EN2202 Pattern Recognition
Assignment 1 - HMM Signal Source

Fernando J. Iglesias García
fjig@kth.se

Bernard Hernández Pérez
bahp@kth.se

September 11, 2012

- Consider the following infinite duration HMM $\lambda = \{\text{q, A, B}\}$

$$q = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}; A = \begin{pmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{pmatrix} B = \begin{pmatrix} b_1(x) \\ b_2(x) \end{pmatrix} \tag{1}$$

where $b_1(x)$ is a scalar Gaussian density function with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$, and $b_2(x)$ is a similar distribution with mean $\mu_2 = 3$ and standard deviation $\sigma_2 = 2$.

  - *What are the values of $P(S_t = j)$, $j \in \{1, 2\}$ for $t = 1, 2, 3...$ theoretically calculated and the corresponding measured values?*

    The initial state probability distribution defines entirely $P(S_1 = j)$, this is

    $$\mathbf{P}_1 = q = \begin{pmatrix} P(S_1 = 1) \\ P(S_1 = 2) \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

    To obtain the probability distribution in the next state,

    $$\mathbf{P}_t = A^T \cdot \mathbf{P}_{t-1} \tag{2}$$

    Using the values of this particular problem it is obtained that

    $$\mathbf{P}_2 = \begin{pmatrix} 0.99 & 0.03 \\ 0.01 & 0.97 \end{pmatrix} \cdot \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

    This probability distribution is still equal to q, the probability of being in state 1 at timestep $t = 2$ is three times larger than being in state 2. This, together with the fact that the elements of A remain constant for all t, implies that the computations of $P(S_t = j)$, $j \in \{1, 2\}$ for $t > 1$ are identical to the ones shown above. Therefore, $P(S_t)$ is constant for all t or, in other words, the HMM is *stationary*.

    Using Markov chain's `rand` function to generate a sequence of length $T = 10\ 000$, the relative frequencies obtained experimentally of occurrence of states $S_t = 1$ and $S_t = 2$ are:

$\hat{P}(S_t = 1)$	$\hat{P}(S_t = 2)$
0.7341	0.2659

  - *What are the values of* $\mathrm{E}[X_t]$ *and* $\mathrm{Var}[X_t]$ *theoretically calculated and the corresponding measured values?*

    $$\begin{aligned} \mathrm{E}[X_t] = \mathrm{E}_{S_t}[X_t|S_t] =& \mathrm{E}[X_t|S_t = 1] \cdot P(S_t = 1) + \\ & \mathrm{E}[X_t|S_t = 2] \cdot P(S_t = 2) = \\ =& 0 \cdot 0.75 + 3 \cdot 0.25 = \boxed{0.75} \end{aligned}$$

$$\text{Var}[X_t] = \text{E}_{S_t}[\text{Var}_{X_t}[X_t|S_t]] + \text{Var}_{S_t}[\text{E}_{X_t}[X_t|S_t]]$$

$$\begin{aligned}\text{E}_{S_t}[\text{Var}_{X_t}[X_t|S_t]] =& \text{Var}[X_t|S_t = 1] \cdot P(S_t = 1) + \\ & \text{Var}[X_t|S_t = 2] \cdot P(S_t = 2) = \\ =& 1 \cdot 0.75 + 4 \cdot 0.25 = \underline{1.75}\end{aligned}$$

$$\begin{aligned}\text{Var}_{S_t}[\text{E}_{X_t}[X_t|S_t]] &= \text{E}_{S_t}[(\text{E}_{S_t}[X_t|S_t] - \text{E}_{S_t}[\text{E}_{S_t}[X_t|S_t]])^2] \\ &= 0.75 \cdot [(0 - 0.75)^2] + 0.25 \cdot [(3 - 0.75)^2] = \underline{1.6875}\end{aligned}$$

$$\text{Var}[X_t] = \text{E}_{S_t}[\text{Var}_{X_t}[X_t|S_t]] + \text{Var}_{S_t}[\text{E}_{X_t}[X_t|S_t]] = \boxed{3.4375}$$

Finally, using HMM's `rand` function to generate a sequence of $T = 10\,000$ output scalar random numbers, the mean and variance computed in MATLAB are

$\hat{\text{E}}[X_t]$	$\hat{\text{Var}}[X_t]$
0.7871	3.5510

– Figure 1 shows 500 contiguous observations generated by the HMM. The output of this HMM is characterized by two levels, one centered around 0 and the other around 3. These levels correspond to the states of the HMM. Also, it can be seen that the observations fluctuate considerably. This is one of the random components of the HMM and it stems from the definition of the observations as Gaussian density functions. Note too that the fluctuation in the lower level is smoother than the one in the higher level. This derives from the fact that the state characterized with a zero mean Gaussian has smaller standard deviation than the other state.

- Now a new HMM is created, identical to the one defined in Equation (1) but with both means equal to zero for the observation distributions of both states, i.e. $\mu_1 = \mu_2 = 0$. Figure 2 shows 500 observations generated by this new HMM. Comparing this plot with the one produced by the former HMM shown in Figure 1, in both of them two different patterns are observed. In other words, both HMMs have in common their number of states, equal to two. On the other hand, the output of the second HMM fluctuates around the same level whereas there are clearly two different levels in the output of the first HMM. This reflects the fact that the two states of the first HMM have Gaussian output distributions with different mean values while the Gaussians of the second HMM are both zero mean.

  Even though it may be probably more difficult to estimate the state sequence $\underline{\text{S}}$ of the underlying Markov chain of the second HMM from the output $\underline{\text{x}}$, it is still possible as long as the output distributions of the states are not exactly the same. In this case, the standard deviations are different.

- Consider the finite-duration Markov chain $\text{M} = \{q, A\}$ defined in (3), where the last column of A denotes the probability to stop from every state of the chain.

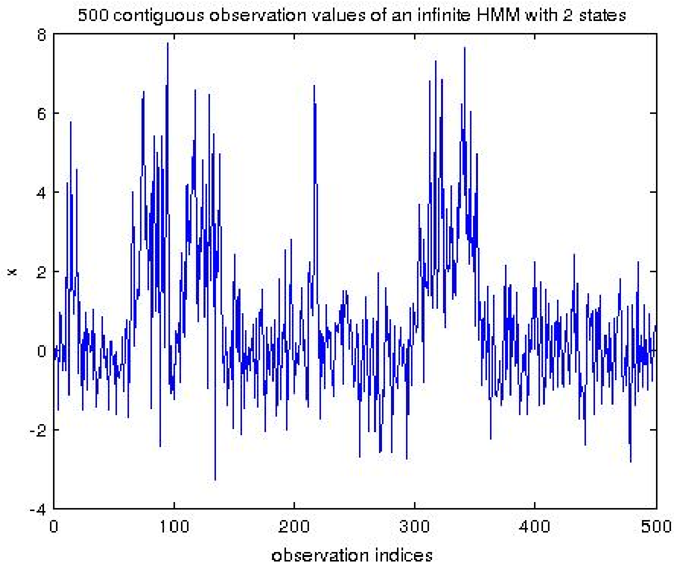

Figure 1: 500 contiguous output samples taken from the HMM defined in Equation (1).

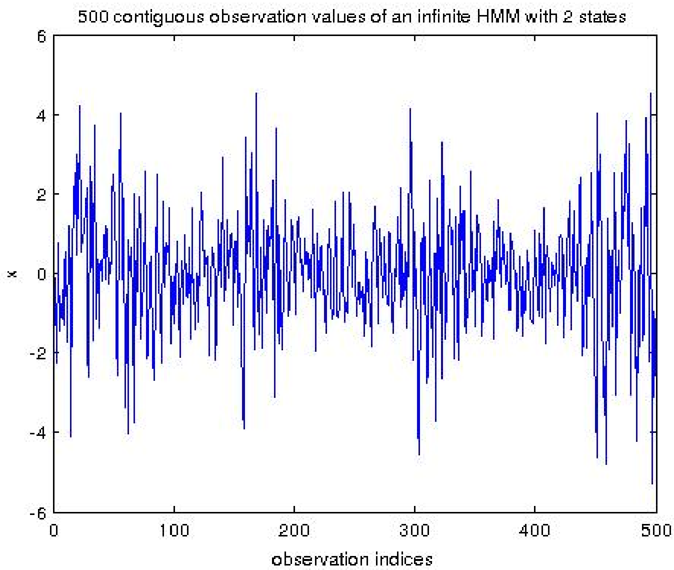

Figure 2: 500 contiguous output samples taken from the HMM defined in Equation (1) modified so both output distributions have zero mean.

$$q = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}; A = \left(\begin{array}{cc|c} 0.98 & 0.01 & 0.01 \\ 0.02 & 0.97 & 0.01 \end{array}\right) \tag{3}$$

Instead of defining a finite-duration HMM, only the Markov chain part has been defined since in this exercise just the length of the output sequences are of interest.

The probability to stop defined in Equation (3) is the same for every state and equal to 0.01 or, in other words, on average one out of a hundred transitions from any of the states the state sequence generated by the Markov chain ends.

The following piece of MATLAB code creates the Markov chain defined in (3), generates N sequences of states and computes their average length. Note that even if a large number is given to Markov chain's **rand** function, the sequence generated is likely to be shorter than that since this is a finite-duration Markov chain.

```
%% Length of sequences from a finite-duration HMM

mc = MarkovChain( [0.75; 0.25], ...
                  [0.98 0.01 0.01; 0.02 0.97 0.01]);

counter = 0;
N = 10;
for i = 1:N
  S   = mc.rand(1e4);
  counter = counter + length(S);
end

fprintf([ 'Average length of the states sequence ' ...
          'generated= %f\n'], counter/N);
```

The output of one execution gives 105.76 as result for the average length, value close to what was expected.

- Finally, the code used to verify that vector-valued output distributions work correctly is shown:

```
%% Verify use of vector-valued output distributions

mc = MarkovChain([0.75; 0.25], [0.99 0.01; 0.03 0.97]);

d = 10;   % dimension
a = 1.5;  % factor for the mean generation
b = 3;    % additive constant for the mean generation

% Generate a symmetric matrix to use as covariance
A = rand(10);
B = A+A';

% distribution for state 1
g1 = GaussD('Mean', a*randn(1,d), 'Covariance', B);
% distribution for state 2
g2 = GaussD('Mean', b+a*randn(1,d), ...
  'Covariance', diag(rand(1,10)));

h = HMM(mc, [g1; g2]);  % the HMM
X = h.rand(1e4);        % data generation
```