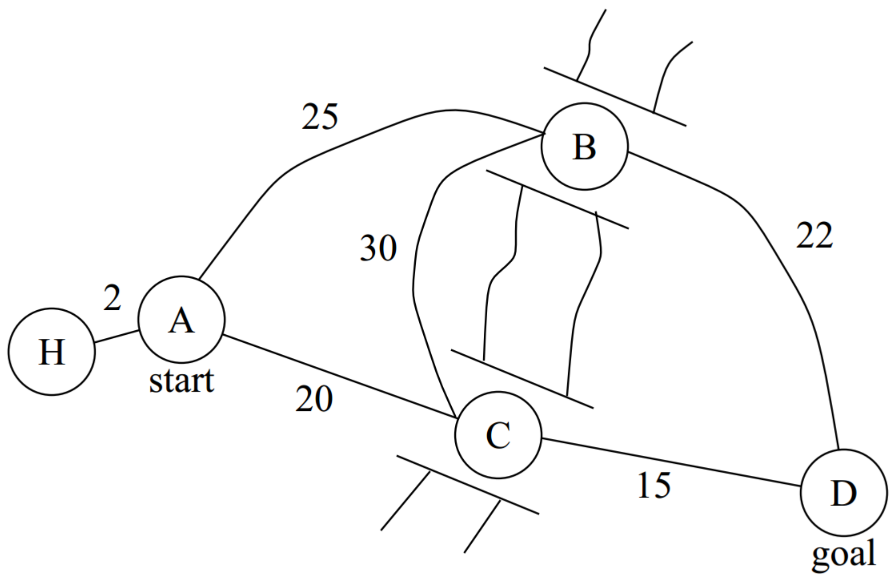
CS360 – Homework #12

Value of Information

1) A mobile robot is trying to get from his current position A to a destination D as quickly as possible. There is a river separating A from D and there are two bridges, B and C, spanning the river. The robot must design a strategy to move from A to D via one of the bridges.



Though the robot knows that one (and only one) of the bridges is inoperable, it is uncertain regarding which one of the two bridges is out. From its start position, position A, the robot can climb a hill to position H and use sensors to obtain information regarding which bridge is out. Collecting this information will take time, since it must go to H and return back to A, and the information gained is uncertain because the sensors will not be able to tell precisely which bridge is out.

The numbers in the above graph give the distance between locations in miles. In similar situations in the past, the robot experienced that 4 out of 5 times bridge C was out and only 1 out of 5 times bridge B was out. The robot has a short-range sensor that tells it with 100 percent reliability whether a bridge is out. The sensor can only be used when the robot is directly in front of the bridge. The long-range sensor of the robot is unreliable. It errs with a probability of 10 percent, that is, suggests that the broken bridge is operable and the other bridge is broken.

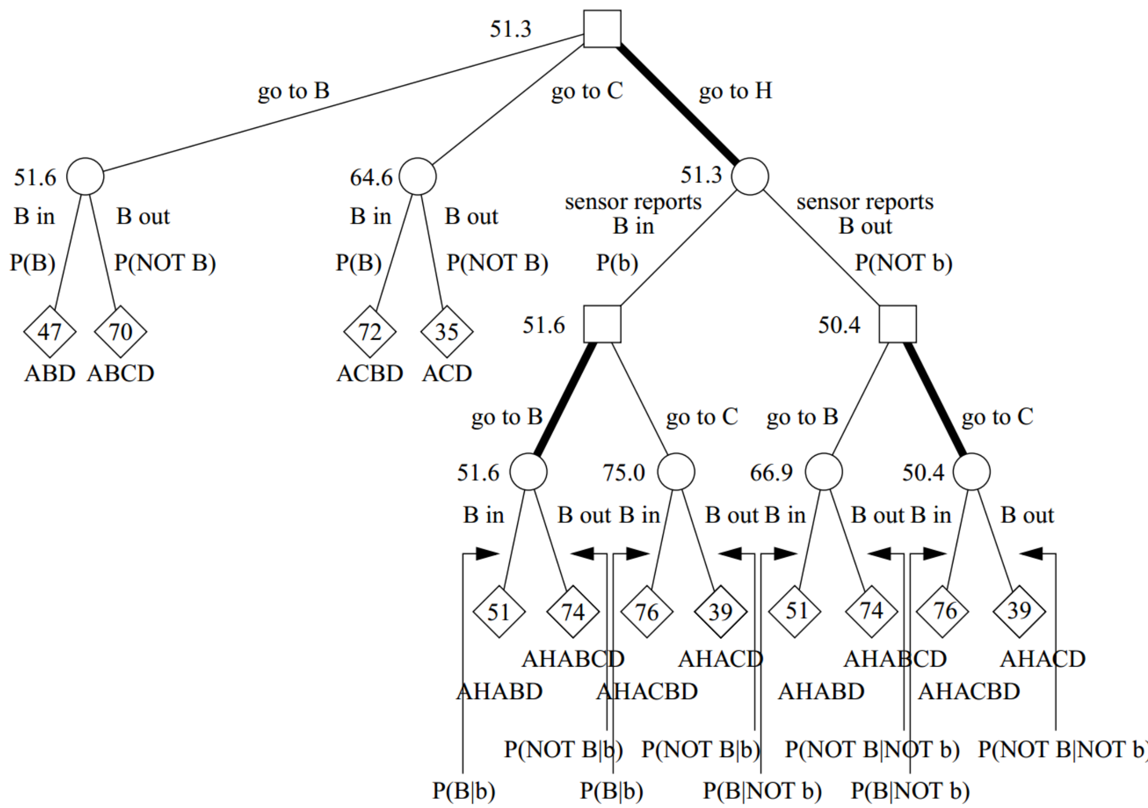
Design a strategy for the robot that minimizes the expected execution time.

The plan that minimizes the expected execution time is the following: the robot should go from A to H and use its long-range sensor. If the long-range sensor reports that bridge B is operational, the robot should then go to B via A and, if the bridge is indeed operational, continue to D. If the bridge is not operational, the robot has to continue to D via C. If the long-range sensor reports that bridge B is not operational, the robot should go to C via A and, if the bridge is indeed operational, continue to D. If the bridge is not operational, the robot has to continue to D via B. The expected execution time of this conditional plan is 51.3 minutes. The value of information of climbing the hill is 51.6 - 51.3 + 4.0 = 4.3 minutes.

The calculations are as follows:

```
B = bridge B is operable (and bridge C is broken)
C = bridge C is operable (and bridge B is broken)
b = long range sensor reports that bridge B is operable (and bridge
C is broken)
c = long range sensor reports that bridge C is operable (and bridge
B is broken)

P(B) = P(NOT C) = 4/5 = 0.8000
P(NOT B) = P(C) = 1/5 = 0.2000
P(NOT b | B) = P(c | B) = 0.1000
P(b | B) = P(NOT c | B) = P(b | NOT C) = P(NOT c | NOT C) = 0.9000
P(b | NOT B) = P(NOT c | NOT B) = P(b | C) = P(NOT c | C) = 0.1000
P(NOT b | NOT B) = P(c | NOT B) = P(NOT b | C) = P(c | C) = 0.9000
P(NOT b | B) = P(c | B) = P(NOT b | NOT C) = P(c | NOT C) = 0.1000
P(b) = P(b AND B) + P(b AND NOT B)
     = P(b | B) P(B) + P(b | NOT B) P(NOT B) = 0.7400
P(NOT b) = 1 - P(b) = 0.2600
P(B | b) = P(b | B) P(B) / P(b) = 0.9730
P(NOT B | b) = 1 - P(B | b) = 0.0270
P(B | NOT b) = P(NOT b | B) P(B) / P(NOT b) = 0.3077
P(NOT B | NOT b) = 1 - P(B | NOT b) = 0.6923
```



Expectation Maximization

2) Similar to the example in our favorite textbook, assume that your favorite Surprise candy comes in two flavors: cherry and lime. The manufacturer has a peculiar sense of humor and wraps each piece of candy in the same opaque wrapper, regardless of flavor. Surprise candies are sold in very large bags, of which there are known to be two kinds - again, indistinguishable from the outside. A bag of Type 1 contains $p_1 \times 100$ percent cherry candies and $(1 - p_1) \times 100$ percent lime candies, while a bag of Type 2 contains $p_2 \times 100$ percent cherry candies and $(1 - p_2) \times 100$ percent lime candies. When buying a bag of Surprise candy, one gets a bag of Type 1 with probability p and a bag of Type 2 with probability $1 - p$.

You buy four bags of Surprise candy at different stores and sample two candies from each bag. The following table shows the results:

Bag	Candy
Bag 1	One Cherry candy and one Lime candy
Bag 2	One Lime candy and one Cherry candy
Bag 3	Two Lime candies
Bag 4	Two Lime candies

a) You learned that bags 1 and 2 are of Type 1 and bags 3 and 4 are of Type

2. You don't know the values of p, p_1 and p_2. Develop a maximum-likelihood algorithm (that uses frequencies to estimate the corresponding probabilities). What estimates of p, p_1 and p_2 does it compute?

$p = 2/4 = 0.5$

$p_1 = 2/4 = 0.5$

$p_2 = 0/4 = 0$

b) You don't know the values of p, p_1 and p_2. Develop an expectation-maximization algorithm and run it until it converges, using $p = 0.1$, $p_1 = 0.2$ and $p_2 = 0.3$ as initial guesses. What estimates of p, p_1 and p_2 does it compute?

We use the notation 'aC + bL' to denote a bag that contains a cherry candies and b lime candies.

```
Iteration: 1

E-step:
Bag 1:
P(Type1, 1C + 1L) = 0.1 x 0.2^1 x 0.8^1 = 0.016
P(Type2, 1C + 1L) = 0.9 x 0.3^1 x 0.7^1 = 0.189
P(1C + 1L) = P(Type1, 1C + 1L) + P(Type2, 1C + 1L) = 0.016 + 0.189 = 0.205
P(Type1 | 1C + 1L) = P(Type1, 1C + 1L) / P(1C + 1L) = 0.016/0.205 = 0.07805
P(Type2 | 1C + 1L) = P(Type2, 1C + 1L) / P(1C + 1L) = 0.189/0.205 = 0.922

Bag 2:
P(Type1, 1C + 1L) = 0.1 x 0.2^1 x 0.8^1 = 0.016
P(Type2, 1C + 1L) = 0.9 x 0.3^1 x 0.7^1 = 0.189
P(1C + 1L) = P(Type1, 1C + 1L) + P(Type2, 1C + 1L) = 0.016 + 0.189 = 0.205
P(Type1 | 1C + 1L) = P(Type1, 1C + 1L) / P(1C + 1L) = 0.016/0.205 = 0.07805
P(Type2 | 1C + 1L) = P(Type2, 1C + 1L) / P(1C + 1L) = 0.189/0.205 = 0.922

Bag 3:
P(Type1, 0C + 2L) = 0.1 x 0.2^0 x 0.8^2 = 0.064
P(Type2, 0C + 2L) = 0.9 x 0.3^0 x 0.7^2 = 0.441
P(0C + 2L) = P(Type1, 0C + 2L) + P(Type2, 0C + 2L) = 0.064 + 0.441 = 0.505
P(Type1 | 0C + 2L) = P(Type1, 0C + 2L) / P(0C + 2L) = 0.064/0.505 = 0.1267
P(Type2 | 0C + 2L) = P(Type2, 0C + 2L) / P(0C + 2L) = 0.441/0.505 = 0.8733

Bag 4:
P(Type1, 0C + 2L) = 0.1 x 0.2^0 x 0.8^2 = 0.064
P(Type2, 0C + 2L) = 0.9 x 0.3^0 x 0.7^2 = 0.441
P(0C + 2L) = P(Type1, 0C + 2L) + P(Type2, 0C + 2L) = 0.064 + 0.441 = 0.505
P(Type1 | 0C + 2L) = P(Type1, 0C + 2L) / P(0C + 2L) = 0.064/0.505 = 0.1267
P(Type2 | 0C + 2L) = P(Type2, 0C + 2L) / P(0C + 2L) = 0.441/0.505 = 0.8733
```

```
Probability that a bag is of a given type:
        Type1   Type2
Bag 1:  7.80%   92.20%
Bag 2:  7.80%   92.20%
Bag 3:  12.67%  87.33%
Bag 4:  12.67%  87.33%

Expected number of candies in each bag, for each type:
        Type1                   Type2
Bag 1:  0.0780 C, 0.0780 L      0.9220 C, 0.9220 L
Bag 2:  0.0780 C, 0.0780 L      0.9220 C, 0.9220 L
Bag 3:  0.0000 C, 0.2535 L      0.0000 C, 1.7465 L
Bag 4:  0.0000 C, 0.2535 L      0.0000 C, 1.7465 L
Total:  0.1561 C, 0.6630 L      1.8439 C, 5.3370 L

M-step:
p1 = 0.1561/(0.1561 + 0.663) = 0.1906
p2 = 1.844/(1.844 + 5.337) = 0.2568
p = (0.07805 + 0.07805 + 0.1267 + 0.1267)/(0.07805 + 0.07805 + 0.1267 + 0.1267
    + 0.922 + 0.922 + 0.8733 + 0.8733) = 0.1024

-----------------------------------
Iteration: 2

E-step:
Bag 1:
P(Type1, 1C + 1L) = 0.1024 x 0.1906^1 x 0.8094^1 = 0.01579
P(Type2, 1C + 1L) = 0.8976 x 0.2568^1 x 0.7432^1 = 0.1713
P(1C + 1L) = P(Type1, 1C + 1L) + P(Type2, 1C + 1L) = 0.01579 + 0.1713 = 0.1871
P(Type1 | 1C + 1L) = P(Type1, 1C + 1L) / P(1C + 1L) = 0.01579/0.1871 = 0.08442
P(Type2 | 1C + 1L) = P(Type2, 1C + 1L) / P(1C + 1L) = 0.1713/0.1871 = 0.9156

Bag 2:
P(Type1, 1C + 1L) = 0.1024 x 0.1906^1 x 0.8094^1 = 0.01579
P(Type2, 1C + 1L) = 0.8976 x 0.2568^1 x 0.7432^1 = 0.1713
P(1C + 1L) = P(Type1, 1C + 1L) + P(Type2, 1C + 1L) = 0.01579 + 0.1713 = 0.1871
P(Type1 | 1C + 1L) = P(Type1, 1C + 1L) / P(1C + 1L) = 0.01579/0.1871 = 0.08442
P(Type2 | 1C + 1L) = P(Type2, 1C + 1L) / P(1C + 1L) = 0.1713/0.1871 = 0.9156

Bag 3:
P(Type1, 0C + 2L) = 0.1024 x 0.1906^0 x 0.8094^2 = 0.06708
P(Type2, 0C + 2L) = 0.8976 x 0.2568^0 x 0.7432^2 = 0.4958
P(0C + 2L) = P(Type1, 0C + 2L) + P(Type2, 0C + 2L) = 0.06708 + 0.4958 = 0.5629
P(Type1 | 0C + 2L) = P(Type1, 0C + 2L) / P(0C + 2L) = 0.06708/0.5629 = 0.1192
P(Type2 | 0C + 2L) = P(Type2, 0C + 2L) / P(0C + 2L) = 0.4958/0.5629 = 0.8808
```

```
Bag 4:
P(Type1, 0C + 2L) = 0.1024 x 0.1906^0 x 0.8094^2 = 0.06708
P(Type2, 0C + 2L) = 0.8976 x 0.2568^0 x 0.7432^2 = 0.4958
P(0C + 2L) = P(Type1, 0C + 2L) + P(Type2, 0C + 2L) = 0.06708 + 0.4958 = 0.5629
P(Type1 | 0C + 2L) = P(Type1, 0C + 2L) / P(0C + 2L) = 0.06708/0.5629 = 0.1192
P(Type2 | 0C + 2L) = P(Type2, 0C + 2L) / P(0C + 2L) = 0.4958/0.5629 = 0.8808

Probability that a bag is of a given type:
        Type1   Type2
Bag 1:  8.44%   91.56%
Bag 2:  8.44%   91.56%
Bag 3:  11.92%  88.08%
Bag 4:  11.92%  88.08%

Expected number of candies in each bag, for each type:
        Type1                   Type2
Bag 1:  0.0844 C, 0.0844 L      0.9156 C, 0.9156 L
Bag 2:  0.0844 C, 0.0844 L      0.9156 C, 0.9156 L
Bag 3:  0.0000 C, 0.2384 L      0.0000 C, 1.7616 L
Bag 4:  0.0000 C, 0.2384 L      0.0000 C, 1.7616 L
Total:  0.1688 C, 0.6455 L      1.8312 C, 5.3545 L

M-step:
p1 = 0.1688/(0.1688 + 0.6455) = 0.2073
p2 = 1.831/(1.831 + 5.354) = 0.2548
p = (0.08442 + 0.08442 + 0.1192 + 0.1192)/(0.08442 + 0.08442 + 0.1192 + 0.1192
     + 0.9156 + 0.9156 + 0.8808 + 0.8808) = 0.1018
```

The values for p, p_1 and p_2 converge as follows:

Itr	p	p1	p2
1	0.1024	0.1906	0.2568
2	0.1018	0.2073	0.2548
3	0.1015	0.22	0.2534
4	0.1013	0.2293	0.2523
5	0.1013	0.2358	0.2516
6	0.1012	0.2404	0.2511
7	0.1012	0.2435	0.2507
8	0.1012	0.2456	0.2505
9	0.1012	0.2471	0.2503
10	0.1012	0.248	0.2502
11	0.1012	0.2487	0.2501
12	0.1012	0.2491	0.2501

13	0.1012	0.2494	0.2501
14	0.1012	0.2496	0.25
15	0.1012	0.2497	0.25
16	0.1012	0.2498	0.25
17	0.1012	0.2499	0.25
18	0.1012	0.2499	0.25
19	0.1012	0.2499	0.25
20	0.1012	0.25	0.25

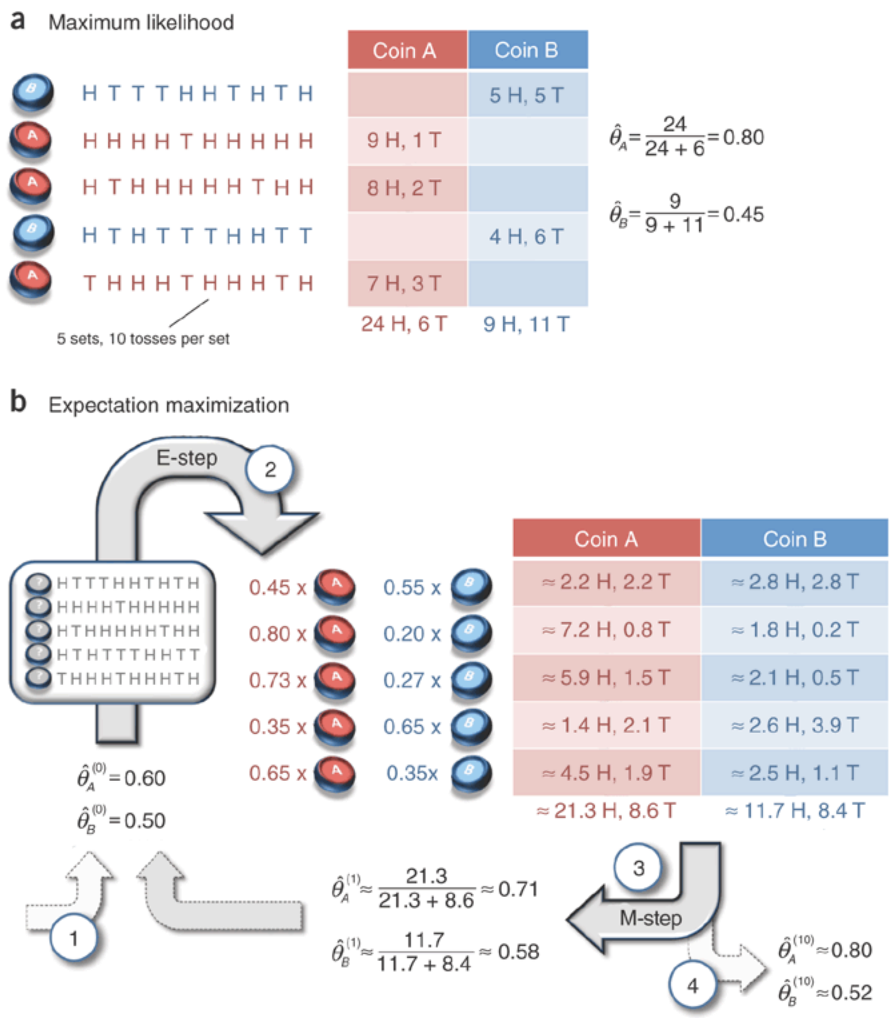
c) You know that $p_1 = 0.4$ but don't know the values of p and p_2. Develop an expectation-maximization algorithm and run it until it converges, using $p = 0.1$ and $p_2 = 0.3$ as initial guesses. What estimates of p, p_1 and p_2 does it compute?

The calculations are similar to b), but we do not update p1. The values for p and p_2 converge as follows:

Itr	p	p1	p2
1	0.1008	0.4	0.2383
2	0.0971	0.4	0.2411
3	0.0933	0.4	0.2416
4	0.08964	0.4	0.242
5	0.08612	0.4	0.2423
6	0.08273	0.4	0.2427
7	0.07948	0.4	0.243
8	0.07635	0.4	0.2433
9	0.07334	0.4	0.2435
10	0.07045	0.4	0.2438
11	0.06767	0.4	0.2441
12	0.065	0.4	0.2443
13	0.06243	0.4	0.2445
14	0.05996	0.4	0.2448
15	0.05759	0.4	0.245
16	0.05531	0.4	0.2452
17	0.05312	0.4	0.2454
18	0.05102	0.4	0.2456
19	0.049	0.4	0.2458
20	0.04706	0.4	0.2459
...			
100	0.00181	0.4	0.2498
1000	2e-19	0.4	0.25
10000	5e-179	0.4	0.25
...			
inf	0	0.4	0.25

3) Consider Figure 1 from "What is the expectation maximization algorithm" by Do and Batzoglou, which appeared in Nature Biotechnology, Volume 26, Number 8, August 2008, pages 897ff:

http://www.nature.com/nbt/journal/v26/n8/full/nbt1406.html



You can read only the following part of the text: "As an example, consider a simple coin-flipping experiment in which we are given a pair of coins A and B of unknown biases, Θ_A and Θ_B, respectively (that is, on any given flip, coin A will land on heads with probability Θ_A and tails with probability $1 - \Theta_A$ and similarly for coin B). Our goal is to estimate $\Theta = (\Theta_A, \Theta_B)$ by repeating the following procedure five times: randomly choose one of the two coins (with equal probability) and perform ten independent coin tosses with the selected coin. Thus, the entire procedure involves a total of 50 coin tosses (Fig. 1a)."

a) Explain the calculation that yields $\hat{\Theta}_A = 24/(24{+}6) = 0.80$ in Figure 1a. Is it important for this calculation that the coins were chosen with equal probability? Is this important anywhere in the maximum likelihood calculation?

It is the frequency of heads in the sets where coin A was used (24 heads, 6 tails, 30 total coin tosses). It is not important whether the coins were chosen with equal probability, because we know which coins are chosen in each set. As we see in b), it is important there to know the probability that the coins were chosen.

b) Explain how you can verify in Figure 1b that the two coins were chosen with equal probability. In this context, also explain how to calculate the value 0.45 shown in Figure 1b.

Given $\hat{\Theta}_A^{(0)} = 0.6$, $\hat{\Theta}_B^{(0)} = 0.5$, the fact that the two coins were chosen with equal probability and the observation "H T T T H H T H T H", we want to figure out the probability that coin A was used in the first set, that is, P(A | "H T T T H H T H T H"). We can derive the following:

P(A, "H T T T H H T H T H") = P(A) P("H T T T H H T H T H" | A) = $0.5 \times (0.6)^5 \times (0.4)^5 = 0.0003981312$

P(B, "H T T T H H T H T H") = P(B) P("H T T T H H T H T H" | A) = $0.5 \times (0.5)^5 \times (0.5)^5 = 0.00048828125$

P("H T T T H H T H T H") = P(A, "H T T T H H T H T H") + P(B, "H T T T H H T H T H") = 0.00088641245

P(A | "H T T T H H T H T H") = P(A, "H T T T H H T H T H") / P("H T T T H H T H T H") = 0.0003981312/0.00088641245 $\sim$ 0.45

This result also verifies that the two coins were chosen with equal probability because, otherwise, the result would have been different.

c) Explain why Figure 1b states that "$0.45 \times A$ $0.55 \times B = 2.2H, 2.2T$ for Coin A and $= 2.8H, 2.8T$ for Coin B." What does this mean? In this context, also explain why $\hat{\Theta}_A^{(1)} = 21.3/(21.3 + 8.6)$.

From the calculation in b), we know that there is a 45 percent chance that coin A was used in the first set (given our assumption that $\hat{\Theta}_A^{(0)} = 0.6$, $\hat{\Theta}_B^{(0)} = 0.5$). Since we observe 5 heads and 5 tails in the first set, we multiply them with 0.45 to get an expected number of 2.2 heads and 2.2 tails for A (and 2.8 heads and 2.8 tails for B) in the first set. When we do this calculation for each experiment, we observe that there is an expected number of 21.3 heads and 8.6 tails for A in total, allowing us to update the value of the bias $\hat{\Theta}_A$ (that is, the probability that A will land on heads) to $\hat{\Theta}_A^{(1)} = 21.3/(21.3 + 8.6)$.