ISSN: 2277-9655
Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

CONTEMPORARY MATHEMATICAL TOOLS OF FORECASTING: OVERVIEW OF METROPOLIS HASTINGS ALGORITHMS

Clement Ayarebillah Ali*, Francis T. Oduro

* Department of Basic Education University of Education, Winneba Winneba, Ghana
Department of Mathematics Kwame Nkrumah University of Science and Technology Kumasi, Ghana

DOI: 10.5281/zenodo.51437

ABSTRACT

This paper used the Metropolis Hastings Markov Chain Monte Carlo algorithms to forecast and monitor the dynamics of HIV/AIDS prevalence rates in Ghana. The study sampled four sites in Upper East Region and explored them with these algorithms. Even though there were many Monte Carlo algorithms, the study discovered that the independence Metropolis-Hastings' were the most suitable and appropriate for this mathematical forecasting and monitoring. We therefore, recommended for the continuous and extensive use of these algorithms in immunological surveys to help modify the continuous use of prevalence rates.

KEYWORDS: mathematical tools; Markov Chain Monte Carlo; Metropolis Hastings.

INTRODUCTION

Smith (2007), Mira & Sargent (2006), and Roberts & Rosenthal (2004) explain the Markov Chain Monte Carlo (MCMC) algorithms as chains having complicated stationary distributions, for which it is important to understand some simulation techniques and the speed of convergence. The idea iteratively and recursively simulate probability distributions that are representative of the magnitude and nature of the prevalence rates of HIV virus forecast.

Chib & Carlin (2007), Lewis et al, (2007) and Krishnan (2004) confirm that MCMC algorithms have been applied successfully to analyze HIV/AIDS data and are widely applicable. This is because they are computationally more efficient and faster to implement than many multinomial methods in analyzing more complex evolutionary models and larger datasets. In particular, Smith (2007), Mira & Sargent (2006), and Roberts & Rosenthal (2004) discovered that the Metropolis-Hastings algorithms have become extremely popular in high dimensional data. However, as supported by Guimarães et al (2009), these methods are quite unpopular in the analysis of HIV/AIDS data in the developing world. We therefore wish to explore the MCMC algorithms as mathematical tools of forecasting and educating people on the dynamics of HIV/AIDS prevalence rates in Ghana.

Markov Chain Monte Carlo Algorithms

Chen (2009), Carter (2008), Walsh (2004), and Hanson (2000) explain the MCMC algorithms as a class of general computational algorithms for sampling from posterior distributions and computing posterior quantities of the random variables X_n. What a good MCME models seeks to do is implore a preceding state (X_{n-1}) to arrive at the final steady state(s) to produce the same expected values.

Gibbs Algorithms

The Gibbs algorithms are special cases of Metropolis-Hastings' algorithms. Even though the Gibbs' are usually faster and easier to use, they are less applicable to approximate joint distributions. It is however, agreed that the characteristics and composition of HIV/AIDS prevalence rates are multifaceted and require multiple distributions to compute and understand the dynamics (Walsh, 2004; Ewens & Grant, 2001; Hanson 2000).

ISSN: 2277-9655
Impact Factor: 3.785

Metropolis Algorithms

The Metropolis algorithms are the methods of computing complex integrals by expressing them as expectations for some distributions X_n, and estimating them as separate expectations. Thus, the Metropolis algorithms generate sequences of chains $(X_0, X_1, ..., X_n)$, as the transition probabilities that do not depend $(X_0, X_1, ..., X_{n-1})$ and obtain the stationary distributions of X_n (Walsh, 2004; Hanson, 2000).

Metropolis-Hasting (MH) Algorithms

The Metropolis–Hastings algorithm is an MCMC method for obtaining a sequence of random samples X_n from a probability distribution for which direct sampling is difficult. Unlike the Gibbs' and Metropolis, the data are probabilistic, results depend on previous, and stability is arrived at the optimum number of iterations. Quite apart, the simulation procedures follow the constructs of Markov chains with the random, and compute its states to the target steady state distribution. (Ali & Oduro, 2012; Malve, 2007; Walsh, 2004).

There are two main MH algorithms, vis-a-vis random-walk metropolis and random-walk metropolis.
In random, the simulation introduces an error term (ε_k) in the models and analyzes this error simultaneously with the main parameters. It is however, tedious and time-consuming because one has to always physically control ε_k, or at least adjust ε_k to an acceptable levels before making any valid conclusions. Thus, the theory of random walk supposes that data must be statistically independent. But this is not case in many forms of HIV/AIDS prevalence rates that are often correlated (Ali& Oduro, 2014; Browne, 2003; Johannes & Polson, 2003).

On the other hand, the Independence Metropolis-Hastings algorithms simulate data, which depends on the previous states. Even though the future states are drawn independently of the previous states, the current states would not be independent, since the acceptance probabilities also depend on previous states. All that one needs is to approximate the initial states in a way that they closely match to certain properties of the steady distribution. This ensures that the algorithms correctly forecast and target the designated final states (Ali& Oduro, 2014; Browne, 2003; Johannes & Polson, 2003).

Computing Independence Metropolis-Hastings Algorithms

[1] We usually start with the initial probabilities as $X_n^{(0)}$, and then iterate it.

[2] We then arrive at the final probabilities as $X_n^{(s)}$ $\Pr\left(\frac{X_n^{(k)}}{X_n^{(s)}}\right)$.

[3] We then calculate ratio of the initial to the final as 'a'.

[4] We analyze as if $a > 1$, then we have attained the steady, otherwise, we continue the iteration or abort our initial statistical assumptions.

Because the mathematical computation is somehow tedious here, we could redirect the algorithms to Markov chains, and outline the initial and final distributions in matrix forms The Markov chains make it easier to test the convergence with autocorrelation lags, Z-tests or cross correlations (Ali& Oduro, 2014; Chen, 2009; Lam, 2009; Browne, Tsurumi, 2005; 2003; Johannes & Polson, 2003).

Statement of the Problem

In many instances, classical hypothesis testing in HIV/AIDS analysis postulates a null hypothesis, and fails to accept the hypothesis, if otherwise. However, the dynamics of the HIV/AIDS in the sentinel sites cannot normally be determined. In many parts of the world, stakeholders and researchers mostly use prevalence rates to report and forecast the HIV/AIDS virus in the sentinel sites. The prevalence rates methods desire lot of modifications and transformations. This is because the methods do not take into considerations the interdependence amongst infected and uninfected

ISSN: 2277-9655
Impact Factor: 3.785

people, migration dynamics of the people, and interdependence of people in many human activities, and other activities that provide potential grounds for people to interact. In effect, one finds it very difficult to use these prevalence rates to predict, forecast and monitor the incidence either within a particular geographical area or between two or more places. Therefore, these algorithms provide this opportunity to simulate to assess the threat of infected people on those uninfected to help in forecasting and monitoring the dynamics of human endeavours. In particular, the independence MH algorithms provide interdependence between the past, present and future rates to forecast long-term phenomena. This study therefore, sought to explore these contemporary mathematical tools of forecasting and monitoring disease phenomena.

METHODOLOGY

Written informed consent and assent letters were obtained from my head of department and distributed to stakeholders of the Ghana Health Services. There were assurances of confidentiality and anonymity of the data. The study sampled the HIV/AIDS data from four sites in the Upper East Region of Ghana, through the Regional HIV/AIDS Coordinator and under the authority of the Regional Health Directorate. The data collection and collation process spanned between September to December 2015. In the data, we designated $X_n^{(0)}$ as the initial, $X_n^{(k)}$ transition and hence the product of $X_n^{(0)} X_n^{(k)}$ designated the various intermediary states. We then used the SPSS to produce the cross correlation graphs, and the Matlab software to simulate the state distributions.

RESULTS AND ANALYSIS

This section contains the autocorrelation graphs of Bawku with Bolgatanga, Builsa and Navrongo, and the algorithms of the state distributions.

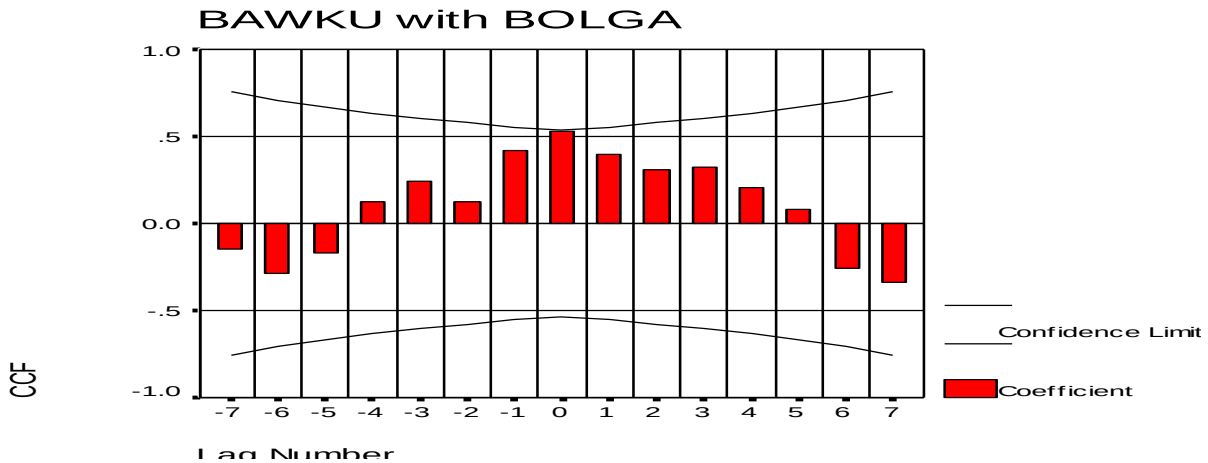

Figure 1: CCF Graph of Bawku with Bolgatanga

Figure 1 above shows the CCF graph between Bawku and Bolgatanga. The lags were both negative and positive as k increases. However, there were more positive lags than the negative ones. This suggests that there were more movement and interaction between infected and uninfected persons in Bawku and Bolgatanga zones.

ISSN: 2277-9655
Impact Factor: 3.785

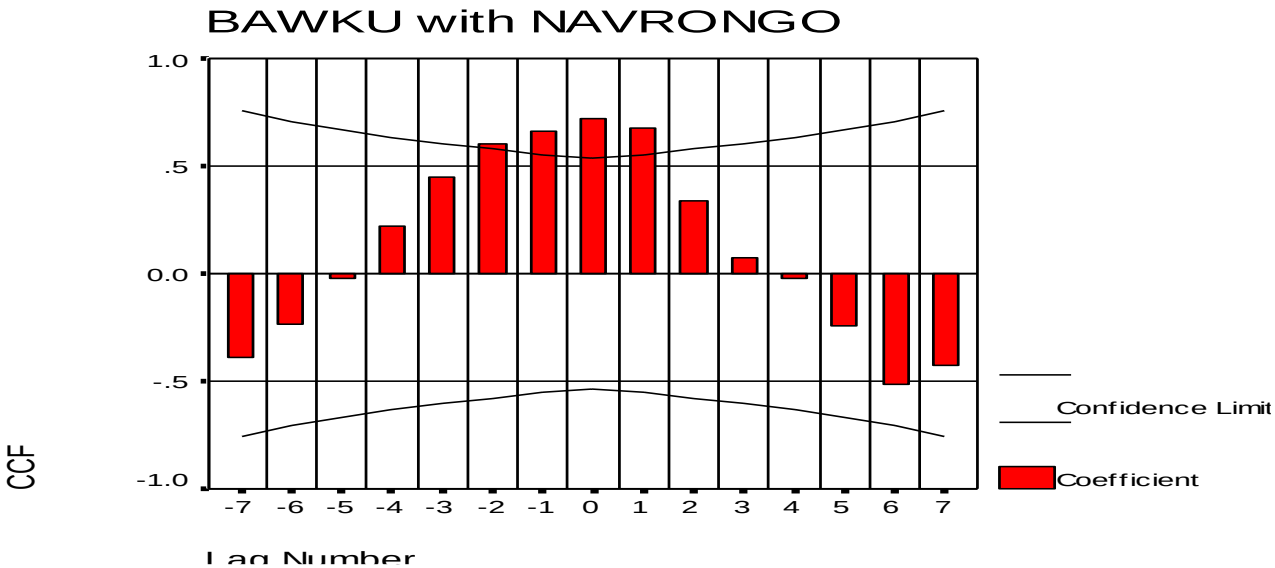

Figure 2: CCF Graphs of Bawku with Navrongo

Figure 2 above shows the CCF graph between Bawku and Navrongo. The lags were both negative and positive as k increases. Here, there were equal positive and negative lags. However, the positives exceeded the threshold of five. Therefore, the degree of interaction was even much more observed in these two sites as compared to Bawku with Bolgatanga, and must be checked and controlled.

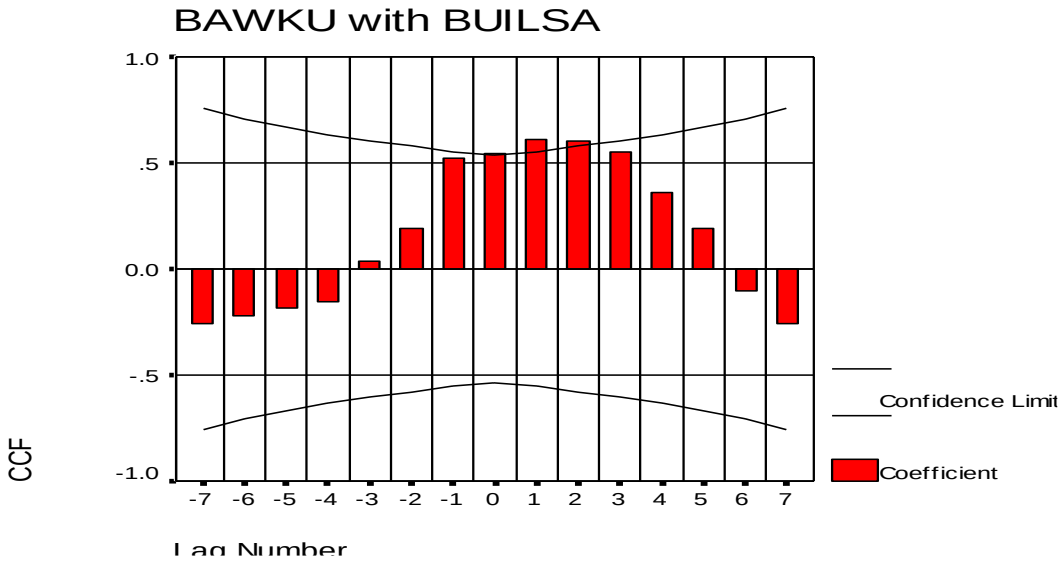

Figure 3: CCF Graphs of Bawku with Builsa

Figure 3 above shows the CCF graph between Bawku and Builsa. As usual, there were both negative and positive lags as k increases. However, there were morel positive than negative lags, and just a few exceeded 5. These also showed intensive interaction between these two sentinel zones.

MCMC algorithms

The CCF graphs have established interactions that sought to explain that infected and uninfected people readily have contacts with one another. So suppose Bawku— B_a , Bolgatanga— B_o , Navrongo— N_a and Builsa-- B_u , and from the initial proportions of infected people from the four zones, we would obtain the initial MCMC transmission matrix as follows:

$$X_n^{(ij)} = \begin{bmatrix} & B_a & B_o & N_a & B_u \\ B_a & 0.4 & 0.3 & 0.2 & 0.1 \\ B_o & 0.3 & 0.4 & 0.2 & 0.1 \\ N_a & 0.2 & 0.3 & 0.3 & 0.2 \\ B_u & 0.1 & 0.2 & 0.3 & 0.4 \end{bmatrix} \quad \text{.....(1)}$$

Data from the various zones show the proportions of people already infected with the virus.

$$X_n^{(0)} = \begin{bmatrix} B_a & B_o & N_a & B_u \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix} \quad \text{.....(2)}$$

This means the highest cases were recorded in Bawku and Bolgatanga. We would obtain the intermediate state by multiplying equations (1) and (2), and so on, until we arrive at the final state as in equation (4) below. A back up check to stability is to compute the powers of equation (1) until each row of the final matrix is similar to equation (4).

$$X_n^{(0)} X_n^{(ij)} = \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{bmatrix} \quad \text{.....(3)}$$

$$X_m^{(s)} = \begin{bmatrix} 0.2671 & 0.3136 & 0.2419 & 0.1774 \end{bmatrix} \quad \text{.....(4)}$$

$$X_n^{(ij)^k} = \begin{bmatrix} & B_a & B_o & N_a & B_u \\ B_a & 0.2670 & 0.3136 & 0.2419 & 0.1774 \\ B_o & 0.2670 & 0.3136 & 0.2419 & 0.1774 \\ N_a & 0.2670 & 0.3136 & 0.2419 & 0.1774 \\ B_u & O.2670 & 0.3136 & 0.2419 & 0.1774 \end{bmatrix}, k \to \infty \quad \text{.....(5)}$$

The equations (1) and (2) represented the initial matrices. The MH simulations obtained the steady states at the equations (4) and (5). We have noticed nonzero entries in (4) and (5). This means that in the end, the HIV/AIDS virus transmissions will stabilize to the rates in (4) in the four sites in the Upper East Region. Stakeholders should use these rates as mathematical tools of forecasting and monitoring the dynamics of HIV/AIDS.

DISCUSSIONS OF RESULTS

ISSN: 2277-9655
Impact Factor: 3.785

We observed that $\begin{bmatrix} B_a & B_o & N_a & B_u \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix}$ were the proportions of persons infected with the HIV virus in the original data. After the independence MH simulations, the rates reached an optimum steady matrix at $\begin{bmatrix} B_a & B_o & N_a & B_u \\ 0.27 & 0.31 & 0.21 & 0.18 \end{bmatrix}$. This represented the long-term rates of persons from Bawku, Bolgatanga, Navrongo and Builsa sites respectively. This means the frequency of its occurrence will level off with time in the region. However, one can infer and envisage that the urban sites of Bawku and Bolgatanga would remain higher cases of the HIV/AIDS than the rural sites of Navrongo and Builsa in the Region. Therefore, prevention, education and supply of logistics should be swifter in areas with higher interactions and dependencies than those with observed lower interactions. That notwithstanding, care must be taken to suppress its spread in within a particular zones, as internal interactions could equally cause more infections.

CONCLUSIONS AND RECOMMENDATIONS

In these methods, we have discovered the theoretical underpinnings of the independence MH algorithms as a mathematical tool of forecasting and monitoring the dynamics of HIV/AIDS prevalence rates. We discovered that while the other MCMC algorithms may have a number of practical problems, independence MH had successfully iterated the HIV/AIDS data. The algorithms and its simulations produced the initial states as well as steady state distributions of the sites. We have also used the CCF convergence graphs to assess the suitability of the MCMC algorithms. Therefore, the independence MH algorithms can mathematically revolutionize the MCMC algorithms as mathematical tools of forecasting and monitoring the dynamics of HIV/AIDS prevalence rates.
We therefore, recommend for the continuous and extensive use of these MH algorithms in immunological surveys to help modify and replace the continuous and over-dependent reliance on prevalence rates for surveillance, forecasting and monitoring the dynamics of HIV/AIDS prevalence rates, and other disease data.

However, the study focused on only secondary reported data because the researchers were unable to perform the tests themselves. The immunological surveys would have really determined the antibody levels of the infected persons within the various sites, how fast persons can spread to enable stakeholders make much more intensive diagnosis of the tools, give insights into the efficacy of the current antiretroviral vaccines in use and the possibility of implementing immunization programmes against HIV/AIDS.

REFERENCES

[1] Ali, C.A. & Oduro, F.T. (2014). Bayesian Models of HIV/AIDS Sentinel Surveillance Transmission of HIV/AIDS in the Upper East Region of Ghana***. International Journal of Advanced Studies in Engineering and Scientific Inventions - Vol. 2, No. 1 January***, *2014[Website: http://www.internationalpolicybrief.org ISSN (electronic): 1741-8771 ISSN (print): 1741-8763]*.
[2] Ali, C.A. & Oduro, F.T. (2012). Markov Chain Models of Age and Gender Structure Dynamics of HIV/AIDS Transmission in the Upper East Region of Ghana. ***Proceedings of Koforidua Polytechnic 5th Annual International Applied Science Research, from 25th to 28th June, 2012***. Paa-WillsPress: Koforidua Polytechnic.
[3] Browne, W.J. (2003). ***MCMC Estimation in MLwiN***. London: Institute of Education, University of London.
[4] Carter Jr, E.F. (2008). ***Random Walks, Markov Chains and the Monte Carlo Method***. [http://www.taygeta.com/rwalks/rwalks.html].
[5] Chen, F. (2009). ***Bayesian Modeling Using the MCMC Procedure***. [http://support.sas.com/resources/papers/proceedings09/257-2009.pdf].
[6] Chib, S. and Carlin, P.B. (2007). ***MCMC Sampling in Hierarchical Longitudinal Models***. [http://portal.acm.org/citation.cfm?id=599297.html].
[7] Ewens, W.J. and Grant, G.R. (2001). ***Statistical Methods in Bioinformatics, An Introduction (2nd edition)***. New York: Springer.
[8] Guimarães et al (2009)***. Close phylogenetic relationship between Angolan and Romanian HIV-1 Subtype F1 Isolates***. [http://www.retrovirology.com/content/6/1/39.html].
[9] Hanson, M.K. (2000). ***Tutorial on Markov Chain Monte Carlo***. [http://www.lanl.gov/home/kmh/talks.pdf].

[10] Johannes, M. and Polson, N. (2003). ***MCMC Methods for Financial Econometrics***. [http://gsbwww.uchicago.edu/fac/nicholas.polson/res.pdf].

[11] Krishman, J.M. (2004). *Phylogenetic Inference of Complex, Evolutionary Models: A Bayesian Approach* [http://etd.Isu.edu/doc/available/etd- 07072004o94725/unrestricted/Krishman_thesis.pdf].

[12] Lam, P. (2009). *Convergence Diagnosis*.[http:/www.people.fas.harvard.edu/convergence/convergence-print.pdf].

[13] Lewis, F., Hughes, G.J., Rambaut, A., Pozniak, A., and Brown, A.J.L. (2007). ***Episodic Sexual Transmission of HIV Revealed by Molecular Phylodynamics***. [www.pnas.org/content/104/47/18566.full.pdf].

[14] Malve, O. (2007). ***Water Quality Prediction for River Basin Management*** [http://lib.tkk.fi/Diss/2007/isbn9789512287505/].

[15] Mathews, J.H. and Fink, K.D. (2004). ***Numerical Methods Using Matlab (4th ed)***. New Jersey: Pearson Prentice Hall, Inc, 600-609.

[16] Mira, A. and Sargent, D.J. (2006). ***A new strategy for speeding Markov chain Monte Carlo algorithms*** [http://www.springerlink.com/content/812605777731h323].

[17] Roberts, G.O. and Rosenthal, J.S. (2004). ***General State Space Markov Chains and MCMC Algorithms***.[www.emis.de/journals/ps/images/getdoc510c.pdfid=35].

[18] Ross, M.S. (2000). ***Introduction to Probability Models, (7th ed)***. San Diego: Academic Press, 137-142.

[19] Smith, B.J. (2007). *BOA:* ***An R Package for MCMC Output Convergence Assessment and Posterior Difference***. [http://www.jstatsoft.org/v21/i11/paper].

[20] Tsurumi, H. (2005). ***Recent Developments of Bayesian Econometrics: Bayesian Inference Using MCMC Algorithms***. [http://21coe.ier.hitu.ac.jp/information/lecturelog/../L0501.pdf].

[21] Walsh, B. (2004). ***Markov Chain Monte Carlo and Gibbs Sampling*** [http://www.maths.surrey.ac.uk/personal/st/S.Brooks/mcmc/.pdf].