Markovian Model of Rainfall Pattern with Application

Lucy Makokha[1], Kennedy Nyongesa[2], Adu Wasike[3], Metrine Chonge[4],
Frankline Tireito[5], Stephen Waswa[6]

[1,2,3,4,,5,6](Department of Mathematicss, Masinde Muliro University of Science and Technology, Kenya)

Abstract: *In this study, we model the occurrence and length of wet, medium wet and dry spells by Markov chain that best describes the rainfall pattern of Bungoma County (Western Kenya).This is achieved by Markov chain theory and estimation of probabilities of the chain by MLE. Also computed is the distribution of the length of each spells; wet, medium wet and dry from which the central moments of the rainfall pattern are computed. The model developed is applied to rainfall data from Bungoma meteorological station. A three by three transition matrix is obtained and used to predict the weather pattern. It is observed that if everything remains constant, prediction can be certain at the twelfth year as the matrix show stationarity. The three states are recurrent, non-null and a periodic hence forming an ergodic chain.*
Keywords: *Markov chain, Wet spell, Medium wet spell, Dry spell, Prediction, Stationary distribution*

I. Introduction

Sport Climate is the weather conditions of a region as temperature, air, pressure, humidity, precipitation (rainfall), sunshine, cloudiness and winds throughout the year. These weather conditions are averaged over a standard period of 30 years to describe the climate of a place (ASAP, 2012). The major component of climate that characterizes agricultural activities of any region is rainfall. Hence there is need to understand its patterns and occurrence by use of suitable mathematical models.

There has been a growing demand for modeling daily rainfall data using various stochastic models because once the rainfall pattern is adequately and appropriately modeled, the model can be used to provide prior knowledge of the structural characteristics of varying rainfall systems which are very much essential for agricultural and hydrological planning, industrial and water recourse management, and climate change studies (see for instance, Mangaraj, 2013). Hydrology and climatological studies have highlighted the significance of the analysis of rainfall pattern because of its influence on all human activities (Rahman, 2000). Such analysis in most cases focuses on two main characteristics of rainfall patterns, wet and dry spells. A dry spell is a period where there has been no rainfall for a long period, shorter than and not severe as drought, while a wet spell occurs when there is an interrupted sequence of rainy weeks. The volume of rainfall in any geographical area depends heavily on the distribution of such spells. (Borgardi, 1988). It is therefore important to investigate the pattern and occurrence of these spells through models that consists of studying statistical properties of two indicators namely; the length of a spell and its frequency. The knowledge on the probability of occurrence of such spells can be used for crop planning and water management decisions.

The occurrence of dry and wet spells cannot be predicted with certainty and hence must be treated as random variables that can be investigated by theories of probability and stochastic (Sharma, 1994). Such theories include the Markov chain theory that led to the development of Markov chain model that has its origin in the work of Gabriel and Neuman (1962). Gabriel and Neuman (1962), used the Markov chain in estimating the distribution of the wet and dry day sequence during the rainy season at Tel-Aviv (Israel). Other studies have also cited that the reliability of meteorological persistence can be described through Markov chain model of proper order (Rahman, 1999a). Markov chain model is also found to be promising in simulating the length of the longest dry and wet spells and largest rainfall amounts (Shamah, 1996). In view of the forgoing background, this study analyzed the pattern of wet, medium wet and dry spells of rainfall by fitting a Markov chain of order one on rainfall data that was obtained from a meteorological station in Bungoma County (Kenya).

II. Methodology

The Markov chain model assumes that the occurrence of a wet, medium wet or dry week depends on the weather condition of the previous week only, this is: If a sequence of random variables $X_t, t = 0,1,2,\ldots\ldots\ldots\ldots$ defined on an algebra δ of the probability space (Ω, ∂, ρ) has the property that:

$$pr(X_{t+1} = k / X_0 = k_0, \ldots\ldots X_t = i)$$

$$= pr(X_{t+1} = k / X_t = i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

That is, the process losses memory of the history and only recognizes the present state.

Suppose a week is classified as wet if the amount of rainfall is greater or equal to *b mm*, and dry if the amount of rainfall received is less or equal to *a mm*. Clearly, a week is medium wet if the amount of rainfall lies between *a* and *b* with *a < b*. The amount of rainfall received per week in a given period is a random variable herein denoted by *Y*. The probability that a week is wet, medium wet and dry respectively are:

$$P = P(Y \geq b) \quad , \quad P = P(a < Y < b) \quad \text{and} \quad P = P(Y < a) \quad \text{...............................(2)}$$

Define the indicator functions: $I_{1i} = \begin{cases} 1 \\ 0 \, otherwise \end{cases}$, $I_{2j} = \begin{cases} 1 \\ 0 \, otherwe \end{cases}$, $I_{3k} = \begin{cases} 1 \\ 0 \, otherwise \end{cases}$

Let the total number of wet, medium wet and dry weeks be denoted by, $x_1 = \sum_{i=1}^{n} I_{1i}$,

$x_2 = \sum_{j=1}^{n} I_{2j}$ and $x_3 = \sum_{k=1}^{n} I_{3k}$.

Hence the total number of weeks in a given period to classify the climate is $x_1 + x_2 + x_3 = n$

The number of wet, medium wet and dry are assumed to be random variables here denoted by (X_1, X_2, X_3), with their joint distribution given by;

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid P\,P\,P\,)\alpha P^{x_1} P^{x_2} P^{n-x_1-x_2} \quad \text{.............................(3)}$$

Equation (3) is then utilized to obtain the estimates of $P_1\,P_2$ and P_3 through MLE method and they are:

$$\hat{P}_1 = \frac{x_1}{n} \quad , \quad \hat{P}_2 = \frac{x_2}{n} \quad \text{and} \quad \hat{P}_3 = 1 - \hat{P}_1 - \hat{P}_2 \quad \text{...(4)}$$

The probability of interest here is the probability that the system can move from one state say i to another j in one step denoted by P_{ij}. The climate of Bungoma oscillates between three states namely, wet, medium wet and dry denoted by S = {i, j, k} where S is called the state space of the process. The transition matrix for these states is,

$$T_3 = \begin{bmatrix} P_{ii}, P_{ij}, P_{ik} \\ P_{ij}, p_{jj}, p_{jk} \\ p_{ki}, p_{kj}, P_{kk} \end{bmatrix} \quad \text{..(5)}$$

The estimation of the entries of T_3 is accomplished by use of MLE. First, we have to arrange our data in pattern form to obey Equation (1) as given in the following observed frequencies table.

Table (1) Observed Frequency Table

Previous week	Current week			
	Wet	Medium wet	Dry	Total
Wet	F_{WW}	F_{WM}	F_{WD}	$F_W *$
Medium	F_{MW}	F_{MM}	F_{MD}	F_{M*}
Dry	F_{DW}	F_{DM}	F_{DD}	F_{D*}

The frequencies in table (1) are vital in the estimation of entries of T_3. The conditional distribution of wet, medium and dry given that the previous week was wet is a quadrinormial model;

$$f(W,M,D \mid W) \quad \alpha \quad (p_{ii}{}^{F_{WW}} P_{ij}{}^{F_{WM}} P_{ik}{}^{F_{WD}})(1 - P_{ii} - P_{ij} - P_{ik})^{F_{W*}-F_{WW}-F_{WM}-F_{WD}} \quad \text{....................(6)}$$

From which the MLE of the entries of the first row of T_3 are easily obtained as:

$$\hat{p}_{ii} = \frac{F_{WW}}{W *}, \quad \hat{p}_{ij} = \frac{F_{WM}}{W *} \quad \text{and} \quad \hat{p}_{ik} = \frac{F_{WD}}{W *}.$$

Similarly, the joint distribution of wet, medium wet and dry given that the previous week was medium wet is;

$$f(W,M,D \mid M) \quad \alpha \quad (P_{ji}{}^{F_{MW}} P_{jj}{}^{F_{MM}} P_{jk}{}^{F_{MD}})(1 - P_{ji} - p_{jj} - P_{jk})^{F_{M*}-F_{MW}-F_{MM}-F_{MD}} \quad \text{....................(7)}$$

From which the MLE of the entries of the second row of T_3 are ;

$$\hat{p}_{ji} = \frac{F_{MW}}{M *} \quad \hat{p}_{jj} = \frac{F_{MM}}{M *} \quad \hat{p}_{jk} = \frac{F_{MD}}{M *}$$

The conditional joint distribution of W,M,D given the previous week is D is

$$f(\text{W,M,D|D}) \quad \alpha \quad (P_{ki}^{F_{DW}} P_{kj}^{F_{DM}} P_{kk}^{F_{DD}})(1 - P_{ki} - P_{kj} - P_{kk})^{F_{D*} - F_{DW} - F_{DM} - F_{DD}} \quad \text{..............................(8)}$$

The MLE of the entries of the third row of T_3 from equation (8) are;

$$\hat{p}_{ki} = \frac{F_{DW}}{D*}, \quad \hat{p}_{kj} = \frac{F_{DM}}{D*} \text{ and } \hat{p}_{kk} = \frac{F_{DD}}{D*}$$

The estimates above are then used in calculating the probability distribution of wet, medium wet and dry spells which are obtained from;

$$P(k_w = k) = (P_{10} + P_{20})(1 - P_{00})P_{00}^{k-1},$$

$$P(k_m = k) = (P_{01} + P_{21})(1 - P_{11})P_{11}^{k-1} \text{ and}$$

$$P(k_d = k) = (P_{02} + P_{12})(1 - P_{22})P_{22}^{k-1}$$

Where k_w, k_m and k_d are lenghs of wet, medium wet and dry spells respectively.

From definition and use of equation (9), the expected value of a wet, medium wet and dry spells are given by;

$$E(k_w) = \sum_{k=1}^{\infty} kP(k_w = k) = [P_{10} + P_{20}]\frac{1}{(1 - P_{00})},$$

$$E(k_m) = \sum_{k=1}^{\infty} kP(k_m = k) = [P_{01} + P_{21}]\frac{1}{(1 - P_{11})}$$

And

$$E(k_d) = \sum_{k=1}^{\infty} kP(k_d = k) = [P_{02} + P_{12}]\frac{1}{(1 - P_{22})} \quad \text{..............................(10)}$$

While their variances can easily be computed from definition and elements of T_3 as;

$$Var(k_w) = E(k_w - E(k_w))^2 = [P_{10} + P_{20}]\frac{P_{00}}{(1 - P_{00})^2},$$

$$Var(k_m) = E(k_m - E(k_m))^2 = [P_{02} + P_{12}]\frac{P_{11}}{(1 - P_{11})^2} \quad \text{and}$$

$$Var(k_d) = E(k_d - E(k_d))^2 = [P_{02} + P_{12}]\frac{P_{22}}{(1 - P_{22})^2} \quad \text{..............................(11)}$$

III. Results And Discussion

The Transition matrix T_3 is estimated from the rainfall data of area of study with help of table (1) as;

$$T_3 = \begin{bmatrix} 0.5491, 0.1525, 0.2984 \\ 0.4698, 0.1256, 0.4047 \\ 0.2567, 0.0982, 0.6451 \end{bmatrix}$$

Observe that T_3 is a stochastic transition matrix. Starting with parameters of T_3, one can obtain the corresponding

parameters after say n years, this are obtained by aid of MATLAB computer programme or Eigen vector

method, for instance the parameters at n=5 are given by;

$$[T_3]^5 = \begin{bmatrix} 0.4017, 0.1236, 0.4747 \\ 0.4007, 0.1234, 0.4760 \\ 0.3997, 0.1230, 0.4788 \end{bmatrix}$$

While the parameters at n=8 are;

$$[T_3]^8 = \begin{bmatrix} 0.4000, 0.1233, 0.4788 \\ 0.4000, 0.1233, 0.4760 \\ 0.3997, 0.1233, 0.4771 \end{bmatrix}$$

The distribution of the chain attain stationarity as $n \to \infty$, this well defined limiting behavior is attained at n = 12 where

$$[T_3]^{12} = \begin{bmatrix} 0.4000, 0.1233, 0.4769 \\ 0.4000, 0.1233, 0.4769 \\ 0.4000, 0.1233, 0.4769 \end{bmatrix}$$

At these point, if π is a row vector such that $\pi = [\pi_1, \pi_2, \ldots \ldots \pi_m]$, where m is the number of states and π^0 the initial distribution of the chain, then

$$\pi = \frac{Lim}{n \to \infty} \pi^0 p^n, \text{ With } \pi_j \geq 0 \text{ and } \sum_{j=1}^{m} \pi_j = 1$$

The stationary distribution exists because the states of the chain are recurrent, non null and aperiodic forming an ergodic chain. For recurrence, utilizing elements of T₃, the probability of return to state w after n steps $f_w^{(n)}$ is;

$$f_W^n = 0.5491 + 0.1525 X 0.4698 (0.1256)^{n-2}$$

For all $n \geq 2$

$$\therefore \sum_{n=0}^{\infty} f_W^n = 0.5451 + 0.1525 X 0.4698 \sum_{n=2}^{\infty} (0.1256)^{n-2}$$

$$= 1.0013$$

$$\approx 1$$

Since $\sum_{n=0}^{\infty} f_W^n = 1, \Rightarrow$ state W is recurrent.

Non-nullity of states is tested by computing the expected number of returns to the same state by the chain $E\left[f_w^{(n)}\right]$ by use of elements of T₃, this is;

$$E\left[f_w^{(n)}\right] = \sum_{n=0} n f_w^n = 0.5491 + \sum_{r=2}^{\infty} n(0.0716)(0.1256)^{n-2}$$

$$= 0.5491 + 0.1973$$

$$= 0.7464$$

$E\left[f_w^{(n)}\right] < \infty \Rightarrow W$ is a non –null state.

From T₃, $p_{ii} > 0$ in $T_3^1, T_3^2, T_3^3, \ldots \ldots \ldots, T_3^n$. Gcd (1,2,3,………n) =1, hence state w is aperiodic. With similar reasoning, states M and D are also recurrent, non-null and aperiodic therefore the chain formed by the states is Egordic.

The central moments of the different spells as computed from the data are $E(k_w) = 1.6, E(k_m) = 1.1, E(k_d) = 2.6, Var(k_w) = 0.20, Var(k_m) = 0.04, Var(k_d) = 0.28$.

Results from expected values indicate that longer dry spells are expected to occur annually in this region as compared to the wet and medium wet spells. There is a clear indication from variance of spells that the medium wet spell was better simulated as compared to the wet and dry spells.

IV. Conclusion

In this study, a Markov chain method was applied on a data set from a rainfall network of Bungoma County (Western Kenya)for a period of 36 years (1972-2013). The study investigated prediction transition probabilities of weekly rainfall with techniques based on Maximum likelihood estimation. The model enabled us to predict stationary distribution of the rainfall events which occurred at the twelfth year. The study results could be used in formulation of important agricultural and water management policy issues in Bungoma County region (Western Kenya).

V. Recommendations

The study concentrated on use of first order Markov chain and it could be interesting for further investigations to be carried out on use of higher order Markov chains then results compared for efficiency. We also recommend inclusion of a suitable technique that can model rainfall amounts for the purpose of water harvesting.

Acknowledgement

References

[1]. Barkotulla M.A.B, Rahmann M.S. (2012). Multi-state Markov chain modeling system for Environmental Impact of climate change. *International Journal of statistical sciences* Vol. **12**: 29 - 46.

[2]. Biamah E. K, Gichuki N, Kaumbutho P.G. (1993). Tillage methods and soil and water conservation in East Africa. *Soil and Tillage Research*, Vol. **27**: 105-123.

[3]. Biamah E. K, Sterk G, Shamah T.C. (2004). Analysis of agricultural drought in Liuni eastern Kenya: Application of Markov model. *Journal of Hydrology Processes* Vol.**19**: 1307-1322.12

[4]. Bogardi J, Duckstein L, Ruhambo O.H. (1988). Practical generation of synthetic rainfall event series in a semi-arid climate zone. *Journal of hydrology* Vol. **103**: 357-313.

[5]. Fischer B. M. C, Mul M.L and Savinije H.H.G. (2012). *Determining spatial variability of dry Spells – a Markov based method, applied to the Makanya catchment*. Tanzania Hydrology And Earth Sci system. 11707 - 11731.

[6]. Gabriel K.R, Neuman J. (1962). A Markov chain model for daily occurrence at Tel Aviv, *Royal Metrological Society*. Great Britain, Vol. **88**: 90 - 95.

[7]. bibitem{Mangaraj} Mangaraj A.K, Sahoo L.N,and Sukia M.K. (2013). \textit{A Markov chain analysis of daily rainfall Occurrence at West Orisa of India}. Journal of reliability and statistical studies Vol. 6: 77-86.

[8]. Ochola W. O, Kerkides P. (2003). A Markov Chain simulation model for predicting critical wet and dry spells in Kenya: Analysing rainfall events in the Kano plains. *Irrigation and Drainage* Vol. **52(4):** 327-342

[9]. Rahman M. S. (1999, a). A stochastic simulated Markov Chain Model for daily rainfall in Bangladesh. *Journal of interdisciplinary Mathematics* Vol. **2(1):** 7-32.

[10]. Rahman M. S.(1999, b). Logistic regression estimation of a simulated Markov Chain Model for daily rainfall in Bangladesh. *Journal of interdisplinary Mathematics* Vol. **2(1):** 33-40

[11]. Rahman M. S. (2000). A rainfall simulation model for agricultural development in Bangladesh. Discrete Dynamics in Nature and society Vol. **5**: 1-7.

[12]. Rahman S, Basher M. M. A. (2002). Stochastic study of asymptotic behavior Of rainfall distribution in Rajsahi. *International journal of statistical sciences* Vol. **1**: 36-48.

[13]. Sharma T.C. (1994). *Stochastic features of drought in Kenya, East Africa*. Kluwer Academic Publishers, The Netherlands, Vol. **1**: 125-135

[14]. Sharma T.C. (1996). Simulation of the Kenyan longest dry and wet spell and the largest rain-sums using a Markov model. *Journal of hydrology*. Vol. **178**: 55-67.