Viterbi Hidden Markov Models

Angela B. Shiflet[1], George W. Shiflet[1], Dmitriy A. Kaplun[1],
Pietro Hiram Guzzi[2], Mario Cannataro[2]
[1]Wofford College, Spartanburg, SC, USA
[2]University "Magna Græcia" of Catanzaro, Catanzaro, Italy
`shifletab@wofford.edu, shifletgw@wofford.edu,`
`kaplunda@email.wofford.edu,`
`hguzzi@unicz.it, cannataro@unicz.it`
© 2017

Prerequisites: "What are the chances?—Hidden Markov Models" (Shiflet et al. 2016)

Introduction

In the module "What are the chances?—Hidden Markov Models" (Shiflet et al. 2016), which is a prerequisite to the current material, we considered Hidden Markov Model (HMM) solutions to likelihood problems, such as the probability of a sequence of observations, often using the hypothetical HMM associated with a howler monkey called Holly (Figure 1). For example, employing the HMM forward algorithm, we determined the likelihood of monitoring equipment registering breathing sounds (B) and then munching noises (M) the next two hours for Holly Howler, $P(v = \text{BMM})$. In the current module, we use the **Viterbi algorithm** to solve another type of HMM problem, **decoding**. In this case, given a sequence of observations, such as $v = \text{BMM}$, we determine the most probable sequence of underlying states, u, to yield v. Thus, we wish to determine the u with maximum $P(u \mid v)$, such as the sequence of three states, u, that yields max($P(u \mid v = \text{BMM})$).

Holly's HMM

State space, or set of possible states, $S = \{\text{E, R}\}$, with elements representing eating and resting/sleeping, respectively

Observation space, or set of possible observations, $O = \{\text{M, B}\}$, with elements representing munching and breathing noises, respectively

Initial state probabilities, $\pi(\text{E}) = 0.30$ and $\pi(\text{R}) = 0.70$

Transition probabilities, $t(\text{E, E}) = 0.6$, $t(\text{E, R}) = 0.4$, $t(\text{R, E}) = 0.2$, and $t(\text{R, R}) = 0.8$, summarized by the following transition matrix:

$$T = \begin{array}{cc} & \begin{array}{cc} \text{E} & \text{R} \end{array} \\ \begin{array}{c} \text{E} \\ \text{R} \end{array} & \left[\begin{array}{cc} 0.6 & 0.4 \\ 0.2 & 0.8 \end{array} \right] \end{array}$$

$u_k \setminus u_{k+1}$

Emission probabilities, $e(\text{B} \mid \text{E}) = 0.2$, $e(\text{B} \mid \text{R}) = 0.9$, $e(\text{M} \mid \text{E}) = 0.8$, and $e(\text{M} \mid \text{R}) = 0.1$, summarized by the following **output**, or **emission**, **matrix**:

hidden\observable	M	B
E	0.8	0.2
R	0.1	0.9

Figure 1   Holly's HMM

In module "What are the chances?—Hidden Markov Models," we learned by joint probability that $P(u, BMM) = P(u \mid BMM) \cdot P(BMM)$ (Shiflet et al. 2016.) Thus, the problem of determining the state sequence u that maximizes $P(u \mid BMM)$ is equivalent to the problem of determining the u that maximizes $[P(u, BMM) / P(BMM)]$. However, because $P(BMM)$ is a constant, the problem simplifies to finding u where $P(u, BMM)$ is maximum. The most obvious method to solve the problem is to determine $P(u, BMM)$ for every possible three-element state sequence, $S^3 = \{RRR, RRE, RER, REE, ERR, ERE, EER, EEE\}$, using a version of the forward algorithm, and then to select the u with the largest probability. However, this solution is exponentially large because for a sequence of n observations with $h = 2$ states, the number of possible n-element state sequences, or the number of elements in S^n, is $h^n = 2^n$. In general, for h number of states, the number of n-element state sequences is h^n. The much faster **Viterbi algorithm** is another dynamic programming algorithm, which has the forward algorithm as its base.

Viterbi Algorithm

The key to the Viterbi algorithm is the following recursive equation for calculating a joint probability (Equation 14 from Shiflet et al. 2016):

$$P(u,v) = P(u_{1,n}, v_{1,n}) = \begin{cases} \pi(u_1) \cdot e(v_1 \mid u_1), & \text{if } n = 1 \\ P(u_{1,n-1}, v_{1,n-1}) \cdot t(u_{n-1}, u_n) \cdot e(v_n \mid u_n), & \text{if } n > 1 \end{cases} \tag{1}$$

Beginning the Viterbi algorithm in the same way as the forward algorithm for observation sequence BMM in Holly's HMM, we employ a 2×3 matrix, G, with first-column elements being $g_{E1} = P(B, E) = \pi(E) \cdot e(B \mid E)$ and $g_{R1} = P(B, R) = \pi(R) \cdot e(B \mid R)$. Figure 2 presents the initialization step, which is identical to that of the forward algorithm, for the following initial Viterbi matrix, G:

$$G = \begin{matrix} & B & M & M \\ E & \begin{bmatrix} 0.06 & \square & \square \\ R & 0.63 & \square & \square \end{bmatrix} \end{matrix}$$
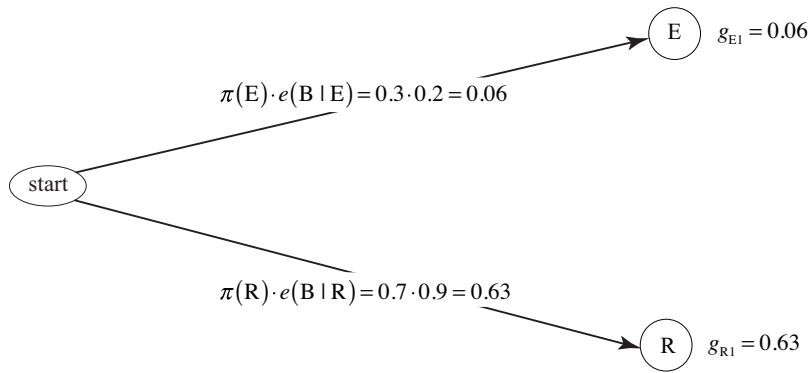
Figure 2   Initialization step of the Viterbi algorithm

Quick Review Question 1 Calculate the first-column elements of the Viterbi matrix to calculate u for max($P(u \mid \text{HHGH})$) using the HMM in Figure 3, which contains the following information:

$S = \{A, B, C\}$ and $O = \{G, H\}$

$\pi(A) = 0.2, \pi(B) = 0.1, \pi(C) = 0.7$

$$
\begin{array}{c c}
u_k \setminus u_{k+1} & \begin{array}{ccc} A & B & C \end{array} \\
T = \begin{array}{c} A \\ B \\ C \end{array} & \left[\begin{array}{ccc} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.3 & 0.2 & 0.5 \end{array} \right]
\end{array}
$$

$$
\begin{array}{c c}
\text{hidden\textbackslash observable} & \begin{array}{cc} G & H \end{array} \\
\begin{array}{c} A \\ B \\ C \end{array} & \left[\begin{array}{cc} 0.9 & 0.1 \\ 0.6 & 0.4 \\ 0.1 & 0.9 \end{array} \right]
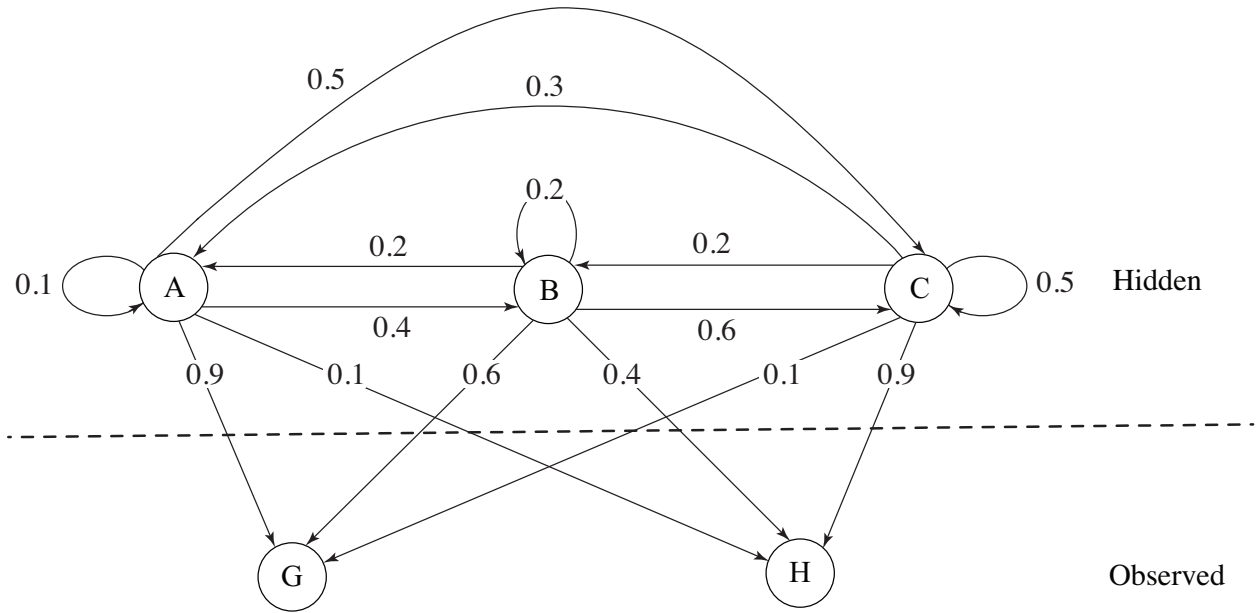\end{array}
$$

Figure 3   HMM diagram for Quick Review Question 1

   Computations of the second column of the Viterbi and the forward matrices also
begin in the same way with calculating the product of a first column element, a transition
value, and an emission value. As in Figure 7 of (Shiflet et al. 2016) for the forward
algorithm, Figure 4 of the current module for the Viterbi algorithm makes the following
computations using first column values, g_{E1} and g_{R1}, that correspond to f_{E1} and f_{R1},
respectively, of the forward matrix:

$$g_{E1} \cdot t(E, E) \cdot e(M \mid E)$$
$$g_{E1} \cdot t(E, R) \cdot e(M \mid R)$$
$$g_{R1} \cdot t(R, E) \cdot e(M \mid E)$$
$$g_{R1} \cdot t(R, R) \cdot e(M \mid R)$$

However, instead of taking the sum of pairs of expressions (first and third transitioning to
E, second and fourth transitioning to R) as we did with the forward algorithm, we take
the maxima, as follows:

$$g_{E2} = \max(g_{E1} \cdot t(E, E) \cdot e(M \mid E), \ g_{R1} \cdot t(R, E) \cdot e(M \mid E))$$
$$g_{R2} = \max(g_{E1} \cdot t(E, R) \cdot e(M \mid R), \ g_{R1} \cdot t(R, R) \cdot e(M \mid R))$$

Figure 4 details these computations with boldface arrows indicating the maxima. The
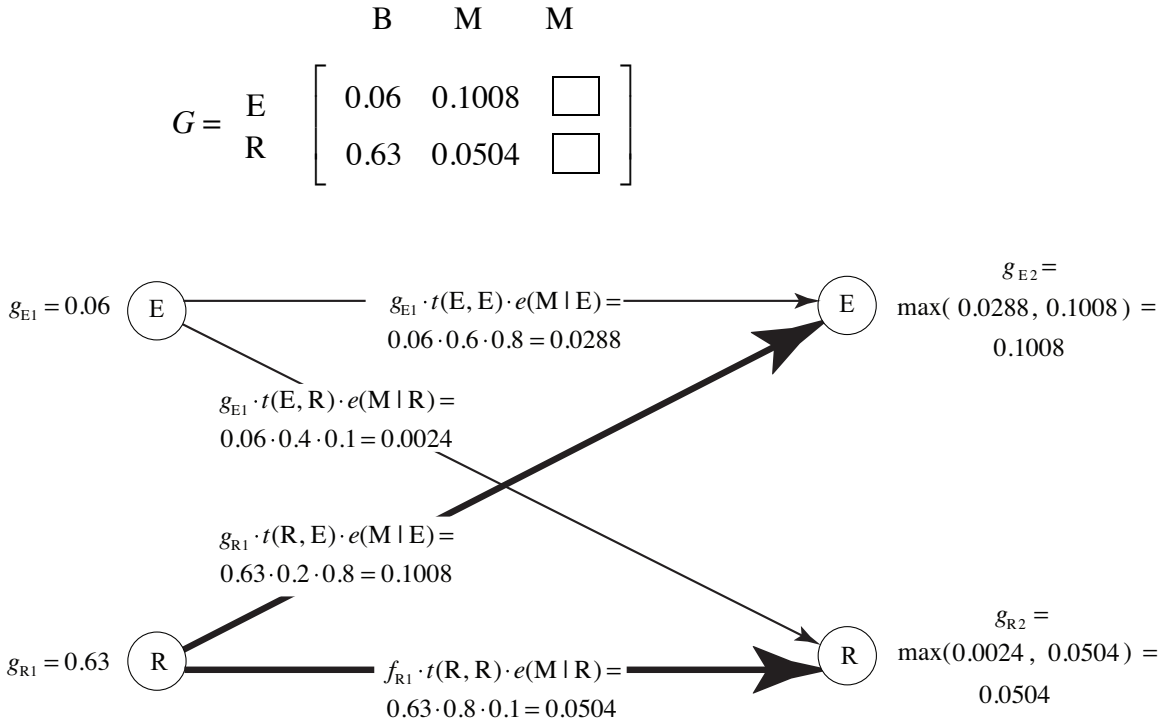following displays the developing Viterbi matrix, G:

$$\begin{array}{cc} & B \quad\quad M \quad\quad M \\ G = \begin{array}{c} E \\ R \end{array} & \left[\begin{array}{ccc} 0.06 & 0.1008 & \square \\ 0.63 & 0.0504 & \square \end{array} \right] \end{array}$$

$g_{E1} = 0.06$ (E)

$g_{E1} \cdot t(E, E) \cdot e(M \mid E) = $
$0.06 \cdot 0.6 \cdot 0.8 = 0.0288$

$g_{E1} \cdot t(E, R) \cdot e(M \mid R) = $
$0.06 \cdot 0.4 \cdot 0.1 = 0.0024$

$g_{R1} \cdot t(R, E) \cdot e(M \mid E) = $
$0.63 \cdot 0.2 \cdot 0.8 = 0.1008$

$g_{R1} = 0.63$ (R)

$f_{R1} \cdot t(R, R) \cdot e(M \mid R) = $
$0.63 \cdot 0.8 \cdot 0.1 = 0.0504$

(E) $g_{E2} = $
max(0.0288, 0.1008) =
0.1008

(R) $g_{R2} = $
max(0.0024 , 0.0504) =
0.0504

Figure 4   Calculation of the second column of Viterbi matrix, G

Quick Review Question 2 Suppose we wish to use the Viterbi algorithm to find the state sequence, u, with maximum $P(u, HHGH)$ for the HMM in Quick Review Question 1. As calculated in that question, the first column of the Viterbi-algorithm matrix G contains $g_{A1} = 0.2$, $g_{B1} = 0.4$, and $g_{C1} = 0.63$.
  a. The calculation of g_{B2} involves three expressions whose values are $0.0032, 0.0032$, and 0.0504. Calculate g_{B2}.
  b. Calculate g_{A2}.

Calculations of subsequent Viterbi-matrix elements for this example proceed in a similar fashion. With observation v_i, we employ the following evaluations for the elements of column i:

$$g_{Ei} = \max(g_{E(i-1)} \cdot t(E, E) \cdot e(v_i \mid E), \ g_{R(i-1)} \cdot t(R, E) \cdot e(v_i \mid E))$$
$$g_{Ri} = \max(g_{E(i-1)} \cdot t(E, R) \cdot e(v_i \mid R), \ g_{R(i-1)} \cdot t(R, R) \cdot e(v_i \mid R))$$

With boldface arrows indicating maximum values, Figure 5 illustrates the calculation of the final column of the Viterbi matrix. Note that there are two paths to R that yield the maximum, 0.004032. The completed Viterbi matrix follows:

$$
G = \begin{array}{c} \\ E \\ R \end{array} \begin{array}{ccc} B & M & M \\ \left[\begin{array}{ccc} 0.06 & 0.1008 & 0.048384 \\ 0.63 & 0.0504 & 0.004032 \end{array} \right] \end{array}
$$

In general, using Viterbi's algorithm for any HMM; for state, x; observation, v_i; and set of states, S, we have the following calculation for a Viterbi matrix element in row x and column i:

$$
g_{xi} = \max_{y \in S} \left(g_{y(i-1)} \cdot t(y,x) \cdot e(v_i \mid x) \right)
$$

$g_{E2} = 0.1008$ (E)

$g_{E2} \cdot t(E, E) \cdot e(M \mid E) =$
$0.1008 \cdot 0.6 \cdot 0.8 = 0.048384$

(E) $g_{E3} =$
max(0.048384, 0.008064) =
0.048384

$g_{E2} \cdot t(E, R) \cdot e(M \mid R) =$
$0.1008 \cdot 0.4 \cdot 0.1 = 0.004032$

$g_{R2} \cdot t(R, E) \cdot e(M \mid E) =$
$0.0504 \cdot 0.2 \cdot 0.8 = 0.008064$

$g_{R2} = 0.0504$ (R)

$g_{R2} \cdot t(R, R) \cdot e(M \mid R) =$
$0.0504 \cdot 0.8 \cdot 0.1 = 0.004032$

(R) $g_{R3} =$
max(0.004032, 0.004032) =
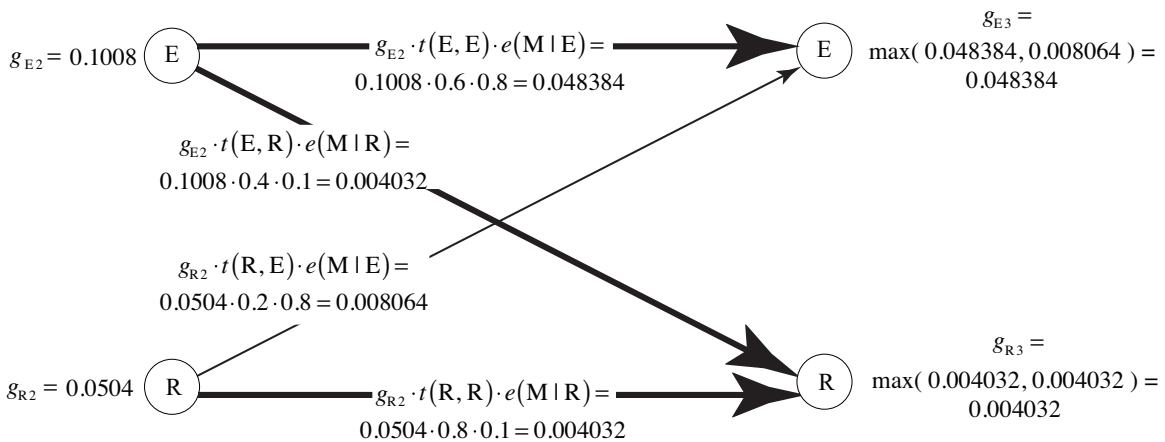0.004032

Figure 5   Step 3 of Viterbi algorithm in calculation of $P(\text{BMM})$

Quick Review Question 3   Suppose we wish to use the Viterbi algorithm to find the state sequence, u, with maximum $P(u, \text{HHGH})$ for the HMM in Quick Review Question 1. Suppose, also, that the third column of the Viterbi matrix G contains $g_{A3} = 0.0765$, $g_{B3} = 0.0340$, and $g_{C3} = 0.0142$. Calculate g_{C4} to four decimal places.

To calculate the probability of the visible sequence with the forward algorithm, we added the probabilities in the final column. However, to calculate the maximum joint probability of a hidden sequence and a given visible sequence using the Viterbi algorithm, we find the maximum of the values in the final column. Thus, for Holly's HMM, we have the following:

$$
\max_{u \in S^3}(u, \text{BMM}) = \max(0.048384, 0.004032) = 0.048384
$$

However, we would like to calculate $\max(P(u \mid \text{BMM}))$ for this u. Recall that

$$
P(u, \text{BMM}) = P(u \mid \text{BMM}) \cdot P(\text{BMM})
$$

Dividing both sides by the factor $P(\text{BMM})$, we have

$$P(u \mid \text{BMM}) = P(u, \text{BMM}) / P(\text{BMM})$$

Moreover, using the forward algorithm in (Shiflet et al. 2016), we discovered $P(\text{BMM}) = 0.080064$. Thus, over all three-element hidden sequences, u,

$$\max(P(u \mid \text{BMM})) = 0.048384 / 0.080064 = 0.604317$$

More important than finding this maximum probability, we would like to discover the particular state sequence that yields this maximum. Fortunately, by backtracking through the Viterbi matrix, we can determine this hidden sequence. Figure 6 summarizes results of Figures 4 and 5 with arrows indicating the expressions generating the maxima. To calculate the state sequence, u, that results in $\max(P(u \mid \text{BMM})) = 0.604317$, we start by finding the maximum in the final column, 0.048384, which is in row E. Backtracking through the path indicated by the arrows, we then go to column 2, row E and finally to column 1, row R. Reading the row values from left to right, we obtain the state sequence $u = \text{REE}$. Thus, given observed sequence BMM, REE is the most likely state sequence, and $P(\text{REE} \mid \text{BMM}) = 0.604317$, which is over 60%.
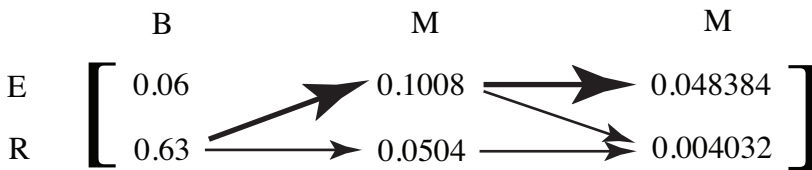

Figure 6   Final Viterbi matrix with arrows indicating the paths

Quick Review Question 4 Suppose we wish to use the Viterbi algorithm to find the state sequence, u, with maximum $P(u \mid \text{HHGH})$ for the HMM in Quick Review Question 1. Suppose, also, that $P(\text{HHGH}) = 0.1028$ and the Viterbi matrix, G, is in Figure 7, with arrows indicating the direction from which maxima came.
  a. Calculate the maximum $P(u, \text{HHGH})$ for hidden state sequence u.
  b. Calculate the maximum $P(u \mid \text{HHGH})$ for hidden state sequence u.
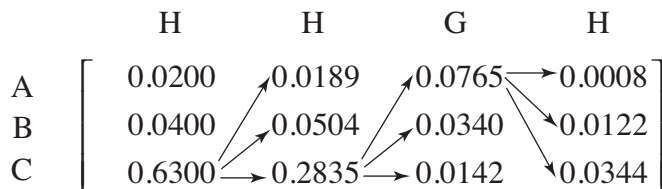 c. Give the u that achieves these maxima.

Figure 7   Viterbi matrix for Quick Review Question 4 with arrows indicating the direction from which maxima came

Detecting CpG Islands

As discussed in section "Probability of a Genomics Sequence" of (Shiflet et al. 2026), an area of greater frequency of the base sequence CG can be an indicator that a gene is to follow. The section presented initial probabilities, emission matrices, and possible transition matrices for samples within and not within such CpG islands, called positive and negative areas, respectively (Durbin *et al*. 1998). Suppose we also have transition probabilities from bases in positive areas (A^+, C^+, T^+, G^+) to bases in negative areas (A^-, C^-, T^-, G^-, respectively) and vice versa. Then, using the Viterbi algorithm, for a given observed sequence of bases from $\{A, C. T, G\}$, we can compute the most likely hidden sequence from the set of states, $S = \{\text{begin/end}, A^+, C^+, T^+, G^+, A^-, C^-, T^-, G^-\}$, where the sign indicates whether the base is probably in a CpG island or not. Project 2 considers such a decoding problem, where we can decode areas of high CpG concentration, containing bases A^+, C^+, T^+, and G^+.

Parallel Viterbi Algorithm

As with the forward algorithm, we can use high performance computing (HPC) to achieve faster results when a decoding problem involves a large number of states and/or observations. Moreover, we can parallelize the Viterbi algorithm similarly to the forward algorithm with OpenMP and threads communicating by reading and writing to the same matrix. Project 1 calculates the speedup that can be achieved using HPC with this algorithm.

Projects

1. **a.** Write a sequential program to calculate the probability of a sequence of observations using the Viterbi algorithm.
    b. Write a parallel version of this program.
    c. For large sequence length, time the parallel version for increasing numbers of threads. Produce a graph of the speedup versus the number of threads.

2. From the website containing this module, download *ProbabilitiesHumanV.txt*, which contains the transition matrix and other data described in the section "Detecting CpG Islands" from (Huson 2008). Also, from (*Homo Sapiens* 2001) download all or part of the DNA sequence on chromosome 19 of the human genome. Using (UCSC 2009), select several subsequences that straddle CpG islands. Subsequences that occur in a CpG island are displayed in green. Using a parallel program, for each downloaded sequence, determine the most likely hidden sequence and its probability. Do your results concur with those of (UCSC 2009)?

Answers to Quick Review Questions

1. $g_{A1} = 0.2, g_{B1} = 0.4,$ and $g_{C1} = 0.63$

2.
    a. $0.0504 = \max(0.0032, 0.0032, 0.0504)$

    b. $0.0189 = \max(0.0002, 0.0008, 0.0189)$

3. 0.0344 because of the following: $g_{A3} \cdot t(A, C) \cdot e(H \mid C) = 0.034425$; $g_{B3} \cdot t(B, C) \cdot e(H \mid C) = 0.01836$; $g_{C3} \cdot t(C, C) \cdot e(H \mid C) = 0.00639$; and the maximum of these expressions is 0.034425.

4.

    a. 0.0344, the maximum in the final column
    b. $0.33463 = 0.0344 / 0.1028$
    c. CCAC

References

Durbin, R., S. Eddy, A. Krogh, and G. Mitchison. 1998. *Biological Sequence Analysis*. Cambridge University Press, Cambridge.

Huson, Daniel. 2008. Chapter 8, "Markov chains and Hidden Markov Models," from course "Algorithms in Bioinformatics," University of Tübingen. https://ab.inf.uni-tuebingen.de/teaching/ss08/gbi/script/chapter08-hmms.pdf Accessed September 3, 2016

Shiflet, A., Shiflet, G., Couch, D., Guzzi, P., and Cannataro, M. 2016. "What are the chances?—Hidden Markov Models" https://wofford-ecs.org/ Accessed April 6, 2016.

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly. 2009. http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr19:571325-583493&hgsid=264592883&knownGene=pack&hgFind.matches=uc002loy.3. Accessed August 19, 2016.