International Journal of Statistics and Applied Mathematics 2023; SP-8(4): 78-83

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(4): 78-83

https://www.mathsjournal.com
Received: 01-05-2023
Accepted: 05-06-2023

Mamata
Ph.D., Research Scholar, Department of Agricultural Statistics, Visva Bharati, Sriniketan, Bolpur, West Bengal, India

Dr. D Bhattacharya
Professor and HOD, Department of Agricultural Statistics, Visva Bharati, Sriniketan, Bolpur, West Bengal, India

Corresponding Author:
Mamata
Ph.D., Research scholar, Department of Agricultural Statistics, Visva Bharati, Sriniketan, Bolpur, West Bengal, India

Use of Markov chain model for studying disparity in rice yield in the major rice producting districts of West Bengal

Mamata and Dr. D Bhattacharya

Abstract
This paper concentrates in studying the disparity among the districts growing rice by analyzing rice yield data of 1991-92 to 2020-21 for the districts of West Bengal. On the basis thirty years of rice yield data the districts have been classified into three classes (states), namely, highly developed (HD), developed (D) and under developed (UD) by using σ classifier. Next, observing the frequency in the states the transition probability matrix and initial probability vectors are obtained. The steady state probability and expected return time to a particular state are also obtained. Stationary probabilities for different states of each district under study have been used to predict the future movement of the district from one classification to other in terms of rice yield. The model developed for disparity study of a crop is quite general and can be applied to any other related studies.

Keywords: Initial probability vector, Markov chain model, production, return time, transition probability matrix

Introduction
Rice (*Oryza sativa*) is the second-most important cereal crop in the world after maize. Rice being a staple food crop in the country plays a key role in minimizing the gap between food grains demand and production. In the most recent harvesting year, about 510 million metric tons of milled rice were produced globally. Rice is the largest cropping system practised in South Asian countries [8]. Asia produced the majority of the world's rice. According to the most current official statistics, China was the world's top producer of rice in 2021, followed by India and Bangladesh with a production output of over 212 million metric tons. India's total production of rice during 2021-22 is estimated at record 127.93 million tonnes it is higher by 11.49 million tonnes than the last five years' average production of 116.44 million tonnes. National Food Security Mission was launched in 2007-08 to increase the production of rice 10 million tonnes. The mission was continued during the 12th Plan with new targets of additional production of 25 million tonnes of food grains comprising of 10 million tonnes of rice. 12th Plan (2017-18 to 2019-20) it was decided to continue the programme with new targets to achieve 13 million tonnes of additional food grains production comprising of rice-5 million tonnes. The target during 2021-22 is for an additional 1.7 million tonne of rice.
West Bengal is the richest reservoir of rice bio-diversity and the rice bowl of the country. Rice is grown all over the state and the grain has become part and parcel of Bengal life. It is not only the essential staple food for the Bengalese, but also a number of cultural festivals and rituals have been intermingled with rice. West Bengal is the largest producer of the rice in the country followed by Uttar Pradesh and Punjab. According to economic survey 2022-23, West Bengal produces 16.76 million tonnes of rice which shares nearly 12.87% of India's rice production.
The four districts of West Bengal with the highest yields of rice are Bardhaman, Birbhum, Nadia and Hooghly, which serve as the basis for considering those districts for the present study.
The objective of the present study is to examine disparities in rice yield (kg/ha) among the four West Bengal districts and forecast future district movement.

In many recent studies, Markov chain has been used to model many economic and social problems like stock market prediction, land use pattern and vegetable market prediction etc [12]. Used Markov chain model to forecast the stock market trend of China [2]. Made an attempt to predict the future market price of potatoes grown in the Nagaon District of Assam [4], constructed the Markov chain model for a daily vegetable price movement in Jaffna [5]. Used Markov chain to strengthen potato production ensuring positive growth in yield [1], predicted the Yield of maize in Iraq using Markov chain model [6]. Predicted the yield of pomegranate trees and palm trees in southern Iran based on the probability of future drought [11], predicted agricultural product prices through Markov chain model [7]. Applied Markov chain approach to forecast cotton yield [10]. Used Markov chain model to assess the Maize Production trend [9]. Applied multiple Markov models and simulation techniques to forecast the sugarcane crop yield of India. In this work Markov model has been used to study disparity for the first time.

Material and Method

A process that alters in an unexpected way as time passes is known as a stochastic process.

In probability theory, a Markov model is a stochastic model used to model randomly changing systems over time. Markov process is a particular category of random process where only the current value of a variable is used to forecast the future and the variables in the previous history are considered with very relevant when the current value is given. In most cases, share prices are observed to follow a Markov process according to [3]. The MC model's most fundamental property is that the occurrence of any future event depends only on the current state. The state-space of a Markov process is the set of all possible values that takes. A Markov process whose state space is discrete is known as a Markov chain.

Markov Chain defined is a Markov process of particular type in which the state of the future events depends on its instantaneous preceding state and not on the previous history. This random process follows memory less property or has a short-term memory.

Markov property or memory less property states that the system's state at time (*t*+1) depends only on the state at time *t* and not on (*t*-1), (*t*-2), 3, 2, 1.

Mathematically, the Markov property is stated as

$$P(X_{t+1} = x_{t+1}/X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_2 = x_2, X_1 = x_1, X_0 = x_0) = P(X_{t-1} = x_{t-1}/X_t = x_t)$$

For all *t*=0,1,2,… and for all the states $x_0, x_1, x_2, \dots x_t, x_{t+1}$

Transition probability

Probability of transition of the process from state *i* to state *j* at 1 step time period, denoted by $p_{ij,}$ is defined as,

$$p_{ij} = P(X_{n+1} = j/X_n = i)$$

Probability of transition from state *i* to state *j* after *k* steps time period is denoted by p_{ij}^k where p_{ij}^k defined as

$$p_{ij}^k = P(X_{n+k} = j/X_n = i)\ k \geq 0, i, j = 1,2,3$$

The matrix describing the Markov chain is called a transition probability matrix (TPM).

The probability $p_{ij}^k = P(X_{n+k} = j/X_n = i) \geq 0\ k \geq 0\ i, j = 1,2,3\ and\ n \geq 0$ are the transition probabilities for going to state *j* from state *i* in *k* steps.

Markov chain model is denoted by $\lambda = (A, \pi)$, where *A* and π are the transition probability matrix and Initial probability vector which are called the parameters of the model.

- The possible states of a MC are used as rows and columns, hence the TPM is always a square matrix.
- $p_{ij} \geq 0\ for\ all\ i\ and\ j$

Here $A = \left((p_{ij})\right)\ i, j = 1,2,3$, where row sum $\sum_{j=1}^{3} p_{ij} = 1$.

Using σ classifier (to be defined later) the districts have been classified into three states, namely highly developed (HD), Developed (D) and Under developed (UD).

A TPM for a three state MC is obtained as

$$A = \begin{bmatrix} P(HD/HD) & P(D/HD) & P(UD/HD) \\ P(HD/D) & P(D/D) & P(UD/D) \\ P(HD/UD) & P(D/UD) & P(D/UD) \end{bmatrix}$$

Where, $P(HD/HD)$ is the probability that districts will be in highly developed state in the current year given that highly developed state has been observed in immediate past year and so on. Hence symbolically the three state TPM (HD, D, UD) is

$$A = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

Since the state space of the present Markov chain model is {HD, D, UD}, therefore the initial probability vector (IPV) consists of three elements p_{01}, p_{02}, p_{03}.

The IPV denoted by $\prod_0$, is of the form $\prod_0 = [\pi_1, \pi_2, \pi_3]$, where for our present being study, π_1 is the probability of a highly developed, π_2 is the probability of developed and π_3 is the probability of under developed state for each district.

The equilibrium situation for the data can be obtained by the higher order TPM. Stationary distributions are associated with the Eigen vectors, where the Eigen value is one. The forecast of long run behaviour of district can be made by using the stationary probability obtained for the districts. Therefore, an attempt has been made in this work to obtain stationary probability for each districts under study.

The TPM is

$$A = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

The long run behaviour of district is observed by determining the higher order TPM of districts as given below

$$A^2 = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

Similarly, calculate the higher order matrix for *k* step until get stationary Vector as $[p1\ p2\ p3]$.

TPM reaches to the equilibrium state or steady state distribution. In the stationary situation, transition matrix remains stable or invariant if we increase the number of steps.

Markov chain is explained through the diagrammatic representation called the state transition diagram or schematic diagram, that is quite same as the TPM but expressed diagrammatically. The state transition diagram of' a Markov chain model λ= (A, π) is a one way directed graph where every vertex depicts the state of the Markov chain model. The parameters of the Markov chain model can well be explained by the state transition diagram presented in Figure 1, where the arrow marks represent probabilities of transition from one class to another class.

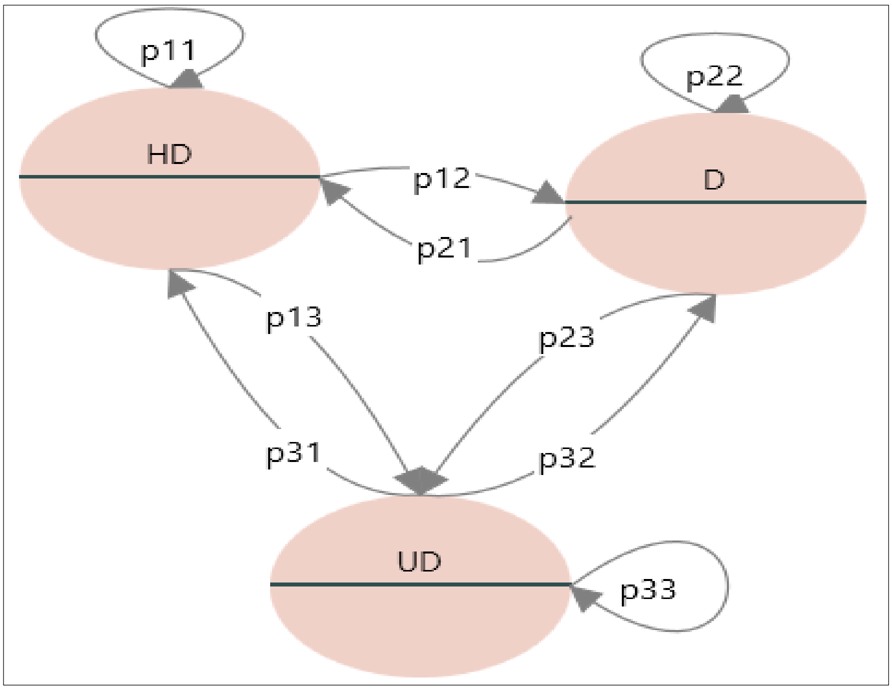

Fig 1: The Schematic diagram of the Markov chain model

The stationary distribution can also be useful for obtaining the expected return time u_{ij} that is the time that the chain visits state j when it left state i. It is simply the reciprocal of probabilities of the stationary vector. Mathematically, the formula for calculating the return time is given as.

$$u_{ij} = \frac{1}{p_j}, i, j = 1,2,3$$

Where $p_j's, j = 1,2,3$ are the row elements of stationary vector. Hoogly, Medinipur, Bankura and Bardhaman districts are ergodic that is TPM reaches to steady state after a large number of transitions. In order words as k tends to ∞, p_{ij}^k tends to a stationary vector whose rows are identical which is independent of initial state i. This property of limiting distribution of p_{ij}^k is known as ergodicity.

Secondary data for the study purpose collected from Statistical abstract, Evaluation wing, Directorate of Agriculture. Government of West Bengal.

Results and Discussion

The 30 years Rice yield data of four districts of West Bengal were divided into three states namely HD, D and UD (Table 1). Without loss of generality let $y = x - \mu$ where x is the production data having average μ and variance σ^2.

Using σ classifier we divide the districts in the following manner.

High developed (HD) when $-\sigma < y < \sigma$,

Developed (D) when $-2\sigma < y < -\sigma \cup \sigma < y < 2\sigma$ and

Under developed (UD) when $-3\sigma < y < -2\sigma \cup 2\sigma < y < 3\sigma$

Table 1: Transition of states over the years for different districts growing rice

Year	District 1 (Bardhaman)	District 2 (Birbhum)	District 3 (Nadia)	District 4 (Hooghly)
1991-92	D	D	HD	HD
1992-93	D	D	D	D
1993-94	D	D	HD	D
1994-95	HD	D	D	HD
1995-96	D	UD	HD	D
1996-97	HD	D	HD	HD
1997-98	HD	HD	HD	HD
1998-99	HD	HD	HD	HD
1999-00	D	HD	HD	D
2000-01	D	HD	HD	UD
2001-02	HD	HD	HD	HD
2002-03	HD	HD	HD	HD
2003-04	HD	HD	HD	HD
2004-05	HD	HD	HD	HD
2005-06	HD	HD	HD	HD
2006-07	HD	HD	HD	HD
2007-08	HD	HD	HD	HD

2008-09	HD	HD	HD	HD
2009-10	HD	HD	HD	HD
2010-11	HD	HD	HD	HD
2011-12	HD	HD	HD	HD
2012-13	HD	HD	HD	HD
2013-14	D	D	HD	HD
2014-15	HD	D	D	HD
2015-16	D	D	D	HD
2016-17	HD	D	D	HD
2017-18	HD	D	D	HD
2018-19	D	HD	D	D
2019-20	UD	D	D	D
2020-21	D	D	D	D

In order to determine the initial state probability vector, the rice yield data of four districts were divided into three states *viz*., highly developed, developed and under developed. The state space is {HD, D, UD}, and state probability is total number of data in a single state. IPV is denoted by $\prod_0 = [\pi_1, \pi_2, \pi_3]$

District 1: Bardhaman
Initial Probability Vector for Bardhaman district
Rice yield data for 30 years are given in Table 1, where HD = 19, D=10 and UD=1.

$\pi_1 = \frac{19}{30}$, $\pi_2 = \frac{10}{30}$, $\pi_3 = \frac{1}{30}$ and state IPV $\prod_0$=(0.6333 0.3333 0.0333).

Transition Probability Matrix for Bardhaman district
To find the elements of the TPM we count the number of transitions from one state to another state. For example, to find the probability that the district is in HD state given that the district is also in HD state in the immediate past year. We count the number of times the district remains in HD state in the consecutive years. Then that number is divided by the total number years the district was in HD state. Transition probability for HD to HD $p_{11} = \frac{14}{19}$=0.7368, similarly calculated remaining elements of TPM $p_{12} = \frac{5}{19}$=0.2631, $p_{13} = 0, p_{21} = \frac{5}{10} = 0.5, p_{22} = \frac{3}{10}$=0.3, $p_{23} = \frac{1}{10} = 0.1$, $p_{31} = 0$, $p_{32} = \frac{1}{1} = 1$ and $p_{33} = 0$ respectively.
The state transition probabilities are summarized in matrix form so it is called as Transition probability matrix (TPM).

$$\text{TPM of Bardhaman district} = \begin{bmatrix} 0.7368 & 0.2631 & 0 \\ 0.5 & 0.3 & 0.1 \\ 0 & 1 & 0 \end{bmatrix}$$

After fourteen years TPM of Bardhaman district attainted stable or stationary state since 2020-21.
Stationary vector for Bardhaman district will be reaming in HD, D and UD is [0.6551 0.31034 0.0344] from stationary vector observed that the probability that there is 65%,31% and 3% likelihood that Bardhaman district will be in highly developed, developed and under developed category respectively for rice yield in near future and in the long run.
The expected return time to the highly developed state is one year, developed state is three years and under developed state is two nine years. Here we have observed that the return time to highly developed state is less as compared to other states and hence we can interpret that highly developed state is occurring more frequently in the process as compared to other states.
Similarly constructed IPV and TPM for rest of districts

District 2: Birbhum
Transition probability matrix for Birbhum district

$$\text{TPM of Birbhum} = \begin{array}{c} \\ \end{array}\begin{matrix} HD & D & UD \end{matrix}$$
$$\text{TPM of Birbhum} = \begin{bmatrix} 0.882353 & 0.117647 & 0 \\ 0.181818 & 0.727273 & 0.090909 \\ 0 & 1 & 0 \end{bmatrix}$$

Initial transition vector for Birbhum district

$\pi_0 = [0.566667\ 0.4\ 0.033333]$

After twenty-five TPM of Birbhum district attainted stable or stationary state since 2020-21
Stationary vector for district Birbhum will be reaming in HD, D and UD is [0.5863 0.3792 0.03446] from stationary vector observed that the probability of remaining in highly developed state is 58%, developed state is 37% and 3% in under developed state for rice yield in near future and in the long run.
The expected return time to the highly developed state is one year, developed state is two years and under developed state is twenty-eight years. Here we have observed that the return time to highly developed state is less as compared to other states and hence we can interpret that highly developed state is occurring more frequently in the process as compared to other states.

District 3: Nadia
Transition probability matrix for Nadia district

$$\begin{matrix} HD & D & UD \end{matrix}$$
$$\text{TPM of Nadia} = \begin{bmatrix} 0.857143 & 0.142857 & 0 \\ 0.25 & 0.75 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Initial transition vector for Nadia district

$\pi_0 = [0.7\ 0.3\ 0]$

After twenty-two years TPM of Nadia district attainted stable or stationary state since 2020-21.
Stationary vector for district Nadia will be reaming in HD, D and UD is [0.636 0.3636 0] from stationary vector observed that the probability that there is 64% likelihood that Nadia district will be in highly developed and 36% in developed state for rice yield in near future and in the long run.
The expected return time to the highly developed state is one year and developed state is two years. Here we have observed that the return time to highly developed state is less as compared to other states and hence we can interpret that highly developed state is occurring more frequently in the process as compared to other states.

District 4: Hooghly

Transition probability matrix for Hooghly district

$$\text{TPM of Hooghly} = \begin{matrix} & HD & D & UD \\ \end{matrix} \begin{bmatrix} 0.818182 & 0.181818 & 0 \\ 0.333333 & 0.5 & 0.166667 \\ 1 & 0 & 0 \end{bmatrix}$$

Initial transition vector for Hooghly district

$$\pi_0 = [0.733333 \quad 0.233333 \quad 0.033333]$$

After ten years TPM of Hooghly district attainted stable or stationary state since 2020-21.

Stationary vector for district Hooghly will be reaming in HD, D and UD is [0.7 0.26 0.04] from stationary vector observed that the probability that the Hooghly will remain in HD state in the long run is 70% and that of D state in 26% and UD state will be 4% in respect of rice yield.

The expected return time to the highly developed state is one year, developed state is three years and under developed state is twenty-three years. Here we have observed that the expected return time to highly developed state is less as compared to other states hence it can be said that highly developed state is occurring more frequently in the process as compared to the other states.

Conclusion

The model developed here is quite general and can be applied in any other study related to disparity for a crop.

It has been found that the chances of districts Bardhaman, Birbhum, Nadia and Hoogly to remain in HD if the districts were in HD in last year are 0.73, 0.88, 0.85 and 0.81 respectively with respect to rice yield. Hence the disparity among the four districts with respect to the chances of remaining in HD state in the current year given that the districts were in HD state in last year has been established. The probabilities of districts Bardhaman, Birbhum, Nadia and Hoogly to remain in D if the districts were in HD in the last year with probability 0.26, 0.11, 0.14 and 0.18 respectively with respect to rice yield, disparity among the four districts with respect to the chances of remaining in D state in the current year given that the districts were in HD state in last year is clear. The chances of districts Bardhaman, Birbhum, Nadia and Hoogly to remain in HD if the districts were in D in last year with probability 0.55, 0.18, 0.25 and 0.33 respectively for rice yield, Hence the disparity among the four districts with respect to the chances of remaining in HD state in the current year given that the districts were in D state in last year has been established. The chances of districts Bardhaman, Birbhum, Nadia and Hoogly districts to remain in D if the districts were in D with probabilities 0.33, 0.72, 0.75 and 0.5 respectively for rice yield, the disparity among the four districts with respect to the chances of remaining in D state in the current year given that the districts were in D state in last year has been established. The probabilities of districts Bardhaman, Birbhum and Hoogly to remain in UD if the districts were in D in the last year with probability 0.11, 0.09 and 0.16 respectively with respect to rice yield, disparity among the three districts with respect to the chances of remaining in UD state in the current year given that the districts were in D state in last year is clear. The chances of Hoogly district to remain in HD if the district was in UD with probability 1 for rice yield. The probabilities of districts Bardhaman and Birbhum to remain in D if the districts were in UD in the last year with probability 1 and 1 respectively with respect to rice yield. Strictly, speaking the disparity between Bardhaman and Birbhum are not that prominent.

Next, Disparity among the districts are observed by comparing the stationary probability of a district to remain in a HD state once it has reached to the HD state. Stationary probability of Bardhaman is 0.63, Birbhum is 0.56, Nadia is 0.7 and Hoogly is 0.73 remain in HD state. Disparity among the districts are observed by comparing the stationary probability of a district to remain in a D state once it has reached to the D state. Stationary probability of Bardhaman is 0.33, Birbhum is 0.4, Nadia is 0.3 and Hoogly is 0.23 remain in D state. Disparity among the districts are observed by comparing the stationary probability of a district to remain in a UD state once it has reached to the UD state. Stationary probability of Bardhaman is 0.03, Birbhum is 0.03 and Hoogly is 0.03 remain in UD state.

This work also helps to predict the number of years required by a state to reach the stationary state. It has been observed that Bardhaman required fourteen years to reach the stationary state. Similarly, Birbhum required twenty-five years, Nadia required twenty-two years and Hoogly required ten years to reach stationary state respectively.

It has been observed that expected return time to HD is one year for all. Similarly, expected return time to D is three years for Bardhaman, two years for Birbhum, two years for Nadia and three years for Hoogly respectively. Expected return time to UD is twenty-nine for Bardhaman, twenty-eight years for Birbhum and twenty-three years for Hooghly.

It has been observed that for some districts transition took place from D (lower) state to HD (upper) state the reason may be that famers are adapting improved agricultural technology (better high yielding tubers, improved fertilizer etc.) as the area under the crop has not changed significantly.

Acknowledgement

We are grateful to the department of Agricultural Statistics, Visva-Bharati University for all sort of assistance provided during this study.

Author's contribution

Conceptualization and designing of the research work (Mamata, D. Bhattacharya); Data collection (Mamata); Analysis of data and interpretation (Mamata, D. Bhattacharya); Preparation of manuscript (Mamata, D. Bhattacharya).

Competing Interest: The authors have declared that no competing interest exists.

References

1. Al-Ani LAF, Alhiyali ADK. Using Markov Chains to Predict YIELD of Maize in Iraq for the Period (2019-2025). Applied Economics and Finance. 2021;8:2332-7308.
2. Aparna B, Sarat CK. Markov Chain Modelling for Prediction on Future Market Price of Potatoes with Special Reference to Nagaon District. Journal of Business and Management. 2017;19:25-31.
3. Hull JC. Options, Future, and Bother Derivatives-Wiener Processes and Ito's Lemma. Pearson Education, New York; c2018.
4. Jasinthan P, Laheetharan A, Satkunanathan N. A Markov Chain Model for Vegetable Price Movement in Jaffna. Sri Lankan Journal of Applied Statistics. 2015;16:2424-6271.

5. Mahendra S. Projection of Potato export from India: A Markov Chain Approach. Potato Journal. 2010;37:48-55.
6. Marzieh M, Tam MP. CA-Markov Model application to Predict Crop Yield using Remote Sensing Indices. Ecological Indicators. 2022;139:108952.
7. Matis JH, Birkett T, Boudreaux D. An application of the Markov Chain Appoarch to Forecasting Cotton Yields from Surveys. Agricultural Systems. 1989;29:357-370.
8. Nawaz A, Farooq M, Nadeem F, Siddique KH, Lal R. Rice-wheat cropping systems in South Asia: Issues, options and opportunities. Crop Pasture Sci. 2019:70(5):395-427.
9. Ramasubramanian V, Lalmohan B. Crop Yield Forecasting by Multiple Markov Chain Models and Simulation. Statistics and Applications. 2014;12:1-13.
10. Saranyadevi MM, Kachi MA. A Markov Chain Model for Maize Production in Tamil Nadu. Infokara Research. 2019;8:1021-9056.
11. Uktam U, Bekzod D. Prediction of prices for agricultural products through Markov chain model. International Journal of Pharmaceutical Research. 2020;24:293-303.
12. Zhang D, Zhang X. Study on Forecasting the Stock Market Trend Based on Stochastic Analysis Method. International Journal of Business and Management. 2009;4:6.