Raul G. Martinez (PID: A12461871)

DSE 210: Probability and statistics Winter 2020

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

1. *(2 points) A die is thrown twice. Let X1 and X2 denote the outcomes, and define random variable X to be the minimum of X1 and X2. Determine the distribution of X.*

   Define the Random Variable X as the min value in $\{X_1, X_2\}$
   The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, $|\Omega| = 36$

   The sample space elements for the Random Variable can be defined pairs (X_1, X_2) as:
   For X = 1 -> ω = {(1, 1),(1, 2),(1, 3),(1, 4),(1, 5),(1, 6),(2, 1),(3, 1),(4, 1),(5, 1),(6, 1)}
   For X = 2 -> ω = {(2, 2),(2, 3),(2, 4),(2, 5),(2, 6),(3, 2),(4, 2),(5, 2),(6, 2)}
   For X = 3 -> ω = {(3, 3),(3, 4),(3, 5),(3, 6),(4, 3),(5, 3),(6, 3)}
   For X = 4 -> ω = {(4, 4),(4, 5),(4, 6),(5, 4),(6, 4)}
   For X = 5 -> ω = {(5, 5),(5, 6),(6, 5)}
   For X = 6 -> ω = {(6, 6)}

   Therefore, the distribution becomes:

   Pr(X = 1) = 11/36
   Pr(X = 2) = 9/36 = 1/4
   Pr(X = 3) = 7/36
   Pr(X = 4) = 5/36
   Pr(X = 5) = 3/36 = 1/12
   Pr(X = 6) = 1/36

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

   The probability of a fair die to roll a six is P(fair die rolls 6) = 1/6
   Then, using the Geometric Distribution we can estimate the expected number of rolls as:

   E(X) = 1/p = 1/(1/6) = 6

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?*

   Since all three events are independent, the probability for all three occurring becomes:

   P(all three events) = P(sunny) * P(nice dinner) * P(bed early) = 0.8 * 0.25 * 0.5 = 0.1

   Again, using the Geometric Distribution we can find the expected number of days as follows:

   E(X) = 1/p = 1/0.1 = 10

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.

Use the Binomial Distribution to solve this question, where k = 1 and p = 1/10:

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\Pr(X = 1) = \binom{n}{1} p^1 (1-p)^{n-1} = n\frac{1}{10}\left(1-\frac{1}{10}\right)^{n-1}$$

$$\mathbf{Pr}(\boldsymbol{X} = \mathbf{1}) = \frac{\boldsymbol{n}}{\mathbf{10}}\left(\frac{\mathbf{9}}{\mathbf{10}}\right)^{\boldsymbol{n}-\mathbf{1}}$$

(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.

Define the random variable X_i for each floor i in {1, 2, 3, …, 10} as:

$$X_i = \begin{cases} 1 & \textit{exactly one person gets out} \\ 0 & \textit{otherwise} \end{cases}$$

And the expectation for $E(X_i) = (1) * (p) + (0) * (1 - p) = p$

Using linearity of expectation have $X = X_1 + X_2 + X_3 + \ldots + X_{10}$

Likewise, the expected value for E(X) becomes:

$E(X) = E(X_1 + X_2 + X_3 + \ldots + X_{10}) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_{10})$

Finally, using the previously found probability in (a) for exactly one person the expected number of floors is:

$$E(X) = \left(\frac{n}{10}\left(\frac{9}{10}\right)^{n-1}\right)_1 + \left(\frac{n}{10}\left(\frac{9}{10}\right)^{n-1}\right)_2 + \cdots + \left(\frac{n}{10}\left(\frac{9}{10}\right)^{n-1}\right)_{10} \quad \text{sum of all 10 probabilities}$$

$$\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{n}\left(\frac{\mathbf{9}}{\mathbf{10}}\right)^{\boldsymbol{n}-\mathbf{1}}$$

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.

   Since each of the events X_i are independent of each other, we can define X as:

   $$\mathbf{X = X_1 + X_2 + X_3 + \ldots + X_m}$$

   (b) What is the expected value of X?

   Use the Binomial Distribution to define the random variable X_i for the number of balls that end up in bin 1:

   $$X_i = \begin{cases} 1 & end\ up\ in\ bin\ 1 \\ 0 & otherwise \end{cases}$$

   And the expectation for $E(X_i) = (1) * (p) + (0) * (1 – p) = p$, where p is bin 1 divided by the total number of bins expressed as 1/n

   Using the linearity of expectation we can define E(X) as follows:

   $E(X) = E(X_1 + X_2 + X_3 + \ldots + X_m) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_m)$

   For m balls the expected value for X becomes:

   $$\mathbf{E(X) = m * (1/n) = m/n}$$

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?*

   Again, this is a Binomial Distribution problem and we can apply the same logic as above to find the expected number of students that end up in their own bed as follows:

   E(X) = np, where p = 1/n (probability for each student to pick among n beds) and n = n (number of students). Therefore:

   $$\mathbf{E(X) = 1}$$

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, . . . , n). X is the number in the first position and Y is the number in the second position.

   X and Y are **Not Independent**. In this case Pr(X) = 1/n and Pr(Y) = 1/(n-1), the product is not equal for the case when they are independent Pr(X) = 1/n and Pr(Y) = 1/n

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

   X and Y are **Not Independent**. Words in text are not independent, therefore the first word sentence will influence 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.

   X and Y are **Independent**. Pr(X) = 4/52 and Pr(Y) = 13/52, where the condition Pr(X∩Y) = Pr(X) Pr(Y) is satisfied.

   (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.

   X and Y are **Independent.** Getting the number nine is independent to all ten-cards being hearts.

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.*

z	1	2	3	4	5	6
Pr(z)	1/8	1/8	1/8	1/8	1/4	1/4

   (a) (Bonus: 1.5 points) You roll the die; let Z be the outcome. What is E(Z) and var(Z)?

$E(Z) = \sum_z zPr(Z = z)$ = (1)(1/8) + (2)(1/8) + (3)(1/8) + (4)(1/8) + (5)(1/4) + (6)(1/4)

= (10/8) + (11/4) = **4**

$var(Z) = \sum_z \left(z - E(Z)\right)^2 Pr(Z = z)$

$= (1-4)^2(1/8) + (2-4)^2(1/8) + (3-4)^2(1/8) + (4-4)^2(1/8) + (5-4)^2(1/4) + (6-4)^2(1/4)$

= (9/8) + (4/8) + (1/8) + 0 + (1/4) + (4/4) = **3** *(b) (Bonus: 1.5 points) You roll the die 10 times, independently; let X be the sum of all the rolls. What is E(X) and var(X)?*

Since all 10 rolls are independent, with the answer from (a) the sum of all rolls becomes:

$$E(X) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_{10}) = (10)\,(4) = \mathbf{40}$$

$$var(X) = var(X_1) + var(X_2) + var(X_3) + \ldots + var(X_{10}) = (10)\,(3) = \mathbf{30}$$

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

When taking the average of n rolls and using the answer of part (a) for a single roll, the expectation and variance for A becomes:

$$E(A) = \frac{E(X_1) + \ E(X_1) + \cdots + \ E(X_n)}{n} = \frac{n\,E(X)}{n} = E(X) = \mathbf{4}$$

$$var(A) = \frac{var(X_1) + \ var(X_1) + \cdots + \ var(X_n)}{n^2} = \frac{n * var(X)}{n^2} = \boldsymbol{\frac{3}{n}}$$

9. *(4 points total, 1 point each) Let X1, X2, . . . , X100 be the outcomes of 100 independent rolls of a fair die.*

(a) What are E(X1) and var(X1)?

Use the formula for the expected value in a single roll of a fair die:

$E(X_1) = (1)*(1/6) + (2)*(1/6) + (3)*(1/6) + (4)*(1/6) + (5)*(1/6) + (6)*(1/6)$

$= \mathbf{3.5}$

Use the calculated expected value to find the variance of X_1:

$var(X_1) = (1 - 3.5)^2 * (1/6) + (2 - 3.5)^2 * (1/6) + (3 - 3.5)^2 * (1/6) + (4 - 3.5)^2 * (1/6) + (5 - 3.5)^2 * (1/6) + (6 - 3.5)^2 * (1/6) = \mathbf{2.9}$

(b) Define the random variable X to be X1 − X2. What are E(X) and var(X)?

$E(X) = E(X_1 – X_2) = E(X_1) - E(X_2) = 3.5 – 3.5 = \mathbf{0}$

$var(X) = var(X_1 – X_2) = var(X_1) + (-1)^2(var(X_2)) = var(X_1) + var(X_2) = 2.9 + 2.9 = \mathbf{5.8}$

(c) Define the random variable Y to be X1 − 2X2 + X3. What is E(Y) and var(Y)?

$E(X) = E(X_1 - 2X_2 + X_3) = E(X_1) - 2E(X_2) + E(X_3) = 3.5 - (2)(3.5) + 3.5 = \mathbf{0}$

$var(X) = var(X_1 - 2X_2 + X_3) = var(X_1) + (-2)^2(var(X_2)) + var(X_3)$

$= var(X_1) + 4var(X_2) + var(X_3) = 2.9 + (4)(2.9) + 2.9 = \mathbf{17.4}$

(d) Define the random variable Z = X1 −X2 +X3 −X4 +···+X99 −X100. What are E(Z) and var(Z)?

From the random variable Z above it seems that odd numbers for X are positive and even numbers for X are negative, since there are 50 odd numbers and 50 negative numbers the expected and variance can be calculated as follows:

$E(Z) = E(X1 - X2 + X3 - X4 + \cdots + X99 - X100)$

$= E(X1 + X3 + X5 + \cdots + X99) + E(-X2 - X4 - X6 - \cdots - X100) = \mathbf{0}$

Similar to previous parts, for variance all the negative random variables in even numbers will become positive when you pull out the constant (-1):

$var(Z) = var(X1 - X2 + X3 - X4 + \cdots + X99 - X100)$

$= var(X1 + X3 + X5 + \cdots + X99) + var(-X2 + X4 - X6 - \cdots - X100)$

$= var(X1 + X2 + X3 + X4 + \cdots + X99 + X100) = 100 * 2.9 = \mathbf{290}$

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

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

Use the binomial distribution to approach this problem where $n_{binomial} = m$ (number of balls), k = 0 (choose zero number of balls) and p = 1/n (probability of hitting bin i):

$$\Pr(X = k) = \binom{n_{\text{binomial}}}{k} p^k (1-p)^{n_{\text{binomial}} - k}$$

$$\Pr(X_{\text{i}} = 0) = \binom{m}{0}\left(\frac{1}{n}\right)^0\left(1 - \frac{1}{n}\right)^{m-0} = (1)(1)^0\left(1 - \frac{1}{n}\right)^m = \left(\mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{n}}\right)^{\boldsymbol{m}}$$

(b) What is Pr(Xi = 1)?

Again, use the binomial distribution to approach this problem where $n_{binomial}$ = m (number of balls), k = 1 (choose zero number of balls) and p = 1/n (probability of hitting bin i):

$$\Pr(X = k) = \binom{n_{\text{binomial}}}{k} p^k (1-p)^{n_{\text{binomial}} - k}$$

$$\Pr(X_{\mathrm{i}} = 1) = \binom{m}{1}\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{m-1} = (m)\left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{m-1} = \boldsymbol{\frac{m}{n}\left(1-\frac{1}{n}\right)^{m-1}}$$

(c) What is E(Xi)?

The expected value for a binomial distribution is equal to the product of $n_{binomial}$ = m (number of balls) and p = 1/n (probability of hitting bin i), therefore:

$$\mathbf{E}(\boldsymbol{X}_{\mathbf{i}}) = \boldsymbol{m\,\frac{1}{n} = \frac{m}{n}}$$

(d) What is var(Xi)?

Likewise, using the binomial distribution for the variance this equals to the product of the number of balls with the probability of hitting bin i, and also times the complement:

$$\mathbf{var}(\boldsymbol{X}_{\mathbf{i}}) = \boldsymbol{m\,\frac{1}{n}\left(1-\frac{1}{n}\right) = \frac{m}{n}\left(1-\frac{1}{n}\right)}$$

11. *(2 points) Give an example of random variables X and Y such that var(X + Y) /= var(X) + var(Y).*

Remember the property for variance when the random variable has a constant (a in this example):

var(aX + b) = a^2var(X)

With this in mind consider the case when X = Y, therefore the following will occur:

var(X + Y) = var (X + X) = var (2X) = 2^2var(X) = 4var(X)

Therefore, **var(X + Y) != var(X) + var(Y) when X = 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?*

Define X as the random variable for the number of kth tosses until the same outcome occurs twice. Because X can only take positive values we have:

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + \cdots$$

$$P(X \geq 2) = P(X = 2) + P(X = 3) + \cdots$$

$$P(X \geq 3) = P(X = 3) + \cdots$$

Then we can estimate the expectation as:

$$E(X) = \sum_{k=1}^{\infty} E(X_k) = \sum_{k=1}^{\infty} P(X \geq \mathrm{k}) = P(X = 1) + 2P(X = 2) + 3P(X = 3) + \cdots$$

$$= 1 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = \mathbf{3}$$

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 …

```
# import libraries
import numpy as np

# define the function
def expectation_die(v):

    # define arrays for the die values and its probabilities
    x_values = np.array([1, 2, 3, 4, 5, 6])
    x_probabilities = np.array(v)

    # return the expectation
    return np.sum(x_values*x_probabilities)
```

(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.

```
# Compute the answer for the following
expectation_die([0.1, 0.2, 0.3, 0.1, 0.1, 0.2])
```

Out[4]: 3.5

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 ∼ Binomial(n, p). In this problem assume p is 0.5 and n = 50.

```
# import libraries
import numpy as np
import matplotlib.pyplot as plt

# Define the values for X
X_values = np.random.binomial(n = 50, p = 0.5, size = 1)
```

(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?

```
# Define figure for subplots
fig = plt.figure(figsize = (7,7))
fig.suptitle('Distribution of X for different sizes', fontsize=20)
i = 1

# Iterate over every size for X
for size_n in [10, 100, 1000, 10000]:

    # define random variable with 50 tosses and fair coin probability
    X_values = np.random.binomial(n = 50, p = 0.5, size = size_n)

    # number of bins for histogram
    mybins = np.linspace(0, 50, 51)
```

```
    # plot histogram
    ax = fig.add_subplot(2,2, i)
    ax.set_title('Size = ' + str(size_n))
    ax.set_xticks(np.linspace(0, 50, 11))
    ax.hist(x = X_values, bins = mybins, density = True)

    i+=1 # move to the next subplot

    # adjust separation of subplots
    fig.tight_layout(rect=[0, 0.03, 1, 0.95])
```

The results agree with the concepts learned in class. As n increases the distribution gets closer to the theoretical values for expectation, variance, and probability for X:
E(X) = np = (50)(1/2) = 25
var(X) = np(1-p) = (50)(1/2)(1 - 1/2) = 12.5
Pr(X = 25) = (50 choose 25) * (1/2)^25 * (1-1/2)^50-25 = 0.11

You can see that when size is 10 the distribution seems bi-modal and close to the expected value but is still unclear with precision where it is. As the size for the number of runs increases from 100, 1000, and 10000 you can clearly see how the binomial distribution starts getting centered at 25 with variance of 12.5, and the probability for the distribution seems to be converging to the theoretical value for 0.11 at X = 25.