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

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.
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{align*}
bag 1 : & \quad \text{(Healthy, \{Apple, Banana\})} \\
bag 2 : & \quad \text{(Tasty, \{Caramel, Dark-Chocolate\})} \\
bag 3 : & \quad \text{(Healthy, \{Apple, Banana\})} \\
bag 4 : & \quad \text{(Tasty, \{Caramel, Dark-Chocolate\})} \\
bag 5 : & \quad \text{(Healthy, \{Apple, Banana\})}
\end{align*}

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 \( p_2 = 0 \).

\[
\lambda = 3/5 \quad p_1 = 12/12 = 1 \quad p_2 = 0/8 = 0
\]

That was too easy, let’s try again:

\begin{align*}
bag 1 : & \quad \text{(Healthy, \{Apple, Banana, Caramel\})} \\
bag 2 : & \quad \text{(Tasty, \{Apple, Caramel, Dark-Chocolate\})} \\
bag 3 : & \quad \text{(Healthy, \{Banana, Caramel\})} \\
bag 4 : & \quad \text{(Tasty, \{Apple, Banana, Dark-Chocolate\})} \\
bag 5 : & \quad \text{(Healthy, \{Apple, Banana\})}
\end{align*}

Estimate \( \lambda, p_1, p_2 \) by maximum likelihood (i.e. counting and normalizing). Hint: Estimate \( p_{1/2} \) or \( 1 - p_{1/2} \) but not both.

**Solution** \( \lambda = 3/5 \) is the same since we still have 3/5 Healthy bags. For \( p_1 \) and \( p_2 \) we have parameter sharing. Thus we can count the number of times we get Apple or Banana and don’t get Caramel or Dark-Chocolate and use that to estimate \( p_1 \) or \( p_2 \) depending on the bag type. Define

\[
p_1 = \frac{|\text{Apple}| + |\text{Banana}| + |\neg \text{Caramel}| + |\neg \text{Dark-Chocolate}|}{4|\text{Healthy bags}|} \quad \text{in Healthy bags}
\]

\[
p_1 = \frac{9}{12}
\]
Similarly for \( p_2 \) we get \( 4/8 \). The hint is referring to not separating out Apple and Banana from Caramel and Dark-Chocolate, as that can lead to an incorrect result.

(c) 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.

\[
\begin{align*}
bag 1 : & \quad (?, \{\text{Apple, Banana, Caramel}\}) \\
bag 2 : & \quad (?, \{\text{Caramel, Dark-Chocolate}\}) \\
bag 3 : & \quad (?, \{\text{Apple, Banana, Caramel}\}) \\
bag 4 : & \quad (?, \{\text{Caramel, Dark-Chocolate}\}) \\
bag 5 : & \quad (?, \{\text{Apple, Banana, Caramel}\})
\end{align*}
\]

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. Hint: You might use conditional probabilities found in last week's problem session (2d) for this problem.

(i) E-step:

**Solution**

To evaluate \( P(Y = T | \{A, B, C\}) \) we plug in the parameter values in the formula in from last week:

\[
P(T|\{A, B, C\}) = \frac{P(A, B, C, \neg D | T)P(T)}{P(A, B, C, \neg D)}
\]

\[
= \frac{p_2p_2(1 - p_2)p_1(1 - \lambda)}{\lambda p_1p_1(1 - p_1)p_1 + (1 - \lambda)p_2p_2(1 - p_2)p_2}
\]

\[= 0\]

To evaluate \( P(Y = T | \{C, D\}) \) we use a similar formula obtaining

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

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. There are $3 + \frac{2}{17}$ Healthy bags out of 5. For $p_1$, each (Healthy, \{A, B, C\}) corresponds to 3 selections with $p_1$ and 1 with $1-p_1$ (probability $p_1p_1(1-p_1)p_1$). Each (Healthy, \{C, D\}) corresponds to 4 selections with $1-p_1$ (probability $(1-p_1)^4$). For $p_2$, each (Tasty, \{C, D\}) corresponds to 4 selections with $1-p_2$ (probability $(1-p_2)^4$). We can ignore the Tasty with 0 weight (why do we get zero weight? $p_2 = 0$ initial guess). The new parameters are:

$$\lambda = \frac{3 + \frac{2}{17}}{5}$$

$$p_1 = \frac{9}{9 + 3 + 4 \times \frac{2}{17}}$$

$$p_2 = 0$$

(d) 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 $X_i = (A_i, B_i, C_i, D_i)$. Note that in the figure below, each $X_i$ represents four separate nodes $A_i, B_i, C_i, D_i$, each with parent $Y_i$ (just like the earlier figure).

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 $p_1, p_2, \lambda, \mu$. 4
Solution  Emission: When $Y_{i+1} = \text{Healthy}$, the probability of observing $X_{i+1} = (1, 1, 1, 0)$ is $p_1 p_1 (1 - p_1) p_1$.

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 $1 - \mu$.

\[
\propto ((1 - f_i)(1 - \mu) + f_i \mu)p_1 p_1 (1 - p_1) p_1
\]

(ii) Write an expression that is proportional to

\[
P(\tilde{Y}_{i+1} = \text{Tasty} \mid X_0 = (1, 1, 1, 0), \ldots, X_i = (1, 1, 1, 0), X_{i+1} = (1, 1, 1, 0))
\]

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) (you don’t need to find it).

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

\[
P(\tilde{Y}_{i+1} = \text{Healthy} \mid X_0 = (1, 1, 1, 0), \ldots, X_i = (1, 1, 1, 0), X_{i+1} = (1, 1, 1, 0))
\]

in terms of $h, t$.

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

\[
h/(h + t)
\]