CSE 312: Foundations of Computing II Autumn 2023

Problem Set 8

Due: Friday, December 1, by 11:59pm.

Instructions

Solutions format, collaboration policy, and late policy. See PSet 1 for further details. The same requirements and policies still apply. Also follow the typesetting instructions from the prior PSets.

Submit your solution via Gradescope as usual.

Task 1 – Chernoff Bound [20 pts]

A certain city is experiencing a terrible city-wide fire. The city decides that it needs to put its firefighters out into the streets all across the city to ensure that the fire can be put out. The city is conveniently arranged into a 50×50 grid of streets. Each street intersection can be identified by two integers (a, b) where $1 \leqslant a \leqslant 50$ and $1 \leqslant b \leqslant 50$. The city only has 600 firefighters, so it decides to send each firefighter to a uniformly random grid location, independent of each other (i.e., multiple firefighters can end up at the same intersection). The city wants to make sure that every 20×20 subgrid (corresponding to grid points (a, b) with $A \leqslant a \leqslant A + 19$ and $B \leqslant b \leqslant B + 19$ for valid A, B) gets more than 12 firefighters (subgrids can overlap).

a) Use the Chernoff bound (in particular, the version presented in class) to compute the probability that a single subgrid gets at most 12 firefighters. (remember to use the version presented in class because there are other versions of the Chernoff bound that might be found elsewhere!)

b) Use the union bound together with the result from above to calculate an upper bound on the probability that the city fails to meet its goal.

Task 2 – Is this Chern-ough? [20 pts]

A pharmaceutical researcher wants to estimate the expected fraction of original potency that is lost after 6 months of storage for randomly chosen batches of medication her company produces. Suppose that the true fractional potency loss for randomly chosen batches is distributed according to some unknown random variable P. She knows for sure that $\mathbb{E}[P] \geqslant 0.01$

To do this she will run n independent trials and measure the fractional loss X_i of the original potency at the end of 6 months for batches $i = 1, \ldots, n$, compute the sum $X = \sum_{i=1}^{n} X_i$, and output the average loss, $\overline{X} = X/n$ as her estimate of $\mathbb{E}[P]$.

Using only the version of the Chernoff bound that we have given in CSE 312, how many independent trials n will she need to *guarantee* that with probability 99%, the average fractional loss $\overline{X}$ that she measures is at most twice the true expected loss fraction $\mathbb{E}[P]$?

Task 3 – Lazy Grader [12 pts]

Prof. Lazy decides to assign final grades in CSE 312 by ignoring all the work the students have done and instead using the following probabilistic method: each student independently will be assigned an A with probability θ, a B with probability 2θ, a C with probability $\frac{1}{2}$, and an F with probability $\frac{1}{2} - 3\theta$. When the quarter is over, you discover that only 10 students got an A, 35 got a B, 40 got a C, and 15 got an F.

Find the maximum likelihood estimate for the parameter θ that Prof. Lazy used. Give an exact answer as a simplified fraction.

Task 4 – The Place That's Best [12 pts]

Let $y_1, y_2, ...y_n$ be i.i.d. samples of a random variable from the family of distributions $Y(\theta)$ with densities

$$f(y;\theta) = \frac{1}{2\theta} \exp\left(-\frac{|y|}{\theta}\right) ,$$

where $\theta > 0$. Find the MLE for θ in terms of $|y_i|$ and n.

Task 5 – Elections [14 pts]

Individuals in a certain country are voting in an election between 3 candidates: A, B and C. Suppose that each person makes their choice independent of others and votes for candidate A with probability θ_1, for candidate B with probability θ_2 and for candidate C with probability $1 - \theta_1 - \theta_2$. (Thus, $0 \leqslant \theta_1 + \theta_2 \leqslant 1$.) The parameters θ_1, θ_2 are unknown.
Suppose that $x_1, \ldots, x_n$ are n independent, identically distributed samples from this distribution. (Let $n_A =$ number of x_i's equal to A, let $n_B =$ number of x_i's equal to B, and let $n_C =$ number of x_i's equal to C.) What are the maximum likelihood estimates for θ_1 and θ_2 in terms of n_A, n_B, and n_C? **Unsimplified answers will not receive full points.**
(You don't need to check second order conditions.)