Problem Set 7

Due: Wednesday, May 25, 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.

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

- Submit a *single* PDF file containing the solution to all tasks in the homework. 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.
- We encourage you to typeset your solution. The homepage provides links to resources to help you doing so using LaTeX. If you do use another tool (e.g., Microsoft Word), we request that you use a proper equation editor to display math (MS Word has one). For example, you should be able to write $\sum_{i=1}^{n} x^i$ instead of $\text{x\^{}1 + x\^{}2 + ... + x\^{}n}$. You can also provide a handwritten solution, as long as it is on a single PDF file that satisfies the above submission format requirements. It is your responsibility to make sure handwritten solutions are readable – we will *not* grade unreadable write-ups.

Task 1 – Knitting Requires Concentration [20 pts]

Bob is slowly knitting a blanket, made of 100 squares. It takes an average of 1 hour for Bob to knit a square, with a standard deviation of 0.4 hours. The time to knit each square is independent. (You should treat time as continuous for this problem.)

a) What is the expectation of the total time to knit the blanket?

b) What is the variance of the total time to knit the blanket?

c) Bob will have 150 hours to knit between now and when he needs the blanket to be finished to stay warm at a football game. Use Markov's Inequality to give a *lower bound* the probability that Bob finishes the blanket before the game.

d) Can we improve the lower bound from **c)** using Chebyshev's inequality?

Task 2 – 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 100×100 grid of streets. Each street intersection can be identified by two integers (a, b) where $1 \leqslant a \leqslant 100$ and $1 \leqslant b \leqslant 100$. The city only has 1000 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 30×30 subgrid (corresponding to grid points (a, b) with $A \leqslant a \leqslant A + 29$ and $B \leqslant b \leqslant B + 29$ for valid A, B) gets more than 10 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 10 firefighters.

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 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 3θ, a C with probability $\frac{1}{2}$, and an F with probability $\frac{1}{2} - 4\theta$. When the quarter is over, you discover that only 2 students got an A, 10 got a B, 60 got a C, and 40 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 – Maximum Likelihood Estimators [24 pts]

a) Let $x_1, \ldots, x_n$ be i.i.d samples that follow a Two(θ) distribution with unknown parameter $\theta \in [0, 1]$, where the probabilities from the family are given by

$$\mathbb{P}(x; \theta) = \begin{cases} (1-\theta)^2 & x = 0 \\ 2\,\theta(1-\theta) & x = 1 \\ \theta^2 & x = 2 \end{cases}$$

Suppose that in the sample there are n_0 0's, n_1 1's, and n_2 2's.

What is the maximum likelihood estimator for θ in terms of n, n_0, n_1, n_2?

b) Let $x_1, \ldots, x_n$ be i.i.d. samples from a random variable that follow a so-called Borel distribution with unknown parameter θ, i.e., a distribution from the family

$$\mathbb{P}(k; \theta) = \frac{e^{-\theta k}(\theta k)^{k-1}}{k!} ,$$

where $0 < \theta \leqslant 1$ is a real number, and $k \geqslant 1$ is an integer.

What is the maximum likelihood estimator for θ?

c) If the samples from the Borel distribution are 5, 4, 10, 2, 9, 5, 6, 13, 9, what is the maximum likelihood estimator for θ? Give an exact answer as a simplified fraction.

Task 5 – Continuous MLE [24 pts]

a) Let $x_1, x_2, \ldots, x_n$ be independent samples from an exponential distribution with unknown parameter λ. What is the maximum likelihood estimator for λ?

b) Suppose that $x_1, \ldots, x_n$ are i.i.d. realizations (aka samples) from the model

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimate for θ.