# FORMULAS FOR PARAMETER ESTIMATES FOR TWO-STATE HIDDEN MARKOV MODELS

JUNKO MURAKAMI

Abstract. This paper considers simple discrete-time hidden Markov models in which both the state and observation spaces are of size two, say $\{0, 1\}$. Here, the exact formula for Bayesian posterior mean of the parameters is presented, which is constructed with a particular type of models in mind; i.e., the models for which one of the two values of the observation sequence (0 in this paper) is dominant than the other; though it also works for other type of models, too, of course. The formula has been vigorously verified empirically, using our own algorithm.

## 1. Introduction

Hidden Markov models (HMMs) consist of a Markov chain of 'hidden' states and an output sequence that can be observed and statistically depends on the states. The models are widely used in diverse areas; for example, traffic modeling (Heffes & Lucantoni, 1986; Wei et al. , 2002; Heyman & Lucantoni, 2003; Scott & Smyth, 2003; Nogueira et al., 2004), bioinformatics (Pavlovic et al., 2002; Sinha et al. , 2003; Cooper & Lipsitch, 2004), event detection (Jana & Dey, 2000), inventory control (Ching, 1997), seismology (Oliveros et al., 2008), and wind time series (Ailliot et al., 2006), to name a few.

One of our previous algorithms, the first version of the algorithm, considers a discrete-time time-homogeneous hidden Markov models, for which both the state and observation spaces are discrete and of any finite size. It finds the exact value of Bayesian posterior mean (BPM) parameter estimates in a polynomial time with respect to the sequence length (Murakami & Taylor, 2006). Another algorithm, which is an extension of the first one, deals with Poisson hidden Markov models (Murakami, 2009). In this paper, we simply present an exact expression for the BPM, which we empirically verified using the algorithms mentioned above. It is obvious the formula can be modified to support higher dimensions of the state and/or observation space than two; however, it is beyond the scope of this paper.

However, this paper does not suggest any practical use of the formula (although there could be some), nor claim that this formula would help speed up the process of finding the Bayesian posterior mean (BPM) parameter estimates. We simply hope it would provide some deeper understanding of HMMs.

The HMM we consider in this paper is defined in Section 2, a basic but not too detailed expression for the BPM is shown in Section 3, and identifiability issues of HMMs that we consider in this paper is briefly explained in Section 4.

To find the BPM, the most time-consuming task is to find the frequency for each possible combination of certain 'counts' (for example, one of these counts is the number of times 1 is observed when the state is 0) that we obtain by considering all the possible Markov chain realizations. The notations and definitions of the quantities that we use to find these frequencies are specified in Section 5; and two functions that we use to compute these frequencies is described in Section 6. The formula is finally described in Section 7, and the conclusions are in Section 8.

As for appendices, the derivation and the details of the BPM formula is given in Appendix A, and the actual ranges of the quantities introduced in Section 5 are shown in Appendix B.

## 2. Hidden Markov Models

Consider a time-homogeneous discrete-time hidden Markov model consists of a Markov chain that we cannot observe $S^{1:n} = (S_1,\ S_2,\ S_3,\ \ldots, S_n)$, $S_i \in \zeta_X$, and an observation sequence $O^{1:n} = (O_1,\ O_2,\ O_3,\ \ldots, O_n)$, $O_u \in \zeta_Y$, for some positive integer $n$ and for $\zeta_X = \zeta_Y = \{0,1\}$. Also, let the transition matrix be $A = \{a_{ij}\}_{i,j\in\zeta_X}$, the emission matrix be $B = \{b_{iu}\}_{i\in\zeta_X,u\in\zeta_Y}$, and $\Pi = \{r_i\}_{i\in\zeta_X}$ be the initial probabilities so that $a_{ij} = P\left(S_{t+1} = j \mid S_t = i\right)$ for $t = 1, 2, \ldots, n-1$, while $b_{iu} = P\left(O_t = u \mid S_t = i\right)$ and $r_i = P\left(S_1 = i\right)$ for $t = 1, 2, \ldots, n$.

Let $\theta = \{A, B, \Pi\}$ be the parameter set, where $\Pi = \{r_i\}_{i\in\zeta_X}$. Then, by the independence in the conditional probabilities of the HMMs, the probability (or the likelihood) of $S^{1:n}$ and $O^{1:n}$ given a parameter set $\theta$ is a product of elements of $\theta$ as shown in below.

$$P(S^{1:n}, O^{1:n} \mid \theta) = r_{_{S_1}} b_{_{S_1\,O_1}} a_{_{S_1\,S_2}} b_{_{S_2\,O_2}} \cdots b_{_{S_{n-1}\,O_{n-1}}} a_{_{S_{n-1}\,S_n}} b_{_{S_n\,O_n}}$$

Putting the factors with the same base together, we can express the above also as

$$P(S^{1:n}, O^{1:n} \mid \theta) = r_{_{S_1}} \prod_{i,j\,\in\,\zeta_X} a_{ij}^{k_{ij}\left(S^{1:n}\right)} \cdot \prod_{\substack{i\,\in\,\zeta_X \\ u\,\in\,\zeta_Y}} b_{iu}^{l_{iu}\left(S^{1:n},\,O^{1:n}\right)} \tag{1}$$

where $k_{ij}$ and $l_{iu}$ are the counts such that

$$k_{ij}\left(S^{1:n}\right) = \sum_{t=1}^{n-1} \mathbf{I}\{S_t = i \text{ and } S_{t+1} = j\}$$

$$l_{iu}\left(S^{1:n},\, O^{1:n}\right) = \sum_{t=1}^{n} \mathbf{I}\{S_t = i \text{ and } O_t = u\}.$$

Let $K_n = \{k_{ij}\}_{i,j\in\zeta_X}$ and $L_n = \{l_{iu}\}_{i\in\zeta_X,u\in\zeta_Y}$. It is clear to see that the counts $K_n$ and $L_n$ are complete data sufficient statistics.

## 3. Basics of BPM Estimates

The BPM of the parameter set $\theta$ is the expected value of $\theta$ given an observation sequence $O^{1:n}$. Denote the range of $\theta = \{A, B, \Pi\}$ as $\Theta = \{\mathcal{R}_A, \mathcal{R}_B, \mathcal{R}_\Pi\}$ where $A$ and $B$ are $2 \times 2$ probability matrices under the restriction $a_{00} \geq a_{11}$, to avoid 'averaging up' the symmetry (see Section 4), and $\Pi$ is a length-2 probability vector; while $\mathcal{R}_A$, $\mathcal{R}_B$, and $\mathcal{R}_\Pi$ are the natural domain of $A$, $B$, and $\Pi$, respectively.

Let $K_n$ and $L_n$ be matrices such that $K_n = \{k_{ij}\}$ and $L_n = \{l_{iu}\}$, $i, j, u, \in \{0, 1\}$. Then, the Bayesian posterior mean is, taking the marginal distribution with respect to the underlying Markov chain $S^{1:n}$,

$$\hat{\theta} = \int_{\theta\in\Theta} \theta P\left(\theta \mid O^{1:n}\right) d\theta = \frac{\sum_{S^{1:n}\in\Omega_n} \int_{\theta\in\Theta} \theta P\left(S^{1:n}, O^{1:n} \mid \theta\right) P(\theta)\, d\theta}{\sum_{S^{1:n}\in\Omega_n} \int_{\theta\in\Theta} P\left(S^{1:n}, O^{1:n} \mid \theta\right) P(\theta)\, d\theta}$$

where $\Omega_n$ is the set of all possible Markov chain realizations of length $n$ (see Appendix A for more details). But, from (1) we see that $P\left(S^{1:n}, O^{1:n} \mid \theta\right)$ can be written as a function of $K_n = K_n(S^{1:n})$ and $L_n = L_n(S^{1:n}, O^{1:n})$. So, define $f$ as

$$f\left(K_n, L_n, \theta\right) = f\left(K_n(S^{1:n}), L_n(S^{1:n}, O^{1:n}), \theta\right) = P\left(S^{1:n}, O^{1:n} \mid \theta\right).$$

Now, we can rewrite the estimate $\hat{\theta}$ above as

$$\hat{\theta} = \frac{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \int_{\theta\in\Theta} \theta f(K_n, L_n, \theta) P(\theta)\, d\theta}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \int_{\theta\in\Theta} f(K_n, L_n, \theta) P(\theta)\, d\theta}. \tag{2}$$

where $\widehat{\Omega}_n$ is the domain of $(K_n, L_n)$ given an observation sequence $O^{1:n}$, and $\mathcal{N}_n(K_n, L_n)$ is the frequency of each $(K_n, L_n)$. Our algorithm is used to find this frequency $\mathcal{N}_n(K_n, L_n)$. For the exact formula for the integrations that appear in (2), see Appendix A.

## 4. Identifiability of HMMs

As for identifiability problem of HMMs, in this paper we only consider so-called 'label-switching' problem, which becomes an issue especially when the estimate is the BPM. It is due to the symmetries in the probability distribution that exist because any permutation of the labels of the states would not change the likelihood of the observation sequence when the prior is set to be symmetric. Many discussions have been made and various methods have been proposed, usually through the context of HMMs as an extension of mixture models (Celeux et al., 2000; Stephens, 2000; Capppé et al., 2005); however, it is beyond the scope of this paper. Here we simply order the diagonal elements of the transition matrix, which alters the otherwise symmetric prior (see Appendix A).

Table 1. Example of Sequences

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $O^{1:11}$ | 0 | 0 | 1 | **0** | **0** | **0** | **1** | 1 | 0 | 1 |
| $Y_p$ | $Y_1 = 3$ | | | $\mathbf{Y_2 = 4}$ | | | | $Y_3 = 1$ | $Y_4 = 2$ | |
| $S^{1:11}$ | 1 | 0 | **0** | **1** | **1** | **0** | **0** | 1 | 1 | 0 |
| | ($\leftarrow n_1 = 0$ ) | | $\uparrow$ $\mathbf{n_2{=}3}$ | | | | $\uparrow$ $n_3{=}7$ | $\uparrow$ $n_4{=}8$ | | $\uparrow$ $n_5{=}10$ |

## 5. Notations for the Quantities That We Use to Find the Frequency of Each Count $(K_n, L_n)$

Consider separating an observation sequence of 0's and 1's as a sequence of subsequences, each of which consists of consecutive 0's followed by a single 1 or a single 1 with no preceding 0. We consider a sequence $(Y_1, Y_2, \ldots, Y_m)$, where $Y_p$, $p = 1, 2, \ldots, m$, is the length of $p$th sub-interval, and $m$ is the total number of 1's in $O^{1:n}$. For example, if $O^{1:4} = (1, 0, 0, 1)$ then we get $(Y_1, Y_2) = (1, 3)$, and if $O^{1:10} = (0, 0, 1, 0, 0, 0, 1, 1, 0, 1)$ then we get $(Y_1, Y_2, Y_3, Y_4) = (3, 4, 1, 2)$ as shown in Figure 1.

Here, for the simplicity of the algorithm, we only consider observation sequences $O^{1:n}$ that end with 1. Furthermore, define $n_p$, as the time when $O_t = 1$ happens for the $p$-th time; in other words, $O_{n_p} = 1$ and the cardinality of the set $\{t : O_t = 1, t \leq n_p\}$ is $p$.

While our previous algorithm finds the frequency $\mathcal{N}_n(K_n, L_n)$ by sequentially processing $O_t$ for each $t = 1, 2, \ldots, n$, the formula here consider $Y_p$ for each $p = 1, 2, \ldots, m$, favoring the observation sequences of dominant 0's and spread-out 1's.

$$n_p = \begin{cases} 0 & \text{if } p = 1 \\ \displaystyle\sum_{j=1}^{p-1} Y_j & \text{if } p > 1 \end{cases}.$$

(Note, by the definition, $n_p + Y_p = n_{p+1}$.) Now, let $\delta_p$ be an indicator such that it is 0 if the number of jumps in the $p$-th interval, which is defined more precisely in below, is even and 1 if it is odd; i.e.,

$$\delta_p = \begin{cases} 0 & \text{if } p = 1 \text{ and } \displaystyle\sum_{t=1}^{Y_1 - 1} \mathbf{I}\{S_t \neq S_{t+1}\} \text{ is even,} \\ 0 & \text{if } p > 1 \text{ and } \displaystyle\sum_{t=n_p}^{n_p + Y_p - 1} \mathbf{I}\{S_t \neq S_{t+1}\} \text{ is even, and} \\ 1 & \text{if otherwise.} \end{cases}$$

For example, in Figure 1, consider the second interval and its previous state $S^{3:7}$ $= (S_3\ \ S_4\ S_5\ S_6\ S_7) = (0\ 1100)$. Since jumps are at $S_3 = 0 \neq S_4 = 1$ and $S_5 = 1$ $\neq S_6 = 0$, the state jumps twice (even times); hence, $\delta_2 = 0$.

In addition, let $\alpha_p$ be the number of times the state changes from 0 to 1 ; let $\beta_p$ be 1 if the state is 1 when the $p$th event occurs (or when $t = n_p + Y_p = n_{p+1}$), and 0 otherwise, which makes $\beta_p = S_{n_p+Y_p}$; and let $\gamma_p$ be the total number of times the state is 1 in the $p$-th subsequence. In other words, for $p = 1, 2, \ldots, m$, let

$$\alpha_p = \begin{cases} \displaystyle\sum_{t=1}^{Y_1-1} \mathbf{I}\{S_t = 0 \text{ and } S_{t+1} = 1\} & \text{if } p = 1 \\ \displaystyle\sum_{t=n_p}^{n_p+Y_p-1} \mathbf{I}\{S_t = 0 \text{ and } S_{t+1} = 1\} & \text{if } p > 1 \end{cases},$$

$$\beta_p = \mathbf{I}\{S_{n_p+Y_p} = 1\} = S_{n_p+Y_p}, \quad \text{and} \quad \gamma_p = \sum_{t=n_p+1}^{n_p+Y_p} \mathbf{I}\{S_t = 1\}.$$

For example, in Figure 1, $\alpha_2 = 1$ since $S_3 = 0$ and $S_4 = 1$ in $S^{3:7}$, $\beta_2 = 0$ since $S_7 = 0$ when $O_7 = 1$, and $\gamma_2 = 2$ since $S_t = 1$ twice at $t = 4$ and 5 in $S^{4:7}$.

Furthermore, as short-hand notations, define a cumulative indicator $\delta^{(p)}$ and cumulative counts $\alpha^{(p)}$, $\beta^{(p)}$, and $\gamma^{(p)}$ as shown below.

$$\delta^{(p)} = \begin{cases} 0 & \text{if } \displaystyle\sum_{i=1}^{p} \delta_i \text{ is even} \\ 1 & \text{if otherwise} \end{cases},$$

$$\alpha^{(p)} = \sum_{i=1}^{p} \alpha_i, \qquad \beta^{(p)} = \sum_{i=1}^{p} \beta_i, \qquad \text{and} \qquad \gamma^{(p)} = \sum_{i=1}^{p} \gamma_i.$$

So, when $p = m$ (= the total number of events), $\delta^{(m)}$ is the indicator whether the number of state changes in $S^{1:n}$ is even or odd; $\alpha^{(m)} = k_{01}$; $\beta^{(m)} = l_{11}$; and $\gamma^{(m)}$ is the total number of 1's in $S^{1:n}$.

Now, a simple observation shows, if $S_1 = 0$ then $S_{n_p+Y_p}$ = the state when the $p$th event occurs, $p = 1, 2, \ldots, m$, is same as $\delta^{(p)}$ (so that $S_n = \delta^{(m)}$), and $\alpha^{(m)}$ is equal to the total number of strings of consecutive 1's in $S^{1:n}$. Also, again if

$S_1 = 0$, we have the following:

$$
\begin{aligned}
k_{01} &= \alpha^{(m)} \\
k_{10} &= \alpha^{(m)} - \delta^{(m)} \\
k_{11} &= \gamma^{(m)} - \alpha^{(m)} \\
k_{00} &= n - 1 - \alpha^{(m)} - \gamma^{(m)} + \delta^{(m)} \\
l_{01} &= m - \beta^{(m)} \\
l_{10} &= \gamma^{(m)} - \beta^{(m)} \\
l_{11} &= \beta^{(m)} \\
l_{00} &= n - m - \gamma^{(m)} + \beta^{(m)}
\end{aligned}
$$

For $p = 1, 2, \ldots, m$, let $V_p = (\delta_1,\ \delta_2,\ \ldots, \delta_p,\ \alpha_1,\ \alpha_2,\ \ldots, \alpha_p,\ \gamma_1,\ \gamma_2,\ \ldots, \gamma_p)$, let $\nu_p = \big(\delta^{(p)},\ \alpha^{(p)},\ \beta^{(p)},\ \gamma^{(p)}\big)$, and let $\mathcal{F}_p(\nu_p)$ be the number of distinct Markov chain realizations $S^{1:n_p+Y_p}$ that produce the given $\nu_p$-value. In particular, when $p = m$, we have $\nu_m = \big(\delta^{(m)},\ \alpha^{(m)},\ \beta^{(m)},\ \gamma^{(m)}\big) = (S_n,\ k_{01},\ l_{11},\ \gamma^{(m)})$, and $\mathcal{F}_m(\nu_m)$ is the number of distinct Markov chain realizations $S^{1:n}$ that produce the given $\nu_m$-value.

As we can see from the equations for conversion, there is a one-to-one correspondence between $\nu_m$ and $(K_n, L_n)$ so that $\mathcal{F}_m(\nu_m)$ is equal to $\mathcal{N}_n(K_n, L_n)$ if $\nu_m$ corresponds to $(K_n, L_n)$, which we express as $\mathcal{N}_n(K_n, L_n) = \mathcal{F}_m(\nu_m)$. The reason we use $\nu_m$ instead of $(K_n, L_n)$ is that we want the algorithm to process the sequence not $n$-unit-time-wise but $m$-event-wise.

Furthermore, as for the case $S_1 = 1$, by simply interchanging the states, 0 and 1, we see

$$
\begin{aligned}
\mathcal{F}_m(\nu_m) &= \mathcal{N}_n(k_{00}, k_{01}, k_{10}, k_{11}, l_{00}, l_{01}, l_{10}, l_{11}) \\
&= \mathcal{N}_n(k_{11}, k_{10}, k_{01}, k_{00}, l_{10}, l_{11}, l_{00}, l_{01})
\end{aligned}
$$

because the frequency does not depend on how we label the hidden states. This is why we can fix $S_1 = 0$, and also why the probability distribution for the likelihood function is symmetric if the prior is symmetric.

Now, in order to find $\mathcal{F}_m(\nu_m)$, we consider all the valid values of a vector $V_m$ (i.e., the ones that result from some $S^{1:n}$ given a particular $O^{1:n}$) that give the same value of $\nu_m = \big(\delta^{(m)},\ \alpha^{(m)},\ \beta^{(m)},\ \gamma^{(m)}\big)$. (Note the value of $\beta^{(m)}$ is determined by $\delta_1,\ \delta_2,\ \ldots, \delta_m$. So, we do not need $\beta_p$'s in $V_m$.) In the expression for the frequency $\mathcal{F}_m(\nu_m)$ that comes later in (4), we first compute the number of Markov chain realizations for each $V_m$ of which the elements sum up to $\nu_m$, then sum the obtained numbers up to find $\mathcal{F}_m(\nu_m)$.

## 6. Defining Two Functions for the Frequency $\mathcal{N}_n(K_n, L_n) = \mathcal{F}_m(\nu_m)$

The two functions that we describe in this section are 'the' key to improve the speed of our previous algorithm. The concept that gives us these two functions is that if we consider the $p$th time interval, the total number of the $p$th Markov chain realizations $S^{(n_p+1):(n_p+Y_p)}$ that satisfy given values of $\delta^{(p-1)}$ (the previous state), $\delta_p$ (indicating the state changes even or odd times), $\alpha_p$ (the number of 01's), and $\gamma_p$ (the number of 1's will be

$$\begin{pmatrix}\text{number of ways to}\\ \text{properly separate 1's}\end{pmatrix} \times \begin{pmatrix}\text{number of ways to}\\ \text{properly separate 0's}\end{pmatrix}.$$

To this end, we define two functions $f_1(y, \alpha, \gamma, \delta)$ and $f_2\left(y, \alpha, \gamma, \widetilde{\delta}, \delta\right)$ as

$$f_1(y, \alpha, \gamma, \delta) = \binom{\gamma - 1}{\alpha - 1}\binom{y - \gamma - 1}{\alpha - \delta}$$

and

$$f_2\left(y, \alpha, \gamma, \widetilde{\delta}, \delta\right) = \binom{\gamma - 1 + \widetilde{\delta}}{\alpha - 1 + \widetilde{\delta}}\binom{y - \gamma - \widetilde{\delta}}{\alpha - \delta}.$$

Here the notation $\binom{a}{b}$ is used as the usual combination function ${}_aC_b = \frac{a!}{b!(a-b)!}$ except that $\binom{-1}{-1}$ is defined to be 1.

Now, let $R_m(\nu_m)$ be the range of $V_m$ given $\nu_m$ (see Appendix B for more details). Then, the frequency $\mathcal{F}_m(\nu_m) = \mathcal{N}_n(K_n, L_n)$ can be expressed as below.

$$\begin{aligned}\mathcal{F}_m(\nu_m) &= \sum_{V_m \in R_m(\nu_m)} \prod_{p=1}^{m} \left(\text{the number of valid } S^{1:p}\text{-values given } V_m\right)\\ &= \sum_{V_m \in R_m(\nu_m)} f_1(Y_1, \alpha_1, \gamma_1, \delta_1) \prod_{p=2}^{m} f_2\left(Y_p, \alpha_p, \gamma_p, \delta^{(p-1)}, \delta^{(p)}\right) \qquad (3)\end{aligned}$$

The function $f_2$, used for $p > 1$, is different from the function $f_1$ for $p = 1$ because we do not consider the state before the initial state $S_1$, while we consider the preceding state $S_{n_p}$ when $p > 1$, and also because the initial state $S_1$ is fixed as 0.

## 7. Explicitly Expressing Frequency $\mathcal{F}_m(\nu_m)$

We now know we could express $\mathcal{F}_m(\nu_m)$ as in (4).

$$
\begin{aligned}
(4)\quad \mathcal{F}_m(\nu_m) &= \mathcal{F}_m\left(\delta^{(m)}, \alpha^{(m)}, \beta^{(m)}, \gamma^{(m)}\right) \\
&= \sum_{\substack{\delta^{(p)}\in\{0,1\},\ \alpha^{(p)}\in R_\alpha^{(p)},\\ \beta^{(p)}\in R_\beta^{(p)},\ \gamma^{(p)}\in R_\gamma^{(p)}\left(\alpha^{(p)}\right)\\ \forall p\in\{1,\dots,m\}}} \left( \sum_{\substack{\delta_1\in\{0,1\},\\ \alpha_1\in R_{\alpha,1},\\ \gamma_1\in R_{\gamma,1}}} \sum_{\substack{\delta_2\in\{0,1\},\\ \alpha_2\in R_{\alpha,2},\\ \gamma_2\in R_{\gamma,2}}} \right. \\
&\qquad \left. \cdots \sum_{\substack{\delta_m\in\{0,1\},\\ \alpha_m\in R_{\alpha,m},\\ \gamma_m\in R_{\gamma,m}}} f_1(Y_1, \alpha_1, \gamma_1, \delta_1) \prod_{p=2}^{m} f_2\left(Y_p, \alpha_p, \gamma_p, \delta^{(p-1)}, \delta^{(p)}\right) \right)
\end{aligned}
$$

The precise ranges of $\alpha_p$ and $\gamma_p$, which are described in Appendix B, are rather complicated mainly because of the restrictions $\gamma^{(m)} = \sum_{p=1}^{m} \gamma_p$ and $\alpha^{(m)} = \sum_{p=1}^{m} \alpha_p$. But, due to our special definition of the notation $\binom{a}{b}$, which appears in the definition of $f_1$ and $f_2$, we could choose simpler restrictions on these variables.

Define sets $R_{\alpha,1}$, $R_{\gamma,1}$, $R_{\alpha,p}(\delta^{(p-1)})$, and $R_{\gamma,p}(\delta^{(p-1)}, \delta_p)$ for $p = 2, 3, \dots, m$ as shown below. Then, $R_{\alpha,1}$ and $R_{\gamma,1}$ cover the ranges of $\alpha_1$ and $\gamma_1$, while $R_{\alpha,p}(\delta^{(p-1)})$ and $R_{\gamma,p}(\delta^{(p-1)}, \delta_p)$ cover the ranges of $\alpha_p$ and $\gamma_p$, respectively.

$$
\begin{aligned}
R_{\alpha,1}(\delta_1) &= \left\{\delta_1, \delta_1+1, \dots, \lfloor Y_1/2 \rfloor\right\} \\
R_{\gamma,1}(\delta_1, \alpha_1) &= \begin{cases} \{0\} & \text{if } \alpha_1 = 0 \\ \{\alpha_1, \alpha_1+1, \dots, Y_1 - \alpha_1 + \delta_1 - 1\} & \text{if } \alpha_1 > 0 \end{cases} \\
R_{\alpha,p}\left(\delta^{(p-1)}\right) &= \begin{cases} \left\{0, 1, \dots, \lceil Y_p/2 \rceil\right\} & \text{if } \delta^{(p-1)} = 0 \\ \left\{0, 1, \dots, \lfloor Y_p/2 \rfloor\right\} & \text{if } \delta^{(p-1)} = 1 \end{cases} \\
R_{\gamma,p}&\left(\delta^{(p-1)}, \delta_p,\ \alpha_p\right) \\
&= \begin{cases} \{0\} & \text{if } \delta^{(p-1)} = 0,\ \alpha_p = 0 \\ \{\alpha_p, \alpha_p+1, \dots, Y_p - \alpha_p + \delta_p\} & \text{if } \delta^{(p-1)} = 0,\ \alpha_p > 0 \\ \{Y_p\} & \text{if } \delta^{(p-1)} = 1,\ \alpha_p = 0, \delta_p = 0 \\ \{\alpha_p, \alpha_p+1, \dots, Y_p - \alpha_p - \delta_p\} & \text{if otherwise} \end{cases}
\end{aligned}
$$

Furthermore, define sets $R_\alpha^{(p)}$, $R_\beta^{(p)}$, and $R_\gamma^{(p)}$ for $p = 1, 2, \ldots, m$ as shown below. Then, $R_\alpha^{(p)}$, $R_\beta^{(p)}$, and $R_\gamma^{(p)}$ cover the ranges of $\alpha^{(p)}$, $\beta^{(p)}$, and $\gamma^{(p)}$, respectively.

$$\begin{aligned} R_\alpha^{(p)} &= \left\{\delta^{(p)}, \delta^{(p)} + 1, \ldots, \left\lfloor \frac{n_{p+1}}{2} \right\rfloor + 1\right\} \\ R_\beta^{(p)} &= \{0, 1, \ldots, p-1\} \\ R_\gamma^{(p)}\left(\alpha^{(p)}\right) &= \left\{\alpha^{(p)}, \alpha^{(p)} + 1, \ldots, n_{p+1} - \alpha^{(p)}\right\} \end{aligned}$$

Again, the exact ranges are shown in Appendix B.

## 8. Conclusions

We presented the exact formula for the BPM of a discrete-time $2 \times 2$ HMM. Since we do not know the exact use of the formula, yet; we decided to simply show it here in this Web site (mathjm.com).

## Acknowledgments

The author wishes to express special thanks to David Vere-Jones and New Zealand Institute of Mathematics and its Applications (NZIMA) for their support, which provided me the opportunity and time to eventually finish this paper.

## Appendix A. Derivation of BPM formulas

Below is a straightforward derivation of formulas that are used for the BPM for the HMMs with $\zeta_X = \zeta_Y = \{0, 1\}$. By the Bayes' formula, we have

$$P\left(\theta \mid O^{1:m}\right) = \frac{P(O^{1:n}, \theta)}{P(O^{1:n})} = \frac{P(O^{1:n} \mid \theta)P(\theta)}{\int_{\theta\in\Theta} P(O^{1:n}, \theta)\, d\theta} = \frac{P(O^{1:n} \mid \theta)P(\theta)}{\int_{\theta\in\Theta} P(O^{1:n} \mid \theta)P(\theta)\, d\theta}.$$

So, as shown in (2),

$$\begin{aligned} \hat{\theta} = E\left(\theta \mid O^{1:n}\right) &= \int_{\theta\in\Theta} \theta P\left(\theta \mid O^{1:n}\right)\, d\theta \\ &= \int_{\theta\in\Theta} \theta \left(\frac{P(O^{1:n} \mid \theta)P(\theta)}{\int_{\theta\in\Theta} P(O^{1:n} \mid \theta)P(\theta)\, d\theta}\right) d\theta \\ &= \frac{\sum_{S^{1:n}\in\Omega_n} \int_{\theta\in\Theta} \theta P\left(S^{1:n}, O^{1:n} \mid \theta\right) P(\theta)\, d\theta}{\sum_{S^{1:n}\in\Omega_n} \int_{\theta\in\Theta} P\left(S^{1:n}, O^{1:n} \mid \theta\right) P(\theta)\, d\theta} \\ &= \frac{\sum_{\omega_n\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \int_{\theta\in\Theta} \theta f(K_n, L_n, \theta)P(\theta)\, d\theta}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \int_{\theta\in\Theta} f(K_n, L_n, \theta)P(\theta)\, d\theta}, \end{aligned}$$

where

$$f\left(K_n, L_n, \theta\right) = P\left(S^{1:n}, O^{1:n} \mid \theta\right).$$

Because $A$ and $B$ are probability matrices, we do not need to estimate all the elements in them. So, we will estimate the probabilities to stay in the state $i$ in the next time unit, $a_{ii}$, and the probabilities to observe 1 when the state is $i$, $b_{i1}$, for $i \in \{0,1\}$, as shown below.

$$A = \begin{pmatrix} a_{00} & 1-a_{00} \\ 1-a_{11} & a_{11} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1-b_{01} & b_{01} \\ 1-b_{11} & b_{11} \end{pmatrix}$$

As for the prior $P(\theta)$, assume Dirichlet distributions for $a_{00}$, $a_{11}$, $b_{01}$, and $b_{11}$, while assume a constant $\frac{1}{2}$ for $r_0$. Then, for some positive constants, $p_{ij}$ and $q_{ij}$, $i,j \in \{0,1\}$,

$$P(\theta) = \frac{1}{2}\frac{a_{00}^{p_{00}-1}(1-a_{00})^{p_{01}-1}}{Z(p_{00},p_{01})} \cdot \frac{a_{11}^{p_{11}-1}(1-a_{11})^{p_{10}-1}}{Z(p_{10},p_{11})} \cdot \prod_{i=0}^{1} \frac{b_{i1}^{q_{i1}-1}(1-b_{i1})^{q_{i0}-1}}{Z(q_{i0},q_{i1})}, \tag{5}$$

where the function $Z$, defined as $Z(a,b) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$, returns the normalizing constant such that each of the factors for $P(a_{ii})$ and $P(b_{i1})$ integrates to unity for $i = 0,1$ (Gelman et al., 1995).

Define $\Phi$ and $\Psi$ as below.

$$\Phi(K_n) = \sum_{i=0}^{k_{10}} \frac{(-1)^i\, k_{10}!}{i\,!\,(k_{10}-i)!\,(k_{11}+i+1)} \cdot \frac{(k_{00}+k_{11}+i+1)!\,k_{01}!}{(n-k_{10}+i+1)!}$$

$$\Psi(L_n) = \prod_{i=0}^{1} \frac{l_{i0}!\; l_{i1}!}{(l_{i0}+l_{i1}+1)!}$$

Then, from (1) and (5),

$$\begin{aligned}
&\int_{\theta\in\Theta} f(\omega_n,\theta)P(\theta)\,d\theta \\
&= \frac{1}{C}\int_{\theta\in\Theta} a_{00}^{k_{00}+p_{00}-1}(1-a_{00})^{k_{01}+p_{01}-1} a_{11}^{k_{11}+p_{11}-1}(1-a_{11})^{k_{10}+p_{10}-1} \\
&\qquad \cdot \prod_{i=0}^{1} b_{i1}^{l_{i1}+q_{i1}-1}(1-b_{i1})^{l_{i0}+q_{i0}-1}\,d\theta \\
&= \frac{1}{C}\int_0^1\int_0^{a_{00}} a_{00}^{\widetilde{k}_{00}}(1-a_{00})^{\widetilde{k}_{01}} a_{11}^{\widetilde{k}_{11}}(1-a_{11})^{\widetilde{k}_{10}} da_{11}da_{00} \\
&\qquad \cdot \prod_{i=0}^{1}\int_0^1 b_{i1}^{\widetilde{l}_{i1}}(1-b_{i1})^{\widetilde{l}_{i0}}\,db_{i1} \\
&= \frac{1}{C}\,\Phi\begin{pmatrix} \widetilde{k}_{00} & \widetilde{k}_{01} \\ \widetilde{k}_{10} & \widetilde{k}_{11} \end{pmatrix} \cdot \Psi\begin{pmatrix} \widetilde{l}_{00} & \widetilde{l}_{01} \\ \widetilde{l}_{10} & \widetilde{l}_{11} \end{pmatrix},
\end{aligned}$$

where $\widetilde{k}_{ij} = k_{ij} + p_{ij} - 1$ and $\widetilde{l}_{iu} = l_{iu} + q_{iu} - 1$; and $C$ is just a constant that will be cancelled out when we compute the expected value. We use here the restriction $a_{00} \geq a_{11}$ to avoid 'averaging out' the symmetry in the probability distribution. (Other choices are possible, of course, such as $b_{00} \geq b_{11}$.)

In the simulations for this paper, we let $p_{ij} = q_{iu} = 1$ for all $i, j, u \in \{0, 1\}$ so that $\widetilde{k}_{ij} = k_{ij}$ and $\widetilde{l}_{iu} = l_{iu}$. Let $E_{ij}$, $i, j, \in \{0, 1\}$ be a $2 \times 2$ matrix such that all the entries are 0 except for the entry on the $i$th row and $j$th column, which is 1. Then, (2) can be rewritten for each element of $\hat{\theta}$ as shown below.

$$\hat{a}_{00} = E\left(a_{00} \mid O^{1:n}\right) = \frac{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n + E_{00}\right) \Psi\left(L_n\right)}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n\right)}$$

$$\hat{b}_{11} = E\left(a_{11} \mid O^{1:n}\right) = \frac{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n + E_{11}\right) \Psi\left(L_n\right)}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n\right)}$$

$$\hat{b}_{01} = E\left(b_{01} \mid O^{1:n}\right) = \frac{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n + E_{01}\right)}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n\right)}$$

$$\hat{b}_{11} = E\left(b_{11} \mid O^{1:n}\right) = \frac{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n + E_{11}\right)}{\sum_{(K_n,L_n)\in\widehat{\Omega}_n} \mathcal{N}_n(K_n, L_n) \cdot \Phi\left(K_n\right) \Psi\left(L_n\right)}$$

## Appendix B. More Precise Ranges of the Elements of $V_m$ and $\nu_m$

Here we list the ranges of the elements of $V_m$ and $\nu_m$. In the algorithm, we use cover sets of these ranges instead. In the descriptions of the ranges below, all the quantities are discrete.

### B.1. **Ranges of the elements of** $V_m = (\delta_1, \ldots, \delta_m, \alpha_1, \ldots, \alpha_m, \gamma_1, \ldots, \gamma_m)$.

We need to consider these because the algorithm processes the $Y_p$ for each $p = 1, 2, \ldots, m$. By the definition, $\delta_p \in \{0, 1\}$. Now, given $\nu_m = (\delta^{(m)}, \alpha^{(m)}, \beta^{(m)}, \gamma^{(m)})$ and $(\delta_1, \ldots, \delta_m)$, we find the ranges for $\alpha_p$ and $\gamma_p$, $p = 1, 2, \ldots, m$, using the following *idea*: As for $\alpha_p$,

$$\alpha_p \leq \min\{\alpha_p \text{ that results when the states are in the form } (\cdots 0, 1, 0, 1, 0, 1 \cdots), \\ \text{the largest } \alpha_p \text{ that satisfies } \alpha^{(p-1)} + \alpha_p = \alpha^{(p)} \leq \alpha^{(m)}\}.$$

As for $\gamma_p$, since, naturallly, the number of 1's (or 0's) should be greater than or equal to the number of $(0,1)$'s in any sequence,

$$\begin{aligned}&\max\{\text{the smallest } \gamma_p \text{ that satisfies } \gamma_p \geq \alpha_p,\\ &\qquad \min(\gamma_p) \text{ that will leave enough 0's for the rest to have } \alpha^{(m)}\}\\ &\quad \leq \gamma_p\\ &\quad \leq \min\left\{\max(\gamma_p) \text{ that will leave enough 1's for the rest to have } \alpha^{(m)},\right.\\ &\qquad\qquad \left.\max(\gamma_p) \text{ that will leave enough 0's in the } p\text{th interval to have } \alpha_p\right\}\end{aligned}$$

So, to perform the necessary adjustments, we define the constants as shown below.

Define $\xi_{\alpha,1,\max}(Y_1,\delta_1)$ and $\xi_{\alpha,p,\max}\left(Y_p,\delta^{(p-1)},\delta_p\right)$ as

$$\xi_{\alpha,1,\max}(Y_1,\delta_1) = \begin{cases}-1 & \text{if } Y_1 \text{ is even and } \delta_1 = 0\\ 0 & \text{otherwise}\end{cases} \qquad \text{and}$$

$$\xi_{\alpha,p,\max}\left(Y_p,\delta^{(p-1)},\delta_p\right) = \begin{cases}1 & \text{if } Y_p \text{ is odd, } \delta^{(p-1)} = 0, \text{ and } \delta_p = 1\\ -1 & \text{if } Y_p \text{ is even, } \delta^{(p-1)} = 1, \text{ and } \delta_p = 1\\ 0 & \text{otherwise}\end{cases}$$

for $p = 2, \ldots, m$.

Also, we define a cumulative indicator $\delta^{(p+)}$ as shown below, which makes $\delta^{(p)}$ $+\delta^{(p+)} = \delta^{(m)}$ for any $p$.

$$\delta^{(p+)} = \begin{cases}0 & \text{if } \sum_{i=p+1}^{m} \delta_i \text{ is even}\\ 1 & \text{if otherwise}\end{cases}$$

Furthermore, define $\xi_{\gamma,p,\min}\left(\delta^{(p)},\delta^{(p+)}\right)$ and $\xi_{\gamma,p,\max}\left(\delta^{(p-1)},\delta_p\right)$ as

$$\xi_{\gamma,p,\min}\left(\delta^{(p)},\delta^{(p+)}\right) = \begin{cases}0 & \text{if } \delta^{(p+)} = 0\\ 1 & \text{if } \delta^{(p)} = 0 \text{ and } \delta^{(p+)} = 1\\ -1 & \text{if } \delta^{(p)} = 1 \text{ and } \delta^{(p+)} = 1\end{cases} \qquad \text{for } p = 1, \ldots, m \text{ and}$$

$$\xi_{\gamma,p,\max}\left(\delta^{(p-1)},\delta_p\right) = \begin{cases}0 & \text{if } \delta_p = 0\\ -1 & \text{if } \delta^{(p-1)} = 0 \text{ and } \delta_p = 1\\ 1 & \text{if } \delta^{(p-1)} = 1 \text{ and } \delta_p = 1\end{cases} \qquad \text{for } p = 2, \ldots, m.$$

Then, straightforward (but rather time-consuming) observations lead to the facts shown in below.

(i) For $p = 1$:

$\delta_1 \le \alpha_1 \le U_{\alpha,1}\left(Y_1, \alpha^{(m)}, \delta_1, \sum_{i=2}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right)$, where

$$U_{\alpha,1}\left(Y_1, \alpha^{(m)}, \delta_1, \sum_{i=2}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right) = \min\left\{\left\lfloor \frac{Y_1}{2}\right\rfloor + \xi_{\alpha,1,\max}(Y_1,\delta_1),\ \alpha^{(m)} - \sum_{i=2}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right\}$$

$u_{\gamma,1}\left(Y_1, \alpha_1, \alpha^{(m)}, \gamma^{(m)}, \delta_1, \delta^{(1+)}\right) \le \gamma_1 \le U_{\gamma,1}\left(Y_1, \alpha_1, \alpha^{(m)}, \gamma^{(m)}, \delta_1\right)$, where

$$u_{\gamma,1}\left(Y_1, \alpha_1, \alpha^{(m)}, \gamma^{(m)}, \delta_1, \delta^{(1+)}\right) = \max\left\{\alpha_1,\ \gamma^{(m)} - n + Y_1 + \alpha^{(m)} - \alpha_1 - \xi_{\gamma,1,\min}\left(\delta_1, \delta^{(1+)}\right)\right\}$$

and

$$U_{\gamma,1}\left(Y_1, \alpha_1, \alpha^{(m)}, \gamma^{(m)}, \delta_1\right) = \begin{cases} 0 & \text{if } \alpha_1 = 0 \\ \min\left\{\gamma^{(m)} - \alpha^{(m)} + \alpha_1,\ Y_1 - \alpha_1 + \delta_1 - 1\right\} & \text{if } \alpha_1 > 0 \end{cases}$$

(ii) For $p = 2, 3, \ldots, m$:

$u_{\alpha,p}\left(\delta^{(p-1)}, \delta_p\right) \le \alpha_p \le U_{\alpha,p}\left(Y_p, \alpha^{(p-1)}, \alpha^{(m)}, \delta^{(p-1)}, \delta_p, \sum_{i=p+1}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right)$, where

$$u_{\alpha,p}\left(\delta^{(p-1)}, \delta_p\right) = \begin{cases} 1 & \text{if } \delta^{(p-1)} = 0 \text{ and } \delta_p = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$U_{\alpha,p}\left(Y_p, \alpha^{(p-1)}, \alpha^{(m)}, \delta^{(p-1)}, \delta_p, \sum_{i=p+1}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right) = \min\left\{\left\lfloor \frac{Y_p}{2}\right\rfloor + \xi_{\alpha,p,\max}\left(Y_p, \delta^{(p-1)}, \delta_p\right),\ \alpha^{(m)} - \alpha^{(p-1)} - \sum_{i=p+1}^m u_{\alpha,i}\left(\delta^{(i-1)}, \delta_i\right)\right\}$$

$u_{\gamma,p}\left(Y_p, \sum_{i=p+1}^m Y_i, \alpha_p, \alpha^{(p)}, \alpha^{(m)}, \gamma^{(p-1)}, \gamma^{(m)}, \delta^{(p-1)}, \delta_p, \delta^{(p+)}\right) \le \gamma_p \le U_{\gamma,p}\left(Y_p, \alpha_p, \alpha^{(p)}, \alpha^{(m)}, \gamma^{(p-1)}, \gamma^{(m)}, \delta^{(p-1)}, \delta_p\right)$, where

$$u_{\gamma,p}\left(Y_p, \sum_{i=p+1}^m Y_i, \alpha_p, \alpha^{(p)}, \alpha^{(m)}, \gamma^{(p-1)}, \gamma^{(m)}, \delta^{(p-1)}, \delta_p, \delta^{(p+)}\right) = \begin{cases} Y_p & \text{if } \alpha_p = 0,\ \delta^{(p-1)} = 1 \text{ and } \delta_p = 0 \\ \max\left\{\alpha_p,\ \gamma^{(m)} - \gamma^{(p-1)} - \sum_{i=p+1}^m Y_i + \alpha^{(m)} - \alpha^{(p)} - \xi_{\gamma,p,\min}\left(\delta^{(p)}, \delta^{(p+)}\right)\right\} & \text{otherwise} \end{cases}$$

and

$$U_{\gamma,p}\left(Y_p, \alpha_p,\ \alpha^{(p)}, \alpha^{(m)}, \gamma^{(p-1)}, \gamma^{(m)}, \delta^{(p-1)}, \delta_p\right)$$
$$= \begin{cases} 0 & \text{if } \alpha_p = 0,\ \delta^{(p-1)} = 0 \\ \min\left\{Y_p - \alpha_p - \xi_{\gamma,p,\max}\left(\delta^{(p-1)}, \delta_p\right), \right. \\ \qquad \left. \gamma^{(m)} - \gamma^{(p-1)} - \alpha^{(m)} + \alpha^{(p)}\right\} & \text{otherwise} \end{cases}$$

B.2. **Ranges of the elements of $\nu_m = \left(\delta^{(m)}, \alpha^{(m)}, \beta^{(m)}, \gamma^{(m)}\right)$.** From the definitions, obviously $\delta^{(m)} \in \{0,1\}$ and $\beta^{(m)} \in \{0, 1, \ldots, m\}$. As for the ranges of the sums $\alpha^{(m)}$ and $\gamma^{(m)}$, with $S_1 = 0$, we have

$$\alpha^{(m)} = k_{01} \in \left\{1, 2, \ldots, \lfloor n/2 \rfloor\right\}$$

and, given a particular $\alpha^{(m)}$, we see $\gamma^{(m)} \in \left\{\alpha^{(m)}, \alpha^{(m)} + 1, \ldots, n - \alpha^{(m)}\right\}$. However, if given both $\alpha^{(m)}$ and $\delta^{(m)}$ (equivalently, given both $k_{01}$ and $S_n$), the range of $\gamma^{(m)}$, the number of 1's in the Markov chain, can be narrowed down to

$$\gamma^{(m)} \in \{\alpha^{(m)}, \alpha^{(m)} + 1, \ldots, \gamma_{\max}\},$$

where

$$\gamma_{\max} = \begin{cases} 0 & \text{if } \alpha^{(m)} = 0 \\ n - \alpha^{(m)} + \delta^{(m)} - 1 & \text{if } \alpha^{(m)} > 0 \end{cases}.$$

## References


Ailliot, P., Monbet, V., Prevosto, & M., 2006. An autoregressive model with time-varying coefficients for wind fields. Environmetrics, 17(2) 107-117.

Albert, P.S., 1991. A two-state Markov mixture model for a time series of epileptic seizure counts. Biometrics, 47 1371-1381.

Baum, L.E., 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities, 3 1-8.

Bulla, J. & Berzel, A., 2007. Computational issues in parameter estimation for stationary hidden Markov models. Computational Statist., 23(1) 1-18.

Baum, L.E., Petrie, T., Soules, G., & Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic function of Markov chains. Ann. of Math. Statist., 41 164-171.

Capppé, O., Moulines, E., & Rydén, T., 2005. Inference in Hidden Markov Models. Springer-Verlag, New York.

Celeux, G., Hurn, M., & Robert, C.P., 2000. Computational and inferential difficulties with mixture posterior distributions. J. American Statist. Assoc., 95 957E70.

Ching, W.K., 1997. Markov-modulated Poisson processes for multi-location inventory problems. Internat. J. Production Econom., 53 217-223.

Chib, S., Chib. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econom., 75 79-97.

Cooper, B., & Lipsitch, M., 2004. The analysis of hospital infection data using hidden Markov models. Biostatistics, 5(2) 223E37.

Dempster, A.D., Laird, N.M., & Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc., Ser. B, 39(1) 1-38.

Gelman, A., Carlin, J., Stern, H., & Rubin, D.B., 1995. Bayesian Data Analysis. Springer-Verlag, New York.

Gilks, W.R., Richardson, S., & Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice, Chapman & Hall, London.

Heffes, H., & Lucantoni, D.M., 1986. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. IEEE J. on Selected Areas in Comm., 4 856-868.

Heyman, D.P., & Lucantoni, D., 2003. Modeling multiple IP traffic streams with rate limits. IEEE/ACM Trans. on Networking, 11(6) 948-957.

Jana, R., & Dey, S., 2000. Change detection in teletraffic models. IEEE Trans. on Signal Processing, 48(3) 846-853.

MacDonald, I.L., & Zucchini, W., 1997. Hidden Markov and Other Models for Discrete-Valued Time Series. Chapman-Hall, London.

Murakami, J., 2009. Bayesian posterior mean estimates for Poisson hidden Markov models. Computational Statist. & Data Analysis, 53(4) 941-955.

Murakami, J., & Taylor, T., 2006. Least square estimation of a hidden Markov chain parameters. Proc. of the 17th Internat. Sympos. on Math. Theory of Networks and Systems (MTNS), July 2006, Kyoto, Japan, 2151-2156.

Nogueira, A., Salvador, P., Valadas, R., & Pacheco, A., 2004. Fitting Self-similar Traffic by a Superposition of MMPPs Modeling the Distribution at Multiple Time Scales. IEICE Trans. on Comm., E87-B(3) 678-688.

Oliveros, A.Q., Carnie, R., Tárraga, M., & Aspinall, W., 2008. On the application of Hidden Markov Model and Bayesian Belief Network to seismic noise at Las Cañadas Caldera, Tenerife, Spain. Chaos Solitons & Fractals, 37(3) 849-857.

Pavlovic, V., Garg, A., & Kasif, S., 2002. A Bayesian framework for combining gene predictions. Bioinformatics, 18(1) 19-27.

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

Robert, C.P., & Titterington, D.M., 1998. Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation. Statist. Computing, 8 145-158.

Sinha, S., Van Nimwegen, E., & Siggia, E.D. 2003. A probabilistic method to detect regulatory modules. Bioinformatics. 19(1) i292-i301.

Scott, S.L., 2002. Bayesian methods for hidden Markov models: Recursive computing in the 21st century. J. Amer. Statist. Assoc., 97 337-351.

Scott, S.L., & Smyth, P., 2003. The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling. Bayesian Statist., 7 671-680, Oxford University Press.

Stephens, M., 2000. Dealing with label switching in mixture models. J. R. Stat. Soc. Ser. B Stat. Methodol., 62(4) 795E09.

Wei, W., Wang, B., and Towsley, D., 2002. Continuous-time Hidden Markov Models for Network Performance Evaluation. Performance Evaluation, 49(1-4) 129-146.