CS 206: Homework 1

16:198:206

Problem 1: Forecasting the Flock (40 points)

You live a peaceful life in the post-apocalypse, collecting mushrooms, playing with your robot cat, and looking for non-radioactive water. It's been years since humanity and the Amazon Delivery Drone Swarm wiped each other out, but you still think sometimes, about hiding behind your couch, wondering if the next package you received would be your last. But what few survivors there are haven't seen any drones, or received any packages, in a very long time.

But something has changed. Somewhere, an automated factory seems to have sprung to life, churning out new drones - you've seen them zipping through the air at a distance, buzzing angrily as they look for someone to deliver things too. To try to gauge the size of the threat, you'd like to know how many drones there are - but it's difficult to count them.

Watching them through an antique telescope you recovered from the remains of the Museum of Natural History, you note that the drones are stamped with a serial number - surely marking the order they were manufactured? You make note of the serial numbers you see:

$$\#743, \#238, \#71, \#912, \#100$$

You assume that every drone is stamped with a serial number, probably starting from #1 and going to some number #N, where N is the total number of drones out there.

1) Assuming no gaps in serial numbers, what's the smallest number of drones that might be out there? **(2 points)**

if the serial number start from 1, then smallest number will be 912. If the serial number start from 71, then smallest number will be 841

2) What is the sample space of possible sets of 5-drone observations? How many possible observations are there? **(3 points)**

sample space will be{ {1,1,1,1,1}, {1,1,1,1,2}…..{N,N,N,N,N}}
possible observation will be N^5

3) If all drones have been deployed, and each is equally likely to be observed, what is the probability of observing the specific drones given above? **(5 points)**

1/(N^5)

4) What number of drones would make the probability of observing these specific drones as large as possible? **(5 points)**

if drone total numbers get smaller, then the probability of observing specific drones will be larger. Thus if serial number start from 1, 912 is the number. Or if serial number start from 71, 841 is the number

5) Intuitively, why is it unreasonable to imagine there are ten million drones in the wild? **(2 points)**

because #912 is the largest number that he saw, if there is ten million drones, then it is reasonable that he saw number that bigger than 912

6) In the general case, if drones $\{n_1, n_2, \ldots, n_k\}$ are observed, where n_i is the serial number of the i-th drone observed, what number of drones makes the probability of observing these specific drones as large as possible? **(3 points)**

if the range of (n1 …. nk) is the total number of drones, where nk is the largest number, then the probability is as large as possible

The answer to Q1.3 above tells you the number of drones that would make the observed data as likely as possible - but this still might not be the true number of drones. If N is the unknown number of drones, we'd like to know what the probability is that $N = n$, for any value n.

7) What is $P(N = 500)$? **(1 point)**

it means we only have 500 drones out there. Then it will be impossible for us to have #743 and #912, since each drones stamped with a serial number. Which means P(N=500) = 0, because it is false.

8) Assume that if n is a feasible value of N (i.e., it is realistic that there might be n drones, given the observations you've made), then $P(N = n)$ is proportional to the probability of observing drones $\{\#743, \#238, \#71, \#912, \#100\}$ when there are n total drones. What is the value of $P(N = n)$ as a function of n? As $n \to \infty$, wha can you say about $P(N = n)$? **(10 points)**

1. P(N=n) =2.767 * 10^12
2. As n -> ∞, P(N=n) will goes to 0

9) Using your answer to Q1.8 - find the smallest $n_{\max}$ such that $P(N < n_{\max}) = 0.95$, i.e., find an upper bound that you are 95% confident the true number of drones is less than. **(9 points)**

923.7708

- **Bonus 1:** How many more drones with serial numbers less than #700 would you have to observe before being 95% confident that there were at most 1000 drones? **(10 points)**
- **Bonus 2:** What assumptions in this analysis aren't necessarily accurate, and how could they be improved? **(5 points)**

Problem 2: Stun and Done (30 points)

Watching all these drones buzz back and forth is giving you eyestrain. But you can't just count every single drone - how could you ever know that you had caught them all?

You instead take the following approach: you build a taser-net that you can use to stun and catch drones (much more fun than just watching), before tagging them and re-releasing them into the wild. Working for a week, you manage to tag 400 drones, releasing them back into the wild. After waiting a week, you start catching drones again. Catching another 400 drones, you find that 270 of the drones you've caught are tagged, meaning you'd caught them before. 130 of your new catch are untagged, meaning this is the first time that you've caught them.

Assume drones were equally likely to be caught at all times. Before catching the second batch of drones, there were 400 tagged drones in the wild, the rest were untagged.

1) After catching the second batch of drones, what is the smallest number of drones N that might be in the wild, in total? **(2 points)** Assume first catch plus the second catch caught every drone, then 400+130 = 530 is the smallest number
2) What was the sample space of possible sets of 400 drones that you might catch in the second batch? How many possible outcomes is this? **(3 points)** 1. The sample space is total comes, Which means N choose 400.
2. N! / (400!(N-400)!)
3) What was the probability of catching exactly 270 tagged drones and 130 untagged drones, in the second batch? **(5 points)**

check the scratch sheet

4) What number of drones N makes your second batch results *as likely as possible*? **(10 points)**

N is the number between 591.59, and 592,59

5) Assuming again that the probability that $N = n$ is proportional to the probability of catching these second batch results with n total drones in the air, find again the smallest $n_{\max}$ so that $P(N \leq n_{\max}) = 0.95$. **(10 points)**

smallest Nmax is between 585.174 - 600.423

Bonus 3: Which method - Problem 1 or Problem 2 - do you think gives more accurate or reliable results? Why? *Be Thorough.* **(5 points)**

I think Problem 1gives us more accurate results, since we found P(N=n) =2.767 * 10^12. By using that, we can find the Nmax, and this Nmax is 95% accurate. However, for problem 2, especially for 2.4 and 2.5, the answer are in range, which is not as accurate as Problem 1.

Problem 3: Killing Time (30 points)

While you are waiting for your nets to fill, you play the following game: you have two coins, one is fair (heads/tails with equal probability), one is biased (heads twice as likely as tails). The coins are otherwise identical. You are going to choose a coin at random, flip it N many times, and then try to decide based on the results whether the coin is the fair coin or the biased coin. You win the game if you correctly identify the coin. (Your robot cat is able to distinguish which coin is which, and scores the game.)

You decide to follow the following strategy: if the majority of the flips are heads, you'll guess it is the biased coin, otherwise you'll guess that it is fair. That is, you're setting a **threshold** of $N/2$ such that if you get that many heads or more, you guess it is the biased coin, otherwise you guess it is the fair coin.

1) Following this strategy, what's the probability you win in a game of 10 flips? 20 flips? **(5 points)**

   P(win 10) = 0.65019 P(win 20) = 0.68713

2) If you want to maximize your probability of winning, what threshold should you set for your strategy in a game of 10 flips? 20 flips? Be thorough as to your methods, and precise. **(5 points)**

   For 10 flips, threshold should be 6. For 20, threshold should be 12

3) If the loaded coin had a probability $p > 1/2$ of giving H and $1 - p$ of giving T, where should you set your threshold for your strategy, to maximize your probability of winning in a game of N flips? **(13 points)**

   Maybe 0.6N, check the scratch sheet, I have two approachs

4) Taking p as in the original problem before, if you could decide in advance how many flips you wanted before guessing - how many flips should you ask for if you wanted as few flips as possible, but still at least a 95% chance of winning? (Assume you use an optimal threshold for deciding if it is the fair or biased coin.) **(7 points)**

Bonus 4: Imagine you have three coins: one is fair, one is biased (twice as likely to be heads as tails), the third is also biased (twice as likely to be tails as heads). You pick one of the three coins at random, and flip it 20 times. How should you decide which coin you are holding, based on the results, to maximize your probability of being correct? *Be Thorough.* **(5 points)**

For 20 flips, we already knows that from 3.2, 12 flips is the maximum chance of finding which one is biased head coin. Since two biased coin both have twice as likely as other. Then if 12 flips can also be the maximum chance of finding the biased tail coin. For the fair coin, if there are only 20-12 > 8, which is 9 flips, to <12, which is 11 flips are head or tail, then we should guess is fair coin. In conclusion, if we have 12 flips that are heads, then guess biased head coin, if we have 12 flips that are tail, then guess biased tail coin, if there are 9-11 (inclusive) head or tail, guess fair.