Notes on this question

This is similar in structure to a real exam question, and it should be fairly realistic, but it hasn't been checked in the way we would check a real question.

One should not assume that real exam questions will always correspond to one session of the course, as this does, although they will normally be reasonably coherent in topic.

The question starts with parts that I would expect everyone to be able to answer (assuming they remember the material). Part (c) requires more thought, and is intended to get progressively harder. The answers also become somewhat more open-ended.

The aim is to write questions where the mean mark is consistent, and where there is a good spread of marks. 20 out of 20 is intended to be unusual, but not impossible. The idea is to have questions where the answers are easy to mark objectively.

I try to set questions which are possibly useful as a learning resource in subsequent years. I would never set a question with a high proportion of bookwork, or which required a mini-essay.

Below, I have added some comments to the answer in bold. These answers are fuller than I would expect from an actual student response.

Answer to (a)

This is not bookwork, but only involves applying the notation from the notes to a straightforward situation, very similar to one in the practical.

The A matrix is the state transition probability matrix. The notation is a_{nm} where n and m are names of hidden states, which here correspond to the actual dice used (F,1,2). These correspond to the probability of moving from state n to state m. We also have a start state, 0, and an end state, E. There is a 4-by-4 matrix, with rows corresponding to start states 0, F, 1 and 2, and columns corresponding to end states F, 1, 2 and E.

$$A = \begin{bmatrix} a_{0F} & a_{01} & a_{02} & - \\ a_{FF} & a_{F1} & a_{FF} & a_{FE} \\ a_{1F} & a_{11} & a_{1F} & a_{1E} \\ a_{2F} & a_{21} & a_{2F} & a_{2E} \end{bmatrix}$$

$$a_{ij} = P(X_t = s_j | X_{t-1} = s_i)$$

B is the emission probability matrix. Terms correspond to the probability of each observation being associated with a particular hidden state.

$$b_i(k_j) = P(O_t = k_j | X_t = s_i)$$

Here, the B matrix has three rows, corresponding to the three hidden states (i.e., the dice), and 6 columns, corresponding to the numbers on the dice. The row corresponding to the fair dice will have values 1/6 for each cell.

$$B = \begin{bmatrix} b_F(1) & b_F(2) & b_F(3) & b_F(4) & b_F(5) & b_F(6) \\ b_1(1) & b_1(2) & b_1(3) & b_1(4) & b_1(5) & b_1(6) \\ b_2(1) & b_2(2) & b_2(3) & b_2(4) & b_2(5) & b_2(6) \end{bmatrix}$$

There is no need to write out every matrix element in an actual exam situation — use . . .

Answer to (b)

Due to shortage of time, I don't include a diagram here as would be expected in a real answer, but just explain what I'd expect a diagram to show.

There are 3 real hidden states, F, 1 and 2, all fully interconnected, all with transitions to themselves, and all emitting observations (i.e., dice rolls, 1, 2 etc) . The transition from F to 1 is labelled a_{F1}, the transition from 1 to F is labelled a_{1F} and the transition from 1 to 1 is labelled a_{11}. There is a start state 0, with transitions to each of F, 1 and 2, and an end state E, with transitions leading to it from F, 1 and 2.

The distinction between hidden and observed should be made clear in the diagram.

Answer to (ci)

This is straightforwardly captured with the HMM: it just corresponds to a zero probability state transition.

$a_{F2} = 0$

$P(X_t = s_2|X_{t-1} = s_F) = 0$

However, a rule of this type would not lead to a 0 probability in an HMM as usually used, because of smoothing.

Answer to (cii)

We cannot deduce much about $a_{FE} = 0$, which corresponds to $P(X_t = s_E|X_{t-1} = s_F) = 0$. Nevertheless, the counts in the training data will reflect the rule described: there will be no cases of transition between 1 and E or between 2 and E, and hence the counts for these events will be 0.

$a_{1E} = 0$

$P(X_t = s_E|X_{t-1} = s_1) = 0$

$a_{2E} = 0$

$P(X_t = s_E|X_{t-1} = s_2) = 0$

Answer to (ciii)

This rule is equivalent to saying that:

$$P(X_t = s_2|X_{t-1} = s_2, X_{t-2} = s_2) = 0$$

A second-order HMM would capture this. This doesn't correspond to anything in the first order HMM, however. All we can say is that, for the unsmoothed count:
$P(X_t = s_2|X_{t-1} = s_2) \leq 0.5$
$a_{22} \leq 0.5$
This follows because the most L2 transitions under this condition will occur if the dice is thrown exactly twice if it is thrown at all. In this case, there will be one L2 L2 transition and one L2 to not-L2 transition each time. This would correspond to:
$P(X_t = s_2|X_{t-1} = s_2) = 0.5$ Hence the observed probability has to be less that 0.5

Answer to (civ)

We have no direct access to this probability in the HMM since the transitions concern hidden states only. The condition could affect the HMM probabilities however, depending on the extent to which the loaded dice probabilities are skewed in favour of 6. For instance, take the extreme case where L1 always gives a 6: the croupier will always switch after throwing L1, and hence:
$P(X_t = s_1|X_{t-1} = s_1) = 0$