STAT 339: HOMEWORK 2A (PROBABILITY BASICS)

DUE VIA GITHUB BY THE START OF CLASS ON WEDNESDAY MARCH 4

Instructions. Create a directory called hw2a in your stat339 GitHub repo. Your main writeup should be called hw2a.pdf.

You may also use any typesetting software to prepare your writeup, but the final document should be a PDF. LATEX is highly encouraged.

This assignment should not require any coding, with the exception of one part of one problem that asks you to do a small simulation, but you don't have to write your own functions for this.

I will access your work by cloning your repository; make sure that any file path information is written relative to your repo – don't use absolute paths on your machine, or the code won't run for me!

You do not need to write up your answers to the problems marked "Warmup". You should probably do them if you are new to probability; if you have some probability background you might skip them.

Date: Last Revised: February 26, 2020.

Warmup Problems.

1. A discrete random variable, Y, has a distribution given by the following table:

y	P(y)
0	0.2
1	0.4
2	0.2
3	0.1
4	0.1

- (a) Calculate $P(1 < Y \leq 3)$.
- (b) Calculate $\mathbb{E}[Y]$.
- (c) Calculate $\mathbb{V}ar[Y]$.
- (d) Let W = 2Y + 3. Calculate $\mathbb{E}[W]$. (Hint: don't start from scratch; use your previous results and the properties of expectation!)
- (e) Calculate $\mathbb{V}ar[W]$. (The same hint applies here!)
- 2. Let X and Y be jointly distributed discrete random variables. Their joint probability distribution is given in the following table:

x			y			p(x)
	1	2	3	4	5	
1	0.02	0.04	0.06	0.08	0.05	
2	0.08	0.02	0.10	0.02	0.03	
3	0.05	0.05	0.03	0.02	0.10	
4	0.10	0.04	0.05	0.03	0.03	
p(y)						

- (a) Find the marginal PMF of X.
- (b) Find the marginal PMF of Y.
- (c) Calculate the conditional probability P(X = 3|Y = 1).

3. Poisson distribution. A taqueria is trying to model the sales of their fish tacos. Assume that the number of customers, Y, that buy fish tacos in a given hour can be modeled by a Poisson distribution with parameter λ , which has the following PMF, defined on non-negative integers.

$$p_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

(where y! is y-factorial, defined for non-negative integers as : $y! := y \times (y-1) \times \cdots \times 2 \times 1$, and 0! = 1).

Suppose $\lambda = 2.7$. Find the probability that, in one hour,

- (a) exactly one customer buys fish tacos?
- (b) more than three customers buy fish tacos?
- (c) between one and four (inclusive) customers buy fish tacos?

Graded Problems.

- 4. Game Show. A game show allows contestants to win money by spinning a wheel. The wheel is divided into four equally sized spaces, three of which contain dollar amounts and one of which says BUST. The contestant can spin as many times as they want, adding any money won to their total, but if they land on BUST, the game ends and they lose everything won so far.
 - (a) Suppose for the moment that the contestant keeps going until they go bust. Let W be the random variable that represents the number of successful spins before going bust. Indicate the valid range for W, and find a formula for its PMF (be sure to specify the domain).
 - (b) The dollar amounts on the non-BUST spaces are \$100, \$500, and \$5000. Let X_n be the random variable denoting the amount of money won or lost on the n^{th} spin. Find $\mathbb{E}[X_1]$ (note that there is nothing to lose on the first spin).
 - (c) Let S_n be the cumulative amount won after n spins. Find $\mathbb{E}[X_{n+1}|S_n]$, as a function of S_n (Hint: Don't try to obtain the conditional probabilities from the joint probabilities. Just find the conditional directly by treatining S_n as a known constant).
 - (d) How large does S_n have to get before the expected gain by spinning again is negative?

5. Triangular density. A "triangular density" on the interval [-1, 1] is given by the piecewise linear expression

$$p_X(x) = \begin{cases} k(1-x), & x \ge 0\\ k(1+x), & x < 0 \end{cases}$$

which yields an isosceles triangle with a peak at 0.

- (a) Find the unique value of k that makes this a valid density (Note: you do not necessarily need to do any calculus a simple geometric argument is sufficient)
- (b) Find $P(|X| > \frac{1}{2})$ (again, a geometric argument is enough).
- 6. Semicircular density. Consider a circle with radius r and center (0,0) on a plane. The points on the circle satisfy, $x^2 + y^2 = r^2$ (where x is the horizontal coordinate and y is the vertical one). If we take just the "upper half" of the circle, so that y > 0, we can solve the circle equation for y in terms of x and r to get $y = \sqrt{r^2 y^2}$. Treating r as a constant, we can write g(x) in place of y to yield the function $g(x) = \sqrt{r^2 x^2}$.
 - (a) For what values of x does g(x) exist (as a real number)? (Hint: the answer will depend on r, which we are treating as constant which is specified in advance)
 - (b) Suppose X is a random variable that takes values in the domain defined in the previous part. For any fixed r, there is a unique constant, k(r), such that p(x) = k(r)g(x) is a valid probability density function. Find k(r). (Again, you can work this out with geometry no calculus is necessary.)
- 7. Mean of a Beta Distribution (Adapted from FCMA Ex. 3.5) A random variable Y whose range is the interval [0, 1] has a **beta distribution** with PDF

$$p_Y(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha - 1} (1 - y)^{\beta - 1},$$

for real parameters $\alpha, \beta > 0$. Find $\mathbb{E}[Y]$ in terms of α and β . You will need the following identity:

$$\Gamma(x+1) = x\Gamma(x).$$

(The gamma function generalizes the factorial function to all real numbers except non-positive integers, where if n is a positive integer, then $\Gamma(n) = (n-1)!$)

Hint: for the beta distribution to be a valid PDF, it must be the case that

$$\int_0^1 r^{\alpha - 1} (1 - r)^{b - 1} \, \mathrm{d}r = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)}$$

4

8. Sum of Two Random Variables. We know how to find $\mathbb{E}[X+Y]$ and and $\mathbb{V}ar[X+Y]$, but we have not yet seen how to find the distribution of the random variable X + Y itself. Let X and Y be arbitrary random variables, and define Z = X + Y. Consider what must happen for Z to take a particular value, z. If X has the value 0 (say), then Z = z if and only if Y = z - 0. If X = 1 then Z = z if Y = z - 1. And so on. In general,

$$p_{X,Z}(x,z) = p_{X,Y}(x,z-x) = p_X(x)p_{Y|X}(z-x|x)$$

Thus by marginalization,

$$p_Z(z) = \sum_x p_X(x) p_{Y|X}(z - x|x)$$

- (a) Suppose X and Y are independent discrete uniform random variables each with range $R = \{1, 2, 3, 4\}$. Find the formula for $p_Z(z) := P(X + Y = z)$.
- (b) Simulate draws from X + Y (there are easy implementations of functions that draw samples from discrete uniform distributions in scipy in Python; R also has one). Do this many times and calculate the proportion of the time that X + Y = z for various Z. Does the simulated distribution match your analytic formula?
- 9. Conditional Expectation. Show that the following holds.

$$\mathbb{E}\left[\mathbb{E}\left[Y|X\right]\right] = \mathbb{E}\left[Y\right]$$

where the inner expectation is taken with respect to the conditional distribution of Y given X, and is a function of X, and the outer expectation is taken with respect to the marginal distribution of X. You can show this for either the discrete or continuous case (up to you); the steps are the same whether you are working with sums and PMFs, or integrals and PDFs.

Hint: You will not be able to reduce either side to a closed form expression with no sums/integrals in it; just show that the left-hand side simplifies to a sum/integral which is equivalent to the definition of the right-hand side.