995: *Statistical Methods in Atmospheric Sciences*, Academic Press, USA.

Wilks, D.S., 1991: Representing serial correlation of meteorological events and forecasts in dynamic decision - analytic model, *Month. W. Rev.*, **119**. pp1640-1662

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis*, Holden-Day

Chaudhuri, S., and S., Chattopadhyay, 2003: Viewing the Relative importance of some surface parameters associated with pre-monsoon thunderstorms by Ampliative Reasoning, *Solstice: An Electronic Journal of Geography and Mathematics*, Institute of Mathematical Geography, Vol.XIV, No, 1 (http://www.imagenet.org/)