Estimation of Malaria Symptom Data Set using Hidden Markov Model

Drinold Mbete[1*], Kennedy Nyongesa[1] and Joseph Rotich[2]

[1]*Department of Mathematics, Masinde Muliro University of Science and Technology, P. O Box 190-50100, Kakamega, Kenya.*
[2]*Vice Chancellor,Laikipia University, P. O Box 1100-20300, Nyahururu, Kenya.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Original Research Article

Abstract

Clinical study of malaria presents a modeling challenge as patients disease status and progress is partially observed and assessed at discrete clinic visit times. Since patients initiate visits based on symptoms, intense research has focused on identification of reliable prediction for exposure, susceptibility to infection and development of severe malaria complications. Despite detailed literature on malaria infection and transmission, very little has been documented in the existing literature on malaria symptoms modeling, yet these symptoms are common. Furthermore, imperfect diagnostic tests may yield misclassification of observed symptoms. Place and Duration of Study: The main objective of this study is to develop a Bayesian Hidden Markov Model of Malaria symptoms in Masinde Muliro University of Science and Technology student population. An expression of Hidden Markov Model is developed and the parameters estimated through the forward-backward algorithm.

*Drinold Mbete: drinoldmbete123@gmail.com:

2019 Statistical Subject Classification: 53C25; 83C05; 57N16.

1 Introduction

The term Malaria was first used by Dr. Fransisco Torti, but it was not until 1880 that scientists discovered that it was a parasitic disease caused by a unicellular protozoan of the genus *Plasmodium* which is transmitted by the Anopheles mosquito. Malaria is an ancient disease that has been affecting people since the beginning of recorded time. It poses serious economic, social and health burdens in tropical and subtropical countries where it is predominantly found [1]. Malaria still remains a huge public health issue regardless of how many years of research has been conducted on how to combat this disease.

According to WHO [2], the latest world malaria report released in November 2017 shows that the number of malaria cases reported in the year 2016 was 216 million up from 211 million cases reported in 2015. The report also shows that malaria death estimates in 2016 stood at 445,000 compared to 446,000 deaths in 2015. The high burden of malaria cases in 2016 was in Africa at 90% with 91% cases of deaths reported in children. According to WHO report on malaria cases in Kenya, malaria is one of the leading causes of morbidity and fatality with about 3.5 million children at risk of developing severe malaria, out of which an estimated 34,000 children under five years die every year. The disease is also responsible for 30% of out-patient visits at health centers, economically, it is estimated that 170 million working hours are lost each year because of malaria illness [3].

The malaria symptoms can be grouped into two; symptoms for uncomplicated malaria (suspected malaria) and symptoms for complicated malaria (severe malaria). In a study by Martins *et al* [4], there are 19 common symptoms associated with malaria disease which were confirmed and assessed by microscopy, namely; fever, chills, sweating, headache, myalgia, arthralgia, abdomial pain, nausea, vomiting, dizzness, cough, diarrhea, weakness, inappetence, bitter mouth, pallor, coryza, sneezing and score throat. Some of these symptoms are observable symptoms in patients. Malaria is considered uncomplicated when symptoms are present but there are no clinical or laboratory signs to indicate severity or vital organ dysfunction. Infection with *Plasmodium falciparum* if not promptly treated can quickly progress to complicated malaria (severe malaria) [5].

2 Literature Review

The Hidden Markov Model (HMM) is a statistical method based on Markov Chain. It is a powerful tool for random processing and modeling which is normally used to predict and classify data. The first step to develop the HMM was taken by Rabinner 1989 after the presentation of the educational article of HMM by revealing the details of the complex models Rabinner. Most of the investigation using HMM have been done in non-medical fields, for instance, Cholewa and Glomb [6] investigated on estimation of the number of critical points in time sequence while Farsi [7] investigated on implementation and optimization of speech recognition system based on HMM using genetic algorithm.

Some of the investigation of the medical fields using HMM include the study by Vimala *et al* [8] which used HMM to identify and classify Electrocardiography (ECG)signals. The results of Vimala confirmed that the HMM could be used as a powerful tool for grouping ECG signals into three signals. Also Li *et al* [9] used HMM to predict the progression of lung cancer among 508 patients in one of the Chinese hospitals from 2010 to 2012. The results showed that HMM was able to predict 0.81 accuracy while Lee *et al* [10] used HMM to classify snoring sounds of 21 patients with sleeping disorders.

HMM has been applied explicitly to modeling one dimensional data and is less used with spatial structure data [11]. Autoregressive HMM (ARHMM) was simplified by focusing on the modeling of continuous observation dependence by Rabinner. Recent applications of these models are study by Barber *et al* for predicting short winds [12], study of Wu *et al* for using a Bayesian non-parametric vector ARHMM for testing robot performance [13] and Tuncel *et al* for using autoregressive forests for model multivariate time series [14]. Therefore as the literature review shows, the development of HMM is more in the field of continuous observation and less attention has been paid to relation of the hidden states in the Discrete Hidden Markov Models (DHMM). Therefore this study aims at using (DHMM)with underlying first order Markov Chain. DHMM can be regarded as a probabilistic generative model such that a sequence of internal hidden states of the model which is not directly visible produces a sequence of discrete observations.

A Hidden Semi-Markov Model (HSMM) is an extension of HMM designed to allow general (i.e. non-geometric or non-exponential) distribution for the state duration. A major drawback with HMM is the inflexibility in describing the time spent in a given state which is geometrically distributed. A discrete Hidden Semi-Markov chain is composed of non observable state process which is a semi-Markov chain and a discrete output process which is an embedded first-order Markov chain representing the transitions between distinct states and discrete state occupancy distribution representing sojourn times in a non-absorbing states [15] An HSMM is constructed by adding a temporal component (duration) into HMM. Unlike a state in a standard HMM, a state in an HSMM generates a sequence of observational as opposed to a single observation in HMM [16].

In many studies of medical treatment, symptoms are measured repeatedly over time in observation called longitudinal observation. Though we cannot observe directly latent variables, we learn about it by measuring symptom. For the longitudinal models, two latent variables govern disease, one for the probability of experiencing a particular symptom and another for the severity of the experienced symptom. Thus the probability of a symptom and the severity of it depends on both latent variables and observed variables [17]. Latent variables are variables that are not directly observed but are inferred through a mathematical model from other variables that are directly observed or measured. A latent variable model is a statistical model that contains latent i.e. unobserved variables. These variables can either be discrete or continuous. Sometimes latent variables corresponds to aspects of physical reality which could in principle be measured but may not be for practical reason thus in this situation the term hidden variable is commonly used. one advantage of using latent variables is that they can serve to reduce the dimensionality of data. Latent variable link observable data in the real world to symbolic data in the model. Bayesian statistics is often used for inferring latent variables, the common method used inferring latent variables in Bayesian statistics are; Hidden Markov Model (HMM), factor analysis, principal component analysis and Expectation Maximization (EM) algorithm [17].

Zammit *et al* [18] developed an intra-individual consistency model using a logistic-type latent variable model. The latent variable in the model was used to represent the propensity of symptoms and intensity of episodes as these could not be observed directly and needed to be estimated through observation of symptoms episodes in hypoglycaemia. The model results showed that their was individual difference in symptom reporting and that adults exhibit distinct intra-individual variability in symptom reporting. Hans *et al* extended on the model developed by Zammit *et al* by allowing for different forms of symptom experiencing thresholds between groups variability when symptoms are classified in groups and performing variable selection to determine a predictive model for the effect of patient characteristics and their interactions on symptom consistency. The study was conducted in several health centers in the United Kingdom and data collected from 381 participants aged between 17-75 years. Bayesian estimation was performed for all coefficients in the developed model without grouped symptoms and with grouped symptoms. The analysis shows that a multiplicative form of symptom propensity and episode intensity provides the most suitable

symptom experiencing threshold and groups of symptoms show distinct propensity and that gender subjects had significant impact on the consistency of symptom reporting.

Xing *et al* [19] developed a Bayesian statistical model using latent semi-Markovian state and state-transition statistics for analysis of the time-evolving properties of influenza-like illness with a particular focus on symptoms. Self-reported data from individual student in a college provided daily over a multiple of months was used. The data corresponded to the strength of various infectious-disease-related symptoms reported separately by each individual student. The computation was performed using Markov Chain Monte Carlo (MCMC) and statistical analysis performed on the daily self-reported symptom scores. The results showed that the weekly pattern (probability of transiting from healthy state to infective state) is typically heightened at either Wednesday or Thursday and tends to be smaller around weekend because of the fact that students are more likely to report symptom during the school week than they are on the weekend.

3 Methodology

3.1 The Study Area

The research was conducted in Masinde Muliro University of Science and Technology (MMUST) located in Kakamega Town, Kakamega County with an altitude of 1561m above the sea level with a student population of approximately 15000. The levels of malaria risk and transmission intensity in MMUST exhibit significant spatial and temporal variability related to variations in amount of rainfall, temperature, altitude, topography and human settlement pattern. In this study area malaria situation is typical of Sub-Saharan Africa making its transmission an all- year -round affair and seasonal variation. The MMUST Health facility records show that between 300-700 cases of malaria are reported each month which constitutes 75% of all out-patient cases. The main malaria vectors in MMUST are *Anopheles gambiae sensu stricto*, *An. Arabiensis* and *An. Funestus*. *Anopheles gambiae* generally increases in density after the start of the long rains, while *An. funestus* density is seen to vary in direct proportion to the proximity of permanent breeding grounds rather than rainfall [20].

The pick period of malaria incidence occurs from April to August following the main rain season. The malaria cases can either be complicated malaria or uncomplicated malaria. For complicated malaria, the following symptoms have been displayed by students; hallucination, prostration, loss of consciousness, hyperparasitaemia, pallor, convulsions, low and high blood pressure, coma, convulsions, low and high pulse beat/min, anaemia and black quarter fever and dark urine. For uncomplicated malaria, the following non-specific symptoms have been displayed by the students; headache, pains (joint, muscle, abdominal), loose stool, fever, rigors, nausea and vomiting. For confirmatory test of malaria, blood slide (BS) for malaria parasite is carried out[20].

Once a student presents himself/herself to a health officer, the following information are recorded in his/her file

  (i) The patient complaints

  (ii) History of the infection.

  (iii) Physical examination for signs and symptoms for both specific and non-specific symptom

  (iv) Impression- decision made by health officer (suspected malaria).

  (v) Investigation of the disease through laboratory test (BS test).

  (vi) Diagnostic - the diagnostic test will result in either mild (+), moderate (++) or severe (+++) infection. The diagnostic can also be recorded as per 200 white blood cell (WBC)

i.e. for 1-10 per 200 WBCs (mild), for 11-100 per 200 WBCs (moderate) and for 1-10 in WBCs (severe)

(vi) Management of the disease (treatment)

3.2 The Model

In a study by Martins *et al* [4], there are 19 common symptoms associated with malaria disease which were confirmed and assessed by microscopy, namely; fever, chills, sweating, headache, myalgia, arthralgia, abdomial pain, nausea, vomiting, dizzness, cough, diarrhea, weakness, inappetence, bitter mouth, pallor, coryza, sneezing and score throat. Some of these symptoms are observable symptoms in patients. A healthy student when he/she is infected with malaria, the disease develops to mild, moderate and final severe depending on the frequency of symptoms he/she has. This description is shown in Figure 3.1.

The states as illustrated in Figure 3.1 are defined as Z_1 for absent of symptom (healthy individual),
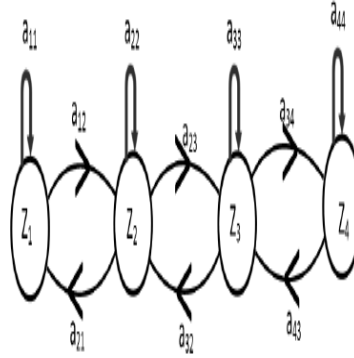

Fig. 1. Malaria transition diagram

Z_2 for mild illness, Z_3 for moderate illness and Z_4 for severe illness.

Therefore in this study, we let $Z = \{Z_1, Z_2, ..., Z_n\}$ where n is the number of possible hidden states an individual can be at any given time point. i.e $Z = \{Z_1, Z_2, Z_3, Z_4\}$

Let x_1 - fever (body temperature), x_2 - chills, x_3 - sweating, x_4 - vomiting, x_5 - diarrhea, x_6 - weakness, x_7 - pallor, x_8 - cough and x_9 - sneezing be observable symptoms used in this study. We denote this observable symptoms by O i.e $O_i = (x_1, ..., x_p)$ where $i = 1, ..., 4$ and p is the number of symptoms. Any of the four hidden states will depict different combinations in O. Upon involving this observation O in Figure 3.1, we obtain Figure 3.2 as shown below.

In Figure 3.2, each hidden state is shown as a circle and state transition represented by directed graph edge between states. The arrow goes from the hidden states Z to observed symptoms O, this is because the state of illness at which an individual is in causes a particular symptom(s) to be observed. The hidden states are interconnected in such a way that any state can be reached from any other state. Thus, the transition from one state to the next state is a Markov process of order
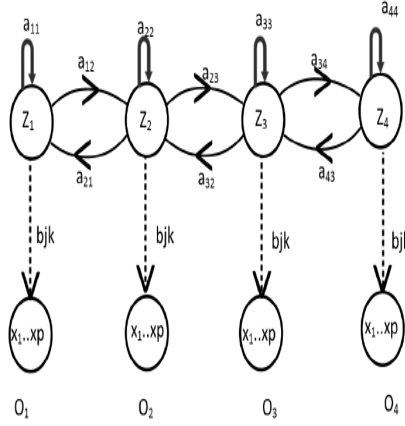
Fig. 2. Observed symptoms with Hidden states

one and the next state depends on the current state and fixed probabilities.

Let a_{ij} be the transition probability of the disease transiting from state i to state j, i.e

$$a_{ij} = p(Z_n = j | Z_{n-1} = i) \tag{1}$$

where $a_{ij} \geq 0$, $\sum_{j=1}^{n} a_{ij} = 1$ and $1 \leq i, j \leq n$.

Let A be the transition probability matrix i.e a set of transition probabilities among states. In this study, malaria disease has four hidden states which are represented by the transition probability matrix shown below;

$$A = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & Z_3 & Z_4 \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \end{array}$$

Let b_{jk} be the probability of observing symptom(s) in each of the hidden states i.e.

$$b_{jk} = P(O = x_k | Z = Z_j) \tag{2}$$

where

$$b_{jk} \geq 0, \qquad \sum_{k=1}^{p} b_{jk} = 1, \qquad 1 \leq k \leq p, \qquad 1 \leq j \leq n$$

Let B be the probability distribution of observation matrix i.e. B=$[b_{jk}]$. In this study, there are 9 different observed symptoms and four different hidden states represented by observation matrix

shown below;

$$
B = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array}
\begin{array}{ccccccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\
\begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} & b_{1,5} & b_{1,6} & b_{1,7} & b_{1,8} & b_{1,9} \\
b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} & b_{2,5} & b_{2,6} & b_{2,7} & b_{2,8} & b_{2,9} \\
b_{3,1} & b_{3,2} & b_{3,3} & b_{3,4} & b_{3,5} & b_{3,6} & b_{3,7} & b_{3,8} & b_{3,9} \\
b_{4,1} & b_{4,2} & b_{4,3} & b_{4,4} & b_{4,5} & b_{4,6} & b_{4,7} & b_{4,8} & b_{4,9} \end{bmatrix}
\end{array}
$$

Let π be the initial state distribution vector; where π is an initial probability for each state of the disease by which the Markov Chain begins to work. π is a edge entering into disease state from state zero (start) which is not shown in figure 1, because we imagine that there is a silent state zero which all states originate from and therefore the system cannot transit to state zero but can only transit out of it.

Therefore the Hidden Markov Model (HMM) is specified using 3 parameters defined as

$$
\mu = (A, B, \pi) \tag{3}
$$

where A is the transition probability matrix, B is the probability distribution of observation matrix and π is the initial state distribution.

In practice we don't observe the state Z but the observation sequence O. The observed sequence in O leads to a particular state in Z, thus there is a relationship between Z and O. This relationship leads to the computation of $P(Z|O)$ as described in the next section. For example when a student displays the following symptoms x_1 = fever ($38°C$), x_2 = chills, x_3 = sweating, this observation sequence leads to observation of state Z_2 where Z_2 is the mild state of malaria disease.

4 Computation of $\mathbf{P}(Z|O)$

Let $Z = Z_1, ..., Z_n$ where n=4 be the hidden states at time point t and $z = \{z_1, ..., z_T\}$ whose element z_t is a state at time point t and T is the length of the sequence. z is a state sequence process and $z_t \in Z$ equals some hidden state Z_n. Since z is a state sequence, it therefore follows the Markov property i.e the conditional probability of current state z_t is only dependent on the previous state z_{t-1}. Upon involving this z in Equation 3.1, the transition probability of the disease transiting from state i to state j becomes;

$$
a_{ij} = P(z_t = Z_j | z_{t-1} = Z_i) \tag{4}
$$

Introducing set of observation to the states shown in Figure 3.2, the model is modified as shown in Figure 3.3 for prediction of state Z.

Each vertical slice in Figure 3 represents a time step. The top node represents the variable z_t where z_t is a state of individual at a particular time point and the bottom node represents the observable variable O_t where O_t is the observation at a particular time point.

We let O = $\{O_1, O_2, ..., O_T\}$ be the sequence of observation where T is the length of the sequence and $\pi_i = P(z_1 = Z_i)$ where i=1,2,3,4 be the initial probability distribution.

Using Figure 3, various conditional independencies can be obtained but the main conditional independency of interest is obtained by conditioning on a single state node, for example conditioning on z_t renders z_{t-1} and z_{t+1} independent while on the other hand conditioning on an observation node does not separate nodes in the graph and therefore will not yield any conditional independencies.
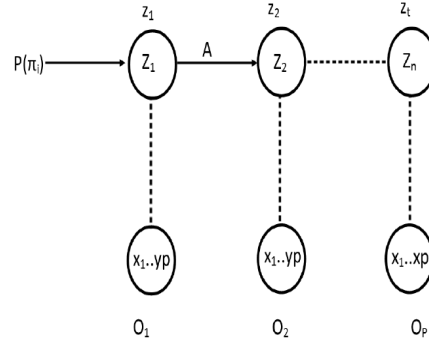
Fig. 3. A representation of HMM as a graphical model

To obtain the conditional independencies at a particular sample point $P(z, O)$, we compute the joint probability by taking a product over the conditional probabilities for a particular sample point as shown below;

Let

$$P(z, O) \quad = \quad P(z_1, z_2, ..., z_T, O_1, O_2, ..., O_T) \tag{5}$$

be the joint probability. Upon simplification the joint probability becomes

$$P(z, O) = P(z_1)[\prod_{t=1}^{T-1} p(z_t | z_{t+1})] \prod_{t=1}^{T} P(O_t | z_t) \tag{6}$$

where $P(z_t | z_{t+1})$ is the transition probability A, $P(O_t | z_t)$ is the obsevation/emission probability and $P(z_1)$ is the initial state distribution.

Using the HMM μ developed and defined in Equation (3.4), Equation (3.8) can be written as;

$$P(z, O) = P(\pi_{z_1})[\prod_{t=1}^{T-1} a_{z_t, z_{t+1}}] \prod_{t=1}^{T} P(O_t | z_t) \tag{7}$$

where

$$a_{z_t, z_{t+1}} \quad \equiv \quad [a_{ij}] z_t^i z_{t+1}^j$$

$$\pi_{z_1} \quad \equiv \quad \prod_{i=1}^{Z} [\pi_i] z_1^i \qquad (z_1 = Z_1)$$

To compute the probability of a hidden state Z given an observation output O, we compute the probability $P(z|O)$ using the defination of conditional probability as follow;

$$P(z|O) = \frac{P(z, O)}{P(O)} \tag{8}$$

The numerator $P(z, O)$ is simplified by substituting it with Equation (3.8) while the denominator $P(O)$ involves computing the sum across all the possible values of the hidden states;

$$P(O) = \sum_z P(z, O) \tag{9}$$

upon simplification and substituting the value of $P(z, O)$, we get

$$P(O) = \sum_{z_1} \sum_{z_2} ... \sum_{z_T} \pi_{z_1} \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^{T} P(O_t|z_t)$$

therefore Equation (3.10) can be written as,

$$P(z|O) = \frac{\pi_{z_1} \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^{T} P(O_t|z_t,)}{\sum_{z_1} \sum_{z_2} ... \sum_{z_T} \pi(z_1) \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^{T} P(O_t|z_t)} \qquad (10)$$

Equation (10) implies that each state node z_t can take on Z values. Since we have T state nodes, this implies that we perform Z^T sums to observe all the hidden states. According to Rabinner [16] calculation of Z^T sums is infeasible, therefore rather than calculating P($z|O$) for the entire sequence z, we should focus on a particular state node z_t and calculate its posterior probability i.e., $P(z_t|O)$

We use a fragment of the Figure shown in 3.2 and modified it to Figure shown 3.3 so as to compute $P(z_t|O)$

To compute $P(z_t|O)$, we apply Bayes rule and the defination of conditional probability as follow;
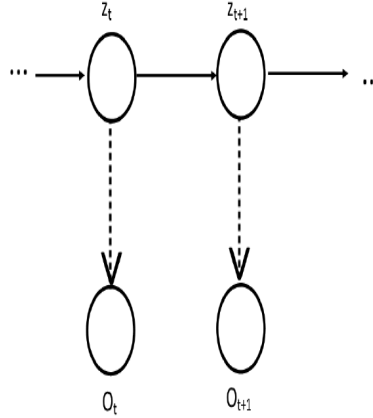
Fig. 4. A fragment of the graphical model representation of an HMM.

$$P(z_{(t)}|O) = \frac{P(z_t, O)}{P(O)} \qquad (11)$$

applying the notion of independence, we get

$$P(z_t, O) = P(z_t|O)P(O) \quad and \quad P(O, z_t) = P(O|z_t)P(O) \qquad (12)$$

applying Bayes rule in Equation (3.13), we get

$$P(z_{(t)}|O) = \frac{P(O|z_{(t)})P(z_{(t)})}{P(O)} \qquad (13)$$

We use Figure 4 to get conditional independence of $P(z_{(t)}|O)$ as shown below;

$$
\begin{aligned}
P(z_{(t)}|O) &= \frac{P(O_1, O_2..., O_t, O_{t+1}, O_{t+2}, ..., O_T|z_{(t)})}{P(O)} \\
&= \frac{P(O_1, ..., O_t|z_{(t)})P(O_{t+1}, ..., O_T|z_t)P(z_t)}{P(O)}
\end{aligned}
\tag{14}
$$

Let $\alpha(z_t) \equiv P(O_1, ..., O_t, z_t)$ be the probability of emitting a partial sequence of output $O_1, ..., O_t$ and ending up in state z_t and $\beta(z_t) \equiv P(O_{t+1}, ..., O_T|z_t)$ be the probability of emitting a partial sequence of output $O_{t+1}, ..., O_T$ given that the system starts in state z_t.
Then Equation 14 can be written as

$$
P(z_{(t)}|O) = \frac{\alpha(z_{(t)})\beta(z_{(t)})}{P(O)}
\tag{15}
$$

where $\alpha(z_t)$ and $\beta(z_t)$ are vector with component $\alpha(z_t^i)$ and $\beta(z_t^i)$.
Given that the sum $P(z_t|O)$ over the components of z_t must equal to one, then we obtain

$$
P(O) = \sum_i \alpha(z_t^i)\beta(q_t^i)
\tag{16}
$$

Let $\gamma(z_t)$ be the posterior probability. Then $\gamma(z_t)$ is defined as

$$
\gamma(z_{(t)}) \equiv \frac{\alpha(z_{(t)})\beta(z_{(t)})}{P(O)}
\tag{17}
$$

where P(O) is computed once as normalization constant for a particular arbitrary choice of t.
Given that $\alpha(z_t)$ depends only on quantities up to time t and using the Markov properties in the model, we obtain recursion between $\alpha(z_t)$ and $\alpha(z_{t+1})$ in figure 4. Upon simplification, the forward recursion is obtained as follow;

$$
\begin{aligned}
\alpha(z_{t+1}) &= P(O_1, ..., O_{t+1}, z_{t+1}) \\
&= P(O_1, ..., O_{t+1}|z_{t+1})P(z_{t+1}) \\
&= P(O_1, ..., O_t|z_{t+1})P(O_{t+1}|z_{t+1})P(z_{t+1}) \\
&= P(O_1, ..., O_t, z_{t+1})P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t, z_t, z_{t+1})P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t|z_t)P(z_t)P(z_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(z_1, ..., O_t|z_t)P(z_{t+1}|z_t)P(_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t, z_t)P(z_{t+1}|z_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} \alpha(z_t)a_{z_t, z_{t+1}}P(O_{t+1}|z_{t+1})
\end{aligned}
\tag{18}
$$

For the beta variable we obtain "a backward" recursion by expressing $\beta(z_t)$ in terms of $\beta(z_{t+1})$.

Upon simplification, the backward recursion is obtained as follows.

$$
\begin{aligned}
\beta(z_t) &= P(O_1, ..., O_{t+1}, z_t) \\
&= \sum_{q_{t+1}} P(O_{t+1}, ..., O_T, z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} P(O_{t+1}, ..., O_T | z_{t+1}, z_t) P(z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} P(O_{t+2}, ..., O_T | z_{t+1}) P(O_{t+1} | z_{t+1}) P(z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} \beta(z_{t+1}) a_{z_t, z_{t+1}} P(O_{t+1} | z_{t+1})
\end{aligned}
\tag{19}
$$

For the alpha recursion, the definition of alpha at the initial step yields

$$
\begin{aligned}
\alpha(z_1) &= P(O_1, z_1) \\
&= P(O_1 | z_1) P(z_1) \\
&= P(O_1 | z_1) \pi_{z_1}
\end{aligned}
\tag{20}
$$

In the next section, we compute the likelihood of an observed sequence O given the model μ i.e given the model and a sequence of observation, we want to evaluate how well the model predicts the observation sequence.

5 Computation of $\mathbf{P}(O|\mu)$

Let $Z = \{Z_1, Z_2, ..., Z_n\}$ be a state sequence as already defined, $\mu = (\pi, A, B)$ be the Hidden Markov Model as defined in Equation (3.4) and $O = \{O_1, O_2, ..., O_T\}$ be a sequence of observations corresponding to state sequence as shown in Figure **??**. Then we define π_{Z_1} as the probability of starting in state Z_1, $b_{Z_1}(O_1)$ as the probability of initially observing O_1 and a_{Z_1, Z_2} as the probability of transiting from state Z_1 to state Z_2.
Therefore P(Z,O) is written as

$$
P(z, O) = \pi_{Z_1}, b_{Z_1}(O_1) a_{Z_1, Z_2}, ..., a_{Z_{n-1}, Z_n} b_{Z_n}(O_T)
\tag{21}
$$

and by defination of B (i.e the probability of the observation sequence given the state sequence) we have

$$
P(O|Z, \mu) = \prod_{t=1}^{T} P(O_t | Z_{(t)}, \mu) = b_{Z_1}(O_1) b_{Z_2}(O_2), ..., b_{Z_T}(O_T)
\tag{22}
$$

and by defination of π and A it follows that the probability of the state sequence is given by

$$
P(Z|\mu) = \pi_{Z_1} a_{Z_1, Z_2} a_{Z_2, Z_3} ... a_{Z_{n-1} Z_n}
\tag{23}
$$

using conditional probability, we have

$$
P(Z|\mu) = \frac{P(Z \cap \mu)}{P(\mu)}
\tag{24}
$$

and

$$
P(O|Z, \mu) = \frac{P(O \cap Z \cap \mu)}{P(Z \cap \mu)}
\tag{25}
$$

and

$$P(O, Z|\mu) = \frac{P(O \cap Z \cap \mu)}{P(\mu)} \tag{26}$$

multiplying Equation(3.26) and Equation (3.27), we get

$$P(O|Z, \mu)P(Z|\mu) = \frac{P(O \cap Z \cap \mu)}{P(Z \cap \mu)} \cdot \frac{P(Z \cap \mu)}{p(\mu)} = \frac{P(O \cap Z \cap \mu)}{P(\mu)} \tag{27}$$

substituting Equation(3.29)in Equation (3.28), we have

$$P(O, Z|\mu) = P(O|Z, \mu)P(\mu) \tag{28}$$

to compute $P(O|\mu)$, we sum over all possible state sequence so as to obtain the likelihood of the observed sequence O i.e

$$P(O|\mu) = \sum_Z P(O, Z|\mu) \tag{29}$$

using Equation (3.30), Equation (3.31) becomes

$$P(O|\mu) = \sum_Z P(O|Z, \mu)P(Z|\mu) \tag{30}$$

(substituting Equation 22 and 23), we obtain

$$P(O|\mu) = \sum_Z \pi_1 b_{Z_1}(O_1) a_{Z_1, Z_2} b_{Z_2}(O_2) ... a_{Z_{n-1}, Z_n} b_{Z_n}(O_T) \tag{31}$$

However, Equation (3.33) is a direct computation which is generally infeasible since it requires $2Tn^T$ multiplications. Since it is not possible to perform z^n sums, a more efficient procedure called Forward-Backward algorithm procedure is required to solve P($O|\mu$) by decreasing computational procedure.

5.1 Forward-Backward Algorithm

Let $\alpha(z_t^i)$ be the joint probability of partial observation sequence $\{O_1, O_2, ..., O_t\}$ at state $z_t = Z_i$ where $1 \le t \le T$ is specified as

$$\alpha(z_t^i) = P(O_1, O_2, ..., O_t, z_t = Z_i|\mu) \tag{32}$$

Multiplying Equation (32) by a_{ij} where a_{ij} is the transition probability from state i to state j and counts for probability of joint event that partial observation sequence exists and state Z_i at time point t is changed to Z_j at time point t+1. Upon simplification via multiplication rule and Markov property, we obtain

$$
\begin{aligned}
\alpha(z_t^i) &= P(O_1, ..., O_t, z_t = Z_i|\mu)P(z_{t+1} = Z_j|z_t = Z_i) \\
&= P(O_1, ..., O_t|z_t = Z_i)P(z_t = Z_i)P(z_{t+1} = Z_j|z_t = Z_i) \\
&= P(O_1, ..., O_t|z_t = Z_i)P(z_{t+1} = Z_j)P(z_t = Z_i) \\
&= P(O_1, O_2, ..., O_t, z_{t+1} = Z_j|z_t = Z_i)P(z_t = Z_i) \\
&= P(O_1, O_2, ..., O_t, z_t = z_i, z_{t+1} = Z_j)(Markov \quad property)
\end{aligned} \tag{33}
$$

Summing product over all n possible states of z_t produces probability of joint event that the partial observation sequence exists and the next state is $z_{t+1} = Z_j$ regardless of the state z_t. By summing product we obtain

$$
\begin{aligned}
\sum_{i=1}^{n} \alpha(z_t^i a_{ij}) &= \sum_{i=1}^{n} P(O_1, O_2, ..., O_t, z_t = Z_i, z_{t+1} = Z_j) \\
&= P(O_1, O_2, ..., O_t, z_{t+1} = Z_j)
\end{aligned}
\tag{34}
$$

The forward variable at time t+1 and state Z_j is calculated as follows using the multiplication rule

$$
\begin{aligned}
\alpha(z_{t+1}^i) &= P(O_1, O_2, ..., O_t, O_{t+1}, z_{t+1} = Z_j | \mu) \\
&= P(O_{t+1} | O_1, O_2, ..., O_t, q_{t+1} = Z_j) P(O_1, O_2, ..., O_t, z_{t+1} = Z_j) \\
&= P(O_{t+1} | z_{t+1} = Z_j) P(O_1, O_2, ..., O_t, z_{t+1} = Z_j) \\
&= b_j(O_{t+1}) \sum_{i=1}^{n} \alpha(z_t^i) a_{ij}
\end{aligned}
\tag{35}
$$

where $b_j(O_{t+1})$ is the probability of an observation O_{t+1} when the markov state is in state Z_j. using the Forward recurrence Equation in (35), we obtain the observation sequence O $=\{O_1, O_2, ..., O_T\}$ of the Forward variable as

$$
\alpha_T(z_t^i) = P(O_1, O_2, ..., O_T, z_T = Z_i | \mu)
\tag{36}
$$

The probability $P(O|\mu)$ is sum of $\alpha_T(z_t^i)$ over all n possible states of z_T specified by

$$
\begin{aligned}
P(O|\mu) &= P(O_1, O_2, ...O_T) \\
&= \sum_{n}^{i=1} P(O_1, O_2, ..., O_T, z_T = Z_i | \mu) \\
&= \sum_{i=1}^{n} \alpha_T(i)
\end{aligned}
\tag{37}
$$

Let $\beta(z_t^i)$ be the Backward variable which is a conditional probability of partial observation sequence $\{O_t, O_{t+1}, ..., O_T\}$ given state $z_t = Z_i$ where $1 \le t \le T$ specified as $\beta(z_t^i) = P(O_{t+1}, O_{t+2}, ..., O_T | z_t = Z_i, \mu)$

multiplying the transition probability a_{ij} and $b_j(O_{t+1})$ the probability of the observation sequence O_{t+1} when the Markov is in state Z_j together with the Backward variable $\beta(z_{t+1}^j)$ at time point $t + 1$ we obtain

$$
\begin{aligned}
a_{ij} b_j(O_{t+1}) \beta(z_{t+1}^j) &= P(z_{t+1} = Z_j | z_t = Z_i) \times P(O_{t+1} | z_{t+1} = Z_j) \\
&\times P(O_{t+2}, O_{t+3}, ..., O_T | z_{t+1} = Z_j, \mu)
\end{aligned}
\tag{38}
$$

because observation $(O_{t+2}, O_{t+3}, ..., O_T)$ are mutually independent, we have

$$
\begin{aligned}
a_{ij} b_j(O_{t+1}) \beta(z_{t+1}^j) &= P(z_{t+1} = Z_j | z_t = Z_i) \\
&\times P(O_{t+1}, O_{t+2}, ..., O_T | z_{t+1} = Z_j \mu)
\end{aligned}
\tag{39}
$$

because partial observation sequence $(O_{t+2}, O_{t+3}, ..., O_T$ is independent from state z_t at time point t, we have

$$
\begin{aligned}
a_{ij} b_j(O_{t+1}) \beta(z_{t+1}^j) &= P(O_{t+1}, O_{t+2}, ..., O_T | z_{t+1} = z_j, \mu) \\
&= P(z_{t+1} = Z_j | z_t = z_i) \\
&\times P(O_{t+1}, O_{t+2}, ..., O_T | z_t = Z_i, z_{t+1} = Z_j, \mu)
\end{aligned}
\tag{40}
$$

due to multiplication rule, we have

$$a_{ij}b_j(O_{t+1})\beta(z_{t+1}^j) \quad = \quad P(O_{t+1}, O_{t+2}, ..., O_T, z_{t+1} = Z_j | z_t = Z_i, \mu)$$

Summing the product $a_{ij}b_j(O_{t+1})\beta(z_{t+1}^j)$ over all n possible states of $z_{t+1} = Z_j$ and applying total probability rule, we have

$$
\begin{aligned}
\sum_{j=1}^{n} a_{ij}b_j(O_{t+1})\beta(z_{t+1}^i) \quad &= \quad \sum_{j=1}^{n} P(O_{t+1}, O_{t+2}, ..., O_T, z_{t+1} = z_j | z_t = Z_i, \mu) \quad (41) \\
&= \quad P(O_{t+1}, O_{t+2}, ..., O_T | z_t = Z_i, \mu) \\
&= \quad \beta(z_t^i)
\end{aligned}
$$

therefore the Backward recurrence equation is specified as

$$\beta(z_t^i) \quad = \quad \sum_{j=1}^{n} a_{ij}b_j(O_{t+1})\beta(z_{t+1}^j) \qquad (42)$$

where $b_j(O_{t+1})$ is the probability of observation O_{t+1} when the Markov state is in state Z_j
In the next section, we develop the Hidden Semi-Markov Model (HSMM) which captures the relationships among the transition state, duration and observation sequence over time.

6 The Hidden Semi-Markov Model (HSMM)

A Hidden Semi-Markov Model (HSMM) is an extension of HMM by allowing the underlying state process to be a semi-Markov chain with variable duration for each state. Therefore in addition to the notation defined for the HMM in Equation (3.4), the duration D of a given state is explicitly defined for the HSMM. By state duration we mean the amount of time an HMM dwells in a state. State duration is a random variable and assumes an integer value in the set $D = \{d_1, d_2, ..., d_T\}$.
Unlike a state in HMM, a state in HSMM genarates a sequence of observation as opposed to a single observation in the HMM. The number of observations produced while in state i is determined by the length of time spent in state i, i.e the duration d. In Figure 6, the first state Z_1 and its duration d_1 are selected according to the transition probability $a_{(Z_0,d_0)(Z_1,d_1)}$ where (Z_0, d_0) is the initial state and duration Z_1 lasts for $d_1 \geq 0$ time units and produces two observations (O_1, O_2) according to emission probability $b_{Z_1,d_1}(O_{1:2})$. It transmits according to the transition probability $a_{(Z_1,d_1)(Z_2,d_2)}$ of state Z_2. Z_2 lasts for $d_2 = 4$ units which produces four observations (O_3, O_4, O_5, O_6) according to emission probability $b_{Z_1,d_2}(O_{3:6})$. The trend continues until the last state which may last beyond time T.

Suppose the current time is t, then we can define the state transition probability from state i having duration h to state j $\neq$ i having duration d by

$$a_{(i,h)(j,d)} \equiv P[Z_{(t+1:t+d)} = j | Z_{(t-h+1:t)} = i] \qquad (43)$$

which is assumed independent of time t for $i, j \in Z, h, d \in D$. and satisfy the condition

$$\sum_{j \in Z} \sum_{d \in D} a_{(i,h)(j,d)} = 1 \qquad (44)$$

for all given $i \in Z$ and $h \in D$ with zero self-transition probabilities i.e $a_{(i,h)(i,d)} = 0$, when a state ends at time t, it cannot transit to the same state at the next time t+1 because the state duration are explicitly specified by same duration other than geometric or exponential distributions.
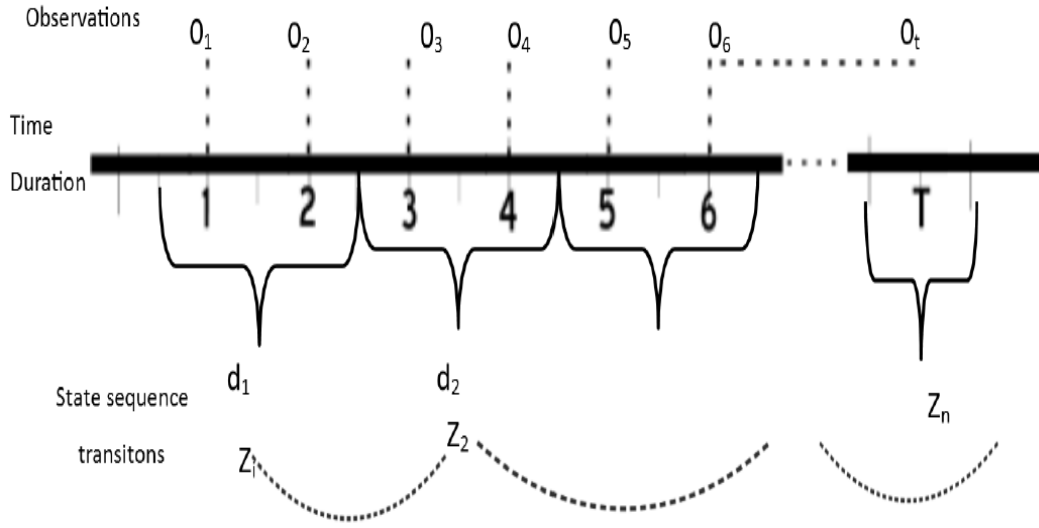
Fig. 5. Hidden semi-Markov Model

In Equation (43), the previous state i stated at time t-h + 1 and ended at time t with duration h. Then it transits to state j having duration d according to the state transition probability $a_{(i,h)(j,d)}$. State j will start at t+1 and end at t+d which means both state and the duration are dependent on both the previous state and its duration while in state j there will be d observations $O_{t+1:t+d}$ being emitted.

Let

$$b_{j,d}(O_{t+1:t+d}) \equiv P[O_{t+1:t+d}|Z_{[t+1:t+d]}]$$ (45)

be the observation probability which is assumed to be independent of t.
Let

$$\pi_{j,d} \equiv P[Z_{(1:d)} = j]$$ (46)

be the initial state distribution of the first state. Then the characterization of HSMM is through its parameters i.e., Initial state duration (π), transition probability (A), Observation probability (B) and state duration (D). Therefore, HSMM can be specified as

$$\lambda = (\pi, A, B, D)$$ (47)

In the next section, compute the likelihood of an observed sequence O given the model λ by incorporating time using the Forward-Backward algorithm i.e given the model and a sequence of observation, we want to evaluate how well the model predicts the observation sequence at a given time.

6.1 Forward -Backward Algorithm for HSMM

Let

$$\alpha_t(j, d) \equiv P[Z_{[t-d+1:t]} = j, O_{1,t}|\lambda]$$

be the forward variable and let

$$\beta_t(j, d) \equiv P[O_{t+1:T}|Z_{[t-d+1:t]} = j, \lambda]$$

be the backward variable.

Based on Markov property, the future observations are dependent on the current state i.e

$$P[O_{t-d+1:t}|Z_{[t-d+1:t]} = j, \lambda] = P[O_{t-d+t:t}|Z_{[t-d+t:t]} = j, \lambda] \tag{48}$$

and

$$P[O_{t-d+1:t}|Z_{[t-d+1:t]} = j, Z_{[t+1:t+h]} = i, \lambda] = P[O_{t+1:T}|Z_{[t+1:t+h]} = i, \lambda] \tag{49}$$

and independent of previous observations.

$$\begin{aligned}
P[Z_{[t-d+1:t]} = j, O_{t-d+1:t}|Z_{[t-d-h+1:t-d]} = i, \lambda] &= \\
P[Z_{[t-d+1:t]} = j, O_{t-d+1:t}|Z_{[t-d-h+1:t-d]} = i, \lambda]
\end{aligned} \tag{50}$$

and

$$P[O_{t+h+1:T}|Z_{[t+1:t+h]} = j, O_{t+1:t+h}, \lambda] = P[O_{t+h+1:T}|Z_{[t+1:t+h]} = i, \lambda] \tag{51}$$

Using Equations (49 - 51), we obtain the Forward-Backward algorithm as follows;

$$\begin{aligned}
\alpha_t(j, d) &= \sum_{i \neq j, h} P[Z_{[t-d-h+1:t+d]} = i, Z_{[t-d+1:t]} = j, O_{1:t}|\lambda] \\
&= \sum_{i \neq j, h} \alpha_{t-d}(i, h) P[Z_{[t-d+1:t]} = j, O_{t-d+1:t}|Z_{[t-d-h+1:t-d]} = i, \lambda] \\
&= \sum_{i \neq j, h} \alpha_{t-d}(i, h) a_{(i,h)(j,d)} P[O_{t-d+1:t} = j, O_{t-d+1:t} = j, d] \\
&= \sum_{i \neq j, h} \alpha_{t-d}(i, h) a_{(i,h)(j,d)} b_{j,d}(O_{t-d+1:t}) \quad for \quad t > 0, d \in D, j \in Z
\end{aligned} \tag{52}$$

$$\begin{aligned}
\beta_t(j, d) &= \sum_{i \neq j, h} P[Z_{[t+1:t+h]} = j, O_{t+1:T}|Z_{[t-d+1:t]} = j, \lambda] \\
&= \sum_{i \neq j, h} a_{(j,d)(i,h)} P[O_{t+1:t+h} = i, \lambda] \\
&= \sum_{i \neq j, h} a_{(j,d)(i,h)} b_{i,h}(O_{t+1,t+h}) P[O_{t+h+1:T}|Z_{[t+1:t+h]} = i, \lambda] \\
&= \sum_{i \neq j, h} a_{(j,d)(i,h)} b_{i,h}(O_{t+1:t+h}) \beta_{t+h}(i, h) \quad for \quad t < T, d \in D, j \in Z.
\end{aligned} \tag{53}$$

7 Durational Measure

In HMM the markov property implies that the value of the hidden state at time t+1 depends exclusively on its value of time t while in HSMM, the probability of transition from state Z_i to state Z_j at time t depends on the duration spent in state i prior to time t+1. Let n be the number of hidden states and individual state at time t as z_t, then the semi-markov property can be written as;

$$P(z_{t+1} = Z_i|z_t = Z_j, ..., z_1 = Z_l) = P(z_{t+1} = i|z_t = j, d_t(j)) \tag{54}$$

where $1 \leq i, j, l \leq n$ and duration variable $d_t(j)$ is defined as the time spent in state Z_j prior to time t.

Let state duration variable d_t be defined as;

$$d_t = \begin{cases} d_t & if \, z_t = Z_j \\ 1 & if \, z_t \neq Z_j \end{cases} \tag{55}$$

then the quantity $d_t(j)$ is calculated inductively from $d_{t-1}(j)$ as

$$d_t(j) = z_t(j).z_{t-1}(j).d_{t-1}(j+1)) \tag{56}$$

where

$$z_t(j) = \begin{cases} 1 & if \, z_t = Z_j \\ 0 & otherwise \end{cases} \tag{57}$$

If we assume that at time t the model is in state Z_i, then we define the duration-dependent matrix as;

$$Ad_t = [a_{ij}(d_t)] \tag{58}$$

where

$$a_{ij}(d_t) = P(z_{t+1} = Z_j | z_t = Z_i, d_t(i)) \quad 1 \leq i, j, l \leq n \tag{59}$$

The matrix Ad_t is then decomposed into two i.e. recurrent transition probabilities and non-recurrent transition probabilities.

7.1 Recurrent Transition Probabilities

Let

$$\begin{aligned}
P(O|A, \pi, q_1 = i) &= \frac{P(O, q_1 = i | A, \pi)}{P(q_1 = i)} \\
&= \frac{\pi_{a_{ii}}^{d-1}(1 - a_{ii})}{\pi} \\
&= a_{ii}^{d-1}(1 - a_{ii}) \\
&= P_{ii}(d)
\end{aligned} \tag{60}$$

where $P_{ii}(d)$ is the recurrent transition probability for the state to remain in the same state for d time instances given the model is in a known state.

Based on $P_{ii}(d)$ in equation (3.82), the expected number of duration is given by;

$$\begin{aligned}
E[d] = \sum_{d=1}^{\infty} d.P_{ii}(d) &= \sum_{d=1}^{\infty} d.a_{ii}^{d-1}(1 - a_{ii}) \\
&= (1 - a_{ii})\frac{d}{da_{ii}}(\sum_{d=1}^{\infty} a_{ii}^d) \\
&= (1 - a_{ii})\frac{d}{da_{ii}}(\frac{a_{ii}}{1 - a_{ii}}) \\
&= \frac{1}{1 - a_{ii}}
\end{aligned} \tag{61}$$

where

7.2   Nonrecurrent Transition Probabilities

The nonrecurrent state transition probabilities $A^\circ = [a_{ij}^\circ]$ is the transitions between two different states and is represented by a $n \times n$ matrix with the diagonal elements equal to zero, defined as

$$A^\circ = [a_{ij}^\circ] = \begin{cases} 0 & if\, i = j \\ P(q_{t+1} = Z_j | q_t = Z_i) & if\, i \neq j \end{cases} \tag{62}$$

where A° is a transition matrix.

8   Results and Discussion

8.1   Estimating the Transition Matrix
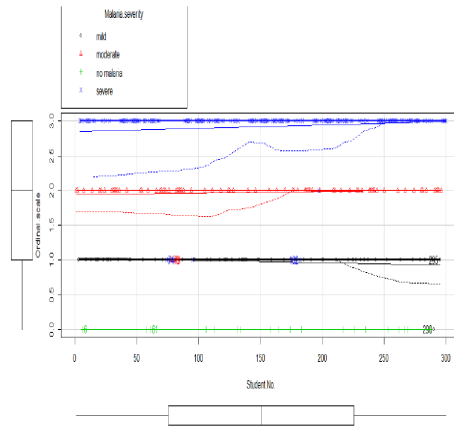
Fig. 6. A plot of observation symptom and Ordinal scale

The results of Figure 6 shows the coding of malaria symptoms to the ordinal scale. The results shows that majority of the students had moderate illness of malaria disease.

To estimate observation matrix, we used *hmmestimate* which corresponds to sequence of states that the model went through to generate sequence (seq). The command takes the emission (observation sequences), seq and states, and returns estimates of the transition and emission matrix. Using initial probability distribution values, we estimate the transition matrix and observation matrix with the help of *hmmtrain* command in MATLAB as shown in Appendix I. The optimal value obtained for the transition probability matrix A between states of malaria disease is shown below;

$$A = \begin{matrix} & \begin{matrix} Z_1 & \quad Z_2 & \quad Z_3 & \quad Z_4 \end{matrix} \\ \begin{matrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{matrix} & \begin{bmatrix} 0.0034 & 0.9966 & 0.0000 & 0.0000 \\ 0.1297 & 0.1511 & 0.7192 & 0,0000 \\ 0.0002 & 0.0645 & 0.4235 & 0.5118 \\ 0.0000 & 0.0542 & 0.1192 & 0.8266 \end{bmatrix} \end{matrix}$$

The results of matrix A shows that the probability of an individual remaining in infectious state after displaying malaria related symptoms is 83% ($a_{44} = 0.8266$) for the case of severe malaria, 42%

($a_{33} = 0.4235$) for the case of moderate malaria and 15% ($a_{22} = 0.1511$) for the case of mild malaria. The results also shows that for the individual student to remain in an healthy state after displaying malaria related symptoms is 0.3% ($a_{11} = 0.0034$). The probability of an individual transiting from healthy state to mild state of the disease is 99% ($a_{12} = 0.9966$), from mild state to moderate state is 72% ($a_{23} = 0.7192$) and from moderate state to severe state is 51% ($a_{34} = 0.5118$). The results also shows that there is 0% ($a_{41} = 0.0000$) transition from severe state of the disease to healthy state. Generally, the results shows that the probabilities lower than 0.1 in the transition matrix A represents a very weak transition whereas the probabilities higher than 0.4 represents a very high transition.

Using the transition matrix A and Equation (3.61), the model approximated the expected duration a student would be infectious after displaying the symptoms conditioned on the initial state;

(i) $a_{22} = \frac{1}{1-0.1511} = 1$ day for the case of mild state

(ii) $a_{33} = \frac{1}{1-0.4235} = 2$ days for the case of moderate state

(iii) $a_{44} = \frac{1}{1-0.8266} = 6$ days for the case of severe state

8.2   Estimating the Observation Matrix

The results in Figure 7 shows the number of observation sequence of symptoms as displayed by different students who visited the health facility. The results shows that each student displayed more than two symptoms with majority having between 4 to 8 symptoms.
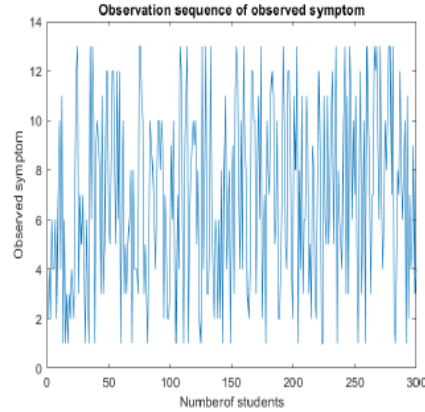

Fig. 7. Observation sequence of symptom

$$B = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ 0.1251 & 0.0723 & 0.0702 & 0.7033 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0260 & 0.0649 & 0.0909 & 0.1299 & 0.0260 & 0.1034 & 0.10390 & 0.0779 & 0.1039 \\ 0.0704 & 0.1549 & 0.1268 & 0.0845 & 0.0563 & 0.0563 & 0.0704 & 0.0563 & 0.0704 \\ 0.0286 & 0.0714 & 0.0857 & 0.1143 & 0.1143 & 0.0429 & 0.0571 & 0.1429 & 0.0571 \end{bmatrix}$$

The symptoms used in the study were; Let x_1 - fever (body temperature), x_2 - chills, x_3 - sweating, x_4 - vomiting, x_5 - diarrhea, x_6 - weakness, x_7 - pallor, x_8 - cough and x_9 - sneezing. The observation matrix B shows the emission (observation) probabilities and their relationship with the hidden state (status of the individual) of the model as provided by symptom dataset. Each emission probability

represent the chance of a particular observation, for instance, the results shows that there is 70% (x_4) chance of observing sweat in healthy state than in coma (0%).

8.3 Computing the Most Likely Sequence

To compute the most likely state to be observed after displaying the malaria related symptoms, we used the function *hmmviterbi* in MATLAB (Appendix I) which uses the Viterbi algorithm to compute the most likely sequence of state that the model would go through to generate the given sequence of observation. using the function;

$likelystates = hmmviterbi(seq, A, B)$

The results shows that the most likely state sequence is 2.

To test the accuracy of *hmmviterbi*, we compute the percentage of the time that the actual sequence states agrees with the sequence of observation by writing the function

$sum(states == likelystates)/300$

The results obtained from running this function in MATLAB is 0.8467 which shows that the most likely sequence of states agrees with the actual sequence by 85%.

8.4 Calculating Posterior Probability

The posterior state probabilities of an emission (observation) sequence are the conditional probabilities that the model is in a particular state when it generates a particular sequence. To compute the posterior state probabilities, we use the function as shown in Appendix II

$PSTATES = hmmdecode(seq, A_{EST}, B_{EST})$

The output $PSTATES$ is an n by T matrix, where n is the number of states and T is the length of sequence (seq). $PSTATES(i, j)$ is the conditional probability that the model is in state i when it generates the jth symbol of sequence. The actual probability of a sequence tends to 0 as the length of the sequence increases, therefore we use the function *hmmdecode* which gives the logarithm of the probability.

$[PSTATES, logpseq] = hmmdecode([13], A_{EST}, B_{EST});$

$exp(logpseq)$

The results of $PSTATES$ is -2.5649 and its probability is 0.0769 which is the logarithm of the probability.

8.5 Calculating the Forward probability

The forward algorithm evaluates how well the model predicts the given observation sequence. Using the MATLAB function shown in Appendix II, the results shows that the model is able to predict 87% (0.8719) of the observation sequence

8.6 Parameter Estimates

Using the training and test data in MATLAB (Appendix III), we estimates the parameters of the model as shown in Figure 8, Figure 9, Figure 10 and Figure 11;
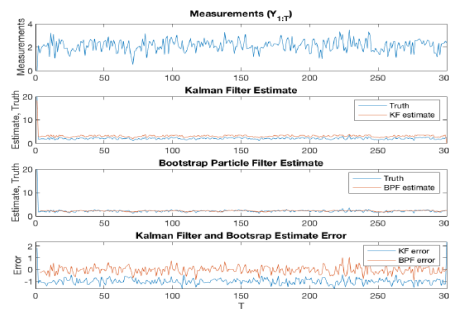
The results shows that the measurement are not far much removed from the hidden states i.e.
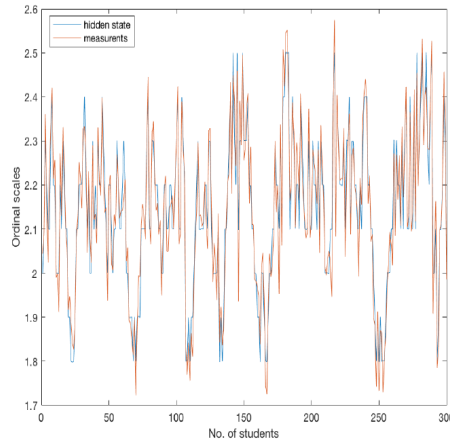
Fig. 8. A plot of ordinal scale measurement and Hidden State

measurements have small error or noise indicating high precision.

The results shows that the filter estimates are close to the true values and the error is small, this is


**Fig. 9. A plot of Kalman Filter estimates, Boostrap Particle Filter estimates and Error**

also true for sequential Monte Carlo (Boostrap filter particle)

The results shows that the Kalman Filter produces an optimal estimate which is shown by the lowest variance of the two filters. The more the number of samples in the Monte Carlo filter,the lower the variance.

The results shows that the filter estimates for parameter β is 0.1020. The actual value of β is 0.1 but the estimate value is 0.1020 which shows that the filter estimates is not far much removed from the actual value. Therefore using measurements only, we can obtain estimate of β which can then be used in the model equation to provide a fit for the data and any other data.
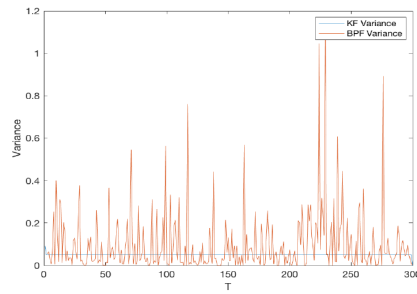
Fig. 10. A plot of Kalman Filter and Monte Carlo variances
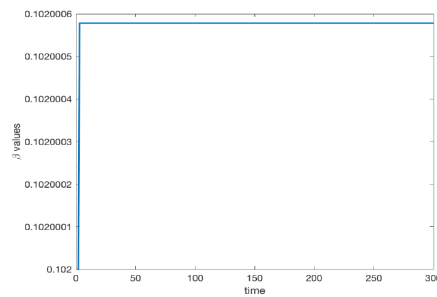
Fig. 11. A plot of estimate of β

Acknowledgement

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria-a review. Malaria Journal. 2011;10:202.

[2] WHO. Malaria Fact Sheet 2017 Report; 2017.

[3] WHO. World Malaria Report; 2012.

[4] Martins, et al. Clustering symptoms of non-severe malaria in semi-immune Amazonian patients. Peer J; 2015.

[5] Dondorp AM, Day NP. The treatment of severe malaria. Trans R Soc Trop Med Hyg. 2007;101:633-634.

[6] Cholewa M, Gomb P. Estimation of the number of states for gesture recognition with hidden markov models based on the number of critical points in time sequence. Recognition Letters. 2013;34(5):574-9.

[7] Farsi H, Saleh R. Implementation and optimization of a speech recognition system based on hidden Markov model using genetic algorithm. Intelligent Systems (ICIS), 2014 Iranian Conference on. IEEE; 2014.

[8] Vimala K. Stress causing arrhythmia detection from ECG signal using HMM. International Journal of Innovative Research in Computer and Communication Engineering. 2014;2(10):6079-85.

[9] Li HM FL Y, Wang P, Yan JZ. Hidden markvo models based research on lung cancer progress modeling. Research Journal of Applied Sciences, Engineering and Technology. 2013;6(13):2470-3.

[10] Lee HK, Lee J, Kim H, Ha JY, Lee KJ. Snoring detection using a piezo snoring sensor based on hidden Markov models. Physiological Measurement. 2013;34(5):41-45.

[11] Liu Ying, Shuang Li, Fuxin Li, Le Song, James M Rehg, Efficient learning of Continuous-time hidden markov models for disease progression. In Advances in Neural Information Processing Systems. 2015;3600-3608.

[12] Barber C, Bockhorst J, Roebber P. Auto-regressive HMM inference with incomplete data for short-horizon wind forecasting. Adv.Neural Inf Process Syst; 2010.

[13] Wu H, Rojai J, Lin H, Harada K. Introspection with bayesian non-parametric vector autoregressive hidden markov model; 2017.

[14] Tuncel KS, Baydogan MG. Autoregressive forest for multivariate time series modeling. Pattern recognition. 2018;73:202-215.

[15] Ferguson JD. Variable duration models for speech in proceeding of the symposium on the application of hmm to text and speech ed. J.D Ferguson, Princeton NJ. 1980;143-179.

[16] Rabinner LR. A tutorial on hidden markov model and selected application in speech recognition. Proceeding of the IEEE. 1989;77:257-285.

[17] Xu J, Zeger S. Joint analysis of longitudinal data comprising repeated measures and time to events. Journal of the Royal Statistical Society series C. Applied statistics 2001;50:375-87.

[18] Zammit, Nicola N, George Streftaris, Gavin J. Gibson, Ian J. Deary, Brian M. Frier. Modelling the consistency of hypoglycaemic symptoms:High variability in diabetes. Diabetes Technology and Therapeutics. 2011;13(5)571-578.

[19] Xing Z, Nicholson B, Jimenez M, Velderman T, Hudson L, Lucas J, Duson D, Zaas A, Woods C, Geofrey G, Carin L. Bayesian modeling of temporal properties of infectious disease in college student population. Journal of Applied Statistics. 2014;41(6):1358-1382.

[20] Masinde Muliro University of Science and Technology Student Health Facility Records; 2018.

APPENDIX I: COMPUTATION OF TRANSITION AND OBSERVATION MATRIX

pi=[0.25 0.25 0.25 0.25]

T=13

N=4

A = [0.25 0.25 0.25 0.25;0.25 0.25 0.25 0.25;0.25 0.25 0.25 0.25; 0.25 0.25 0.25 0.25];

B = [1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13;

1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13; 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13; 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13 1/13];

$[seq, states] = hmmgenerate(300, A, B);$

$likelystates = hmmviterbi(seq, A, B);$

$sum(states == likelystates)/300$

$[A_{EST}, B_{EST}] = hmmestimate(seq, states)$

$[A_{EST2}, B_{EST2}] = hmmtrain(seq, A_{EST}, B_{EST})$

$hmmtrain(seq, A_{EST}, B_{EST},' maxiterations', maxiter)$

$hmmtrain(seq, A_{EST}, B_{EST},' tolerance', tol)$

$PSTATES = hmmdecode(seq, A_{EST}, B_{EST})$

$[PSTATES, logpseq] = hmmdecode(seq, A_{EST}, B_{EST})$

$[PSTATES, logpseq] = hmmdecode([13], A_{EST}, B_{EST});$

$exp(logpseq)$

APPENDIX II: FORWARD ALGORITHM

initialization

T=13

N=4

A=[0.0034 0.9966 0.0000 0.0000; 0.1297 0.1511 0.7192 0.0000; 0.0002 0.0645 0.4235 0.5118; 0.0000 0.0542 0.1192 0.8266]

B=[0.0366 0.0732 0.0488 0.0732 0.0976 0.1098 0.1341 0.0610 0.0854 0.0244 0.0488 0.0854 0.1220; 0.0260 0.0649 0.0909 0.1299 0.0260 0.1039 0.0390 0.1039 0.1169 0.0779 0.0779 0.0390 0.1039; 0.0704 0.1549 0.1268 0.0845 0.0563 0.0563 0.0704 0.1268 0.0282 0.0563 0.0704 0.0704 0.0282;0.0286 0.0714 0.0857 0.1143 0.0429 0.1286 0.0714 0.0429 0.1143 0.0429 0.0571 0.1429 0.0571];

pi=[0.2500 0.2500 0.2500 0.2500]

O=[1 1 1 1 1 1 1 1 1 1 1 1 1]

for i=1:N

alpha(1,i)=pi(i)*B(i,O(1));

end

induction

for t=1:T-1

for j=1:N

u=0.8;

for i=1:N

u=u+alpha(t,i)*A(i,j);

end

alpha(t+1,j)=u*B(j,O(t+1));

end

end

Evaluatingtheprobability

p=0.8;

for i=1:N

p=p+alpha(T,i);

end

APPENDIX III: PARAMETER ESTIMATES USING MATLAB FUNCTIONS

```
[x] = csvread('Ordinalscale.csv');
Filters
beta = 0.1;
C = 1;
Q = sqrt(0.1);
R = sqrt(0.1);
rng(1);
T = 300;
X = zeros(1, T);
Y = zeros(1, T);
X_0 = Age(1);
X(1) = Age(1);
for t = 2 : T
X(t) = beta * Age(t - 1) + Q * randn;
Y(t) = C * X(t) + R * randn;
end
X_hat = zeros(1, T);
Xpred_hat = zeros(1, T);
P_hat = zeros(1, T);
Ppred_hat = zeros(1, T);
Xpred_hat(1) = X_0;
P_0 = 1;
Ppred_hat(1) = P_0;
for t = 2 : T
correction step
K = Ppred_hat(t - 1) * C'/(C * Ppred_hat(t - 1) * C' + R^2);
X_hat(t - 1) = Xpred_hat(t - 1) + K * (Y(t - 1) - C * beta * Xpred_hat(t - 1));
P_hat(t - 1) = Ppred_hat(t - 1) - K * C * Ppred_hat(t - 1);
prediction step
Xpred_hat(t) = beta * Age(t - 1);
Ppred_hat(t) = beta * P_hat(t - 1) * beta' + Q^2;
end
M = 1000;
XX_hat = zeros(M, T);
XX_hat(:, 1) = X_0 * ones(M, 1);
XX_mean = zeros(1, T);
XX_mean(1) = mean(XX_hat(:, 1));
weights = ones(M, 1)/M;
e = ones(1, M);
for t = 2 : T
[ indx ] = resample Multinomial(weights);
for j = 1 : M
XX_hat(j, t - 1) = XX_hat(indx(j), t - 1);
end
XX_{hat(:,t)} = beta * Age(t - 1) * ones(M, 1) + Q * randn(M, 1);
weightsU = normpdf(Y(t - 1), C * XX_{hat(:,t-1)}, R);
weightsU = exp(weightsU -max(weightsU));
weights = weightsU/sum(weightsU);
XX_{mean(t)} = weights' * XX_{hat(:,t)};
figure(1)
plot(1 : t, X(1 : t), 1 : t, XX_mean(1 : t));
drawnow
end
figure(1)
subplot(4, 1, 1);
plot(1 : T, Y(1 : T));
title('Measurements(Y_{1:T})')
```

```
axis([1 T -1.5 1.5])
ylabel('Measurements')
subplot(4, 1, 2);
plot(1 : T, X(1 : T), 1 : T, X_hat(1 : T));
title('KalmanFilterEstimate')
legend('Truth',' KFestimate')
axis([1 T -1.5 1.5])
ylabel('Estimate, Truth')
subplot(4, 1, 3);
plot(1 : T, X(1 : T), 1 : T, XX_{mean(1:T)});
title('BootstrapParticleFilterEstimate')
legend('Truth',' BPFestimate')
axis([1 T -1.5 1.5])
ylabel('Estimate, Truth')
subplot(4, 1, 4);
plot(1 : T, X − X_hat, 1 : T, X − XX_mean);
title('KalmanFilterandBootsrapEstimateError')
legend('KFerror',' BPFerror')
axis([1 T -1 1])
xlabel('T'); ylabel('Error')
figure(2)
plot(1 : T, P_{hat}, 1 : T, (X − XX_{mean}).^2);
title('KalmanFilterandBootsrapVariances')
legend('KFVariance',' BPFVariance')
axis([1 T 0 1.5])
xlabel('T'); ylabel('Variance')
parametere stmation
X_0 = Numberofsymptoms(1);
initial_{beta} = 0.102;
Y = Numberofsymptoms;
XX_{hat} = zeros(2, T);
XXpred_{hat} = zeros(2, T);
PP_{hat} = zeros(2, 2, T);
PPpred_{hat} = zeros(2, T);
XXpred_hat(1 : 2, 1) = [X_0; initial_{beta}];
P_0 = XXpred_{hat}(1 : 2, 1) ∗ XXpred_{hat(1:2,1)'};
PPpred_{hat(1:2,1:2,1)} = P_0;
bbeta = zeros(1, T);
bbeta(1) = initial_{beta};
for t = 1 : T
prediction step
XXpred_hat(1 : 2, t) = [bbeta(t) ∗ Age(t); bbeta(t)];
PPpred_hat(1 : 2, 1 : 2, t) = bbeta(t) ∗ PP_hat(1 : 2, 1 : 2, t) ∗ bbeta(t)' + Q^2 ∗ [1, 0; 0, 0.0000001];
correction step
K = PPpred_hat(1 : 2, t) ∗ C'/(C ∗ PPpred_hat(1, t) ∗ C' + R^2);
XX_hat(1 : 2, t + 1) = XXpred_hat(1 : 2, t) + K ∗ (Y(t) − C ∗ bbeta(t) ∗ XXpred_hat(1, t));
PP_hat(1 : 2, 1 : 2, t + 1) = PPpred_hat(1 : 2, 1 : 2, t) − (K ∗ ones(1, 2)) ∗ C ∗ PPpred_{hat(1:2,1:2,t)};
bbeta(t + 1) = XX_hat(2, t + 1);
plot(1 : t, bbeta(1 : t))
end
figure(1)
plot(1 : t, bbeta(1 : t),' LineWidth', 2)
title('Estimatesofbeta')
xlabel('time')
ylabel('βvalues')
Resample multinomial
function[indx] = resampleMultinomial(w)
M = length(w);
```

```
Q = cumsum(w);
Q(M) = 1;
i = 1;
while(i <= M),
sample = rand(1, 1);
j = 1;
while(Q(j) < sample),
j = j + 1;
end;
indx(i) = j;
i = i + 1;
end
simulation
beta = 0.1;
T = 300;
x = Age';
z = zeros(T, 1)';
y = zeros(T, 1)';
R = 0.05;
Q = 0.001;
fork = 1 : 300
z(k) = x(k) * beta + Q * randn;
y(k) = z(k) + R * randn;
plot(1 : k, z(1 : k), 1 : k, y(1 : k));
end
figure(1)
plot(1 : k, z(1 : k), 1 : k, y(1 : k));
title('Aplotofmeasurementsandthehiddenstate')
legend('hiddenstate', 'measurents', 'Location', 'best')
xlabel('No.ofstudents')
ylabel('Ordinalscales')
X_hat = zeros(1, T);
Xpred_hat = zeros(1, T);
P_hat = zeros(1, T);
Ppred_hat = zeros(1, T);
Xpred_hat(1) = x(1);
P_0 = 1;
Ppred_hat(1) = Q;
C = 1;
fort = 2 : T
correction step
K = Ppred_hat(t - 1) * C'/(C * Ppred_hat(t - 1) * C' + R^2);
X_hat(t - 1) = Xpred_hat(t - 1) + K * (y(t - 1) - C * Xpred_hat(t - 1));
P_hat(t - 1) = Ppred_hat(t - 1) - K * C * Ppred_hat(t - 1);
prediction step
Xpred_hat(t) = beta * X_hat(t - 1);
Ppred_hat(t) = beta * P_hat(t - 1) * beta' + Q^2;
figure(2)
plot(1 : t, x(1 : t), 1 : t, X_hat(1 : t));
end
M = 1000;
xx = zeros(M, T);
xx(:, 1) = x(1) * ones(M, 1) + 1 * randn(M, 1);
xxmean = zeros(1, T);
xxmean(1) = mean(xx(:, 1));
weights = ones(M, 1)/M;
M_eff = 1/sum(weights.^2);
e = ones(1, M);
fort = 2 : T
```

Resample (multinomial)
$if M_eff < 0.7 * M$
$indx = randsample(M, M, true, weights);$
$for j = 1 : M$
$xx(j, t - 1) = xx(indx(j), t - 1);$
end
end
propagate the particles
$rd = randn(M, 1); rd = rd - mean(rd);$
$xx(:, t) = beta * xx(:, t - 1) + Q * rd;$
$weights = weights. * exp(-((xx(:, t - 1)).^2/2 - xx(:, t - 1) * y(t))/R);$
$weights = weights/sum(weights);$
$M_eff = 1/sum(weights.^2);$
$xxmean(t) = weights' * xx(:, t);$
$figure(3)$
$plot((1 : t), x(1 : t), (1 : t), xxmean(1 : t));$
$drawnow$
end