This problem set covers material from Week 3, dates 2/27- 3/01. Unless otherwise noted, all problems are taken from the textbook. Problems can be found at the end of the corresponding subsection.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Unless otherwise stated, you must confirm that your critical point is indeed a maximum for full credit!

Tuesday 2/27

1. 7.10: Problem 7(a).

2. Let $X_1, \dots, X_n|\theta \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$. In class, we showed $\hat{\theta}_{MLE} = \bar{X}$ given $X_1, \dots, X_n$. Now define the following random variable:

$$Y_i = \begin{cases} 0 & \text{if } X_i = 0 \\ 1 & \text{if } X_i > 0 \end{cases}$$

   We seek to find am MLE for θ using the $Y_1, \dots, Y_n$.

   (a) Find the PMF for Y_i and the likelihood function for θ given $Y_1, Y_2, \dots, Y_n$. (*Hint*: it should have a similar structure to the Bernoulli likelihood function.)

   (b) Show that

$$\hat{\theta}_{MLE} = \log\left(\frac{n}{\sum(1-y_i)}\right)$$

   is the MLE estimate for θ given $y_1, \dots, y_n$ and provided that not all the y_i's are equal to 1.

   (c) What does the likelihood function equal if $y_i = 1$ for $i = 1, \dots, n$? Use the likelihood to explain why the MLE for θ does not exist if all of the y_i's are equal to 1.

   (d) For fixed n, what the probability that all of the y_i's are equal to 1? What happens to that probability as $\theta \to \infty$? What does that say about the probability that MLE does not exist?

   (e) For a fixed θ, what happens to the probability that the MLE does not exist as $n \to \infty$?

3. **Hardy-Weinberg equilibrium**. The Hardy-Weinberg equilibrium is a principle stating that the genetic variation in a population will remain constant from one generation

to the next in the absence of disturbing factors. In the simplest case, there are two kinds of alleles denoted A and a with frequencies θ and $1-\theta$, respectively. Organisms inherit two alleles (one from the father and one from the mother), and the possible genotypes (pairs of alleles) an offspring can have are AA, Aa, and aa. The different ways to form these genotypes from a male and female parent are as follows, where the quantity in the brackets [ ] denotes the proportion/frequency under the assumption of Hardy-Weinberg equilibrium:

		Female	
		A $[\theta]$	a $[1-\theta]$
Male	A $[\theta]$	AA $[\theta^2]$	Aa $[\theta(1-\theta)]$
	a $[1-\theta]$	Aa $[\theta(1-\theta)]$	aa $[(1-\theta)^2]$

Note that the genotype frequencies sum to one: $\theta^2 + 2\theta(1-\theta) + (1-\theta)^2 = 1$.

In a random sample of n observations, let n_{AA} denote the number of observations with the AA genotype, with analogous definitions for n_{Aa} and n_{aa}, and so $n = n_{AA} + n_{Aa} + n_{aa}$. Under the assumption of Hardy-Weinberg equilibrium (i.e. that n_{AA} occurs with probability θ^2, n_{Aa} with probability $2\theta(1-\theta)$ and n_{aa} with probability $(1-\theta)^2$), obtain the MLE of θ for this two-allele case.

Thursday 2/29

4. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\alpha, \beta)$ with corresponding CDF

$$F(x|\alpha,\beta) = \Pr(X_i \le x|\alpha,\beta) = \begin{cases} 0 & \text{if } x < 0 \\ (\frac{x}{\beta})^\alpha & \text{if } 0 \le x \le \beta \\ 1 & \text{if } x > \beta, \end{cases}$$

where the parameters α and β are positive and unknown. Find the MLE of $\boldsymbol{\theta} = (\alpha, \beta)$.

5. 7.5: Problem 12 *Note: the implied statistical model is a Multinomial. The sum-to-one constraint on the proportions is important! Dig deep into your Calculus classes and try to re-learn the method of LaGrange Multipliers. Also, you do not need to verify that the critical point is indeed a maximum; believe me that it is!*

Friday 3/01

6. 7.6: Problem 4 *(Hint: what is the statistical/sampling model for X, and how does that relate to T and/or β?)*

7. In a sample from the Chinese population of Hong Kong in 1937, blood types occurred with the following frequencies, where A and a are erythrocyte antigens:

AA	Aa	aa	Total
342	500	187	1029

Suppose we are interested in the true proportions of each of these three blood types among the Chinese population in Hong Kong at that time. That is, we would like to estimate the true proportions of AA, Aa, and aa blood types.

(a) Assuming Hardy-Weinberg equilibrium, obtain maximum likelihood estimates of the three proportions from these data. Provide brief justification of how you obtained your MLEs, if applicable.

(b) Without assuming Hardy-Weinberg equilibrium, obtain maximum likelihood estimates of the three proportions from these data. Provide brief justification of how you obtained your MLEs, if applicable.

(c) How do the sets of estimates in (a) and (b) compare? Do you believe that there is Hardy-Weinberg equilibrium for these blood types? Why or why not?

8. Suppose $X_1, \ldots, X_n | \theta \overset{\text{iid}}{\sim} \text{Unif}[0, \theta]$. Show that the sequence of MLEs for θ, $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$ is a consistent sequence of estimators for θ.

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.