# Markov chain model and its application to annual rainfall distribution for crop production

## Abubakar Usman Yusuf, Lawal Adamu, Muhammed Abdullahi

Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria

**Email address:**
abbkruy@yahoo.com (A. U. Yusuf)

**Abstract:** A stochastic process with a first order dependence in discrete state and time is described as Markov chain. This principle was used to formulate a four state model for annual rainfall distribution in Minna with respect to crop production. The model is designed such that if given any of the four state in a given year, it is possible to determine quantitatively the probability of making transition to any other three states in the following year(s) and in the long-run. The model was used to study the data of annual rainfall in Minna. The results show that in the long run 14% of annual rainfall shall be low rainfall, 34% annual rainfall will be moderate rainfall also well spread, 47% of the annual rainfall shall be high rainfall and 5% of the annual rainfall shall be moderate rainfall not well spread respectively. The model provides some information about rainfall in relation to crops cultivation that could be used by the farmers and the government to plan strategy for high crop production in Minna and the immediate environment.

**Keywords:** Markov Chain, Rainfall, Crop Production, Transition Probability

## 1. Introduction

Rainfall modelling and prediction are essential input into Agriculture and management of water resource because bumper harvest and learn years depend on rainfall variability and quantity. The development of a rainfall model is increasing in demand, not only for data-generation purposes, but also to provide some useful information in various applications, including water resource management and the hydrological and agricultural sectors. The generation of rainfall and other climate data needs a range of models depending on the time and spatial scales involved. Gabriel and Neumann (2) studied the sequence of daily rainfall occurrence. They found that the daily rainfall occurrence for Tel Aviv data was successfully fitted with the first-order Markov chain model. Meanwhile, Kottegoda et al. (6) reported that a first order Markov chain model was found to fit the observed data in Italy successfully. The model was based on the assumption that there is a dependency of the daily rainfall occurrence to that of the previous day. Akintunde et al (1), modified the Chapman-Kolmogorov Equation (CKE) and applied it to model the daily precipitation data of Abeokuta, Ogun State, Nigeria. This gave the best fit for precipitation pattern which is relevant in the development of new growth and yield

models of major crops such as maize, sorghum and soya bean; enabling farmers to estimate the distribution of crop yield as the growing season progressed. Jimoh and Webster (4) determined the optimum order of a Markov chain model for daily rainfall occurrences at 5 locations in Nigeria using AIC and BIC. It was concluded that caution is needed with the use of AIC and BIC for determining the optimum order of the Markov model and the use of frequency duration curves can provide a robust alternative method of model identification. Jimoh and Webster (5) also investigated the intra-annual variation of Markov chain parameters for 7 sites in Nigeria. They found that there was a systematic variation in the probability of a wet day following a dry day as one moves northwards and limited regional variation. A general conclusion is that a first order Markov model is adequate for many locations but second or higher order model may be required at other locations or during some times of the year. This paper therefore, considers a four state model in discrete time to predict annual rainfall pattern for growing of crops in Minna and its immediate environment. It is an extension of the result reported in Abubakar et al (13) where the model was considered in continuous time space. In practice the two models complement one another for providing quantitative predictions for the annual rainfall for the purpose of crop

cultivation.

## 2. Theoretical Background

Following Ross(12), we consider a stochastic process $\{X_n, n = 0,1,2....\}$ that takes on a finite or countable number of possible values. Unless otherwise mentioned, this set of possible values of the process will be denoted by

the set of non negative integers $\{0,1,2,.....\}$. If $X_n = i$, then the process is said to be in state $i$ at time n. we suppose that whenever the process is in state $i$, there is a fixed probability $p_{ij}$ that it will next be in state j. that is, we suppose that

$$p\{X_{n+1} = j \mid X_n = i, \ X_{n+1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0\} = p_{ij} \tag{1.1}$$

for all states $i_0, i_1, ...., i_{n-1} i, j$ and all $n \ge 0$ such a stochastic process is known as a Markov chain. Equation (1.1) may be interpreted as stating that, for a Markov chain, the conditional distribution of any future state $X_n$, is independent of the state $X_0, X_1, ..., X_{n-1}$ and the present state $x_n$ is independent of the past state and depends only on the present state.

The value $p_{ij}$ represents the probability that the process will, when in state $i$, next make a transition into state $j$ .since probabilities are non negative and since the process must make a transition into some state, we have that $p_{ij} \ge 0, i, j \ge 0; \ \sum_{j=0}^{\infty} p_{ij} = 1, i = 0,1,....$

Chapman – kolmogorov equations
We have already defined the one step transition probabilities $p_{ij}$. We now define the n-step transition probabilities $p_{ij}^n$ to be the probability that a process in state $i$ will be in state $j$ after $n$ additional transitions. That is,

$$p_{ij}^n = p\{X_{n+m} = j \mid X_m = i\}, \ n \ge 0, \ all \ i, j \ge 0$$

Of course $p_{ij}^1 = p_{ij}$ . The Chapman kolmogorov equations provide a method for computing these n-step transition probabilities. These equations are

$$p_{ij}^{n+m} = \sum_{k=0}^{\infty} p_{ik}^n p_{kj}^m \ \ for \ all \ \ n, m \ge 0, \ all \ i, j \tag{2.1}$$

and are most easily understood by nothing that $p_{ik}^n p_{kj}^m$ represents the probability that starting in $i$ the process will go to state $j$ in $n + m$ transitions through a path which takes it into state $k$ at $n^{th}$ transition. Hence, summing over all intermediate states  k  yield the probability that the process will be in state $j$ after n + m transitions. Formally, we have

$$p_{ij}^{n+m} = p\{X_{n+m} = j \mid X_n = i\}$$
$$= \sum_{k=0}^{\infty} p\{X_{n+m} = j, X_n = k \mid X_0 = i\}$$
$$= \sum_{k=0}^{\infty} p\{X_{n+m} = j \mid X_n = k, X_0 = i\} p\{X_n = k \mid X_0 = i\}$$
$$= \sum_{k=0}^{\infty} p_{kj}^m p_{ik}^n$$

If we let $p^{(n)}$ denotes the matrix of n-step transition probabilities $p_{ij}^n$ , then equation (2.1) asserts that

$$p^{n+m} = p^{(n)} \cdot p^{(m)}$$

Where the dot product represents matrix multiplication hence, in particular

$$p^{(2)} = p^{(1+1)} = p^2$$

And by induction

$$p^{(n)} = p^{(n-1+1)} = p^{n-1} \cdot p = p^n$$

That is, the $n$-step transition matrix may be obtained by multiplying the matrix p by itself n times

### *Limiting Probabilities*

When iterations are performed, it seems that $p_{ij}^n$ is converging to some value (as n → ∞) which is the same for all i. In other words, there exists a limiting probability that the process will be in state j after a large number of transitions, and this value is independent of the initial state.

### *Theorem*

For an irreducible ergodic Markov chain $\lim_{n \to \infty} p_{ij}^n$ exists and is independent of $i$ . Furthermore, letting

$$\pi_j = \lim_{n \to \infty} p_{ij}^n, \ j \ge 0$$

Then $\pi_j$ is the unique nonnegative solution of

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}, \quad j \geq 0$$

$$\sum_{j=0}^{\infty} \pi_j = 1. \tag{4.1}$$

***Remarks***

(i) Given that $\pi_j = \lim_{n \to \infty} p_{ij}^n$ exists and is independent of the initial state $i$, it is not difficult to (heuristically) see that $\pi^{,s}$ must satisfy equation (4.1). For let us derive an expression for $p\{X_{n+1} = j\}$ by conditioning on the state at time n. that is,

$$p\{X_{n+1} = j\} = \sum_{i=0}^{\infty} p\{X_{n+1} = j \mid X_n = i\} p\{X_n = i\}$$

$$= \sum_{i=0}^{\infty} p_{ij} p\{X_n = i\}$$

Letting $n \to \infty$, and assuming that we can bring the limit inside the summation, leads to

$$\pi_j = \sum_{i=0}^{\infty} p_{ij} \pi_i$$

(ii) It can be shown that $\pi_j$, the limiting probability that the process will be in state $j$ at time $n$, also equals the long-run proportion of time that the process will be in state $j$.

(iii) In the irreducible, positive recurrent, periodic case we still have that $\pi_j, j \geq 0$, are the unique non negative solution of

$$\pi_j = \sum_i \pi_i p_{ij},$$

$$\sum_j \pi_j = 1$$

But now $\pi_j$ must be interpreted as the long-run proportion of time that the Markov chain is in state $j$.

## 3. The Model

A finite Markov chain is a discrete time parameter stochastic process in which the future state of the system is dependent only on the present state and is independent of the past history where the number of states are finite or countably infinite Cox and Miller (7). The daily rainfall in Minna and generally in Nigeria is highly unpredictable. The daily rainfall forms the basis of the weekly, monthly and eventually the annual rainfall. In Nigeria, the annual rainfall decreases from the south to the North. Minna is located in the middle belt and thus makes it possible to produce all forms of crops that can be grown in both the North and the South Iloeje (11). This is the reason why Niger state (Minna) is simply described as the food basket of Nigeria. It is because of the relevance of Niger State (Minna) and the environment to the agricultural and particularly Crop production in Nigeria, that this study is undertaken in an attempt to contribute to the high crop production in the region and Nigeria at large.

It has been observed over the years that the annual rainfall requirement for bumper harvest of major crops grown in Minna Niger State Nigeria, and the environment is within the range of 1000mm-1200mm. The annual rainfall below or above this range that is significant in value may lead to poor harvest, however, some crops like legumes and vegetables do better below this range and also rice do better above this range NSADP(9).

Now, suppose that the amount of annual rainfall in Minna in a year is considered as a random variable $\{X\}$. The collection of this random variable over the years (n) constitutes a stochastic process $X_n, n = 0, 1, 2, 3 \ldots$

It is assumed that this stochastic process satisfies Markov property of first order dependence.

A first order Markov chain could be defined as a sequence $X_0, X_1, \ldots$ of random variables with the property that the conditional probability distribution of $X_{n+1}$ given $X_0, X_1, \ldots X_n$ depends only on the value of $X_n$ but not further on $X_0, X_1, \ldots X_{n-1}$. That is, for any set of values h, i.. . j in discrete state space $P(x_{n+1} = j | x_0 = h \ldots x_n = i) = P(x_{n+1} = j | x_n = i) = P_{ij}$  i,j = 1,2,3… Howard (10).

Let the annual rainfall be modeled by four State, Markov model.

State1: Low rainfall (Annual rainfall below 1000mm)

State2: Moderate rainfall (Annual rainfall within the range 1000mm-1201mm) also well spread

State3: High rainfall (Annual rainfall above 1201mm)

State4: Moderate rainfall within the range (1000mm-1201mm) but not well spread

It is observed that moderate annual rainfall also well spread (evenly spread across the months/raining season) is the rainfall most favorable to crops cultivation.

The transition between the states is described by the transition diagram in figure1 and probability Matrix P.

$$p = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ 0 & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & 0 \\ P_{41} & P_{42} & 0 & 0 \end{bmatrix}$$

The matrix P is homogeneous transition stochastic matrix because all the transition probabilities $p_{ij}$ are fixed and independent of time. The transition probabilities must satisfy

$$P_{ij} \geq 0, \; i,j = 1, 2, 3, 4. \text{ and } \sum_{j=1}^{4} P_{ij} = 1, \quad i = 1, 2, 3, 4.$$
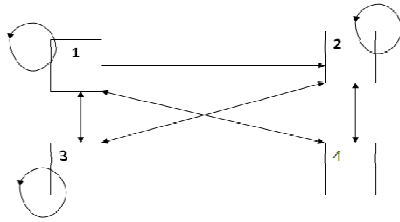
**Figure 1.** *Transition diagram for the Model.*

Following Cox and Miller (7), Let $P^{(n)}n = 0,1,2,................$ be the probability state vectors of the Markov chain, where $P_i^{(n)}$ is the probability that the annual rainfall is in the $i^{th}$ state at the $n^{th}$ observation. In particular, $P^{(0)}$ is the initial state vector of the Markov chain.

Then we can write

$$p^{(n+1)} = P^{(n)}P \qquad (1)$$

Where P is our transition probability matrix and $p^{(n+1)}$ is the state vector at the $(n+1)^{th}$ observation. On iteration, we have

$$p^{(n)} = p^{(0)}P^n \qquad (2)$$

Thus the initial state vector $P^{(0)}$ and the transition matrix P determine the state vector $P^{(n)}$ at the $n^{th}$ year Jain (3).

## 4. Limiting State Probabilities

The state occupation probabilities are independent of the starting state of the process if the number of the state transition is large, thus the process reaches a steady state after a sufficiently large period of time Howard (10). This is equilibrium distribution

$$\pi = (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4)$$

If we let $n \to \infty$ in equation (2) we have

$$\pi = \pi p \qquad (3)$$

and $\pi = \sum_{i=1}^{4} \pi_i = 1$

## 5. Application

**Table 1.** *A summary of annual rainfall in Minna between 1970-2010 and states distribution.*

| Annual rainfall in mm | Frequency | State |
|---|---|---|
| Above 1201 | 18 | 3 |
| 1000-1201 (WS) | 14 | 2 |
| 1000-1201(NS) | 3 | 4 |
| Below 1000 | 6 | 1 |

WS: means the annual rainfall that spread evenly across the rainy season
NS: means annual rainfall that does not spread evenly across the rainy season. *(*Source; NMAM (8)*)*

The transition count matrix is shown below.

$$M = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 0 & 6 & 6 & 1 \\ 3 & 4 & 11 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix} \qquad (4)$$

From equation (4), using the maximum likelihood estimator, we obtained the transition probability matrix P

$$P = \begin{bmatrix} 0.333 & 0.333 & 0.167 & 0.167 \\ 0 & 0.462 & 0.462 & 0.077 \\ 0.167 & 0.222 & 0.611 & 0 \\ 0.333 & 0.667 & 0 & 0 \end{bmatrix}$$

## 6. The n-Step Transition Probability

From equation (1), we have on iteration

$$P^2 = \begin{bmatrix} 0.194 & 0.413 & 0.311 & 0.081 \\ 0.102 & 0.367 & 0.495 & 0.035 \\ 0.157 & 0.293 & 0.503 & 0.044 \\ 0.110 & 0.419 & 0.363 & 0.106 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.138 & 0.357 & 0.451 & 0.053 \\ 0.139 & 0.337 & 0.476 & 0.047 \\ 0.145 & 0.339 & 0.464 & 0.050 \\ 0.133 & 0.351 & 0.464 & 0.051 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.141 & 0.342 & 0.467 & 0.050 \\ 0.142 & 0.343 & 0.468 & 0.050 \\ 0.141 & 0.342 & 0.467 & 0.050 \\ 0.142 & 0.342 & 0.468 & 0.050 \end{bmatrix}$$

$$P^{16} = \begin{bmatrix} 0.142 & 0.343 & 0.468 & 0.050 \\ 0.142 & 0.344 & 0.469 & 0.050 \\ 0.142 & 0.343 & 0.468 & 0.050 \\ 0.142 & 0.343 & 0.469 & 0.050 \end{bmatrix} \qquad (5)$$

$$P^{16} = \begin{bmatrix} 0.14 & 0.34 & 0.47 & 0.05 \\ 0.14 & 0.34 & 0.47 & 0.05 \\ 0.14 & 0.34 & 0.47 & 0.05 \\ 0.14 & 0.34 & 0.47 & 0.05 \end{bmatrix} . \quad corrected \; to \; 2 \; decimal \; places$$

## 7. The Limiting State Probabilities

As $n$ increases $P^n$ gets closer and closer to (5), that is, $n \geq 16$ the transition probabilities stabilizes to (5) and from equation (2) with the initial state probability vector $(1 \quad 0 \quad 0 \quad 0)$ we have

$$P^0 P^n = (1 \quad 0 \quad 0 \quad 0)\begin{bmatrix} 0.142 & 0.343 & 0.468 & 0.050 \\ 0.142 & 0.344 & 0.469 & 0.050 \\ 0.142 & 0.343 & 0.468 & 0.050 \\ 0.142 & 0.343 & 0.469 & 0.050 \end{bmatrix} = (0.142 \quad 0.343 \quad 0.468 \quad 0.050)$$

From equation (3) and (5) the limiting state probability vector is given by

$$\pi = \pi P = (0.14 \quad 0.34 \quad 0.47 \quad 0.05)$$

## 8. Discussion of Result

The result shows that the probabilities to have a low rainfall, Moderate rainfall (well spread), High rainfall, and Moderate rainfall (not well spread) in the first year, given that it is a low rainfall at present are 0.333,0.333,0.167 and 0.167 respectively. The first and the last of these probabilities dropped slowly to 0.14 and 0.05 in about sixteen years. However, the probabilities of Moderate rainfall (well spread) and High rainfall increased to the maximum of 0.34 and 0.47 respectively within the same period of time. This shows that in the long-run 14% of the annual rainfall in Minna will be Low rainfall, 34% will be Moderate rainfall also well spread, 47% will be High rainfall and 5% moderate annual rainfall but not well spread. These equilibrium probabilities are independent of the initial state. That is be it high rainfall, low rainfall and the others. The model could therefore be used to make a forecast of the annual rainfall pattern. This could provide some information for the farmers and the government that could be used to plan strategies to boost crop production in Minna and the environment.

## 9. Conclusion

This paper demonstrates the application of Markov chain model to study the annual rainfall data in Minna in the North Central geo-political zone of Nigeria with respect to crop production. In view of the uncertainty of annual rainfall and crop production in the region, the annual rainfall of the present year could be used to make a forecast for the following year(s) and in the long run.

## References

[1] Akintunde A.A, Asiribo O.E. Adebanji A.O. Adelakun A.A Agwuegbo S.O.N. (2008) Stochastic modelling of daily precipitation in Abeokuta, proceedings of the third Conference on science and National Development pp108-118 www.unaab.edu.ng/journal/index.php/COLNAS/article/.../150/153. Accessed 15/10/2012.

[2] Gabriel, K. R. and Neumann, J. (1962) A Markov chain model for daily rainfall occurrences at Tel Aviv. Quart. J. Roy. Met. Soc. 88:90-95.

[3] Jain S. (1986), A Markov Chain Model and its Application, Comp.Biomed. Res. 19. 374-378

[4] Jimoh, O. D. and Webster, P. (1996). Optimum order of Markov chain for daily rainfall in Nigeria. Journal of Hydrology 185: 45-69.

[5] Jimoh, O. D. and Webster, P. (1999). Stochastic modelling daily rainfall in Nigeria: intra-annual variation of model parameters. Journal of Hydrology 222:1- 17.

[6] Kottegoda, N. T., Natale, L. and Raiteri, E. (2004); Some considerations of periodicity and persistence in daily rainfalls, J. Hydrol. 296:23–37.

[7] Cox D.R., and Miller H.D. (1984). The Theory of Stochastic Processes. Chapman and Hall London.

[8] NMAM (2011), Unpublished records of the Nigerian Meteorological Service Department, Federal Ministry of Aviation Minna.

[9] NSADP (2011). Unpublished document of the Niger State Agricultural Development Project. A Department in the ministry of Agriculture and Rural development.

[10] Howard R.A.(1971), Dynamic Probabilistic Systems, Vols.1and2. John Wiley, New York.

[11] Iloeje N.P. (1981), A New Goegraphy of Nigeria, Longman Nigeria Limited.

[12] Ross S.M. (1989), Introduction to Probability Models Academic Press, Inc. Ltd , London.

[13] Abubakar U.Y., Lawal A., Muhammed A. (2013). The Use of Markov Model in Continuous Time for the Prediction of Rainfall for Crop Production, International Organization for Scientific Research, (IOSR) Vol.7, Issue1, pp 38 – 45.