Data 102 Homework 5

Due: 11:59pm Friday, April 7, 2023

Overview

Submit your writeup, including any code, as a PDF via gradescope.[1] We recommend reading through the entire homework beforehand and carefully using functions for testing procedures, plotting, and running experiments. Taking the time to reuse code will help in the long run!

Data science is a collaborative activity. While you may talk with others about the homework, please write up your solutions individually. If you discuss the homework with your peers, please include their names on your submission. Please make sure any handwritten answers are legible, as we may deduct points otherwise.

Optimal Mean Estimation via Concentration Inequalities (cont.)

Recall the setup of Question 3 on Homework 4:

Suppose we observe a sequence of i.i.d. random variables $X_1, \dots, X_n$, each distributed according to an unknown distribution with a known variance σ^2 and an unknown mean μ that we would like to learn more about. In particular, we would like to estimate the true value of mean μ from the observations we have.

In that question, we spent some time analyzing the sample mean, and how well it functions as an estimator for mean μ. In part (c), we came to a startling observation: when we strive to achieve a high confidence (i.e setting $1 - \delta$ closer and closer to 1) in our estimate of μ, the number of samples n that we need to collect dramatically increases in situations where we cannot assume that X is bounded. This makes our usage of the sample mean a liability: since we don't know X's distribution, we cannot assume that X is bounded, and we'll have to collect thousands of samples just to reach our desired level of confidence.

Since we would ideally like to take advantage of Hoeffding's bound, we'll replace the sample mean with another estimator, one which involves the construction of bounded random variables. To construct this estimator, we'll start by considering m groups of X_i, each with fixed size n_0. We'll compute the sample mean for each group, and call these sample means $S^{(1)}, \dots, S^{(m)}$. Then, we'll use the median of all these group means as our estimate for the mean μ. The diagram below summarizes our approach.

$$\underbrace{\underbrace{X_1, X_2, \dots X_{n_0}}_{\text{Sample mean } S^{(1)}} \quad \underbrace{X_{n_0+1}, X_{n_0+2}, \dots, X_{2n_0}}_{\text{Sample mean } S^{(2)}} \quad \cdots \quad \underbrace{X_{n-n_0+1}, X_{n-n_0+2}, \dots, X_n}_{\text{Sample mean } S^{(m)}}}_{\textbf{Median } S_n}$$

We do this because even though one such sample mean $S^{(i)}$ might be far from the true mean μ, we hope (and will show) that the median of all of them is more likely to be close to

[1] In Jupyter, you can download as PDF or print to save as PDF

the true mean μ.

(a) (2 points) Fix a sample size $n_0 = \lceil \frac{4\sigma^2}{\epsilon^2} \rceil$. For each of the group means i, we define a binary random variable Z_i:

$$Z_i = \mathbb{1}(|S^{(i)} - \mu| \geq \epsilon).$$

In other words, Z_i is 0 if the corresponding group mean is close to the true mean μ (within ϵ), and 1 otherwise.

Show that $\mathbb{E}[Z_i] \leq 1/4$.

Hint: Z_i is a Bernoulli random variable.

(b) (2 points) We set $S_{\text{Med}} := \text{Median}(\{S^{(1)}, \ldots, S^{(m)}\})$. This is called the *median-of-means estimator*. Explain in words why having $|S_{Med} - \mu| \geq \epsilon$ implies that $\sum_{i=1}^m Z_i \geq \frac{m}{2}$.

Hint: If y is the median of m numbers, it means that $\lceil m/2 \rceil$ of the numbers are greater than or equal to y, and similarly $\lceil m/2 \rceil$ of the numbers are less than or equal to y.

(c) (2 points) By taking probabilities, part (b) implies

$$\mathbb{P}(|S_{\text{Med}} - \mu| \geq \epsilon) \leq \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m Z_i \geq \frac{1}{2}\right).$$

If we combine this fact with the result of (a), we can show that

$$\mathbb{P}(|S_{\text{Med}} - \mu| \geq \epsilon) \leq \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m \left(Z_i - \mathbb{E}[Z_i]\right) \geq \frac{1}{4}\right).$$

Now use Hoeffding's inequality to compute what number m is sufficient to ensure that $|S_{\text{Med}} - \mu| \leq \epsilon$ with probability at least $1 - \delta$. What is the final number of samples of X required?

Simulation Study of Bandit Algorithms

In this problem, we evaluate the performance of two algorithms for the multi-armed bandit problem. The general protocol for the multi-armed bandit problem with K arms and n rounds is as follows: in each round $t = 1, \ldots, n$ the algorithm chooses an arm $A_t \in \{1, \ldots, K\}$ and then observes reward X_t for the chosen arm. The bandit algorithm specifies how to choose the arm A_t based on what rewards have been observed so far. In this problem, we consider a multi-armed bandit for $K = 2$ arms, $n = 50$ rounds, and where the reward at time t is $X_t \sim \mathcal{N}(A_t - 1, 1)$, i.e. $\mathcal{N}(0, 1)$ for arm 1 and $\mathcal{N}(1, 1)$ for arm 2.

(a) (4 points) Consider the multi-armed bandit where the arm $A_t \in \{1, 2\}$ is chosen according to the explore-then-commit algorithm (below) with $c = 4$. Let $G_n = \sum_{t=1}^n X_t$ denote the total reward after $n = 50$ iterations. Simulate the random variable G_n a total of $B = 2000$ times and save the values $G_n^{(b)}$, $b = 1, \ldots, B$ in a list. Report the (empirical) average pseudoregret $\frac{1}{B}\sum_{b=1}^B \left(50\mu^* - G_n^{(b)}\right)$ (where μ^* is the mean of the best arm) and plot a normalized histogram of the rewards.

Algorithm 1 Explore-then-Commit Algorithm

input: Number of initial pulls c per arm
for $t = 1, \ldots, cK$: **do**
| Choose arm $A_t = (t \bmod K) + 1$
end
Let $\hat{A} \in \{1, \ldots, K\}$ denote the arm with the highest average reward so far.
for $t = cK + 1, cK + 2, \ldots, n$: **do**
| Choose arm $A_t = \hat{A}$
end

(b) (4 points) Consider the multi-armed bandit where the arm $A_t \in \{1, 2\}$ is chosen according to the UCB algorithm (below) with $c = 4$, $n = 50$ rounds. Repeat the simulation in Part (a) using the UCB algorithm, again reporting the (empirical) average pseudoregret and the histogram of $G_n^{(b)}$ for $b = 1 \ldots B$ for $B = 2000$. How does the pseudoregret compare to your results from part (a)?
Note: If $T_A(t)$ denote the number of times arm A has been chosen (up to and including time t) and $\hat{\mu}_{A,t}$ is the average reward from choosing arm A (up to and including t), then use the upper confidence bound $\hat{\mu}_{A,T_A(t-1)} + \sqrt{\frac{2\log(20)}{T_A(t-1)}}$. Note also that this algorithm is slightly different than the one used in lab and lecture as we are using an initial exploration phase.

Algorithm 2 UCB Algorithm

input: Number of initial pulls c per arm
for $t = 1, \ldots, cK$: **do**
| Choose arm $A_t = (t \bmod K) + 1$
end
for $t = cK + 1, cK + 2 \ldots$: **do**
| Choose arm A_t with the highest upper confidence bound so far.
end

(c) (1 point) Compare the distributions of the rewards by also plotting them on the same plot and briefly justify the salient differences.