CSE 312: Foundations of Computing II — Summer 2022

Problem Set 8 (due Friday, August 19, 11:59pm)

Directions: *For each problem, explain/justify how you obtained your answer, as correct answers without an explanation may receive **no credit**. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. Unless you are asked to, you should leave your answer in terms of factorials, combinations, etc., for instance* 26^7 *or* $26!/7!$ *or* $26 \cdot \binom{26}{7}$.

***Submission**: You must upload a **pdf** of your solutions to Gradescope under "Final Pset". The use of LaTex is highly recommended, and we have provided a template. (Note that if you want to hand-write your solutions, you'll need to scan them. If we cannot make out your writing, your work may be ungradable, so make sure it is legible.) There is no coding for this homework.*

Instructions as to how to upload your solutions to Gradescope are on the course web page.

*Remember that you must tag your written problems on Gradescope, or you will potentially receive **no credit** as mentioned in the syllabus. Please put each numbered problem on its own page in the pdf (this will make selecting pages easier when you submit), and ensure that your pdfs are oriented correctly (e.g. not upside-down or sideways).*

***Collaboration**: This pset must be submitted **individually**. You are welcome and encouraged to discuss approaches with your fellow students, but everyone **must write up their own solutions.** Failure to do so is an instance of academic dishonesty. **While this pset is released, there will be no office hours and the staff will not be allowed to give you any help**. If you have clarifying questions about the problems, or any logsitics, please ask privately on Edstem.*

For this homework, there is no need to use continuity correction for CLT problems. You may have practiced this in section, but don't worry about it here. There is also need to do second-derivative tests for MLE.

1. Miscounting (15 points)

Consider the question: what is the probability of getting a **7-card** poker hand (order doesn't matter) that contains at least two 3-of-a-kind (3-of-a-kind means three cards of the same rank). For example, this would be a valid hand: ace of hearts, ace of diamonds, ace of spaces, 7 of clubs, 7 of spades, 7 of hearts and queen of clubs. (Note that a hand consisting of all 4 aces and three of the 7s is also valid.)

Here is how we might compute this:

> Each of the $\binom{52}{7}$ hands is equally likely. Let E be the event that the hand selected contains at least two 3-of-a-kinds. Then
>
> $$\Pr(E) = \frac{|E|}{\binom{52}{7}}$$
>
> To compute $|E|$, apply the product rule. First pick two ranks that have a 3-of-a-kind (e.g. ace and 7 in the example above). For the lower rank of these, pick the suits of the three cards. Then for the higher rank of these, pick the suits of the three cards. Then out of the remaining $52 - 6 = 46$ cards, pick one. Therefore
>
> $$|E| = \binom{13}{2} \cdot \binom{4}{3} \cdot \binom{4}{3} \cdot \binom{46}{1} \qquad \text{and hence} \qquad \Pr(E) = \frac{\binom{13}{2} \cdot 4^2 \cdot 46}{\binom{52}{7}}.$$

Explain what is wrong with this solution. If there is over-counting in $|E|$, say by how much. If there is under-counting in $|E|$, say by how much as well. In either case, remember to justify how you found this solution.

Also, give the correct answer for $\Pr(E)$.

2. Blood types (20 points)

As you may remember from basic biology, the human A/B/O blood type system is controlled by one gene for which 3 variants ("alleles") are common in the human population unsurprisingly called A, B, and O.
As with most genes, everyone has 2 copies of this gene, one inherited from the mother and the other from the father, and everyone passes a randomly selected copy to each of their children (probability 1/2 for each copy, independently for each child). Focusing only on A and O, people with AA or AO gene pairs have type A blood; those with OO have type O blood. (A is "dominant", O is "recessive".) Suppose Xena and both of her parents have type A blood, but her sister Yvonne has type O.

Give exact answers as simplified fractions, and in your answers, use the following notation: G_X, G_Y, G_Z, G_C represent the genotypes of Xena, Yvonne, Zachary, and their child respectively. Ph_X, Ph_Y, Ph_Z, Ph_C represent the phenotypes of Xena, Yvonne, Zachary and their child respectively. Hence, to represent the probability, for example, that Zachary has AO genes given that he has type A blood, we would write $P(G_Z = AO|Ph_Z = A)$. Keep in mind that we won't necessarily use all of these.

(a) What is the probability that Xena carries an O gene? In other words, find $P(G_X = AO|Ph_X = A)$

(b) Xena marries Zachary, who has type O blood. What is the probability that their child will have type O blood? In other words, find $P(G_C = OO|Ph_X = A \cap G_Z = OO)$

(c) If Xena and Zachary's child had type A blood, what is the probability that Xena carries an O gene? In other words, find $P(G_X = AO|Ph_C = A, Ph_X = A)$

3. Duck Hunt (20 points)

Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability 0.6, use the law of total expectation to compute the expected number of ducks that are hit. Assume that the number of ducks in a flock is a Poisson random variable with mean 6.

4. Joint Graph (15 points)

Suppose (X, Y) have the joint continuous distribution with joint density

$$f_{X,Y}(x, y) = \begin{cases} cx^6 e^{-y} & x^2 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

(a) Sketch the joint range $\Omega_{X,Y}$, and label the boundaries with the equations of the formulae. (You may attach a screenshot from WolframAlpha's RegionPlot - type in "RegionPlot[x>0 && y >0 && y<1-x]" for example).

(b) Give an expression for the value of c. Carefully set up the integrals with clearly defined integrands and limits of integration. Then additionally give your answer to 4 decimal places either solving it yourself or using WolframAlpha. Google search for "WolframAlpha double integral calculator".

(c) Compute the marginal densities $f_X(x)$ and $f_Y(y)$. Be sure to use piecewise functions for $f_X(x)$ and $f_Y(y)$ and for each identify the ranges Ω_X and Ω_Y.

(d) Are X and Y independent? Justify your answer with the **definition of independence**, and either prove it or give a counterexample. **An intuitive or English-only answer will not receive credit.**

(e) Give an expression for $E[\log|\tan^{-1}(X+Y)|]$. That is the log of the absolute value of the arctan of $X+Y$. This may not be a particularly relevant value, but you should apply what you know and set up the integral. Carefully set up the integrals with clearly defined integrands and limits of integration. You do **NOT** need to evaluate your answer for this part.

(f) Give an expression for $P(X > Y)$. Carefully set up the integrals with clearly defined integrands and limits of integration. Then additionally give your answer to 4 decimal places either solving it yourself or using WolframAlpha. (Hint: Draw/visualize this region on your joint range from part (a)!)

5. Perfectly Normal Way to Grade (15 points)

Alex computes final course grades in a large CSE 312 class as arbitrary real numbers, with most of them in the interval [0.0, 4.0]. He computes them so that the students' grades are normally distributed with some mean μ and variance σ^2, parameters that he keeps secret out of fear of a student revolution. If 6.68% of students get a grade of 3.8 or greater and 59.87% of students get a grade less than 2.8, what proportion of students got a grade less than 2.0? Show your work and give your answer rounded to 4 decimal places. (To help you avoid calculation errors, as a check of your calculation we will tell you that μ and σ are each integer multiples of 0.1.)

6. Double Max (15 points)

Suppose that $x_1, \ldots, x_n$ are independent random samples from a normal distribution with mean μ_1, suppose that $y_1, \ldots, y_n$ are independent random samples from a normal distribution with mean μ_2, and suppose that $z_1, \ldots, z_n$ are independent random samples from a normal distribution with mean $\mu_1 + \mu_2$. Assuming that all three distributions have the same variance σ^2, find the maximum likelihood estimators of μ_1 and μ_2. (Hint: You can assume that σ^2 is non-zero).
(You don't need to check second order conditions.)