CS221 Problem Workout Solutions

Week 8

1) [CA session] Problem 1: The Bayesian Bag of Candies Model

You have a lot of candy left over from Halloween, and you decide to give them away to your friends. You have four types of candy: **A**pple, **B**anana, **C**aramel, **D**ark-Chocolate. You decide to prepare candy bags using the following process.

- For each candy bag, you first flip a (biased) coin Y which comes up heads ($Y = \mathrm{H}$) with probability λ and tails ($Y = \mathrm{T}$) with probability $1 - \lambda$.
- If Y comes up heads ($Y = \mathrm{H}$), you make a **H**ealthy bag, where you:
  (a) Add one **A**pple candy with probability p_1 or nothing with probability $1 - p_1$;
  (b) Add one **B**anana candy with probability p_1 or nothing with probability $1 - p_1$;
  (c) Add one **C**aramel candy with probability $1 - p_1$ or nothing with probability p_1;
  (d) Add one **D**ark-Chocolate candy with probability $1 - p_1$ or nothing with probability p_1.
- If Y comes up tails ($Y = \mathrm{T}$), you make a **T**asty bag, where you:
  (a) Add one **A**pple candy with probability p_2 or nothing with probability $1 - p_2$;
  (b) Add one **B**anana candy with probability p_2 or nothing with probability $1 - p_2$;
  (c) Add one **C**aramel candy with probability $1 - p_2$ or nothing with probability p_2;
  (d) Add one **D**ark-Chocolate candy with probability $1 - p_2$ or nothing with probability p_2.

For example, if $p_1 = 1$ and $p_2 = 0$, you would deterministically generate: **H**ealthy bags with one **A**pple and one **B**anana; and **T**asty bags with one **C**aramel and one **D**ark-Chocolate. For general values of p_1 and p_2, bags can contain anywhere between 0 and 4 pieces of candy.

Denote A, B, C, D random variables indicating whether or not the bag contains candy of type **A**pple, **B**anana, **C**aramel, and **D**ark-Chocolate, respectively.

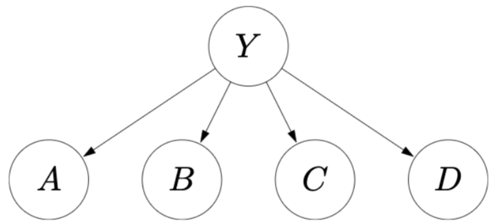

Figure 1: Bayesian network for a single candy bag.

a. (*1 point*)

(i) Draw the Bayesian network corresponding to process of creating a single bag.

Solution Solution for part (i) is shown in Figure 1.

(ii) What is the probability of generating a **H**ealthy bag containing **A**pple, **B**anana, **C**aramel, and not **D**ark-Chocolate? For compactness, we will use the following notation to denote this possible outcome:

$$(\mathbf{H}\text{ealthy}, \{\mathbf{A}\text{pple}, \mathbf{B}\text{anana}, \mathbf{C}\text{aramel}\}).$$

Solution By definition, we create a **H**ealthy bag with probability λ, and include the candies with probability $p_1 p_1 (1 - p_1) p_1$, so the result is

$$\lambda p_1 p_1 (1 - p_1) p_1$$

(iii) What is the probability of generating a bag containing **A**pple, **B**anana, **C**aramel, and *not* **D**ark-Chocolate?

Solution The bag could be **H**ealthy or **T**asty. We have computed the probability for the **H**ealthy case above. For a **T**asty one, a similar computation gives

$$(1 - \lambda) p_2 p_2 (1 - p_2) p_2$$

so the result is:

$$\lambda p_1 p_1 (1 - p_1) p_1 + (1 - \lambda) p_2 p_2 (1 - p_2) p_2$$

(iv) What is the probability that a bag was a **T**asty one, given that it contains **A**pple, **B**anana, **C**aramel, and *not* **D**ark-Chocolate?

Solution Using the definition of conditional probability, we get:

$$\frac{(1 - \lambda) p_2 p_2 (1 - p_2) p_2}{\lambda p_1 p_1 (1 - p_1) p_1 + (1 - \lambda) p_2 p_2 (1 - p_2) p_2}$$

b. (*1 point*)

You realize you need to make more candy bags, but you've forgotten the probabilities you used to generate them. So you try to estimate them looking at the 5 bags you've already made:

$$\begin{aligned} &bag\ 1: & (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\}) \\ &bag\ 2: & (\textbf{T}\text{asty}, \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\ &bag\ 3: & (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\}) \\ &bag\ 4: & (\textbf{T}\text{asty}, \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\ &bag\ 5: & (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\}) \end{aligned}$$

Estimate λ, p_1, p_2 by maximum likelihood.

Solution Out of 5 bags, 3 are **H**ealthy, so $\lambda = 3/5$. To estimate p_1, we only consider the 3 healthy bags. For a **H**ealthy bag, the probability of adding **A**pple,**B**anana, not **C**aramel, and not **D**ark-Chocolateis $(p_1)^4$. For the three bags, the probability becomes $(p_1)^{12}$, which is maximized for $p_1 = 1$. Equivalently, to generate 3 **H**ealthy bags, we flip a (biased) coin of parameter p_1 12 times. Since we observe 12 "heads", the maximum likelihood estimate is $p_1 = 1$. To generate 2 **T**asty bags, we flip a (biased) coin of parameter p_2 8 times. Since we observe 0 "heads", the maximum likelihood estimate is $p_2 = 0$.

- $$\lambda = 3/5$$
- $$p_1 = 12/12 = 1$$
- $$p_2 = 0/8 = 0$$

Estimate λ, p_1, p_2 by maximum likelihood, using Laplace smoothing with parameter 1.

Solution We just need to increment the counts in the previous solution by 1.

- $$\lambda = 4/7$$
- $$p_1 = 13/(13 + 1)$$
- $$p_2 = 1/(1 + 9)$$

c. **(**_1 point_**)** You find out your little brother had been playing with your candy bags, and had mixed them up (in a uniformly random way). Now you don't even know which ones were **H**ealthy and which ones were **T**asty. So you need to re-estimate λ, p_1, p_2, but now without knowing whether the bags were **H**ealthy or **T**asty.

$$\begin{aligned}
&\textit{bag } 1: && (?\ ,\ \{\textbf{A}\text{pple}, \textbf{B}\text{anana}, \textbf{C}\text{aramel}\}) \\
&\textit{bag } 2: && (?\ ,\ \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\
&\textit{bag } 3: && (?\ ,\ \{\textbf{A}\text{pple}, \textbf{B}\text{anana}, \textbf{C}\text{aramel}\}) \\
&\textit{bag } 4: && (?\ ,\ \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\
&\textit{bag } 5: && (?\ ,\ \{\textbf{A}\text{pple}, \textbf{B}\text{anana}, \textbf{C}\text{aramel}\})
\end{aligned}$$

You remember the EM algorithm is just what you need. Initialize with $\lambda = 0.5, p_1 = 0.5, p_2 = 0$, and run one step of the EM algorithm.

(i) E-step:

Solution To evaluate $P(Y = T \mid \{A, B, C\})$ we plug in the parameter values in the formula in (a),(iv), obtaining $P(Y = T \mid \{A, B, C\}) = 0$. To evaluate $P(Y = T \mid \{C, D\})$ we use a similar formula obtaining

$$P(Y = T \mid \{C, D\}) = \frac{(1-\lambda)(1-p_2)^4}{\lambda(1-p_1)^4 + (1-\lambda)(1-p_2)^4} = \frac{16}{17}$$

The resulting weighted dataset is:

- $(\textbf{H}\text{ealthy}, \{A, B, C\})$, 1×3
- $(\textbf{T}\text{asty}, \{A, B, C\})$, 0
- $(\textbf{H}\text{ealthy}, \{C, D\})$, $1/17 \times 2$
- $(\textbf{T}\text{asty}, \{C, D\})$, $16/17 \times 2$

(ii) M-step:

Solution Now we just do counts like in part (b). There are $3 + 2/17$ **H**ealthy bags out of 5. For p_1, each $(\textbf{H}\text{ealthy}, \{A, B, C\})$ corresponds to 3 "heads" and 1 "tail" (probability $p_1 p_1 (1-p_1) p_1$). Each $(\textbf{H}\text{ealthy}, \{C, D\})$ corresponds to 4 "tails" ($(1-p_1)^4$). For p_2, each $(\textbf{T}\text{asty}, \{C, D\})$ corresponds to 4 "tails" ($(1-p_2)^4$). The new parameters are:

$$\lambda = (3 + 2/17)/5$$
$$p_1 = 9/(9 + 3 + 4 * 2/17)$$
$$p_2 = 0$$

d. (*1 point*)

You decide to make candy bags according to a new process. You create the first one as described above. Then with probability μ, you create a second bag of the same type as the first one (Healthy or Tasty), and of different type with probability $1-\mu$. Given this type, the bag is filled with candy as before. Then with probability μ, you create a third bag of the same type as the second one (Healthy or Tasty), and of different type with probability $1-\mu$. And so on, you repeat the process M times. Denote Y_i, A_i, B_i, C_i, D_i the variables at each time step, for $i = 0, \ldots, M$. Let $X_i = (A_i, B_i, C_i, D_i)$.

Now you want to compute:

$$\mathbb{P}(Y_i = \textbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0))$$

exactly for all $i = 0, \ldots, M$, and you decide to use the forward-backward algorithm.

Suppose you have already computed the marginals:

$$f_i = \mathbb{P}(Y_i = \textbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0))$$

for some $i \geq 0$. Recall the first step of the algorithm is to compute an intermediate result *proportional* to

$$\mathbb{P}(Y_{i+1} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

(i) Write an expression that is **proportional** to

$$\mathbb{P}(Y_{i+1} = \textbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of f_i and the parameters p_1, p_2, λ, μ.

Solution Emission: When $Y_{i+1} = \textbf{H}\text{ealthy}$, the probability of observing $X_{i+1} = (1,1,1,0)$ is $p_1 p_1 (1-p_1) p_1$ as in part (a),(ii).
Transition: There are two cases: either $Y_i = \textbf{H}\text{ealthy}$, in which case we transit to $Y_{i+1} = \textbf{H}\text{ealthy}$ with probability μ, or $Y_i = \textbf{T}\text{asty}$, in which case we transit to $Y_{i+1} = \textbf{H}\text{ealthy}$ with probability $1-\mu$.

$$\propto ((1-f_i)(1-\mu) + f_i\mu) p_1 p_1 (1-p_1) p_1$$

(ii) Write an expression that is **proportional** to

$$\mathbb{P}(Y_{i+1} = \textbf{T}\text{asty} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of f_i and the parameters of the model p_1, p_2, λ, μ. The proportionality constant should be the same as in (i).

Solution (Similar to the previous question)
Emission: When $Y_{i+1} = \textbf{T}\text{asty}$, the probability of observing $X_{i+1} = (1,1,1,0)$ is $p_2 p_2 (1-p_2) p_2$.
Transition: There are two cases: either $Y_i = \textbf{H}\text{ealthy}$, in which case we transit to $Y_{i+1} = \textbf{T}\text{asty}$ with probability $1-\mu$, or $Y_i = \textbf{T}\text{asty}$, in which case we transit to $Y_{i+1} = \textbf{T}\text{asty}$ with probability μ.

$$\propto ((f_i)(1-\mu) + (1-f_i)\mu) p_2 p_2 (1-p_2) p_2$$

(iii) Let h be the answer for part (i), and t for part (ii). Write an expression for

$$\mathbb{P}(Y_{i+1} = \textbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of h, t and the parameters of the model p_1, p_2, λ, μ.

Solution Since h and t have same proportionality constant, we get the true value of the probability by normalization:

$$h/(h+t)$$

2) **Problem 2**

You are the president of the small nation of Inferencia, and you have been charged with choosing which of your country's two rival soccer teams - the Bayesians or the Markovians - should represent Inferencia at the upcoming Olympics. You'd like to send whichever team is more popular, so you decide to model the monthly evolution of the two teams' fanbases during the months leading up to the Olympics using a dynamic Bayesian network.

Let B_t denote the number of fans that the Bayesians have in month t, and let M_t denote the number of fans that the Markovians have in month t. You have no way of observing these quantities directly, but you can observe two other quantities which they influence: let J_t denote the number of jerseys sold by the Bayesians in month t, and let A_t denote the attendance of the monthly exhibition game between the Bayesians and the Markovians in month t.

The fanbases of the two teams evolve according to the following model, where each month a fan is either gained or lost with equal probability:

$$\Pr(M_{t+1}|M_t) = \begin{cases} \frac{1}{2} & \text{if } M_{t+1} = M_t - 1 \\ \frac{1}{2} & \text{if } M_{t+1} = M_t + 1 \\ 0 & \text{otherwise} \end{cases} \qquad \Pr(B_{t+1}|B_t) = \begin{cases} \frac{1}{2} & \text{if } B_{t+1} = B_t - 1 \\ \frac{1}{2} & \text{if } B_{t+1} = B_t + 1 \\ 0 & \text{otherwise} \end{cases}$$

The Bayesian fans are big spenders - almost every fan buys a jersey each month! We model the fanbase size's influence on jersey sales by:

$$\Pr(J_t|B_t) = \begin{cases} 0.3 & \text{if } J_t = B_t \\ 0.25 & \text{if } J_t = B_t - 1 \\ 0.2 & \text{if } J_t = B_t - 2 \\ 0.15 & \text{if } J_t = B_t - 3 \\ 0.1 & \text{if } J_t = B_t - 4 \\ 0 & \text{otherwise} \end{cases}$$

Lastly, because most fans attend each monthly exhibition (although sometimes more, and sometimes fewer), we model the influence of the fanbase sizes on the exhibition attendance by:

$$\Pr(A_t|B_t, M_t) = \begin{cases} 0.14 & \text{if } A_t = B_t + M_t \\ 0.13 & \text{if } |A_t - (B_t + M_t)| = 1 \\ 0.11 & \text{if } |A_t - (B_t + M_t)| = 2 \\ 0.09 & \text{if } |A_t - (B_t + M_t)| = 3 \\ 0.06 & \text{if } |A_t - (B_t + M_t)| = 4 \\ 0.04 & \text{if } |A_t - (B_t + M_t)| = 5 \\ 0 & \text{otherwise} \end{cases}$$

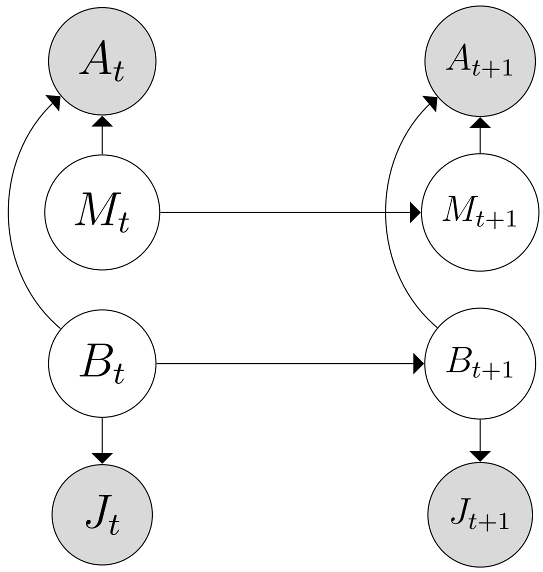

Figure 2: The changing fanbases process modeled as a dynamic Bayesian network. The unshaded nodes correspond to the latent/hidden fanbase counts, and the shaded nodes correspond to the observable emissions.

Note that the assumptions and inferences made in individual parts (i.e. **(a)**, **(b)**, etc.) of this problem do *not* carry over from one to the next; the only assumptions you may make in a given part are those which are explicitly stated in that part's description.

e. (*6 points*) **(Conditional) Independences**

Mark each of the following as True or False.

(i) [1 point] $B_t \perp\!\!\!\perp J_{t+1}$

Solution False

(ii) [1 point] $B_t \perp\!\!\!\perp J_{t+1} \mid B_{t+1}$

Solution True

(iii) [1 point] $B_t \perp\!\!\!\perp M_t$

Solution True

(iv) [1 point] $B_t \perp\!\!\!\perp M_t \mid A_t$

Solution False

(v) [1 point] $A_t \perp\!\!\!\perp M_{t+1}$

Solution False

(vi) [1 point] $A_t \perp\!\!\!\perp M_{t+1} \mid M_t$

Solution True

f. (*10 points*) **Inference**

Suppose the Bayesian's manager took a nationwide poll in month t that concluded they had exactly 75 fans. Suppose additionally that in month $t+2$, the Bayesians sell 73 jerseys. What is the probability that in month $t+2$ the Bayesians have 77 fans?

$\Pr(B_{t+2} = 77 | B_t = 75, J_{t+2} = 73) =$

Solution By Bayes rule, we have:

$$\Pr(B_{t+2} = 77 | B_t = 75, J_{t+2} = 73) = \frac{\Pr(J_{t+2} = 73 | B_t = 75, B_{t+2} = 77)\Pr(B_{t+2} = 77 | B_t = 75)}{\Pr(J_{t+2} = 73 | B_t = 75)}$$

We'll begin with the first term in the numerator; because J_{t+2} is conditionally independent of B_t given B_{t+2}, we have $\Pr(J_{t+2} = 73 | B_t = 75, B_{t+2} = 77) = \Pr(J_{t+2} = 73 | B_{t+2} = 77)$. This is simply given by our jersey sales model; the probability that the Bayesians sell four fewer jerseys than they have fans is 0.1.

We turn next to the second term in the numerator; if there are 75 fans in month t, then with equal probability there are either 74 or 76 fans in month $t+1$. If there were 74 in month $t+1$, then there would be either 73 or 75 in month $t+2$ with equal probability, and if there were 76 in month $t+1$, then there would be either 75 or 77 in month $t+2$ with equal probability. Thus, we have that $\Pr(B_{t+2} = 73 | B_t = 75) = \Pr(B_{t+2} = 77 | B_t = 75) = 0.25$, and $\Pr(B_{t+2} = 75 | B_t = 75) = 0.5$.

Now, to compute the denominator, we simply sum the expression in the numerator across all possible values for B_{t+2}:

$$\Pr(J_{t+2} = 73 | B_t = 75) = \sum_x \Pr(J_{t+2} = 73 | B_t = 75, B_{t+2} = x)\Pr(B_{t+2} = x | B_t = 75)$$

Following the same reasoning as we used for the numerator, this evaluates to:

$$= 0.25 \cdot \Pr(J_{t+2} = 73 | B_{t+2} = 73) + 0.5 \cdot \Pr(J_{t+2} = 73 | B_{t+2} = 75) + 0.25 \cdot \Pr(J_{t+2} = 73 | B_{t+2} = 77)$$

$$= 0.25 \cdot 0.3 + 0.5 \cdot 0.2 + 0.25 \cdot 0.1 = 0.075 + 0.1 + 0.025 = 0.2$$

So altogether, we have:

$$\Pr(B_{t+2} = 77 | B_t = 75, J_{t+2} = 73) = \frac{0.1 \cdot 0.25}{0.2} = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

g. (*4 points*) **Gibbs Sampling**

Inference is exhausting; you decide that you'd be satisfied with simply being able to draw samples from distributions rather than specifying them exactly. In particular, you want to sample joint assignments to the variables $\{B_t, M_t, A_t, J_t\}_{t=1}^{T}$ for some time horizon T. You decide to implement Gibbs sampling for this purpose, but something's not right! What additional information, beyond what we've given you, would allow you to perform Gibbs sampling? Briefly explain.

Solution (The following argument applies identically to M_t as well as B_t): In order to sample B_t, we need to have first assigned a value to B_{t-1}; but in order to have sampled a value for B_{t-1}, we need to have first assigned a value to B_{t-2}, and so on. Continuing in this way, we realize that we must have a way of assigning a value to B_1 in order to perform Gibbs sampling. But to do this, we would either need to specify a fixed value for B_1, or specify a prior distribution $\Pr(B_1)$ from which to sample.

h. (*12 points*) **Exact Filtering**

You now want to begin making inferences as to the sizes of the teams' fanbases given only observations of attendances and jersey sales. Recall that exact inference of this kind in dynamic Bayesian networks can be achieved using a dynamic programming approach - for example, in the context of Hidden Markov Models, we used the forward-backward algorithm to do filtering and smoothing.

Give recursive expressions for the following filtering queries. Leave your expressions in terms of known probabilities.

(i) [4 points] Let's start by making inferences based only on observed jersey sales. Denote $F_t(b_t) = \Pr(B_t = b_t | J_1 = j_1, \dots, J_t = j_t)$. Give a recursive expression for $F_t(b_t)$ assuming that you've already computed $F_{t-1}(b_{t-1})$ for all b_{t-1}.

Solution This is exactly the "forward" computation in an HMM. We can compute the unnormalized quantity, which we'll denote $\tilde{F}_t(b_t)$, using the standard forward update:

$$\tilde{F}_t(b_t) = \sum_{b_{t-1}} F_{t-1}(b_{t-1}) \cdot \Pr(B_t = b_t | B_{t-1} = b_{t-1}) \cdot \Pr(J_t = j_t | B_t = b_t)$$

and can subsequently produce the required probability by normalizing:

$$F_t(b_t) = \frac{\tilde{F}_t(b_t)}{\sum_{b'_t} \tilde{F}_t(b'_t)}$$

(ii) [8 points] Let's bring in the observed attendances as well! Now, denote $F_t(b_t, m_t) = \Pr(B_t = b_t, M_t = m_t | J_1 = j_1, \dots, J_t = j_t, A_1 = a_1, \dots, A_t = a_t)$. Give a recursive expression for $F_t(b_t, m_t)$ assuming that you've already computed $F_{t-1}(b_{t-1}, m_{t-1})$ for all b_{t-1} and all m_{t-1}.

Solution This closely mirrors the "forward" computation in an HMM, but now we must account for the dynamics of both hidden states, as well as the probabilities of both observed emissions. We can compute the unnormalized quantity, which we'll denote $\tilde{F}_t(b_t, m_t)$, using the following forward update:

$$\tilde{F}_t(b_t, m_t) =$$

$$\sum_{b_{t-1}, m_{t-1}} F_{t-1}(b_{t-1}, m_{t-1}) \cdot \Pr(B_t = b_t | B_{t-1} = b_{t-1}) \cdot \Pr(M_t = m_t | M_{t-1} = m_{t-1}) \cdot$$

$$\Pr(J_t = j_t | B_t = b_t) \cdot \Pr(A_t = a_t | B_t = b_t, M_t = m_t)$$

and can subsequently produce the required probability by normalizing:

$$F_t(b_t, m_t) = \frac{\tilde{F}_t(b_t, m_t)}{\sum_{b'_t, m'_t} \tilde{F}_t(b'_t, m'_t)}$$

i. (*10 points*) **Particle Filtering**

Throughout this problem, you are free to leave quantities in terms of unevaluated expressions (i.e. you may write $0.75 \cdot 0.5$ instead of 0.375).

Computing all of those terms exactly seems tedious, so you instead decide to employ particle filtering to quickly and painlessly provide you with approximate solutions. You're fine with a (very) crude approximation, so you only use two particles.

(i) [2 points] Suppose you begin with the two particles $(B_1 = 80, M_1 = 75)$ and $(B_1 = 82, M_1 = 74)$. You then observe that $J_1 = 79$ and $A_1 = 154$. Compute the weights that you should assign to the two particles based on this evidence.

Solution For the first particle, we have $\Pr(A_1 = 154 | B_1 = 80, M_1 = 75) = 0.13$ and $\Pr(J_1 = 79 | B_1 = 80) = 0.25$. Thus, the first particle should get a weight of $0.13 * 0.25 = 0.0325$.

Similarly, for the second particle, we have $\Pr(A_1 = 154 | B_1 = 82, M_1 = 74) = 0.11$ and $\Pr(J_1 = 79 | B_1 = 82) = 0.15$. Thus, the second particle should get a weight of $0.11 * 0.15 = 0.0165$.

(ii) [2 points] Using these weights, we now resample two new particles. Provide this sampling distribution.

Probability of sampling a new particle to be $(B_1 = 80, M_1 = 75) =$

Solution $\frac{0.0325}{0.0325+0.0165}$

Probability of sampling a new particle to be $(B_1 = 82, M_1 = 74) =$

Solution $\frac{0.0165}{0.0325+0.0165}$

(iii) [3 points] Suppose both of our new particles are sampled to be $(B_1 = 80, M_1 = 75)$. We now extend these particles using our dynamics models. What is the probability that a particular one of these two particles is extended to:

$(B_1 = 80, M_1 = 75, B_2 = 78, M_2 = 76)$?

Solution Zero. Under the given model for $\Pr(B_{t+1}|B_t)$, the only possible values for B_2 are 79 and 81.

$(B_1 = 80, M_1 = 76, B_2 = 79, M_2 = 75)$?

Solution Zero. The value assigned to M_1 cannot change upon extending the particle.

$(B_1 = 80, M_1 = 75, B_2 = 79, M_2 = 76)$?

Solution $\Pr(B_2 = 79|B_1 = 80) \cdot \Pr(M_2 = 76|M_1 = 75) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

(iv) [3 points] Suppose now that you have access to a large number of particles which are approximating the distribution over $(B_1, \ldots, B_n, M_1, \ldots, M_n)$. The Olympics are happening in 6 months, but you have to decide now which team to send so that they can start preparing! You decide to make predictions of B_{n+6} and M_{n+6} in order to send whichever team you predict to be more popular during the month in which the Olympics will be held. Explain in a few sentences how you would use your particles for making this decision.

Solution Propagate each particle through the two dynamics models six times in order to sample values of $B_{n+1}, \ldots, B_{n+6}$ and $M_{n+1} \ldots, M_{n+6}$ for each particle. Compute the average values of B_{n+6} and M_{n+6} across all of the particles, and send whichever team has the larger average value.