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: Apple, Banana, Caramel, Dark-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 ($Y = H$) with probability $\lambda$ and tails ($Y = T$) with probability $1 - \lambda$.

- If $Y$ comes up heads ($Y = H$), you make a Healthy bag, where you:
  (a) Add one Apple candy with probability $p_1$ or nothing with probability $1 - p_1$;
  (b) Add one Banana candy with probability $p_1$ or nothing with probability $1 - p_1$;
  (c) Add one Caramel candy with probability $1 - p_1$ or nothing with probability $p_1$;
  (d) Add one Dark-Chocolate candy with probability $1 - p_1$ or nothing with probability $p_1$.

- If $Y$ comes up tails ($Y = T$), you make a Tasty bag, where you:
  (a) Add one Apple candy with probability $p_2$ or nothing with probability $1 - p_2$;
  (b) Add one Banana candy with probability $p_2$ or nothing with probability $1 - p_2$;
  (c) Add one Caramel candy with probability $1 - p_2$ or nothing with probability $p_2$;
  (d) Add one Dark-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: Healthy bags with one Apple and one Banana; and Tasty bags with one Caramel and one Dark-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 Apple, Banana, Caramel, and Dark-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 Healthy bag containing Apple, Banana, Caramel, and not Dark-Chocolate? For compactness, we will use the following notation to denote this possible outcome:

\((\text{Healthy}, \{\text{Apple, Banana, Caramel}\})\).

**Solution**  By definition, we create a Healthy bag with probability \(\lambda\), 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 Apple, Banana, Caramel, and not Dark-Chocolate?

**Solution**  The bag could be Healthy or Tasty. We have computed the probability for the Healthy case above. For a Tasty 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 Tasty one, given that it contains Apple, Banana, Caramel, and not Dark-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:

<table>
<thead>
<tr>
<th>Bag</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Healthy, {Apple, Banana})</td>
</tr>
<tr>
<td>2</td>
<td>(Tasty, {Caramel, Dark-Chocolate})</td>
</tr>
<tr>
<td>3</td>
<td>(Healthy, {Apple, Banana})</td>
</tr>
<tr>
<td>4</td>
<td>(Tasty, {Caramel, Dark-Chocolate})</td>
</tr>
<tr>
<td>5</td>
<td>(Healthy, {Apple, Banana})</td>
</tr>
</tbody>
</table>

Estimate $\lambda, p_1, p_2$ by maximum likelihood.

**Solution** Out of 5 bags, 3 are Healthy, so $\lambda = 3/5$. To estimate $p_1$, we only consider the 3 healthy bags. For a Healthy bag, the probability of adding Apple, Banana, not Caramel, and not Dark-Chocolate is $(p_1)^4$. For the three bags, the probability becomes $(p_1)^{12}$, which is maximized for $p_1 = 1$. Equivalently, to generate 3 Healthy 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 Tasty 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 $\lambda, p_1, p_2$ by maximum likelihood, using Laplace smoothing with parameter $\lambda$.

**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 Healthy and which ones were Tasty. So you need to re-estimate $\lambda, p_1, p_2$, but now without knowing whether the bags were Healthy or Tasty.

```
bag 1 : (? , {Apple, Banana, Caramel})
bag 2 : (? , {Caramel, Dark-Chocolate})
bag 3 : (? , {Apple, Banana, Caramel})
bag 4 : (? , {Caramel, Dark-Chocolate})
bag 5 : (? , {Apple, Banana, Caramel})
```

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:

- (Healthy, \{A, B, C\}), 1 $\times$ 3
- (Tasty, \{A, B, C\}), 0
- (Healthy, \{C, D\}), 1/$17$ $\times$ 2
- (Tasty, \{C, D\}), 16/$17$ $\times$ 2

(ii) M-step:

**Solution** Now we just do counts like in part (b). There are 3 + 2/$17$ Healthy bags out of 5. For $p_1$, each (Healthy, \{A, B, C\}) corresponds to 3 “heads” and 1 “tail” (probability $p_1p_1(1-p_1)p_1$). Each (Healthy, \{C, D\}) corresponds to 4 “tails” ($(1-p_1)^4$). For $p_2$, each (Tasty, \{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 $\mu$, 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 $\mu$, 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:

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

exactly for all $i = 0, \ldots, M$, and you decide to use the forward-backward algorithm. Suppose you have already computed the marginals:

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

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

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

in terms of $f_i$ and the parameters $p_1, p_2, \lambda, \mu$.

**Solution**  

Emission: When $Y_{i+1} = \text{Healthy}$, the probability of observing $X_{i+1} = (1, 1, 1, 0)$ is $p_1p_1(1 - p_1)p_1$ as in part (a),(ii).

Transition: There are two cases: either $Y_i = \text{Healthy}$, in which case we transit to $Y_{i+1} = \text{Healthy}$ with probability $\mu$, or $Y_i = \text{Tasty}$, in which case we transit to $Y_{i+1} = \text{Healthy}$ with probability $1 - \mu$.

$$\propto ((1 - f_i)(1 - \mu) + f_i \mu)p_1p_1(1 - p_1)p_1$$
(ii) Write an expression that is proportional to

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

in terms of $f_i$ and the parameters of the model $p_1, p_2, \lambda, \mu$. The proportionality constant should be the same as in (i).

**Solution** (Similar to the previous question)

**Emission:** When $Y_{i+1} = \text{Tasty}$, 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 $Y_i = \text{Healthy}$, in which case we transit to $Y_{i+1} = \text{Tasty}$ with probability $1 - \mu$, or $Y_i = \text{Tasty}$, in which case we transit to $Y_{i+1} = \text{Tasty}$ with probability $\mu$.

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

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

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

in terms of $h, t$ and the parameters of the model $p_1, p_2, \lambda, \mu$.

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

$$h/(h + t)$$