CSE 312: Foundations of Computing II Summer 2024

Homework 5

Due: Wednesday, July 31st, by 11:59pm

Instructions

See the instructions and FAQ for homeworks on the course website for important notes on the submission format!

Solutions submission. You must submit your solution via Gradescope. In particular:

- Submit a *single* PDF containing the solution to all of Tasks 1-5 to Gradescope under "**HW5 [Written]**".
- Each numbered task should be solved on its own page (or pages). Follow the prompt on Gradescope to link tasks to your pages.
- Do not write your name on the individual pages – Gradescope will handle that.
- Submit your code for Task 6 to Gradescope under "**HW5 [Coding]**".

Task 0 – Collaborators [0 pts]

List the full names of anyone you collaborated with on this homework. If you did not collaborate with anyone, write "None" in this section.

Task 1 – To Bake or To Burn [13 pts]

Claris is determined to learn how to not burn cookies! The average baking time for a batch of brownie cookies is 13 minutes. However, the baking times can very because of various factors such as oven and room temparature, and small variations in the dough. A pro-baker tells us that the baking time for a batch of brownie cookies is normally distributed with mean 13 and variance 4.

a) (5 points) What is the probability that the brownie cookies take more than 15 minutes?

b) (5 points) What is the probability the baking time for the brownie cookies is between 8 and 15?

c) (3 points) This week (7 days), we will make a batch of brownie cookies every day, independent of each other. What is the probability that the baking time was more than 15 minutes on **all** 7 days?

Task 2 – Minute Details [12 pts]

Amalia, a user experience researcher, is evaluating the average time users spend interacting with a new app feature. Each survey response provides an estimate of the interaction time with an unknown mean of T minutes and a variance of 4 minutes. To ensure that the average time spent estimate is accurate within ± 0.5 minutes with 95% confidence, how many user interactions should be surveyed?

Task 3 – 312's Entrepreneur [18 pts]

a) (8 points) In your startup, you've created a new computer vision algorithm with an average processing time of 10 seconds and a standard deviation of 3 seconds per image. You plan to run this algorithm on 50 images with each image being independent. What is the probability that the total time for all 50 images will be less than 8 minutes (i.e., 480 seconds)?

b) (10 points) After a few unsuccessful attempts to sell homemade bookmarks made out of cut-out printer paper, you've decided to try investing in a new line of personalized dog collars. You have 20 different collar designs to invest in, with each design costing \$15. Your research suggests that each collar design has a probability p of yielding a gross return of \$40 (resulting in a net profit of \$25), and a probability $1-p$ of resulting in no gross return (resulting in a \$15 loss).

To secure a small loan from your local bank, you need to show that the probability of your total net return being positive is at least 90%. What is the condition on p that needs to be true in order to secure the loan?

You should treat the amount of revenue you'll get from these investments as discrete (since the net return will only increase in multiples of \$25 and -\$15).

How is this important to us as computer scientists? If you're working with user studies or polls (a lot of human-computer interaction (HCI) research!), you might often want to figure out how many people you need to survey to draw a certain conclusion. You can estimate other conclusions about data once you have collected data from a representative sample. Or, you might want to analyze the performance of a certain algorithm that has some randomness. Notice how all the above questions can be tweaked to answer questions like what's described here!

Task 4 – Shear Patience [18 pts]

You own a salon where you have approximately 2 customers per hour (60 minutes). The timing of customers is independent of each other. You want to model the amount of time you wait between two customers, so you have a better idea of how much time you have to clean up and prepare between customers. Let X_i be the number of **minutes** you wait between the $i-1$'th customer leaves and i'th customer.

a) What distribution does X_i follow, and what are the parameters to this distribution? Remember to justify your answer. (3 points)

b) What is the value of m such that $\mathbb{P}(X_i < m) = \mathbb{P}(X_i > m)$? (in other words, at what time point is it equally likely that the customer would have already come and the customer is yet to come?) (6 points)

c) What is the probability that you need to wait for more than 1 hour till your first customer? (4 points)

d) What is the expected number of minutes you wait till your 10th customer? (5 points)

How is this important to us as computer scientists? Notice how we can use similar techniques to what you did in task 4, if you want to analyze to time between certain requests coming into a system or being sent out of it, the time till and between failures of an instance, etc.

Task 5 – Whale Watching [20 pts]

Avery is going whale watching for H hours, where H is a random variable, equally likely to be 1, 2 or 3. The number of whales W she sees is random and depends on how long she is in the area for. We are told that

$$\mathbb{P}(W = w \mid H = h) = \frac{c}{h}, \qquad \text{for } w = 1, \ldots, h,$$

for some constant c.

a) Compute c. You may want to use one of the axioms of probability. (3 points)

b) Find the joint distribution of W and H using the chain rule. (4 points)

c) Find the marginal distribution of W. (4 points)

d) Find the conditional distribution of H given that $W = 1$ (i.e., $Pr(H = h \mid W = 1)$ for each possible h in 1,2,3). Use the definition of conditional probability and the results from previous parts. (4 points)

e) Suppose that we are told that Avery saw either 1 or 2 whales. Find the expected number of hours she spent whale watching conditioned on this event. Use the definition of conditional expectation and conditional probability theorems. (5 points)
We will cover the content in this part in lecture on Friday (7/26).

Task 6 – Distinct Elements [Coding] [15 pts]

Start by going through the Ed lesson here, describing the distinct elements application. Let us know if you have any questions!

For the coding, we have set up an edstem lesson. However, you are required to upload your final solution to Gradescope (see instructions above).

a) Implement the functions UPDATE and ESTIMATE in the MinHash class of min_hash.py.

b) Implement the functions UPDATE and ESTIMATE in the MultMinHash class of min_hash.py using the improved estimator.