Homework 2

DSE 210

Guidelines for Homework Submission

1. Each HW is released on the day of the lecture. Students are expected to submit their solutions for the respective HW until 11:59 PM, the day before the next lecture.

2. Each HW is divided into multiple worksheets based on the concepts discussed in the class. Students are expected to attempt all the worksheets in a given HW.

3. HW has to be done individually. NO group work is allowed.

4. The solutions for the HW can be handwritten or typed. (Note: If the handwriting is illegible or if the pages are not scanned properly then the student shall receive 0 points for the respective question)

5. The solutions have to be uploaded to Gradescope in PDF format only. (Note: Do not forget to map question numbers and the pages containing the respective answers while uploading on Gradescope)

6. Some questions in the HW have the tag "***Programming Question***". Students are expected to implement the solution for the respective question in Python. PDF version of the source code has to be uploaded to Gradescope. (Note: Students should combine the theoretical solutions and Python source codes into one PDF file and then upload it to Gradescope)

DSE 210: Probability and statistics **Winter 2020**

Worksheet 4 — Random variable, expectation, and variance - Distributions

1. (2 points) A die is thrown twice. Let X_1 and X_2 denote the outcomes, and define random variable X to be the minimum of X_1 and X_2. Determine the distribution of X.

2. (2 points) A fair die is rolled repeatedly until a six is seen. What is the expected number of rolls?

3. (3 points) On any given day, the probability it will be sunny is 0.8, the probability you will have a nice dinner is 0.25, and the probability that you will get to bed early is 0.5. Assume these three events are independent. What is the expected number of days before all three of them happen together?

4. (4 points total, 2 points each) An elevator operates in a building with 10 floors. One day, n people get into the elevator, and each of them chooses to go to a floor selected uniformly at random from 1 to 10.

   (a) What is the probability that exactly one person gets out at the ith floor? Give your answer in terms of n.

   (b) What is the expected number of floors in which exactly one person gets out? *Hint:* let X_i be 1 if exactly one person gets out on floor i, and 0 otherwise. Then use linearity of expectation.

5. (4 points total, 2 points each) You throw m balls into n bins, each independently at random. Let X be the number of balls that end up in bin 1.

   (a) Let X_i be the event that the ith ball falls in bin 1. Write X as a function of the X_i.

   (b) What is the expected value of X?

6. (2 points) There is a dormitory with n beds for n students. One night the power goes out, and because it is dark, each student gets into a bed chosen uniformly at random. What is the expected number of students who end up in their own bed?

7. (4 points total, 1 point each) In each of the following cases, say whether X and Y are independent.

   (a) You randomly permute $(1, 2, \ldots, n)$. X is the number in the first position and Y is the number in the second position.

   (b) You randomly pick a sentence out of *Hamlet*. X is the first word in the sentence and Y is the second word.

   (c) You randomly pick a card from a pack of 52 cards. X is 1 if the card is a nine, and is 0 otherwise. Y is 1 if the card is a heart, and is 0 otherwise.

   (d) You randomly deal a ten-card hand from a pack of 52 cards. X is 1 if the hand contains a nine, and is 0 otherwise. Y is 1 if *all* cards in the hand are hearts, and is 0 otherwise.

8. (Bonus: 5 points total) A die has six sides that come up with different probabilities:

$$\Pr(1) = \Pr(2) = \Pr(3) = \Pr(4) = 1/8,\ \Pr(5) = \Pr(6) = 1/4.$$

   (a) (Bonus: 1.5 points) You roll the die; let Z be the outcome. What is $\mathbb{E}(Z)$ and $\mathrm{var}(Z)$?

(b) (Bonus: 1.5 points) You roll the die 10 times, independently; let X be the *sum* of all the rolls. What is $\mathbb{E}(X)$ and $\text{var}(X)$?

(c) (Bonus: 2 points) You roll the die n times and take the average of all the rolls; call this A. What is $\mathbb{E}(A)$? What is $\text{var}(A)$?

9. (4 points total, 1 point each) Let $X_1, X_2, \ldots, X_{100}$ be the outcomes of 100 independent rolls of a fair die.

   (a) What are $\mathbb{E}(X_1)$ and $\text{var}(X_1)$?

   (b) Define the random variable X to be $X_1 - X_2$. What are $\mathbb{E}(X)$ and $\text{var}(X)$?

   (c) Define the random variable Y to be $X_1 - 2X_2 + X_3$. What is $\mathbb{E}(Y)$ and $\text{var}(Y)$?

   (d) Define the random variable $Z = X_1 - X_2 + X_3 - X_4 + \cdots + X_{99} - X_{100}$. What are $\mathbb{E}(Z)$ and $\text{var}(Z)$?

10. (4 points total, 1 point each) Suppose you throw m balls into n bins, where $m \geq n$. For the following questions, give answers in terms of m and n.

    (a) Let X_i be the number of balls that fall into bin i. What is $\Pr(X_i = 0)$?

    (b) What is $\Pr(X_i = 1)$?

    (c) What is $\mathbb{E}(X_i)$?

    (d) What is $\text{var}(X_i)$?

11. (2 points) Give an example of random variables X and Y such that $\text{var}(X + Y) \neq \text{var}(X) + \text{var}(Y)$.

12. (Bonus: 5 points) Suppose a fair coin is tossed repeatedly until the same outcome occurs twice in a row (that is, two heads in a row or two tails in a row). What is the expected number of tosses?

13. ***Programming Question:*** (10 points total, 5 points each)

    (a) Write a function that calculates the expectation of a loaded dice. Your function should take the probabilities of the numbers on the dice and return the expectation of this dice. For example, when the input [0.1, 0.2, 0.3, 0.4, 0, 0] is given, the output should be 3. Here 0.1 stands for the probability of getting a 1 when the dice rolled, 0.2 stands for the prob. of 2, 0.3 3 ...

    (b) After you write your function, expectation_die(v), find the answer for expectation_die([0.1, 0.2, 0.3, 0.1, 0.1, 0.2]) and show the answer.

14. ***Programming Question:*** (14 points)

    (a) Let X be the number of heads when a fair coin is tossed n times. We know that X $\sim$ Binomial(n, p). In this problem assume p is 0.5 and n = 50.

    (b) Run this experiment for size = 10, 100, 1000, and 10000 runs (Here, the -size- is not n, it is the size of random runs, please take a look at np.random.binomial function). Record your outcomes from these different experiments. Then using the recorded outcomes plot the distribution of X for size = 10, 100, 1000, and 10000 on different plots. Comment on your results and describe how the plots change as n increases. Does this agree or contradict with a concept that we learned in class? Can you explain the results?