Homework 5: Jointly distributed random variables

Due: Thursday, April 2 (Your answers are due at the beginning of the class on the due date. The preferred mode of submission is as a *PDF file* via Gradescope. Please see the instructions on the course webpage. These instructions will tell you how to create a Markdown document in R, and how to "knit" it into a PDF file. *As per the new course guidelines, you are allowed to submit up to a week late without any penalty*)

Note. As per course policy, discussing the concepts behind the problems is OK, but you must write up the answers entirely by yourself. Please see the course policy document (on Canvas or the course webpage) for details. **You need to show all the steps involved in obtaining the answer.**

.

Q1. The following data comes from the Behavioral Risk Factor Surveillance System (BRFSS) survey (run by the CDC) from about 10 years ago. This is a joint probability table for the proportions of survey respondents who smoke and who have had heart attacks.

	Smoker	Non-smoker
Heart attack	0.03	0.03
No heart attack	0.44	0.50

(a) [10 pts] Let S be a 0/1 random variable indicating if a person smokes, and let H be a random variable indicating if the person has had a heart attack. If we choose a person at random from the dataset, what is $P(S = 0)$? [Recall that this is called the marginal distribution of S]

(b) [5 pts] What is $P(H = 0)$?

(c) [20 pts] Compute and compare the quantities $P(H = 1|S = 0)$ and $P(H = 1|S = 1)$. What do they signify intuitively?

(d) [5 pts] Are the variables H and S independent?

.

Q2. Suppose we choose a point uniformly at random from a unit disk (radius = 1) centered at the origin. Let X, Y be the x and y coordinates of the chosen point respectively.

(a) [10 pts] Write down an expression for the joint probability density function $p_{X,Y}(s, t)$. [*Hint:* it must be uniform within the disk, and the integral over the area of the disk must be 1.]

(b) [10 pts] Let p_X denote the marginal p.d.f. for X. Which is larger: $p_X(0)$ or $p_X(1)$? Provide clear explanation.

(c) [10 pts] What is $\mathbb{E}[X + Y]$? [*Hint:* Are X and Y symmetric about some point?]

(d) [10 pts] Bob suggests the following way of sampling a uniformly random point X, Y from the unit disk: first X is sampled uniformly from $[-1, 1]$; then, Y is sampled uniformly from the interval $[-\sqrt{1 - X^2}, \sqrt{1 - X^2}]$. Is this a valid way to sample uniformly from the disk? (Please provide a detailed explanation of why or why not.)

.

Q3. The aim of this question is to see a typical use-case for the linearity of expectation. Consider an experiment in which we toss a biased coin (probability of heads = p) n times. Let Y be the random variable that is the number of heads. Also, let X_i be the 0/1 random variable that is 1 if the ith toss was heads and 0 otherwise.

(a) [10 pts] Prove (using the definiton of the expected value) that:

$$\mathbb{E}[Y] = \sum_{i=0}^{n} i \binom{n}{i} p^i (1 - p)^{n-i}.$$

(b) [5 pts] Let $1 \leq i \leq n$ be any integer. What is $\mathbb{E}[X_i]$?

(c) [15 pts] Find a closed form expression for $\mathbb{E}[Y]$ using part (b). [*Hint:* observe the relationship between Y and the X_i and use the linearity of expectation.]