Homework 9: 11/7 - 11/18

STA 335 Due 11:59pm Monday, November 21 Name:

Instructions: Write-up complete solutions to the following problems and submit answers on Gradescope. 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 rubric for homework problems appears on the final page of this assignment.

- Unless otherwise noted, problem numbers are taken from the 2nd edition of Blitzstein and Hwang's *Intro to Probability*.

Monday 11/7

Chapter 7

65, 67

Additional Problem

AP1. (This is inspired by, but slightly different from, Problem 7.71). In humans (and many other organisms), genes come in pairs. Consider a gene of interest, which comes in two types (*alleles*): type a and type A. The *genotype* of a person for that gene is the type of the two genes in the pair: AA, Aa, or aa (aA is equivalent to Aa). According to the Hardy-Weinberg law, for a population in equilibrium the frequencies of AA, Aa, aa will be $p^2, 2p(1-p), (1-p)^2$, respectively, for some p with $0 < p < 1$. Suppose that the Hardy-Weinberg law holds, and that n people are drawn randomly from the population, independently with replacement. Let X_1, X_2, X_3 be the number of people in the sample with genotypes AA, Aa, aa respectively.

(a) What is the joint PMF of X_1, X_2, X_3?

(b) Find the correlation between X_1 and X_2.

(c) Suppose that our observations can't distinguish between AA and Aa. That is, we can only determine whether or not someone has at least one A (in genetics terms, AA and Aa have the same *phenotype*, and we are only able to observe phenotypes, not genotypes). Let Y be the number of people in the sample with AA or Aa. What is the distribution of Y? Give the name of the distribution as well as its paraemters.

(d) Suppose our sample consists of an even number of people, with $n = 2m$, and that either through chance or design, we have $X_3 = m$. What is the conditional distribution of X_1 given $X_3 = m$?

(e) Finally, suppose that instead of using a sample of size n, the size of the sample is $N \sim \text{Pois}(\lambda)$. Find the joint distribution of X_1, X_2, X_3.

Wednesday 11/16

Chapter 8

1, 15, 17

Additional Problem

AP2. Suppose U and V are iid $\text{Expo}(\lambda)$. In this problem, we will find the joint distribution of $U + V$ and $\frac{U}{U+V}$, as well as the marginal distribution of $\frac{U}{U+V}$.

(a) Define a function $g : \mathbb{R}^2 \to \mathbb{R}^2$ by $(x, y) = g(u, v) = \left(u + v, \frac{u}{u+v}\right)$. Find a formula for the inverse transformation $(u, v) = g^{-1}(x, y)$.

(b) Calculate the Jacobian $J_g(u, v)$ and make the substitution $(u, v) = g^{-1}(x, y)$.

(c) Let $X = U + V$ and $Y = \frac{U}{U+V}$. Use the change-of-variables formula to express the joint PDF $f_{X,Y}$ of X, Y in terms of the joint PDF $f_{U,V}$ of U, V.

(d) Based on your previous answer, are X and Y independent?

(e) Find a formula for the marginal PDF of $Y = \frac{U}{U+V}$. What named distribution is this?

Friday 11/18

Chapter 8

Additional Problem

AP3. Let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$ be independent. Show that $X + Y \sim \text{Gamma}(a + b, \lambda)$ in two ways:

(a) Using continuous LotP (Theorem 7.1.18)

(b) Using MGFs

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.