CSE 312: Foundations of Computing II
Assignment #8
March 2, 2020
due: Monday, March 9, 2020, 10:00 p.m.

Instructions:

- **Answers:** When asked for a short answer (such as a single number), also *show and explain your work* briefly. Simplify your final formula algebraically as much as possible, without using your calculator, and only then use your calculator, paying attention to the exercise's instructions for the form of your final answer. If it asks for a simplified fraction, this means an integer or a fraction a/b, where $\gcd(a, b) = 1$. If it asks for an answer to (say) 3 significant digits, this means that, when written in scientific notation, your answer would look like $a.bc \times 10^d$, where a, b, c are digits and d is an integer. Highlight or box your final answer to each exercise part.

- **Turn-in:** Do not write your name on your pages (your Gradescope account will identify you to us) and do not include a copy of the exercise's question in what you turn in. You must use Gradescope to upload your homework solutions. You will submit a single PDF file containing your solutions to all the exercises in the homework. Each numbered homework question must be answered on its own page (or pages). You must follow the Gradescope prompts that have you link exercise numbers to your pages. You may typeset your solutions on a computer or you can handwrite them, take a picture of (or scan) each handwritten page, and convert the pictures into a single PDF file. You are responsible for making sure that your solution is easily readable. Popular choices for typesetting mathematics are Microsoft Word (be sure to use its Equation Editor and save as PDF) and LaTeX.

1. Maestro 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 4θ, a C with probability $\frac{1}{2}$, and an F with probability $\frac{1}{2} - 5\theta$. When the quarter is over, you discover that only 5 students got an A, 25 got a B, 100 got a C, and 90 got an F. Find the maximum likelihood estimator for the parameter θ that Maestro Lazy used. Give an exact answer as a simplified fraction.

2. Maestro Lazy decides to assign final grades in CSE 312 by using a more sophisticated probabilistic method: each student independently will be assigned an A with probability θ_1, a B with probability $\theta_2 - \theta_1$, a C with probability $\frac{1}{2}$, and an F with probability $\frac{1}{2} - \theta_2$. When the quarter is over, you discover that only 5 students got an A, 25 got a B, 100 got a C, and 90 got an F. Find the maximum likelihood estimators for the parameters θ_1 and θ_2 that Maestro Lazy used. Give exact answers as simplified fractions. For this problem, you do not have to demonstrate that your estimators are

local maxima, which turns out to be more challenging than just computing second derivatives.

3. (a) Let $x_1, x_2, \ldots, x_n$ be independent samples from a geometric distribution with unknown parameter p. What is the maximum likelihood estimator for p?

   (b) If the samples from the geometric distribution are 5, 4, 3, 2, 9, 5, 3, 9, what is the maximum likelihood estimator for p? Give an exact answer as a simplified fraction.

4. 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? (If you do Exercises 3a and 4 correctly, you should see a very natural relationship between the estimators.)

5. (a) Suppose $x_1, x_2, \ldots, x_n$ are the numbers of independent requests arriving at a web server during each of n successive minutes. Under the assumption that these numbers are independent samples from a Poisson distribution with unknown rate λ per minute, what is the maximum likelihood estimator for λ? (If you do Exercises 4 and 5a correctly, you should see a very natural relationship between the estimators.)

   (b) Is your estimator unbiased? Justify your answer.

6. I have played 12273 games against the devilish Schnapsen program *Doktor Schnaps* and have won 5926 of them. Let's suppose that each game's outcome X_i is $\mathrm{Ber}(p)$, where p is my probability of winning a single game, and that the outcomes are all independent. The goal of this exercise is to estimate my long-term winning probability p against *Doktor Schnaps*. Since *Doktor Schnaps* and I are nearly evenly matched, you can use the estimate $\mathrm{Var}(X_i) = 0.25$ (which would be exact if $p = \frac{1}{2}$). Give all your answers to 4 significant digits.

   (a) What is the maximum likelihood estimate $\hat{p}$ of p for the 12273 games I have played? You do not need to rederive the formula for this estimator, since we did it in lecture. Just substitute the numbers into the formula to arrive at a numerical estimate.

   (b) Approximate the 99% confidence interval for p. Give your answer as the least value of Δ such that $\mathrm{P}(\hat{p} - \Delta \le p \le \hat{p} + \Delta) \ge 0.99$. (For consistency for the grader's sake, when you use the standard normal distribution table, choose the entry that barely makes this inequality true.) Once you have provided Δ, also specify the confidence interval $[\hat{p} - \Delta, \hat{p} + \Delta]$ to 4 significant digits.

   For your interest, *Doktor Schnaps* actually makes use of these results: the player's score in the last column of the leaderboard at `http://schnapsen.realtype.at/index.php?page=gesamtwertung` is the lower boundary $\hat{p} - \Delta$ of the 99% confidence interval. This choice prevents a player who has only played a few lucky games from being ranked too high, since Δ decreases as the number of games played increases. You can see examples of how this affects ranking in the leaderboard table, where the next-to-last column is $\hat{p}$ but the table is sorted on the last column $\hat{p} - \Delta$.

7. (a) Let $x_1, x_2, \ldots, x_n$ be independent samples from the continuous uniform distribution on $[0, \theta]$, where θ is the unknown parameter. What is the maximum likelihood estimator $\hat{\theta}$ for θ?

   Hint: It might help to roughly sketch out the shape of the likelihood function as a function of θ, that is, θ on the horizontal axis and likelihood L on the vertical axis. Start your sketch by asking what happens to L as θ goes to infinity. Then figure out the shape as you get closer to $\theta = 0$. From your sketch, at what value of θ do you maximize L?

   (b) Does your answer to part (a) seem as though it would produce the most accurate estimator for θ? Explain.

   (c) In the remaining parts of this exercise, you will quantify your answer to part (b) by computing the bias of the estimator $\hat{\theta}$. Begin by computing the cumulative distribution function for the random variable $\hat{\theta}$. Recall that this is simply the function $F(x) = \mathrm{P}(\hat{\theta} < x)$. Focus first on the interval $0 \leq x \leq \theta$, but when you're done with that, don't forget to also define $F(x)$ on the rest of the real numbers.

   (d) From your answer to part (c), compute the probability density function $f(x)$ of $\hat{\theta}$.

   (e) From your answer to part (d), compute $\mathrm{E}[\hat{\theta}]$. Is $\hat{\theta}$ an unbiased estimator of θ? Was your intuition in part (b) consistent with your answer to this part?

   (f) Starting from the value of $\mathrm{E}[\hat{\theta}]$ you computed in part (e), determine an unbiased estimator of θ and show that it is unbiased. Although there are other unbiased estimators of θ, you are to derive yours by modifying your answer from part (e).