
A. Bayaga

Abstract—This research provides a technical account of estimating Transition Probability using Time-homogeneous Markov Jump Process applying by South African HIV/AIDS data from the Statistics South Africa. It employs Maximum Likelihood Estimator (MLE) model to explore the possible influence of Transition Probability of mortality cases in which case the data was based on actual Statistics South Africa. This was conducted via an integrated demographic and epidemiological model of South African HIV/AIDS epidemic. The model was fitted to age-specific HIV prevalence data and recorded death data using MLE model. Though the previous model results suggest HIV in South Africa has declined and AIDS mortality rates have declined since 2002 – 2013, in contrast, our results differ evidently with the generally accepted HIV models (Spectrum/EPP and ASSA2008) in South Africa. However, there is the need for supplementary research to be conducted to enhance the demographic parameters in the model and as well apply it to each of the nine (9) provinces of South Africa.


I. INTRODUCTION

THE current study addresses among others: The Theoretical Basis of Time-Homogeneous Markov Jump Process (MJP) and Comparative Models. The study evaluates the impact of HIV/AIDS in South Africa (2002-2013), and its application to current study Model which was modelled via the MLE approach; this was also compared with the application of other models. A discussion as well as interpretations and summary based on the study’s motivation are provided.

II. MOTIVATION FOR STUDY

Considering the understanding of MJP and MLE and mainly motivated by financial applications, a stochastic analysis via Itô-formula approach developed semimartingales. Additionally, there exists a coupling approach for diffusion processes and stochastic volatility models which was developed by [1]. The work of [2] addresses approximation method which gives some comparison results for Lévy processes and processes with independent increments. Several examples and applications to - stable processes are discussed.

However, for further understanding of scope, properties and applications on stochastic orders, readers may refer to [3]. To add to the body of knowledge of understanding of the scope, properties and applications on stochastic, the current study looks at Markov processes. The author establishes based on verifiable criteria which allow the results to be applied to general class of model of MLE. Using the general frame of Markov processes theory author are able to apply time-homogeneous Markov processes and MLE to given data set.

III. MARKOV CHAIN MODELLING ANALYSIS OF HIV/AIDS PROGRESSION-RATES OF SICKNESS OR PREVALENCE NUMBER OF PEOPLE LIVING WITH HIV

The current research is based on the release of the cohort-component methodology estimate of the 2013 mid-year population of South Africa; which covers estimates of residents of South Africa at the 2013 mid-year. The Midyear population of South Africa in 2013 by Statistics South Africa [4] is estimated 52, 98 million. Fifty-one per cent (approximately 27, 16 million) of the population is female. The data also suggest that Life expectancy at birth for 2013 is estimated 57, 7 years for males and 61, 4 years for females. Infant mortality rate for 2013 is estimated at 41, 7 per 1 000 live births [5]-[7]. Estimated overall HIV prevalence rate is approximately 10%. The total number of people living with HIV is estimated at approximately 5, 26 million in 2013. For adults aged 15 - 49 years, an estimated 15, 9% of the population is HIV positive [8], [9]. It was estimated that the Median time from HIV infection to death in line with the United Nations Programme on HIV/AIDS (UNAIDS) Reference Group recommendation was 10,5 years for men and 11,5 years for women. While ratio of new infections for Adult HIV incidence is disaggregated into female and male incidence by specifying the ratio of new female infections to new male infections. The report assumed a ratio of female to male prevalence for those aged 15-49 of 1, 5 by 2013.

In terms of HIV prevalence (widespread), Table I (Appendix) shows the prevalence estimates and the total number of people living with HIV from 2002 to 2013 [4]. The total number of persons living with HIV in South Africa increased from an estimated 4 million in 2002 to 5, 26 million by 2013. For 2013 an estimated 10% of the total population is HIV positive [4], [10] estimated the HIV prevalence for 2008 at 10, 9%. Approximately seventeen percent of South African women in their reproductive ages are HIV positive [11]. Table II (Appendix) shows assumptions about fertility and mortality
levels, 2002–2013. The HIV and AIDS estimates for 2013 notes that Number of people living with HIV is approximately 6,300,000 [6,000,000 - 6,500,000]; Adults aged 15 to 49 prevalence rate is 19.1% [18.1% - 19.9%]. Adults aged 15 and up living with HIV is 5,900,000 [5,700,000 - 6,200,000]. Women aged 15 and up living with HIV is 3,500,000 [3,300,000 - 3,700,000], while Children aged 0 to 14 living with HIV is 360,000 [320,000 - 390,000]. The Deaths due to AIDS is 200,000 [170,000 - 220,000] and Orphans due to AIDS aged 0 to 17 is 2,400,000 [810,000 - 2,600,000]. There have been number of studies internationally which have addressed Markov chain modelling analysis of HIV/AIDS [12]-[19]. The work of [1] used a Markov chain analysis to model the progression of the disease among vulnerable people, infective people and AIDS cases for two races separately [3]. Their study based on the Markov model, predicted that the number of African American people living with AIDS diagnosis and HIV infection and dead due to HIV/AIDS will be 662.2, 1225.3 and 62.9 in 2015 and 794.9, 1566.5 and 79.2 in 2030, respectively. As one of stochastic processes, Markov process is a system which changes in random manner between different states at regular or irregular intervals. It was applicable in this study as noted by the authors since Markov chain modelling was used to predict and estimate random or uncertain events associated with specific probabilities of occurrence.

Using MATLAB, the authors shown that, a Monotonic increases in the rates of living with AIDS diagnosis and HIV infection are forecasted at a rate of 662.2 of living with AIDS diagnosis and the rate of 1225.3 of HIV infection by the year of 2015. They also indicate that the number of HIV/AIDS diagnoses and HIV/AIDS related deaths each year is quite consistent, while the number of people living with HIV/AIDS among the infected population is increasing. It is important to note though that, various technical challenges still present in their model. This is because, the Markov model was modelled based on the statistics in the HIV surveillance report only from 2006 to 2009 due to unavailability of previous years’ data and particularly in United States of America [20] suggesting that a more accurate historical data was not available to model estimation and prediction. Additionally they modelled a HIV/AIDS progression with a Markov homogeneous model whose transition probabilities are stationary and constant, changes in transition behaviour and characteristics could not be represented with this model [21] which is the focus of the current study. Though earlier studies modelled disease progress and transmission dynamics using Markov models, few have focused on Estimating forces of transition in a time-homogeneous Markov Jump Process. The prediction of computing Transition Probability using Time-Homogeneous MJP will help in planning and calibrating adequate surveillance systems, as well as in allocating public health resources and in targeting intervention and treatment plans.

A. Analysis and Application of Current Study Model: The Maximum Likelihood Estimator Model and Analysis

As shown earlier in this section, the two-state model can be extended to any number of states, with arbitrary transitions between them, including increments and repeated transitions [1], [22]. Consider again the illness-death model, which has three states: health (H), sick (S) and dead (D):

Using lower case for the observed samples, it is easily shown that the likelihood for the four parameters \( \mu, \sigma, \rho \) given the data is proportional to:

\[
L(\mu, \sigma, \rho) = e^{-(\mu+\sigma)t} e^{-(\theta+\rho)t} \mu \sigma \rho^{t-1} \tag{1}
\]
The likelihood function $L(\mu, \theta, \sigma, \rho)$ for the $i$th life reflects:

- The probability of the life remaining in the healthy state for total time $V_i$ and in the sick state for time $w_i$ given the factors $e^{-(\mu + \sigma)V_i}$ and $e^{-(\theta + \rho)w_i}$ respectively,
- The probability of the life making the relevant number of transitions between states giving the factors $\mu^{d_{ij}}u^{a_{i1}} \rho^{v_{i2}}$.

The likelihood factorises into functions of each parameter of the form $e^{-\mu V_i}$. i.e.

$$L(\mu, \theta, \sigma, \rho) = e^{-(\mu + \sigma)V_i} X e^{-(\theta + \rho)w_i} X e^{-\sigma V_i} X e^{-\rho w_i}$$

Therefore the log-likelihood is:

$$\log L = -(u + \sigma)V_i - (v + p)w_i + d \log u + u \log v + s \log \sigma + r \log \rho$$

Differentiating (3) with respect to each of the four parameters gives:

$$\frac{\partial \log L}{\partial \nu} = -w + \frac{u}{v}$$
$$\frac{\partial \log L}{\partial \mu} = -v + \frac{u}{d}$$
$$\frac{\partial \log L}{\partial \sigma} = -v + \frac{s}{\sigma}$$
$$\frac{\partial \log L}{\partial \rho} = -w + \frac{r}{\rho}$$

Since each of these derivatives equal to 0 and solving the resulting equations, we get

$$u = \frac{d}{v}, \quad v = \frac{u}{w}, \quad \sigma = \frac{s}{v}, \quad p = \frac{r}{w}$$

where $n_{ij}$ is the number of transitions from state I to state j and $t_i$ is the total waiting time or total holding time in state $i$.

During a large study into rates of sickness using the Stats [9]:

- 185 prevalence cases and 14 incidence cases
- 7080869 total number of deaths and 3050631 total number of AIDS deaths
- For the whole group, the periods prevalence and incidence totalled 11 years respectively. Estimate the forces of transition of prevalence and incidence becomes; Solution

$$t_H = 11, \quad t_S = 11, \quad n_{HS} = 185, \quad n_{SH} = 14, \quad n_{HD} = 7080869, \quad n_{SD} = 3050631$$

The MLEs of the transition intensities are therefore:

$$\hat{\sigma} = \frac{S}{V} = \frac{n_{HS}}{t_H} = 185 / 11 = 16.8$$
$$\hat{\rho} = \frac{R}{W} = \frac{n_{SH}}{t_S} = 14 / 11 = 1.3$$

Thus the forces of transition are:

$$\hat{\sigma} = \frac{S}{V} = \frac{n_{HS}}{t_H} = 185 / 11 = 16.8$$
$$\hat{\rho} = \frac{R}{W} = \frac{n_{SH}}{t_S} = 14 / 11 = 1.3$$

IV. DISCUSSION AND INTERPRETATIONS

The current study's results which are estimate; $\hat{\sigma} = \frac{S}{V} = \frac{n_{HS}}{t_H} = 185 / 11 = 16.8$; $\hat{\rho} = \frac{R}{W} = \frac{n_{SH}}{t_S} = 14 / 11 = 1.3$ demonstrate forces of transition between the prevalence state and incidence. Earlier studies have shown that younger women are predominantly and biologically inclined to HIV compared with older ones. Consistent with one such study is that of [23] where the researchers controlled for a large number of known risk factors for HIV transmission and found that each 10-year increase in age correlated with 60%
reduction in the rate of HIV acquisition [23]. But a contrasting results from an East Africa country (Kenya) suggests that HIV transmission probabilities per act of sex decreased by 3.5% per year of age, equivalent to a 30% reduction in HIV risk [22]-[25]. The various reasons ascribed to high risk of HIV acquisition are (1) a high prevalence of cervical ectopy in adolescence and early adulthood (2) it could be other sources of variation in HIV susceptibility.

It is also suggested that the variation in susceptibility could be due to variation in genetic factors or levels of immune activation, which have been linked to the risk of HIV acquisition [21] - [24].

A. Assumptions and Guiding Principles

In this current study, the model assumed the following;

- As sample increases the distribution of the MLE (its probability density or frequency function) takes the shape of very narrow spike centred on the true value of the parameter. Thus the probability of the MLE differs from when the true value of the parameter approaches zero as the sample size increases. Thus MLE is a consistent estimator of the true parameter.

- Bias disappears as the sample size gets large, this suggest that MLE need not be unbiased. Indicating that the MLE is asymptotically unbiased.

- The distribution shape of the MLE is normally distributed as the sample size gets large, thus MLE is asymptotically normal.

The estimators are asymptotically independent; the same argument as in the two-state model shows that

- The vector (μ̂, β̂, σ̂, ρ̂) has an asymptotically multivariate Normal distribution;

- Each component has a marginal asymptotic distribution of the same form as before;

μ̂ ~ Normal (μ, V̂)

Asymptotically the components are uncorrelated and so independent (being normal).

V. CONCLUDE REMARKS

The current study provided algorithms for computing Transition Probability using time-homogeneous Markov Jump Process via MLE. Noting the significance of the practical necessity of similar problems for Markov process, it is anticipated that our methods will prove useful additions to the methods available for analysing continuous-time chains. On the basis of assumptions made and the estimated levels of HIV/AIDS prevalence in different types of relationship, it was possible to demonstrate forces of transition between the prevalence state and incidence as; σ̂ = 16.8 and β̂ = 1.3.

APPENDIX

### TABLE I
HIV PREVALENCE ESTIMATES AND THE NUMBER OF PEOPLE LIVING WITH HIV, 2002-2013

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>15.9</td>
<td>15.1</td>
<td>13.6</td>
<td>8.7</td>
<td>4.00</td>
<td>1.26</td>
<td>4.00</td>
<td>1.26</td>
<td>4.00</td>
<td>1.26</td>
<td>4.00</td>
</tr>
<tr>
<td>2003</td>
<td>16.0</td>
<td>15.1</td>
<td>12.8</td>
<td>8.9</td>
<td>4.10</td>
<td>1.25</td>
<td>4.10</td>
<td>1.25</td>
<td>4.10</td>
<td>1.25</td>
<td>4.10</td>
</tr>
<tr>
<td>2004</td>
<td>16.1</td>
<td>15.1</td>
<td>12.0</td>
<td>8.9</td>
<td>4.18</td>
<td>1.28</td>
<td>4.18</td>
<td>1.28</td>
<td>4.18</td>
<td>1.28</td>
<td>4.18</td>
</tr>
<tr>
<td>2005</td>
<td>16.2</td>
<td>15.1</td>
<td>11.4</td>
<td>9.0</td>
<td>4.25</td>
<td>1.32</td>
<td>4.25</td>
<td>1.32</td>
<td>4.25</td>
<td>1.32</td>
<td>4.25</td>
</tr>
<tr>
<td>2006</td>
<td>16.4</td>
<td>15.2</td>
<td>10.9</td>
<td>9.1</td>
<td>4.34</td>
<td>1.30</td>
<td>4.34</td>
<td>1.30</td>
<td>4.34</td>
<td>1.30</td>
<td>4.34</td>
</tr>
<tr>
<td>2007</td>
<td>16.5</td>
<td>15.3</td>
<td>10.5</td>
<td>9.2</td>
<td>4.46</td>
<td>1.21</td>
<td>4.46</td>
<td>1.21</td>
<td>4.46</td>
<td>1.21</td>
<td>4.46</td>
</tr>
<tr>
<td>2008</td>
<td>16.7</td>
<td>15.4</td>
<td>10.1</td>
<td>9.3</td>
<td>4.59</td>
<td>1.12</td>
<td>4.59</td>
<td>1.12</td>
<td>4.59</td>
<td>1.12</td>
<td>4.59</td>
</tr>
<tr>
<td>2009</td>
<td>16.9</td>
<td>15.5</td>
<td>9.7</td>
<td>9.5</td>
<td>4.74</td>
<td>1.03</td>
<td>4.74</td>
<td>1.03</td>
<td>4.74</td>
<td>1.03</td>
<td>4.74</td>
</tr>
<tr>
<td>2010</td>
<td>17.1</td>
<td>15.6</td>
<td>9.3</td>
<td>9.6</td>
<td>4.88</td>
<td>0.98</td>
<td>4.88</td>
<td>0.98</td>
<td>4.88</td>
<td>0.98</td>
<td>4.88</td>
</tr>
<tr>
<td>2011</td>
<td>17.2</td>
<td>15.7</td>
<td>9.0</td>
<td>9.8</td>
<td>5.01</td>
<td>0.95</td>
<td>5.01</td>
<td>0.95</td>
<td>5.01</td>
<td>0.95</td>
<td>5.01</td>
</tr>
<tr>
<td>2012</td>
<td>17.3</td>
<td>15.8</td>
<td>8.7</td>
<td>9.9</td>
<td>5.13</td>
<td>0.87</td>
<td>5.13</td>
<td>0.87</td>
<td>5.13</td>
<td>0.87</td>
<td>5.13</td>
</tr>
<tr>
<td>2013</td>
<td>17.4</td>
<td>15.9</td>
<td>8.5</td>
<td>10.0</td>
<td>5.26</td>
<td>0.85</td>
<td>5.26</td>
<td>0.85</td>
<td>5.26</td>
<td>0.85</td>
<td>5.26</td>
</tr>
</tbody>
</table>

### TABLE II
ASSUMPTIONS ABOUT FERTILITY AND MORTALITY LEVELS, 2002-2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Crude birth</th>
<th>Total fertility rate (TFR)</th>
<th>Life expectancy at birth</th>
<th>Infant mortality rate (IMR)</th>
<th>Under 5 mortality</th>
<th>Crude death rate</th>
<th>Rate of natural increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Total</td>
<td>Male</td>
<td>Female</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>24.5</td>
<td>2.71</td>
<td>50.0</td>
<td>55.2</td>
<td>52.7</td>
<td>63.5</td>
<td>92.9</td>
</tr>
<tr>
<td>2003</td>
<td>24.2</td>
<td>2.68</td>
<td>49.5</td>
<td>54.4</td>
<td>52.1</td>
<td>62.6</td>
<td>91.9</td>
</tr>
<tr>
<td>2004</td>
<td>23.6</td>
<td>2.61</td>
<td>49.3</td>
<td>53.9</td>
<td>51.7</td>
<td>60.1</td>
<td>89.3</td>
</tr>
<tr>
<td>2005</td>
<td>23.1</td>
<td>2.56</td>
<td>49.4</td>
<td>53.6</td>
<td>51.6</td>
<td>58.0</td>
<td>85.4</td>
</tr>
<tr>
<td>2006</td>
<td>22.8</td>
<td>2.53</td>
<td>50.2</td>
<td>54.6</td>
<td>52.5</td>
<td>55.6</td>
<td>80.9</td>
</tr>
<tr>
<td>2007</td>
<td>22.6</td>
<td>2.53</td>
<td>51.7</td>
<td>56.1</td>
<td>54.0</td>
<td>53.6</td>
<td>76.7</td>
</tr>
<tr>
<td>2008</td>
<td>22.5</td>
<td>2.52</td>
<td>53.3</td>
<td>57.6</td>
<td>55.5</td>
<td>50.8</td>
<td>72.3</td>
</tr>
<tr>
<td>2009</td>
<td>22.3</td>
<td>2.51</td>
<td>54.6</td>
<td>58.8</td>
<td>56.8</td>
<td>49.1</td>
<td>68.5</td>
</tr>
<tr>
<td>2010</td>
<td>22.2</td>
<td>2.50</td>
<td>55.5</td>
<td>59.5</td>
<td>57.6</td>
<td>47.1</td>
<td>65.2</td>
</tr>
<tr>
<td>2011</td>
<td>21.6</td>
<td>2.44</td>
<td>56.1</td>
<td>60.0</td>
<td>58.1</td>
<td>45.1</td>
<td>62.1</td>
</tr>
<tr>
<td>2012</td>
<td>21.0</td>
<td>2.39</td>
<td>56.8</td>
<td>60.5</td>
<td>58.7</td>
<td>43.5</td>
<td>59.5</td>
</tr>
<tr>
<td>2013</td>
<td>20.5</td>
<td>2.34</td>
<td>57.7</td>
<td>61.4</td>
<td>59.6</td>
<td>41.7</td>
<td>56.6</td>
</tr>
</tbody>
</table>
REFERENCES


