IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

Maximum Likelihood Estimation Of Hidden Markov Model: Application To Markers of Infectious Disease Progression

R. U. Okafor [1*], E. B. Nkemnole [1]

1. Department of Statistics, Faculty of Science, University of Lagos, Nigeria.
* Corresponding author: ukagbans@gmail.com*, enkemnole@unilag.edu.ng

Article Info
Received: 13 May 2024 Revised: 03 June 2024
Accepted: 07 August 2024 Available online: 30 August 2024

Abstract

Hidden Markov models describe the probabilistic dependence between the latent state and the observed variable of a system. It is a stochastic model with a sequence of observable events where the underlying process that generates these events is unobserved. Hidden Markov models could be used to analyze the history of various diseases, including infectious disease progression. These models in life experiments describe the disease evolution, estimate the transition rates, and evaluate the therapy effects on progression. In many cases, the states characterize the markers of the diseases. Parameter estimation is indispensable when using the hidden Markov model to model any dataset. In this work, the hidden Markov model is used to analyze the dataset of HIV-infected patients undergoing antiretroviral treatments at a university teaching hospital in Nigeria with different compliance levels. The model's parameters were estimated using the maximum likelihood estimation (MLE) method. The variables are the CD4 counts and viral load results, often clinically characterized as markers of infectious diseases. The transition probabilities provide insights into the stability and dynamics of the hidden states, which is crucial for understanding the underlying processes modeled by the HMM. The results indicate that stage 1 has a high probability of staying on that stage with ART treatment, whereas stage 2 has a higher chance of sliding to stage 3. The results also indicate a high chance of remaining in stage 3 once a patient is diagnosed with AIDS. The results show that keeping the CD4 count up with antiretroviral treatments holds off symptoms and complications of the Human Immunodeficiency Virus (HIV) and helps patients live longer. These highlight the importance of maintaining an undetectable viral load with ART to ensure a healthy life for HIV-infected individuals. Consequently, patient compliance in completing the treatment regimes is optimal.

Keywords: Infectious Diseases; Hidden Markov; Human Immunodeficiency Virus; Maximum Likelihood Estimation; CD4 Count; Viral Load.
MSC2010: 13P25, 62H12.

1 Introduction

The hidden Markov model is assumed to be a Markov process with hidden states, where the probability of moving from one state to another depends only on the current state and not on

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

any previous states. In a HMM, the system comprises a set of states and a set of observable emissions. The states switch by transition probabilities, in which the likelihood at which each state emits an observable symbol relies exclusively on that state. HMMs have many applications in speech recognition, bioinformatics, natural language processing, and other fields where hidden structures depend on observable data for prediction. For a general introduction to the model, see [1] and [2]. Researchers in different disciplines have utilized Markov dynamic models for their works in the past. Some examples in public health include modeling of infectious diseases [3], modeling of immunological and virological states in HIV-1 infected patients [4], and dynamic logistic state space prediction model for clinical decision-making [5]. Other applications include speech recognition [6], neuroimaging [7], computational engineering [8], DNA composition [9], Medical Prognosis [10], and Modeling of CD4 cell decline in HIV patients [11]. [12] presented an approach to predict the stock market price trend based on high-order HMM different from the commonly used first-order HMM. By introducing a dimension reduction method to transform the high-dimensional state vector of high-order HMM into a single one, CSI 300 and S&P 500 index illustrate that high-order HMM has preferable ability to identify market price trend better than the first-order model.

The model parameters are mostly estimated using the Expectation Maximization (EM) algorithms. The algorithm uses an iterative process for maximum likelihood estimation (MLE) in statistical models with latent variables. [13] demonstrated that the estimator generated by maximizing likelihood would be consistent and asymptotically normally distributed under the regularity assumption. [14] suggested fitting a finite mixture model for obtaining parameter estimates suitable for the initial iterative calculations required for evaluating the exact maximum likelihood estimates of the Hidden Markov model. [15] proposes a more precise characterization of the estimation error processes using the L-mixing process approach. [16] used a hidden Markov model to explore the dependencies among four main cryptocurrency log-returns using maximum-likelihood estimation for the model parameters. A heterogeneous correlation structure is estimated, and evidence of structural medium term trend in the correlation of Bitcoin with the other cryptocurrencies was detected.

The immune system, known as the body's natural defense system against pathogens and infections, consists of some special cells known as the CD4 cells. These are white blood cells, called T cells, moving throughout the body to find and destroy bacteria, viruses, and other invading germs. The CD4 count is a test that measures how many CD4 cells are in one's blood. The test results help doctors check the extent of damage to the immune system and the likely consequences if the patient fails to start antiretroviral treatment (ART). Antiretroviral therapy is the treatment for the human immunodeficiency virus (HIV) infection using a combination of antiretroviral (ARV) drugs. ARV drugs do not kill HIV but prevent the virus from multiplying and destroying infection-fighting CD4 (soldiers of the body) cells. Doctors recommend that everyone with HIV take ART regardless of whether the CD4 count is high or low. Keeping the CD4 count up with ART holds off symptoms and complications of HIV and helps patients live longer. Studies have found that people with HIV who stick to their treatments can live as long as people who do not have HIV. Viral load gives an idea of how much of the HIV is in the body. The test measures the number of HIV copies in a milliliter of blood. See [17]. HIV viral load predicts how fast the disease will progress, while other tests, like the CD4 count, indicate how much damage the virus has already caused. Keeping viral load low will keep the immune system healthy, make complications of HIV less likely, and help to live longer. An undetectable viral load will also make it less likely to transmit HIV to others. HIV damages the immune system by targeting CD4 cells. The consequences are fewer and fewer HIV-free, working CD4 cells. HIV destroys entire families of CD4 cells, and the germs these cells fight have easy access to the body. The resulting illnesses are referred to as opportunistic infections (OIs) because they take advantage of the body's lack of defense. See [18].

Dynamic state models like the hidden Markov models, have been used to model chronic, viral, or infectious diseases. Frequently, to monitor the health status and disease progression of HIV-infected patients, CD4 count is used as a marker. In particular, Markov process models of HIV based on the discretization of values of CD4 cells play an essential part in AIDS modeling [19, 20]. This class of models characterizes the course of HIV progression in terms of transition rates between a certain

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

number of states, which represent various stages of the evolution of HIV infection. In this work, the hidden Markov model is utilized to analyze the effects of antiretroviral therapy on the markers of infectious disease progression dynamics. By characterizing the disease progression through three stages, the analysis shows ART treatment's effect on the disease progression of the three stages of HIV infection, which emphasizes the importance of continual administration of ART treatments to curb the disease progression.

2 Materials and Methods

Hidden Markov models are generative models in which the joint distribution of observations and hidden states, or equivalently both the prior distribution of hidden states (transition probabilities) and conditional distribution of observations given states (the emission probabilities) is modeled. Hidden Markov models are characterized by three fundamental problems in real-life application: the likelihood problem; which is the probability of an observed sequence, the decoding problem; which is the most likely series of states to generate an observed sequence and the learning problem; which is how to learn the values for the HMM parameters A and B given some data.

2.1 Hidden Markov Model

A hidden Markov model (HMM) is a stochastic model with a sequence of observable events where the underlying process that generates these events is not directly known. The Markov chain computes the probability of the visible events, but in many cases, the events are hidden and not observed directly. A hidden Markov model allows for both observed and unobserved events that are causal factors in the probabilistic model. As with the Markov chain, the HMM assumes that predicting the future in the sequence requires only the current state. The HMM consists of a set of mathematical representations that describe the probability distributions of the underlying process and the observed events:

(i) Number of hidden states: The underlying process that generates the observed events is modelled using a set of hidden states, which are represented as a discrete-time Markov chain. The set of hidden states is denoted by $Q = \{q_1, q_2, ..., q_N\}$, where N is the number of possible hidden states.

(ii) Observation symbols: At each time step, the HMM emits an observation symbol, which is associated with the current hidden state. The observation symbols is denoted by $O = o_1, o_2, ...o_T$ as a sequence of T observations, each one drawn from a vocabulary $V = \{v_1, v_2, ..., v_V\}$.

(iii) Transition probabilities: The probability of transitioning from one hidden state to another is represented by transition probabilities. The transition probability matrix A is defined as: $A = a_{11}, a_{12}, ...a_{ij}, ..., a_{NN}$, each a_{ij} representing the probability of moving from state i to state j, such that $\sum_{j=1}^{N} a_{ij} = 1 \; \forall i$.

(iv) Emission probabilities: This is a sequence of observation likelihoods, each expressing the probability of an observation o_t(drawn from a vocabulary $V = \{v_1, v_2, ..., v_V\}$) being generated from a state q_i. Emission probabilities represent the probability of emitting an observation symbol given a hidden state. The emission probability matrix B is defined as: $B = b_i(o_t)$, each expressing the probability of an observation $\{o_t\}$ being generated from a state, i.

(v) Initial state probability: The probability of starting in a particular hidden state is represented by a set of initial state probabilities. The initial state probability vector π is defined as: $\pi = \pi_1, \pi_2, ..., \pi_N$. $\{\pi_i\}$ is the probability that the Markov chain will start in state i, and some states j may have $\pi_i = 0$, meaning that they cannot be initial states. Also $\sum_{i=1}^{N} \pi_i = 1$.

A first-order Hidden Markov model instantiates two assumptions:

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

(i) As with a first-order Markov chain, the probability of a particular state depends only on the previous state:
(Markov Assumption):

$$P(q_i \mid q_1, q_2, ..., q_{i-1}) = P(q_i \mid q_{i-1}). \tag{2.1}$$

(ii) The probability of an output observation $\{o_i\}$ depends only on the state that produced the observation $\{q_i\}$ and not on any other states or any other observations:

(Output Independence):

$$P(o_i \mid q_1, ..., q_i, q_N, o_1, ...o_i, ...o_T) = P(o_i \mid q_i). \tag{2.2}$$

The HMM parameters are described using the transition probability matrix A, the emission probability matrix B, and the initial state distribution vector π.

2.1.1 Transition probability

The state transition probability $A = [a_{ij}]$, is the probabilities of transition to the next state which are conditional upon the current state.The state transition probability of a 3-state Markov process in matrix form can be expressed thus:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

2.1.2 The initial state probability

The state probability vector $\{\pi_t\}$ is the unconditional probability of being in a certain state at time t. For the 3-regime Markov random variable q_t, the state probability distribution π_t, is given by the following 3-element vector :

$$\pi_t = P(q_t = 1), P(q_t = 2), P(q_t = 3).$$

2.1.3 Emission Probability Matrix:

The probability of emitting or generating an observation symbol v_i given a hidden state at the corresponding time is represented by a set of emission probabilities. The emission probabilities $p(y_t|q_t)$ specify the probability distributions for the data, y_t, given that the model is in state q_t at time t. This distribution can be chosen arbitrarily (e.g., a normal distribution), and will have its own parameters to estimate.

2.2 Modeling HMM: The Baum-Welch Algorithm

Modeling the parameters of HMMs involves estimating the state transition probabilities A and the output emission probabilities B that make an observed sequence most likely. However, since distributions of HMM do not have a closed-form solution, the Expectation Maximization (EM) algorithm is utilized. The algorithm uses an iterative process for maximum likelihood estimation in models with hidden variables. See [21].

It is generally cumbersome to obtain the MLE for hidden Markov models. The algorithm for training HMMs is the forward-backward, or Baum-Welch algorithm, which is a case of the Expectation-Maximization algorithm. EM is an iterative algorithm that computes an initial estimate for the probabilities, then uses those estimates to determine a better estimate, and so on, iteratively improving the probabilities that it learns. The algorithm trains both the transition and the emission probabilities of the HMM.

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

For the general idea of the Baum-Welch algorithm, let's describe the forward and backward algorithms. See [22]. The forward algorithm is a dynamic programming algorithm that computes the observation probability by summing the probability of all possible hidden state paths that could generate the observation sequence.

To deduce the likelihood of moving from one hidden state to another, let's denote the forward probability as α. Each cell of the forward algorithm $\alpha_t(j)$ represents the probability of being in state j after seeing the first t observations, given the hidden Markov model denoted as λ. The value of each cell $\alpha_t(j)$ is computed by summing over the probabilities of every path that could lead to this cell. The equation for this probability is:

$$\alpha_t(j) = P(o_1, o_2, ...o_t, q_t = j \mid \lambda), \tag{2.3}$$

where $q_t = j$ is the t^{th} state in the sequence of states is state j. For a given state q_t at time t, the value $\alpha_t(j)$, which is the forward probability is computed as:

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(O_t) \tag{2.4}$$

where $\alpha_{t-1}(i)$ is the previous forward path probability from the previous time step a_{ij} is the transition probability from previous state q_i to current state q_j and $b_j(O_t)$ is the state observation likelihood of the observation symbol o_t given the current state j

The forward algorithm can be computed iteratively as follows:

1. Initialization:
$$\alpha_1(j) = \pi_j b_j(O_1), 1 \leq j \leq N. \tag{2.5}$$
2. Recursion:
$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(O_t); \ 1 \leq j \leq N, 1 \leq t \leq T. \tag{2.6}$$
3. Termination
$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{2.7}$$

The backward probability denoted as β is the probability of seeing the observations from time $(t+1)$ to the end, given that we are in state i at time t and given the model λ. This can be expressed as:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, ..., o_T | q_t = i, \lambda). \tag{2.8}$$

The backward probability can be computed iteratively in a similar manner to the forward algorithm as:

1. Initialization:
$$\beta_T(i) = 1, \quad 1 \leq j \leq N. \tag{2.9}$$
2. Recursion:
$$\beta_t(i) = \sum_{i=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad 1 \leq i, j \leq N, 1 \leq t < T. \tag{2.10}$$
3. Termination
$$P(O|\lambda) = \sum_{i=1}^{N} \pi_j(j) b_j(o_1) \beta_1(j). \tag{2.11}$$

The forward and backward probabilities compute the transition probability a_{ij} and observation probability $b_i(o_t)$ from an observation sequence O and and a vocabulary of potential hidden states Q, though the actual path taken through the model is hidden.

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

2.2.1 Transition probabilities

To estimate the transition probability $\{\hat{a}_{ij}\}$, let's consider a variant of maximum likelihood estimation method:

$$\hat{a}_{ij} = \frac{\text{expected number of transitions from state } i \text{ to } j}{\text{expected number of transitions from state } i} \tag{2.12}$$

To compute the expected number of transitions of being in state i at time t and state j at time $t+1$, which is the the joint probability of being in state i at time t and state j at time $t+1$.

Let ξ_t denote the probability of being in state i at time t and state j at time $t+1$, given the observation sequence and the HMM as:

$$\xi_t(i,j) = P(q_t = i, q_{t+1} = j|O, \lambda) \tag{2.13}$$

where $P(q_t = i, q_{t+1} = j|O, \lambda)$ is the probabilities for the model estimation.
Let φ_t denote the joint probability of being in state i at time t and state j at time $t+1$ which is the marginal probability of the observations given the model. See [23]. This is stated as:

$$\varphi_t(i,j) = P(q_t = i, q_{t+1} = j,\ O|\lambda). \tag{2.14}$$

Using the law of conditionally independent; $P\{X = a_i \ \& \ Y = b_j|Z = c_k\} = P\{X = a_i|Z = c_k\} \cdot P\{Y = b_j|Z = c_k\}$.
Hence

$$\varphi_t(i,j) = \alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j), \tag{2.15}$$

where α is the forward probability, β is the backward probabilities, $\{a_{ij}\}$ is the transition probability and $b_j(O_{t+1})$ is the observation probability.

To compute ξ_t from φ_t using the conditional probability density function defined as:

$$P\{X|Y,Z\} = \frac{P\{X,Y,Z\}}{P\{Y|Z\}}. \tag{2.16}$$

Hence

$$\xi_t(i,j) = P(q_t = i, q_{t+1} = j|O, \lambda) = \frac{\varphi_t}{P(O|\lambda)}. \tag{2.17}$$

But the probability of the observation given the model $P(O|\lambda)$ is the forward or backward probability. This is given as:

$$P(O|\lambda) = \sum_{j=1}^{N} \alpha_t(j)\beta_t(j). \tag{2.18}$$

The expected number of transitions from state i to state j is the sum over all t of ξ.

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \tag{2.19}$$

For the estimate of $\hat{a}_{ij}$, the total expected number of transitions from state i is also computed. To find the total expected number of transitions from state i, we sum over all transitions out of state i.

Thus, the estimate of transition probability $\{a_{ij}\}$ is given as:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)} \tag{2.20}$$

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

2.2.2 Observation or Emission probabilities

If $\hat{b}_j(v_k)$ is the probability of a given symbol v_k from the observation vocabulary V , given a state j, then;

$$\hat{b}_j(v_k) = \frac{\text{expected number of times in state j and observing symbol } v_k}{\text{expeected number of times in state j}} \tag{2.21}$$

Let's denote $\gamma_t(j)$ as the probability of being in state j at time t, given the observations and the model, then $\gamma_t(j)$ is expressed as;

$$\gamma_t(j) = P(q_t = j|O, \lambda) \tag{2.22}$$

Again, compute this by including the observation sequence in the probability:

$$\gamma_t(j) = \frac{P(q_t = j|O, \lambda)}{P(O|\lambda)}. \tag{2.23}$$

$P(q_t = j, O|\lambda)$ is just the product of the forward probability and the backward probability:

$$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{P(O|\lambda)}. \tag{2.24}$$

Therefore, to calculate $\hat{b}_j(v_k)$ using $\gamma_t(j)$, sum $\gamma_t(j)$ for all time steps t in which the observation O_t is the symbol v_k for the numerator and sum $\gamma_t(j)$ over all time steps t for the denominator. The result is the percentage of the times that the system is in state j and saw symbol v_k. Thus;

$$\hat{b}_j(v_k) = \frac{\sum_{t=1 s.t. O_t=v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{2.25}$$

In the expectation step, the transition probability determines the expected state occupancy count γ, while the emission probability determines the expected state transition count ξ. In the maximization step, γ and ξ re-compute the new transition and emission probabilities. The equations $\hat{a}_{ij}$ and $\hat{b}_j(v_k)$ re-estimate the transition probability and emission probability from an observation sequence, assuming a known previous estimate of transition probability and emission probability. These re-estimations are the pivot of the iterative forward-backward algorithm used in the maximum likelihood parameter estimation method.

2.2.3 Initial state probabilities

The initial probability distributions are also generated. The forward-backward algorithm starts with some initial estimate of the HMM parameters $\lambda = (A, B)$, and then iteratively runs the expectation and the maximization step. Although in principle the forward-backward algorithm performs unsupervised learning of the A and B parameters, in practice, the algorithm is often given extra information for the initial conditions.

3 Results and Discussion

The HIV viral load predicts how fast the disease progresses, while the CD4 count indicates how much damage the virus has already caused. A CD4 cell count is a test that measures the number of CD4 T cells in the blood. The dataset consists of 330 HIV-infected tuberculosis patients who registered for TB treatment at different time intervals and were already undergoing ART between 2004 and 2014 at a university teaching hospital in Nigeria. The CD4 count and viral load results, often clinically checked as markers of infectious diseases, are used for the analysis. Firstly, Regression analysis to establish a possible correlation between their blood CD4 T-cell counts as the dependent variable and the viral load tests as the independent variables. Secondly, analysis and inference with discrete time and space hidden Markov model was done using R studio as follows:

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

(i) Interpreting the emission probabilities in an HMM examines how likely each observed symbol is to be generated from each hidden state. This information provides insights into the relationships between observed data and underlying states.

(ii) Interpreting the transition probabilities in an HMM examines the likelihood of moving between hidden states over time. The transition matrix provides insights into the stability and dynamics of the states, which is crucial for understanding the underlying processes modeled by the HMM.

(iii) (iii) Interpreting the initial probability distribution in an HMM examines the likelihood of the system starting in each hidden state. It provides insights into the expected starting conditions of the system, which is crucial for initializing the model correctly and understanding the initial dynamics of the process being modeled.

3.1 Descriptive statistics

The descriptive statistics of the data are in Table 1. The data appears not normally distributed and is Log transformed to make the data approximate the Normal distribution. A log transform of variables with this kind of distribution tends to make the residuals more approximately normal and the variance less dependent on the mean.

Table 1: Descriptive Statistics of HIV Markers

Basic Statistics	CD4 Count	Viral Load	Log(CD4 Count)	Log(Viral Load)
Minimum	4.00	20	1.39	3.00
Maximum	1640.00	10000000	7.40	16.12
Mean	227.77	449252	4.82	9.11
Variance	57859.92	1.493078e+12	1.57	19.28
Std. deviation	240.54	1221916	1.25	4.39
Skewness	2.00	5.45	-0.53	-0.34
Kurtosis	5.59	34.4	-0.32	-1.51

The descriptive statistic of the data summarizes it by revealing the distribution, variability and central tendencies of the data.

3.2 Relationship between the variables

Checking the correlation between the CD4 count and the viral load shows a negative relationship with the CD4 count decreasing by 0.0087 for every one-unit change in the viral load. It means that as the Viral load increases, the CD4 count decreases as was anticipated. Maintaining an undetectable viral load is compatible with having a normal or near-normal life span. Continuing to take one's medicine as prescribed to keep the virus undetectable is very important. When the HIV viral load is undetectable, there is little to no risk of infecting others. Anyone who is HIV-positive should take antiretroviral therapy medications regardless of their CD4 count. The CD4 count stays steady or goes up with efficient treatment. Clinically Normal CD4 count is from 500 to 1,400 cells per cubic millimeter of blood. CD4 counts go down over time if a patient does not take ART. The CD4 levels behave differently depending on the stage of HIV. At levels below 200 cells per cubic millimeter, a patient is likely to get a wide variety of OIs, many of which can be deadly.

The benefit of administering ART is to lower the level of HIV in the blood (the viral load) to an undetectable level. That is, the level of HIV in the blood is so low that it doesn't show up in tests. Antiretroviral therapy (ART) reduces the viral load, which allows the immune system to produce more CD4 cells. Subsequently, CD4 levels begin to grow. These help fight infections and HIV-related cancers. A patient becomes far less likely to pass HIV on to a sexual partner because people who maintain an undetectable viral load with ART have almost no chance of passing HIV

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

on to others through sex. With careful ART treatment, many people can go on for decades or more without progressing to the third and most serious stage of HIV infection. That's the stage known as acquired immunodeficiency syndrome (AIDS).

Dependent variable: CD4 count
Independent variable: Viral load

Figure 1 shows the relationship between the dependent and the independent variables used for the analysis.

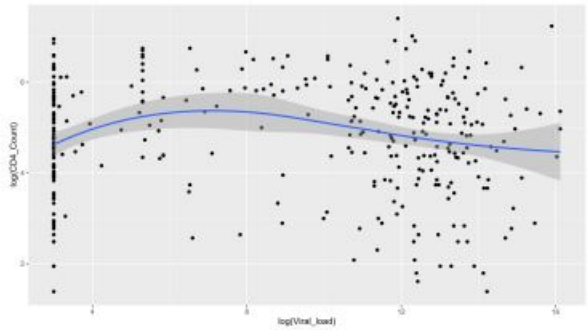

Figure 1: Relationship between the CD4 count and Viral load

3.3 Analysis and inference with discrete time and space Hidden Markov Models

An analysis of a three-state switching process is adopted to model the data using the hidden Markov model. The aim is to model the dependent as the outcome of an unobserved 3-state Markov chain where the states are independent of the measurement error for all time. The parameters characterizing the hidden Markov model are:

3.3.1 Hidden States

In a Hidden Markov Model, the state at any time t is unknown but the state will emit an observation. The Markov process is assumed to have three unobservable states corresponding to stages of disease progression and based on viral load result. The assumed three hidden states used for the analysis are in Table 2. In this work, the set of hidden states is:

Thus, $S = \{stage1, stage2, stage3\}$

Table 2: The Hidden State

State	Viral load Intervals	Patients
stage1 (undetectable)	viral load <= 20	79
stage2 (low)	viral load > 20 and <= 100000	104
stage3 (high)	viral load > 100000	150

Stage 1(Chronic HIV Infection): HIV is still active but reproduces much more slowly. ART treatment helps keep a patient in this stage for many decades and maintains CD4 at healthy levels, sometimes indefinitely. A viral load that can't be detected is less than 20 copies and is always the goal of HIV treatment. Unfortunately, the virus still survives in various cells in the body.

Stage 2 (acute HIV infection): HIV is reproducing in large amounts and destroying CD4 cells. A low HIV viral load is between 20 and 100,000 copies. CD4 levels typically fall quickly initially, but as the immune system responds to ART, the viral load begins to fall while CD4 levels rise

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

again. The virus probably isn't actively reproducing as fast, and damage to the immune system may be slowed, but this is not optimal.

Stage 3 (AIDS): The immune system is damaged badly enough to allow opportunistic infections of different types. Doctors might diagnose AIDS because of these opportunistic infections, but they can also diagnose this stage by low CD4 levels. A high viral load is generally considered around 100,000 copies, but one could have more than 1 million.

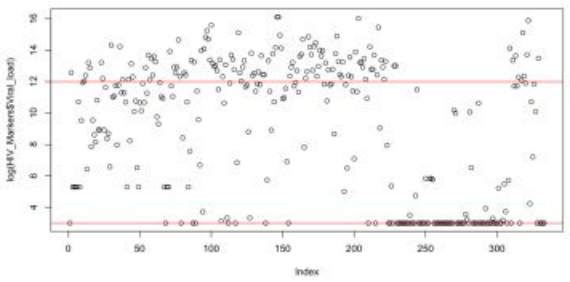

Figure 2: Viral Load(log) Intervals

Figure 2 is the plot of viral load Intervals where the height between 3 and 12 is that of state 1 and state 2.

3.3.2 Observations

At each time step, the system emits an observation that creates a sequence of observations where each observation depends on the state that generated it, not on the neighboring observations. The set of observations in this work is CD4 counts of HIV patients. The clinically normal CD4 count range is from 500 to 1500 $cells/mm^3$. The intervals considered in this work are in Table 3.

The observation symbols are:
Symbols = (Risky, Short, Normal)

Table 3: The Observations

Observation	CD4_Count Intervals	Patients
Risky	CD4_Count < 200	197
Short	CD4_Count >= 200 and < 500	97
Normal	CD4_Count >= 500	39

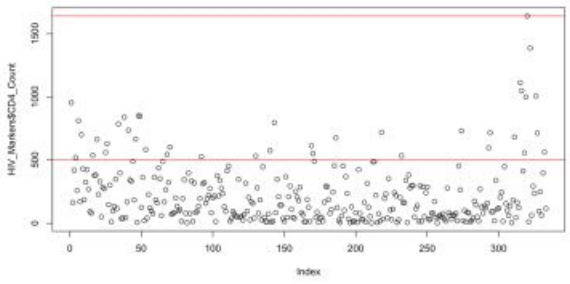

Figure 3: CD4 Count Intervals

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

Figure 3 is the CD4 count Intervals of the patients where the height 500 and above is the normal range interval.

To define the structure of HMM, maximum likelihood estimation (MLE) is a used to estimate the parameters of a distribution based on the observed dataset. Modelling, analysis and inference with discrete time and discrete space Hidden Markov Models was done using R studio. The stochastics of the HMM is fully described by the initial starting probabilities of the states, the transition probabilities between states and the emission probabilities of the states. The first part is to estimate the transition probabilities A and the emission probabilities B when the sequence of hidden states are known.

3.3.3 Transition Probability

The probability of changing from one state to another state is called the transition probability. These probabilities are a critical component of the model, as they govern the dynamics of the hidden state sequence over time. The Baum-Welch algorithm is a case of Expectation Maximization that is used to estimate the transition and emission probabilities of a HMM. The expectations-maximization algorithm uses a filtering-smoothing algorithm, to propose the path of the unobserved variable and then uses maximum likelihood, given the current regime, to estimate the model parameters and alternates the two steps until convergence. The transition probabilities for the data are in Table 4.

Table 4: The transition probabilities

	stage1	stage2	stage3
stage1	0.62820513	0.2435897	0.1282051
stage2	0.18269231	0.4230769	0.3942308
stage3	0.06666667	0.2733333	0.6600000

The transition matrix is represented as a $M \times M$ matrix where the rows sum to 1.Therefore, the transition probability matrix is:

$$A = \begin{pmatrix} 0.628 & 0.244 & 0.128 \\ 0.183 & 0.423 & 0.394 \\ 0.067 & 0.273 & 0.660 \end{pmatrix}$$

These probabilities help to understand the persistence of states and the frequency of state changes. High diagonal values (a_{ii}) suggest that the system tends to stay in the same state once it enters that state. From the results; If the current state is S_1, there is a 63% chance of staying in S_1, 24% chance of moving to S_2 and 13% chance of moving to S_3. If the current state is S_2, there is a 42% chance of staying in S_2 and a 18% chance of moving to S_1 and 39% chance of moving to S_3. If the current state is S_3, there is a 66% chance of staying in S_3 and a 7% chance of moving to S_1 and 27% chance of moving to S_2. This results show that once a patient is diagnosed to be in Stage 3, the likelihood of remaining in that stage is very high with no chance of going back to Stage 1 and a very low chance of moving to Stage 2.

3.3.4 Emission Probability

The emissions probability shows the probability of emitting a particular symbol given state j and only one symbol from a state at each time step. The emission symbols for the intervals are Normal, Risky, and Short. The emission contingency and the estimated emission probabilities are in Tables 5 and 6.

Table 5 is the statistics of the patients depicting the three states with their corresponding observations.

Table 6 is the emission probabilities. The emission probability defines the likelihood of observing a particular symbol given a specific hidden state. By examining the emission matrix, one can infer how each hidden state is related to the observed symbols.

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

Table 5: The Emission Contingency Table

	normal	risky	short
stage1	8	54	17
stage2	17	47	40
stage3	14	96	40

Table 6: Emission Probabilities

	normal	risky	short
stage1	0.10126582	0.6835443	0.2151899
stage2	0.16346154	0.4519231	0.3846154
stage3	0.09333333	0.6400000	0.2666667

The estimated emission matrix is

$$B = \begin{pmatrix} 0.101 & 0.684 & 0.215 \\ 0.163 & 0.452 & 0.385 \\ 0.093 & 0.640 & 0.267 \end{pmatrix}$$

From the results: $b_{11} = 0.1$ is the probability of observing v_1 given hidden state S_1. $b_{12} = 0.7$ is the probability of observing v_2 given hidden state S_1. $b_{21} = 0.2$ is the probability of observing v_1 given hidden state S_2. $b_{22} = 0.5$ is the probability of observing v_2 given hidden state S_2. $b_{31} = 0.1$ is the probability of observing v_1 given hidden state S_3. $b_{32} = 0.6$ is the probability of observing v_2 given hidden state S_3. Higher emission probabilities indicate a stronger likelihood that a particular observation is produced by a given state. This is not the optimal estimate of the emission probabilities for this system. The optimal values are obtained at the second part of HMM estimation on Section 3.4.

3.3.5 Initial Probability Distribution

The initial state probability estimates the probability of starting at a particular state in the observed data. Interpreting the initial probability distribution results helps to understand the starting conditions of the system being modeled. It shows the likelihood of the system starting in each hidden state at the beginning of the observation sequence. The initial probabilities results for the three states are:

Table 7:

stage1	stage2	stage3
0.2372372	0.3123123	0.4504505

$\pi = [0.24, 0.31, 0.45]$

Higher initial probabilities indicate a greater likelihood of the system starting in that particular state. The system has a 20% chance of starting in state S_1, 30% chance of starting in state S_2 and 50% chance of starting in state S_3. This initial probability distribution suggests that it is most likely to be on state 3 (Stage 3) at the beginning of the observation period.

The initial distribution provides insights into the expected starting conditions of the system, which is crucial for initializing the model correctly and understanding the initial dynamics of the process being modeled. The second part of HMM estimation is to estimate the optimal transition

IJMSO
International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

probabilities A and the emission probabilities B when the sequence of hidden states are unknown using this initial probability distribution.

3.4 HMM Learning

Hidden Markov Models are generally an unsupervised learning process where the hidden states are unknown and only the observed symbols are visible. The number of hidden states is specified while training the data. Therefore, given an observation sequence and the set of possible states in the HMM, estimate the HMM parameters A and B. This problem is to adjust the model parameters to maximize the likelihood of the observed sequence. The Baum-Welch algorithm which uses the maximum likelihood estimation was adopted to determine the parameters most likely to have produced the observed sequence. A better model at final convergence with the predicted accurate state percentage of 0.345 is derived.

The results of the final transition and emission probabilities are in Table 7 and 8 respectively.

Table 8: Optimal Transition Probabilities

	stage1	stage2	stage3
stage1	0.5097120	0.2682922	0.2219957
stage2	0.2620134	0.3376100	0.4003767
stage3	0.1855485	0.3434532	0.4709983

Table 9: Optimal Emission Probabilities

	normal	risky	short
stage1	0.3106939	0.4460639	0.24324219
stage2	0.6551841	0.2648307	0.07998517
stage3	0.7737402	0.1835605	0.04269928

The Optimal Transition Matrix is:

$$A = \begin{pmatrix} 0.51 & 0.27 & 0.22 \\ 0.26 & 0.34 & 0.40 \\ 0.19 & 0.34 & 0.47 \end{pmatrix}$$

The result shows that the probability of transiting from state 1 to state 1 is highest at 0.51. The results clearly show that maintaining an undetectable level of the viral load is optimal with ART. Again, the result reveals that the probability of transiting from state 2 to state 3 is highest at 0.40. It shows that without ART, a patient already in stage 2 will easily transit to stage 3. Also, the result shows that the probability of transiting from state 3 to state 3 is highest at 0.47. It means that when a patient is already in stage 3, the chance of staying at that stage is higher than moving back to stage 2 or stage 1. Consequently, the analysis clearly shows that maintaining an undetectable viral load with ART is crucial to living a healthy life for HIV-infected people.
The Optimal Emission Matrix is:

$$B = \begin{pmatrix} 0.31 & 0.45 & 0.24 \\ 0.66 & 0.26 & 0.08 \\ 0.77 & 0.18 & 0.04 \end{pmatrix}$$

These emission results show that the probability of emitting the 'short' symbol is highest at 0.45 when a person is in stage 1. At stage 2, the probability of emitting the 'risky' symbol is highest at 0.66. At stage 3, the probability of emitting the 'risky' symbol is also highest at 0.77. It clearly shows that the probability of emitting 'normal' is lowest amongst the symbols in the three states. Therefore, maintaining a low viral load keeps the immune system healthy, makes complications of HIV less likely, and helps patients live longer.

IJMSO

International Journal of Mathematical Sciences and Optimization: Theory and Applications
10(3), 2024, Pages 154 - 168
https://doi.org/10.5281/zenodo.13626109

4 Conclusion

The regression analysis shows that the CD4 count decreases by 0.0087 for every one-unit change in the viral load. The purpose of ART is to lower the level of HIV in the blood or the viral load to an undetectable level. Firstly, while ART lowers the viral load, the immune system makes more CD4 cells. These cells help to fight infections and HIV-related cancers. Secondly, it becomes far less likely to pass HIV on to sexual partners while maintaining an undetectable viral load with ART. With careful ART treatment, many people can go on for decades or more without progressing to the third and most serious stage of HIV infection. That's the stage known as acquired immunodeficiency syndrome, or AIDS. Without ART, the HIV levels grow while the CD4 levels fall. A Normal CD4 count is from 500 to 1, 400 cells per cubic millimeter of blood, and levels below 200 cells per cubic millimeter, a patient is more likely to get a wide variety of OIs, many of which are deadly.

Authors' Conflicts of interest

None

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Chung-Ming, K. (2002). Lecture on the Markov switching model. Institute of Economics Academia Sinica. https://www.sinica.edu.tw/as/ssrc/ckuan.

[2] Leonard, A. (2020). HMM Weather Example. https://rstudio-pubs-static.s3.amazonaws.com

[3] Twumasi, C., Asiedu, L. and Nortey, E. N. N. (2019). Markov chain modeling of HIV, Tuberculosis, and Hepatitis B transmission in Ghana. I nterdisciplinary Perspectives on Infectious Diseases, https://doi.org/10.1155/2019/9362492.

[4] Mathieu, E.; Loup, P.; Dellamonica, P.; Daures, J.P. and Biom, J. (2005). Markov modelling of immunological and virological states in HIV-1 infected patients. 47(6):833-846. doi:10.1002/bimj.200410164. PMID: 16450856

[5] Jiang, J. Yang, W. Schnellinger, E.M. Kimmel, S.E. and Guo, W. (2023). Dynamic logistic state space prediction model for clinical decision making. Biometrics, 79(1):73-85.

[6] Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257-286.

[7] Degras, D.; Ting, C. M. and Ombao, H. (2022). Markov-switching state-spacemodelswith applications to neuroimaging. Computational Statistics & Data Analysis, 174(C).

[8] Mor, B.; Garhwal, S. and Kumar, Ajay. (2020). A systematic review of hidden Markov models and their applications. Archives of Computational Methods in Engineering. 3, 1429-1448.

[9] Churchill, G.A. (1989). Stochastic models for heterogeneous DNA sequences. Bull.Math, Biol. 51, 79-94.

[10] Pauker, M.D.; Beck J.; and Robert, S. (1993). The Markov Process in Medical Prognosis. Medical Decision Making. 3, 419-458.

[11] Guihenneuc-Jouyaux, C.; Richardson, S. and Longini Jr, I. M. (2000). Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. Biometrics, 56(3)733-741. https://doi.org/10.1111/j.0006-341X.2000.00733.x

[12] Zhang, M., Jiang, X., Fang, Z., Zeng, Y. and Xu, K. (2019). High-order Hidden Markov Model for trend prediction in financial time series. Physica A, Stastistical Mechanics and its Application, 517:1-12.

[13] Lindgren, G. (1978). Markov regime models for mixed distributions and switching regressions. Scand. J. Statist. 5, 81-91.

[14] Leroux, B.G. and Puterman, M. (1992). Maximum-penalized-likelihood estimation for independent and Markov dependent mixture models. Biometrics. 48, 545-558.

[15] Gerencsér, L., Molnar-Saska, G., Michaletzky, G., Tusnady, G. and Vágó, Z. (2002). New methods for the statistical analysis of Hidden Markov Models. Proceedings of the 41th IEEE Conference on Decision & Control, 2: 2272 - 2277. doi:10.1109/CDC.2002.1184870

[16] Pennoni, F., Bartolucci, F., Forte, G. and Ametrano, F. (2021). Exploring the dependencies among main cryptocurrency log-returns: A hidden Markov model. Economic Notes, 51(1). https://doi.org/10.1111/ecno.12193.

[17] Ogunmodimu, M. O., Enock, E. P., Kenyatta, A. P., Affognon, S. B. and Onwuegbuche, F. C. (2024). A Mathematical Model for the Prevention of HIV/AIDS in the Presence of Undetectable Equals Untransmittable Viral Load. International Journal of Mathematical Sciences and Optimization: Theory and Applications, 10 (2): 36-57.

[18] Mufutau, R. A. and Akinpelu, F. (2020). Sensitivity Analysis of Mathematical Modelling of Tuberculosis Disease With Resistance to Drug Treatments. International Journal of Mathematical Sciences and Optimization: Theory and Applications, 6 (2): 940 - 955.

[19] Gentleman, R. C.; Lawless, J. F.; Lindsey, F. C. and Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. Statistics in Medicine. 13, 805-821.

[20] Longini, I. M.; Clark, W. S.; Byers, R. H.; Ward, J. W.; Darrow, W. W.; Lemp, G. P. and Hethcote, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. Statistics in Medicine. 8, 831- 843.

[21] Matan, L. (2020), The path from Maximum Likelihood Estimation to Hidden Markov Models. https://www. towardsdatascience.com

[22] Jurafsky,D. and Martin, J.H. (2021). Speech and Language Processing. https://www.stanford.edu.

[23] Vidyasagar, M. (2011). Hidden Markov Processes: Theory and Applications to Biology. Princeton University Press Princeton and Oxford. https://personal.utdallas.edu/ m.vidyasagar/Fall-2011/Biostat/MV-text.pdf.