Probabilistic Modelling and Reasoning
Tutorial Sheet 7

School of Informatics, University of Edinburgh

Instructor: Dr Chris Williams

November 2011

1. A Hidden Markov Model problem.
Consider a HMM with 3 states ($M = 3$) and 2 output symbols, with a left-to-right state transition matrix

$$A = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.0 & 0.6 & 0.4 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}$$

an output probabilities matrix

$$B = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \\ 0.8 & 0.2 \end{pmatrix}$$

and an initial state probabilities vector $\pi = (0.9\ 0.1\ 0.0)$. Given that the observed symbol sequence $\mathbf{X}$ is 122, compute

(i) $P(\mathbf{X})$

(ii) $p(\mathbf{z}_2|\mathbf{X})$. [As there are 3 observations the HMM will have three time slices—you are asked to compute the posterior distribution of the state variable in the second time slice, numbering the times 1, 2, 3.] You can check this calculation by setting up the HMM in JavaBayes.

2. [Chatfield Ex 3.1] Consider the MA(2) model

$$x_t = w_t + 0.75w_{t-1} - 0.2w_{t-2}.$$

Compute the variance and covariance function for this process assuming that $\text{var}(w) = 1$.

3. Use the Kalman filtering equations given in section 13.3.1 of Bishop (and in the lecture notes) to derive the example of the simple Kalman filter given in the lecture slides, i.e. that for

$$z_{t+1} = z_t + w_t \qquad w_t \sim N(0, 1)$$
$$x_t = z_t + v_t \qquad v_t \sim N(0, 1)$$
$$p(z_1) = N(0, \sigma_0^2)$$

in the limit $\sigma_0^2 \to \infty$ we find that

$$\mu_3 = \frac{5x_3 + 2x_2 + x_1}{8}.$$

[Hint: show that $V_1 = 1$, $V_2 = 2/3$, $V_3 = 5/8$.]

4. **(Bonus question)** Modify question 2 so that $z_{t+1} = z_t$ (i.e. that w_t has zero variance). Using the same initial conditions and limit show that

$$\mu_3 = \frac{x_3 + x_2 + x_1}{3}.$$

Think about this result in relation to the recursive method of calculating averages

$$\overline{x}_n = \frac{n-1}{n}\overline{x}_{n-1} + \frac{1}{n}x_n,$$

initialized at $\overline{x}_1 = x_1$.

5. **(Bonus question)** Suppose the matrix A in the HMM (qu 1.) had its rows all equal to the initial probabilities vector π. In this case the HMM reduces to a simpler model—what is it?

6. **(Bonus question)** Show that if a transition probability a_{ij} in a HMM is set to zero initially, then it will remain at zero throughout training.

7. **(Bonus question)**. Read about the Viterbi algorithm for finding the best state sequence given a sequence of observations, and apply it to the model in question 1.