DS 100: Principles and Techniques of Data Science **Date: March 6, 2019**

Homework #4

Name:

1. (a) Let p denote the probability that a particular item A appears in a simple random sample (SRS). Suppose we collect 5 independent simple random samples, i.e., each SRS is obtained by drawing from the entire population. Let X denote the random variable for the total number of times that A appears in these 5 samples. What is the expected value of X, i.e., $\mathbb{E}[X]$? Your answer should be in terms of p.

   (b) What is $Var(X)$? Again, your answer should be in terms of p.

2. Show that if two random variables X and Y are independent, then $Var(X - Y) = Var(X) + Var(Y)$. You may not use the fact that $Var(X + Y) = Var(X) + Var(Y)$ if X and Y are independent. Instead, use linearity of expectations and the definition of variance. *Hint*: If two random variables are independent, then their covariance is 0 and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

3. Consider rolling (independently) one fair six-sided die and one loaded six-sided die.

   Let X_1 and X_2 denote, respectively, the number of spots from one roll of the fair die and one roll of the loaded die. Suppose the distribution for the loaded die is

$$\begin{aligned} \Pr(X_2 = 1) = \Pr(X_2 = 2) &= \frac{1}{16} \\ \Pr(X_2 = 3) = \Pr(X_2 = 4) &= \frac{3}{16} \\ \Pr(X_2 = 5) = \Pr(X_2 = 6) &= \frac{4}{16}. \end{aligned}$$

   Let $Y = X_1 X_2$ denote the product of the two numbers of spots.

   (a) What is the expected value of Y.

   (b) What is the variance of Y.

(c) Estimate the sampling distribution of Y by simulating 10,000 rolls of the pair of dice. Provide a graphical display of the distribution. Compare the mean and variance from this estimate to the values you computed above.