Covenant Journal of Physical & Life Sciences (CJPL) Vol. 6 No. 2, Dec. 2018

An Open Access Journal Available Online

Modeling Tuberculosis (TB) Using Higher Order Markov Model

Adewara Johnson Ademola* & Gbadegeshin Monsurat

Distance Learning Institute, University of Lagos
Mathematics Dept, University of Lagos
*adewaraja@gmail.com

Abstract: This paper focused on the higher-Order Markov Model whose number of states and parameters are linear with respect to the order of the model and as well as classifying it. Model for Efficient estimation methods of the parameters was developed and the model was applied to solve the application of DOTS in the treatment of tuberculosis health problem. Numerical examples with applications are given to illustrate the power of our proposed model. It was discovered that the second order Markov model was best fit base on the values of the AIC and BIC result obtained.

Keywords: Directly Observed Treatment (DOT), transition matrix, AIC, BIC, optimization, Higher Order Markov Model

1.0 Introduction

Research has shown that Tuberculosis (TB) remains a major global health problem. It causes ill health among millions of people and it is ranked as the second leading cause of death from an infectious disease worldwide after human immunodeficiency virus (HIV) [1]. Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* and can affect different body organs [2, 3]. TB can result from a rapidly progressive disease following recent infection with Mycobacterium tuberculosis or from reactivation of a past latent TB infection. TB is largely transmitted by aerosols produced from coughing in individuals with active pulmonary disease. TB surveillance and preventing further spread of the disease requires full understanding of the biological factors affecting TB, and also finding mathematical patterns explaining the mechanism of TB transmission through the community [4]. The objectives of this study is to apply the use of Directly Observed Treatment short- course (DOTS) to monitor and control the epidemic, the probability of being in a given state at a given point in time, the expected number of transitions between states and finally find the best model for DOTS applications. This strategy was expected to bring about a major change in controlling the disease. Directly Observed Treatment short-course (DOTS) has been found to be an effective means of administering anti-TB drugs, significantly reducing the rates of relapse and drug resistance as well as improving the treatment compliance rate [5, 6].

Many statistical models have been explored to control the TB epidemic. Markov processes are not suitable for modeling all disease types and answering all disease related questions due to the complexity that is involved in the modeling of some diseases [7]. Evaluation of treatment outcome is central to the assessment of effectiveness of tuberculosis [8]. Treatment outcomes for TB patients (excluding patients treated for RR-TB or MDR-TB, that is multi-drug resistance to tuberculosis drugs) are classified as successful (cure or treatment completed) or poor (default, treatment failure or death) as defined by the [9].

Statistical methods like regression techniques, time series analysis, statistical process control and Bayesian methods have been used to monitor the epidemiologic surveillance of infectious diseases [10]. [11] introduces a warning threshold for detecting the unexpected incidences of Tuberculosis (TB) using a Hidden Markov Model (HMM) and it was concluded that the warning threshold constructed based on the Periodic Autoregressive Model can be regarded as a useful alternative for HMM in detection of the weeks with unexpected incidence of TB, therefore it was suggested for monitoring TB surveillance. Treatment outcomes of patients are classified as successful (cure or treatment completed) or poor (default, treatment failure or death) as defined by [12]. Markov chain concerns about a sequence of random variables, which correspond to the states of a certain system, in such a way that the state at one time epoch depends only on the one in the previous time epoch. Higher order markov model is used to model treatment outcome of tuberculosis as DOTS application.

Higher Order Markov Model (HOMM) has been used in analysis and prediction of time series demonstrating the effectiveness of Markov chain model and it has been applied to price and sales volume for beef prediction problem by Tie Liu. Markov chains have been used to model categorical data sequences and this can be found in [13] and [14]. Markov chain model of order higher than one that involves only one parameter for each extra log variable was suggested by [15]. This was extended to qth order marginalized transition model (MTM) by [16] and [17].

[18] generalized the [15] model by allowing $Q = \{q_{ij}\}$ to vary with different lag and then developed effective method for parameter estimation. Higher-order Markov Model, each data point $X(n)$ in a categorical data sequence takes value in the set $M \equiv \{1, 2, \ldots, m\}$ and m is finite i.e. a sequence has m possible categories or states. The total number of independent parameter to be estimated in kth order Markov chain is $m^k(m-1)$.The number of independent parameters increases geometrically as the order increases, thereby

discourages people from using higher order Markov chain. See Table 1 Raftery proposed a higher order Markov chain model which involves one additional parameter for each lag. The model can be written as

$$P\left(X^{(n)} = j_0 \mid X^{(n-1)} = j_1, \dots \dots, X^{(n-k)} = j_k\right) = \sum_{i=1}^{k} \lambda_i q_{ij} \qquad 1.3$$

Where,

$$\sum_{i=1}^{k} \lambda_i = 1 \qquad 1.4$$

And $Q = [q_{ij}]$ is transition matrix with column sum equals one, such that

$$0 \le \sum_{i=1}^{k} \lambda_i q_{j_0 j_i} \le 1, \qquad j_0 j_i \in M \qquad 1.5$$

A more general higher-order Markov chain model is obtained by allowing Q to vary with different lags. Here we assume that the weight λ_i is non-negative. It should be noted that it can be written as

$$X^{(n+k+1)} = \lambda_i Q X^{(n+k+1-i)} \qquad 1.6$$

Where $QX^{(n+k+1-i)}$ is the probability distribution of the states at time$(n+k+1-i)$. Using (1.4) and Q is a transition probability matrix, we consider each entry of $X^{(n+k+1)}$ which is between 0 and 1 and also parameter λ_i is non-negative. The additional constraint should be added to guarantee that $X^{(n+k+1)}$ is the probability distribution of the states.

Raftery's model can be generalized as follows:

$$X^{(n+k+1)} = \lambda_i Q_i X^{(n+k+1-i)}$$

The total number of independent parameters in the new model is $k + km^2$. We note that if $Q_1 = Q2 = \cdots = Q_K$, then the above equation is just the Raftery's model.

In the model we assume that $X^{(n+k+1)}$ depends on $X^{(n+1)}, (i = 1,2, \dots, k)$ via the matrix Q_i and weight λ_i. One may relate Q_i to the i-step transition matrix of the process and this can be used to estimate Q_i. It is assumed that each Q_i is a non-negative stochastic matrix with column sums equal to one.

2.1 Estimation of the model Parameters

Given an observed data sequence $\{X^n\}$, where $\{X_{(n)}\}$ can be written in a vector form as:

$$\{X^{(1)}, X^{(2)}, X^{(3)}, \dots \dots \dots \dots \dots, X^T\}$$

Where T is the length of the sequence and $X_I \in$ DOM (A). DOM (A) is categorical data if it finite and unordered.

One can find the transition frequency F_{ij} in the sequence by counting the number of transitions from state *i* to state *j* in one step. Also one can construct the one step transition matrix or the sequence $\{X_{(n)}\}$ as follows:

$$F = \begin{pmatrix} F_{11} & \dots & F_{1m} \\ \vdots & \vdots & \vdots \\ Fm_1 & \dots & Fmm \end{pmatrix} \qquad 2.1$$

From F, one can get the estimates for P_{ij} as follows:

$$P = \begin{pmatrix} P11 & \cdots & P1m \\ \vdots & \vdots & \vdots \\ Pm1 & \cdots & Pmm \end{pmatrix} \qquad 2.2$$

Where,

$$p(x) = \begin{cases} \frac{F_{ij}}{\sum_1^m F_{ij}}, & if \sum_1^m F_{ij} > 0 \\ 0, & if \sum_1^m F_{ij} = 0 \end{cases}$$

And the estimators in p(x) satisfies

$$E(q_{ij}^k = q_{ij}^k E(f_{ij}^k)$$

The following proposition by [19], helps in estimating the parameters in HOMM.

Proposition 1. The matrix P has an eigenvalue equal to one and all the eigenvalues must have modulus less than or equal to one.

Proposition 2. (Perron-Frobenius theorem). Let A be a non-negative and irreducible square matrix of order m. Then

1. A has positive real eigenvalue, λ, equal to its spectral radius i.e. $\lambda = max|\lambda_k(A)|$, where $\lambda_k(A)$ denotes k^{th} eigenvalue of A

2. To λ there corresponds an eigenvector **x** its entries being real and positive, such that $Ax = \lambda x$

3. λ is a simple eigenvalue of A

using these two propositions, one can see that there exists a positive vector $x = [x_1, x_2, \dots \dots \dots \dots \dots \dots \dots \dots x_m]^T$ such that $Px = x$ if P is irreducible. The vector **x** in normalized form is called the stationary probability of vector P. Therefore, x_i is the probability that the system is in state *i*.

As a result of the proposition (1) and (2) above $Px = x$ and $Q = \sum_{i=0}^n \lambda_i P_i$, then the proposition (1) gives a sufficient condition for sequence $\{X^n\}$ to converge to stationary distribution X. As $X^n \to \bar{X}$ as n goes to infinity, then $\bar{X}$ can be estimated from sequence $\{X^{(n)}\}$. Therefore, the proportionality of occurrence of each state can then be denoted by $\hat{X}$ i.e.

$$\sum_i^k \lambda_i Q_i \hat{X} \approx \hat{X}$$

And this is one of ways to estimate parameter λ.

$\lambda = (\lambda_1, \lambda_2, \dots \dots \dots, \lambda_k)$.

One of the ways to estimate the parameter λ is to consider the minimization problem through certain vector norm, $\| . \|$, using $\| . \|_1$ which leads linear programming equation, $\| . \|_2$ which leads to quadratic programming problem and $\| . \|_\infty$ which avoids gross discrepancies with data as much as possible.

Definition of Matrix-Norm

From Wolfram Mathworld,

Let A be a square or real matrix, a matrix norm $\|A\|$ is a non-negative number associated with A with the following properties:

1. $\|A\| > 0$ *when* $A \neq 0$ *and* $\|A\| = 0$ *iff* $A = 0$

2. $\|kA\| = |k| \|A\|$ *for any scalar* k

3. $\|A + B\| \leq \|A\| + \|B\|$

4. $\|AB\| \leq \|A\| \, \|B\|$

Let $\lambda_1, \lambda_2, \dots \lambda_n$ be the eigenvalues of A, then

$$\frac{1}{\|A^{-1}\|} \leq |\lambda| \leq \|A\|$$

The matrix *p*-norm is defined for a real number $1 \leq p \leq \infty$ and a matrix A by

$$\|A\|_p = \max_{X \text{ s.t. } |X|_p = 1} |Ax|_p$$

where $|x|_p$ is a vector norm. It is tasking to compute *p*-norm for $p > 1$ because it is a non-linear optimization problem with constraints so mathematical software are used.

The maximum absolute column sum norm $\|A\|_1$ is defined as

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|.$$

The spectral-norm, $\|A\|_2$, which is the square root of the maximum eigenvalue of $A^H A$, where A^H is the conjugate transpose of A. therefore

$$\|A\|_2 = (maximum\ eigenvalue\ of A^H A)^{1/2}$$

and this is always referred to as matrix norm.

The maximum absolute row sum norm is defined by

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

$\|A\|_1, \|A\|_2, \|A\|_\infty$ Satisfies the inequality $\|A\|_2^2 \le \|A\|_1 \|A\|_\infty$ [20]

Let us consider $\| \, . \, \|_1$ which leads to solve linear programming problem

$$\min_\lambda \sum_i^k \|\lambda_i Q_i \hat{X} - \hat{X}\|$$

s.t =

$$\begin{cases} \sum_i^k \lambda_i = 1 \\ \lambda \ge 0 \quad i = 1, \ldots\ldots, n \quad \forall i \end{cases}$$

And considering $\| \, . \, \|_\infty$ norm, will lead to

$$\min_\lambda \max_i |[\textstyle\sum_i^k \lambda iQi\hat{X} - \hat{X}]|_i$$

s.t =

$$\begin{cases} \sum_i^k \lambda_i = 1 \\ \lambda \ge 0 \quad i = 1, \ldots\ldots, n \quad \forall_i \end{cases}$$

where $[.]_i$ denotes the *ith* entry of the vector and the optimization problem leads existence of stationary distribution X, while minimization problem can be formulated as a linear programming problem as follows:

$\min_\lambda w$ subject to

$$\begin{pmatrix} w \\ w \\ \vdots \\ w \end{pmatrix} \ge \hat{X} - [\hat{Q}_1\hat{X} \mid \hat{Q}_2\hat{X} \mid \ldots \mid \hat{Q}_n\hat{X}] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}$$

$$\begin{pmatrix} w \\ w \\ \vdots \\ w \end{pmatrix} \ge -\hat{X} + [\hat{Q}_1\hat{X} \mid \hat{Q}_2\hat{X} \mid \ldots \mid \hat{Q}_n\hat{X}] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix}$$

$$w \ge 0, \sum_i^k \lambda_i = 1, \ and \ \lambda_i \ge 0, \ \ \forall_i$$

Materials and Methods

The data was collected from a survey on "Appraisal of Directly Observed Treatment short- course (DOTS) and Tuberculosis Eradication in Secondary Healthcare facility in Southwest, Nigeria" at West African Post Graduate College of Pharmacist, Yaba, Lagos. The questionnaire contains 42 items which was divided into 4 subsections (A – D). Section A (made up of 9 items) consisted of questions on socio-demographic characteristics of individuals – age (years), sex, marital status, religion, education. Section B was on the knowledge of DOTS. Sections C was on application of DOTS while Section D was on impact of DOTS. This study focuses on the application of DOTS with state of the patients (success, failure). The individuals attending the Out-patients Department of the hospital for tuberculosis treatment were used for the study. Convenient sampling technique was used. Data was collected three times a week for the period of one and a halved month. R programming language was used to analyze the data using Higher Order Markov Model.

Results

This chapter presents the analysis and interpretation of the data on "Appraisal of Directly Observed Treatment short course (DOTS) strategy and Tuberculosis Eradication in a Secondary Healthcare Facility in Southwest, Nigeria" used in the study. The result of the analysis is presented using tabular presentations. Data were collected on 250 patients suffering from Tuberculosis. There are two possible states in the Markov chain, which are 1 and 2. States 1 and 2 represent success and failure respectively. Lambda: [1] = 1. The lambda is one showing adequacy of order 1. The steady state vector v satisfies the equation vP = v. That is, it is an eigenvector for the eigenvalue λ = 1. If the probability in P remain the same over a long run, it will get to a stage where the vector will be stable i.e. no change occur and it will be in equilibrium stage. At this stage, the system is said to be in steady state and the steady state vector is called stationary vector.

vP = v

The steady state probability is of the form:

1	2
0.1732523	0.8267477

Table 1: The probability transition matrix of the different states summing up to one for success and failure for order 1. The probability of moving from success to success is 0.56, the probability of moving from success to failure is 0.09, and the probability of moving from failure to success is 0.44 while the probability of moving from failure to failure is 0.91. Lambda: [1] = 1. The lambda is one showing adequacy of order 1. The Table 3 shows that the mean transition from state one to one is 34.6,.the mean transition from state one to two is 5.80, the mean transition from state two to one is 26.70, while the mean transition from state two to two is 56.5. The higher expectation from state one (success) to state two (failure) is due to relapsed in the application of DOTS process during application of treatment i.e. the conditional expectation of success given failure.

The Table 4 is the probability transition matrix of the different states summing up to one for success and failure for order 2. The probability of moving from success to success is 0.47, the probability of moving from success to failure is 0.11, and the probability of moving from failure to success is 0.53 while the probability of moving from failure to failure is 0.89. [1] 0.5 0.5 .The lambda is one showing adequacy of order 2.

Table 5 is the mean transition from state one to one is 28.8, the mean transition from state one to two is 7.0, the mean transition from state two to one is 32.9, while the mean transition from state two to two is 54.8. The higher expectation from state one (success) to state two (failure) is due to relapsed in the application of DOTS process during application of treatment i.e. the conditional expectation of success given failure.

Calculating the AIC and BIC with the following log likelihood:

$LL^{(1)} = -95.66233$

$LL^{(2)} = -62.3243$

$LL^{(3)} = -45.4162$

The Table 6 & 7 shows AIC and BICwhich indicates that the best model that appears to give an excellent fit is the second order Markov model since it has a lower AIC and BIC when compared with the first order. This is because the third order was invalid due to the inability of the lambda value to sum up to one. Hence, the best model is the second order Markov model with AIC (132.649) and BIC (134.219). The model has parameters $\lambda 1$ = 0.5 for the first lag, $\lambda 2$ = 0.5 for the second lag.

Discussion

This study focuses on the application of DOTS with state of the patients (success, failure). Directly Observed Treatment short-course (DOTS) has been found to be an effective means of administering anti-TB drugs, significantly reducing the rates of relapse and drug resistance as well as improving the treatment compliance rate [5, 6]. [11] introduces a warning threshold for detecting the unexpected incidences of Tuberculosis (TB) using a Hidden Markov Model (HMM) and it was concluded that the warning threshold constructed based on the Periodic Autoregressive Model can be regarded as a useful alternative for HMM in detection of the weeks with unexpected incidence of TB, therefore it was suggested for monitoring TB surveillance. This research uses higher order markov model on the application of DOTS with state of the patients (success, failure). This will help to the determine future condition of patients and the efficient control of Tuberculosis by concentrating on the initial conditions of TB patients and focus on other factors that can improve the condition of patients because the conditional probability of being in the current state depends on the previous state. This will also help in reducing cost and making decision on policies on DOTS.

5.0 Conclusion

The Higher Order Markov Model has helped in obtaining vital information on the observable states of patients and has served as an efficient and effective tool for classifying the patients based on their observable states. The information obtained from this model can be used by the health organization to enable them operate maximally in combating the menace and deadly effect of tuberculosis. A quick look at the results for different models reveals that with order increasing for a specific model the number of estimated parameters increases rapidly. It is not surprising, therefore, that higher order models do a comparatively good job in fitting data structures and stating the best model with a lower AIC and BIC in the second order when compared with the first order. Hence, the best fitted model is the second order Markov model with AIC (132.649) and BIC (134.219). The model has parameters λ_1 = 0.5 for the first lag, λ_2 = 0.5 for the second lag.

6.0 References

[1]. World Health Organization (2017). Global Tuberculosis Report. Available on www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf

[2] Nasehi M. and Mirhaghghani L. Guidelines for combat with TB, Disease Management Center of Health Ministry. Arjemand publication; 2010. pp. 6–44.

[3] Ignatova A., Dubiley S., Stepanshina V. and Shemyakin I. (2006) Predominance of multi-drug-resistant LAM and Beijing family strains among Mycobacterium tuberculosis

isolates recovered from prison inmates in Tula Region, Russia. *J. Med. Microbiol*. **55** (Pt 10):1413–8. doi: 10.1099/jmm.0.46575-0.

[4] Aparicio J.P. and Castillo-Chavez C. (2009) Mathematical modelling of tuberculosis epidemics. *Math. Biosci. Eng*. **6**, 209–237.

[5] Erhabor G. E., H. S. Aghanwa, M. Yusuph, R. A. Adebayo, F. A. Arogundade and A. Omidiora (2000). Factors influencing compliance in patients with tuberculosis on directly observed therapy at Ile-Ife, Nigeria. *East Afr. Med. J.* 77(5): 235-239.

[6] Erhabor, G., O. Adewole, A. Adisa and O. Olajolo. (2003) Directly observed short course therapy for tuberculosis – a preliminary report of a three-year experience in a teaching hospital. *J Natl. Med. Assoc.* **95**(11): 1082-1088.

[7] Marais, C. (2010). Markov Processes in Disease Modelling - Estimation and Implementation (M.Sc. (Department of Statistics)), University of Pretoria, Pretoria.

[8] Lisha, P.V, P.T. James and C. Ravindran (2012) Morbidity and mortality at five years after initiating category 1 treatment among patients with new sputum smear positive pulmonary. *Indian J. Tuberc.* ***59***(2): 83-91.

[9] World Health Organization (2013). Definitions and reporting framework for tuberculosis – 2013 revision (updated December 2014. http://apps.who.int/iris/bitstream/handle/10665/79199/9789241505345_eng.pdf?sequence=1 tuberculosis.

[10] Unkel S., Farrington C., Garthwaite P.H., Robertson C. and Andrews N. (2011) Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J. Roy. Statist. Soc. Ser. A.* **175**(1): 49–82.

[11] Ali R., Einollah, P. and Roohangiz, J. O. (2015) A warning threshold for monitoring tuberculosis surveillance data: an alternative to hidden Markov model. *Trop. Med. Int. Health*. **20**(7): 919-29.

[12] Federal Ministry of Health (2012) National Tuberculosis and Leprosy Control Program. Models for the Training of General Health Care Workers on Tuberculosis Control. 4th ed. Abuja: Sound Prints; 2012. pp. 8-51

[13] Ching W., Fung E. and Ng, M. (2002). A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA J. Manage. Math*. **13**, 87–199.

[14] MacDonald I.L. and Zucchini, W. (1997). Hidden Markov and other models for discrete-valued time series. (Vol. 110). CRC Press. London: Chapman & Hall.

[15] Raftery, A., and Tavare, S. (1994). Estimating and modeling repeated patterns in higher order Markov chains with mixture transition distribution model. *Appl. Stat.* **43**(1): 179-199.

[16] Heagerty, P. J. and Zeger, S. L. (2000) Marginalized multilevel models and likelihood inference. *Stat. Sci.* **15**, 1–19.

[17] Heagerty, P. J. (2002) Marginalized transition models and likelihood inference for longitudinal categorical data. Biometrics **58**, 342–351.

[18] Ching W. K., Ng M. K. and Wong, K. K. (2004) Hidden Markov Models and their Applications to Customer Relationship Management. *IMA J. Manage. Math*. **15**(I): 13-24.

[19] Ching, W., Ng, M. and Fung, E. (2008) Higher-order multivariate Markov chains and their applications. *Linear Algebra Its Appl.* **428**, 492–507.

[20] Wolfram Web Resources (2019) Matrix-Norm. Available from http://mathworld.wolfram.com/ MatrixNorm.html

Table 1: Number of Independent parameters by order in HOMC

Number of State(m)	Order(k)	Markov Chain
2	1	2
	2	4
	3	8
	4	16
3	1	6
	2	18
	3	54
	4	162
4	1	20
	2	100
	3	500
	4	2500

Table 2: The probability transition matrix of order 1

	1	2
1	0.5555556	0.09313725
2	0.4444444	0.90686275

Table 3: Mean transitions of order1

	1	2
1	34.58334	5.797794
2	27.66666	56.452206

Table 4: The probability transition matrix of order 2

	1	2
1	0.4666667	0.1133005
2	0.5333333	0.8866995

Table 5: Mean transitions of order 2

	1	2
1	28.81667	6.996306
2	32.93333	54.753694

Table 6: Model selection with AIC

Order	Log likelihood	k	AIC
1	-95.6623	2	195.3247
2	-62.3243	4	132.6486
3	-45.4162	8	106.8324

Table 7: Model selection with BIC

Order	Log likelihood	2Loglikelihood	k	klog(n)	BIC
1	-95.6623	191.3247	2	4.792399	196.1170587
2	-62.3243	124.6486	4	9.570788	134.2193878
3	-45.4162	90.8324	8	19.14158	109.9739756