Supervision Assignments in MLRW - Set 2

Dionysis Manousakas

March 17, 2017

1 Dependence of p-values on irrelevant information -or why we should become Bayesians :-) (from McKay's book)

1.1 In an expensive laboratory, Dr. Bloggs tosses a coin labelled a and b twelve times and the outcome is the string

$$aaabaaaabaab,$$

which contains three bs and nine as. What evidence do these data give that the coin is biased in favour of a? Is it significant at the level of 5% ?

1.2 Dr. Bloggs pays careful attention to the calculation of 1.1, and responds *'no, no, the random variable in the experiment was not the number of bs: I decided before running the experiment that I would keep tossing the coin until I saw three bs; the random variable is thus the total number of tosses, n'*. A different calculation is required in order to assess the 'significance' of the result $n = 12$. Now, the probability distribution of n given $\mathcal{H}_0$ is the probability that the first $n - 1$ tosses contain exactly $r - 1$ bs and then the nth toss is a b. What evidence do these data give that the coin is biased in favour of a in this case? Is the evidence significant at the level of 5% ?

Answer 1.1. Dr. Bloggs consults his sampling theory friend who says 'let r be the number of bs and $n = 12$ be the total number of tosses; I view r as the random variable and find the probability of r taking on the value $r = 3$ or a more extreme value, assuming the null hypothesis $p_a = 0.5$ to be true'. She thus computes

$$P(r \leq 3|n = 12, \mathcal{H}_0) = \sum_{r=0}^{3} \binom{n}{r} \frac{1}{2^n} = 0.07$$

and reports 'at the significance level of 5%, there is not significant evidence of bias in favour of a'. Or, if the friend prefers to report p-values rather than simply compare p with 5%, she would report 'the p-value is 7%, which is not conventionally viewed as significantly small. If a two-tailed test seemed more appropriate, she might compute the two-tailed area, which is twice the above probability, and report the p-value is 15%, which is not significantly small'.

1.2. Now the sampling theorist has to compute the following probability :

$$P(n|\mathcal{H}_0, r) = \binom{n-1}{r-1} \frac{1}{2^n}$$

For Dr.Bloggs' experiment she gets

$$P(n \geq 12|r = 3, \mathcal{H}_0) = 0.03$$

She reports back to Dr. Bloggs, the p-value is 3% – there is significant evidence of bias after all!
Conclusion: The p-values of sampling theory do depend on the stopping rule.

2 HMM (from CMU 10-701, Fall 2015 Homework)

2.1 Let's use the notation $I < X, Y, Z >$: "X is conditionally independent of Z given Y". Assuming the notation of HMM in which x_t is the hidden state at time t and o_t is the observation at time t, which of the following are true of all HMMs?

1. $I < x_{t+1}, x_t, x_{t-1} >$
2. $I < x_{t+2}, x_t, x_{t-1} >$
3. $I < x_{t+1}, x_t, x_{t-2} >$
4. $I < o_{t+1}, o_t, o_{t-1} >$
5. $I < o_{t+2}, o_t, o_{t-1} >$
6. $I < o_{t+1}, o_t, o_{t-2} >$

2.2 Consider a HMM with 6 states (plus a start and end states) and an alphabet $\{A, C, G, T\}$. Table 1 lists the transition and emission probabilities, and Figure 1 shows the state diagram.

	0	S_1	S_2	S_3	S_4	S_5	S_6	f	A	C	G	T
0	0	1	0	0	0	0	0	0				
S_1	0	0	1	0	0	0	0	0	0.2	0.3	0	0.5
S_2	0	0	0	0.3	0	0.7	0	0	0.6	0.1	0.2	0.1
S_3	0	0	0	0	1	0	0	0	0.7	0	0.1	0.2
S_4	0	0	0	0	0	0	0	1	0.2	0.3	0.4	0.1
S_5	0	0	0	0	0	0	1	0	0.3	0.3	0.3	0.1
S_6	0	0	0	0	0	0	0	1	0.5	0.3	0	0.2

Table 1: The transition and emission probabilities.

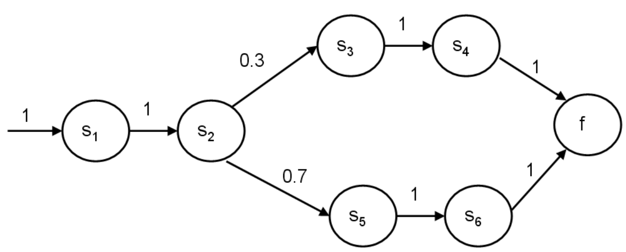


Figure 1: The state diagram of the HMM.

Let z denote latent variables and x denote observed variables. Place $<$, $>$, or $=$ between the two components of each of the following pairs. Justify your answer.

1. $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_3 = S_3, z_4 = S_4)$
   $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4)$
2. $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T)$
   $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4)$

3. $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_1 = S_1, z_2 = S_2)$
   $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_1 = S_1, z_2 = S_2)$

4. $P(x_1 = T, x_2 = C, x_3 = A, x_4 = T)$
   $P(x_1 = T, x_2 = A, x_3 = A, x_4 = G)$

2.3 Prove that $p(x_1, \ldots, x_i, z_i) = p(x_i|z_i) \sum_{z_{i-1}} p(x_1, \ldots, x_{i-1}, z_{i-1}) p(z_i|z_{i-1})$.

Answer

2.1 1, 2, 3, **TRUE**, 4,5,6 **FALSE**

2.2 1. Answer: $<$

$$\begin{aligned}
&P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_3 = S_3, z_4 = S_4) \\
&= P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4) P(z_3 = S_3, z_4 = S_4), \\
&P(z_3 = S_3, z_4 = S_4) = 0.3 < 1.
\end{aligned} \tag{1}$$

2.Answer: $<$

$$\begin{aligned}
&P(x_1 = T, x_2 = C, x_3 = A, x_4 = T) \\
&= P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_3 = S_3, z_4 = S_4) \\
&+ P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_3 = S_5, z_4 = S_6) \\
&= P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4) P(z_3 = S_3, z_4 = S_4) \\
&+ P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_5, z_4 = S_6) P(z_3 = S_5, z_4 = S_6) \\
&= (0.7 \times 0.1) \times 0.3 + (0.3 \times 0.2) \times 0.7.
\end{aligned} \tag{2}$$

Note that

$P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4) = 0.07 > P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_5, z_4 = S_6) = 0.06.$

From the above we conclude that

$$P(x_1 = T, x_2 = C, x_3 = A, x_4 = T) < P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_3 = S_3, z_4 = S_4)$$

3.

Answer:=

$$\begin{aligned}
&P(x_1 = T, x_2 = C, x_3 = A, x_4 = T, z_1 = S_1, z_2 = S_1) \\
&= P(x_1 = T, x_2 = C, x_3 = A, x_4 = T | z_1 = S_1, z_2 = S_2) P(z_1 = S_2, z_1 = S_2), \\
&P(z_1 = S_2, z_1 = S_2) = 1.
\end{aligned} \tag{3}$$

4. Answer: $<$ Since the first and third letters are the same, we only need to worry about the second and fourth. The left hand side is: $0.1(0.3 \times 0.1 + 0.7 \times 0.2) = 0.017$ while the right hand side is: $0.6(0.7 \times 0 + 0.3 \times 0.4) = 0.072$

2.3 From the definition of the HMM we have that:

$$P(x_1, ..., x_i, z_1, ..., z_i) = P(x_i|z_i) P(z_i|z_{i-1}) P(x_1, .., x_{i-1}, z_1, ..., z_{i-1}))$$

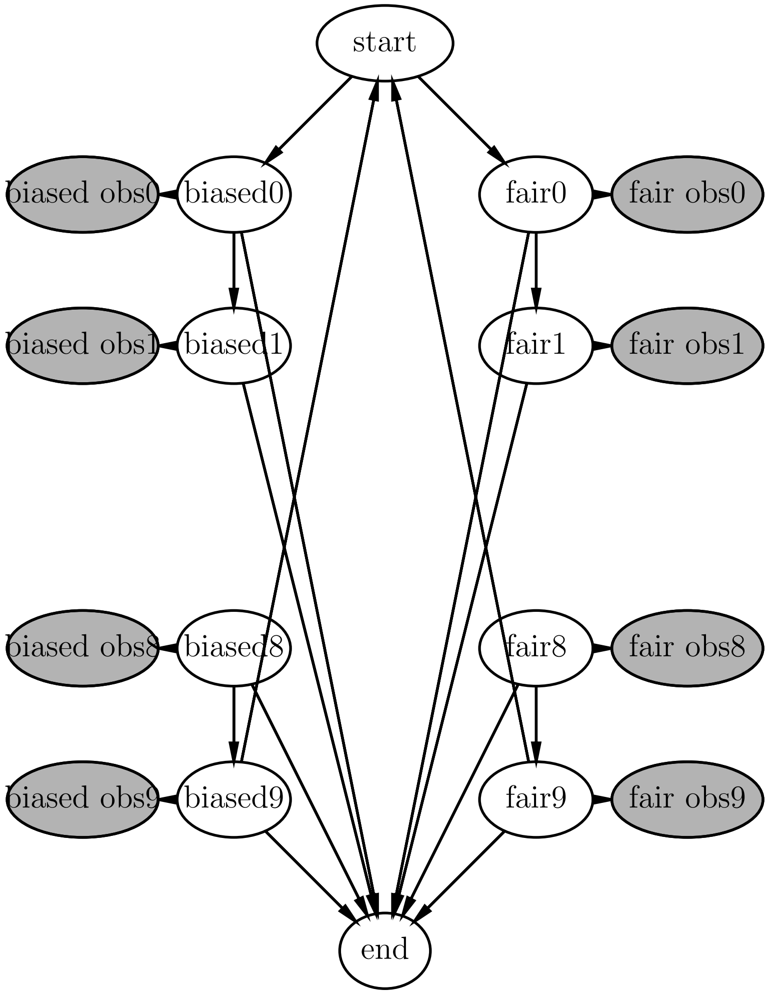
We can now marginalize out the variables $z_1, ..., z_{i-1}$ to obtain the following:

$$
\begin{aligned}
P(x_1, ..., x_i, z_i) &= \sum_{z_1,\dots,z_{i-1}} P(x_1, ..., x_i, z_1, ..., z_i) \\
&= \sum_{z_1,\dots,z_{i-1}} P(x_i|z_i)p(z_i|z_{i-1})P(x_1, .., x_{i-1}, z_1, ..., z_{i-1})) \\
&= P(x_i|z_i) \sum_{z_1,\dots,z_{i-1}} P(z_i|z_{i-1})P(x_1, .., x_{i-1}, z_1, ..., z_{i-1})) \\
&= P(x_i|z_i) \sum_{z_{i-1}} P(z_i|z_{i-1}) \sum_{z_1,\dots,z_{i-2}} P(x_1, .., x_{i-1}, z_1, ..., z_{i-1}) \\
&= P(x_i|z_i) \sum_{z_{i-1}} P(z_i|z_{i-1})P(x_1, .., x_{i-1}, z_{i-1})
\end{aligned} \tag{4}
$$

3 Weary dealer (from Pevzner's book)

To avoid suspicion, the dealer in the Fair Bet Casino keeps every coin for at least ten tosses, regardless of whether it is fair or biased. Describe the corresponding HMM.

Answer



4 HMM Complexity (from CMU 10-701 2003 Final Exams)

(True/False) In general when are trying to learn an HMM with a small number of states from a large number of observations, we can almost always increase the training data likelihood by permitting more hidden states.

Answer **True**: To model any finite length sequence, we can increase the number of hidden states in an HMM to be the number of observations in the sequence and therefore (with appropriate parameter choices) generate the observed sequence with probability 1. Given a fixed number of finite sequences (say n), we would still be able to assign probability $1/n$ for generating each sequence. This is not useful, of course, but highlights the fact that the complexity of HMMs is not limited.