Homework 3

S1201, Summer 2022

100 points

Due July 30, 23:59PM EST

Notes: Please read the following instructions.

(i) Please submit a **single** PDF file on courseworks. You may scan your written solutions or directly write the solutions on your tablet.

(ii) Late submission will lead to some penalty: 20 points off for within 24 hours late, 40 points off for over 24 hours late, all points off for submission after the solutions are posted on courseworks.

(iii) You can discuss the problems with others. But a direct plagiarism will lead to zero point for this assignment, and this will be reported to the university.

(iv) There is one extra problem with 8 bonus points, which is optional. Therefore all points add up to 108, but the maximum score of this homework is 100.

1. **$2 \times 4 = 8$ points** TRUE/FALSE questions. No explanations are needed.

   (a) When random variables X and Y are not independent, we still have $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$; T

   (b) When random variables X and Y are not independent, we still have $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$; F

   (c) For any continuous r.v. X, $\mathbb{P}(X > c) = \mathbb{P}(X < -c)$ for any constant $c > 0$. F

   (d) When we have enough data, most of sample-version summary characteristics (like sample mean, sample variance, sample median and relative frequency etc.) can be good approximation of the their counterparts in the population universe (like mean, variance, median and pmf etc.). T

2. **2 × 4 = 8 points** Prove the following conclusions we covered in class.

   (a) (Lecture 8, expected value of a function of a r.v.) X is a discrete random variable and $\{x_1, \dots, x_m\}$ are all possible values it can be. g is some function of the number line. Then the expectation of $Y = g(X)$ equals

   $$\mathbb{E}Y = \mathbb{E}g(X) = \sum_{j=1}^{m} g(x_j)\mathbb{P}(X = x_j).$$

   Hint: Write down the pmf of Y then use the definition of expectation.

   (b) (Lecture 8, "linearity" of the variance, a simpler version) X is a r.v. Suppose a and b are two constants (fixed numbers, not r.v.'s). Then

   $$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X), \quad \mathrm{SD}(aX + b) = |a|\mathrm{SD}(X).$$

   Hint: Use the definition of $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$.

Solutions: (a) $Y = g(X)$, let $y_j = g(x_j)$

By definition of $\mathbb{E}Y$:

$$\mathbb{E}Y = \sum_{j=1}^{m} y_j \times \mathbb{P}(Y = y_j) = \sum_{j=1}^{m} g(x_j) \times \mathbb{P}(X = x_j).$$

(b) By linearity of the expectation:

$$\mathbb{E}(aX+b) = a\mathbb{E}X + b.$$

$$\Rightarrow \mathrm{Var}(aX+b) = \mathbb{E}\left[(aX+b) - \mathbb{E}(aX+b)\right]^2$$

$$= \mathbb{E}\left[a^2(X-\mathbb{E}X)^2\right]$$

by linearity of the expectation

$$= a^2\,\mathbb{E}(X-\mathbb{E}X)^2$$

$$= a^2\,\mathrm{Var}(X)$$

$$\Rightarrow \mathrm{SD}(aX+b) = \sqrt{\mathrm{Var}(aX+b)} = |a|\cdot\mathrm{SD}(X). \quad \square$$

3. **$3 \times 3 = 9$ points** Two fair six-sided dice are tossed independently. Let M = the maximum of the two tosses. For example, $M(1,3) = 3$, and $M(4,4) = 4$ etc.

   (a) What is the pmf of M? Write it as a pmf table.

   (b) Find the cdf of M and graph it.

   (c) Find $\mathbb{E}M$, $\mathbb{E}M^2$, $\text{Var}(M)$

   Hint: We studied another example of tossing two dice in Lecture ~~8~~ 7, where we care about the sum of two dice. Here we want to know the maximum.

Solution: (a)

Die 2

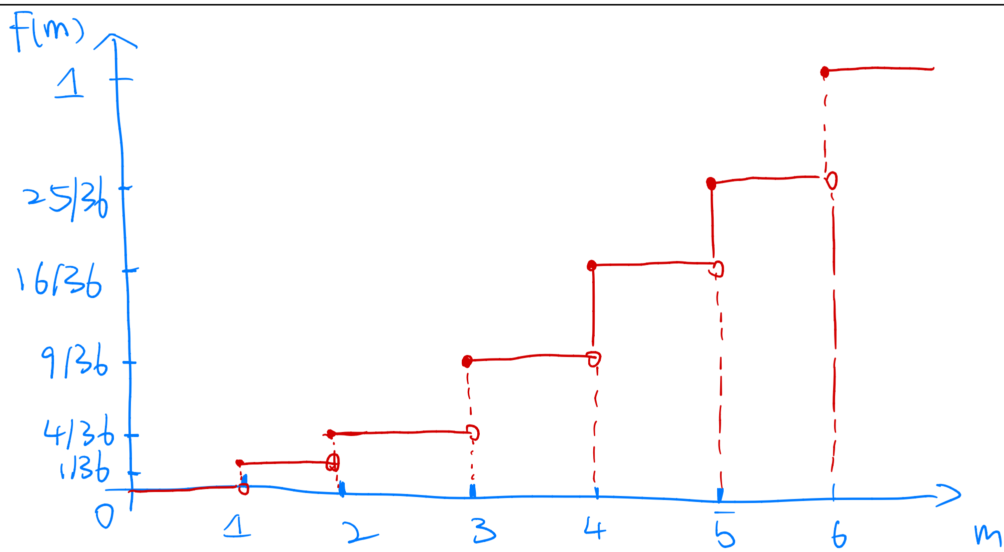
Die 1 \ Die 2	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

pmf of M:

m	1	2	3	4	5	6
P(m)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

(b)

$$F(m) = \mathbb{P}(M \le m) = \begin{cases} 0, & m < 1 \\ 1/36, & 1 \le m < 2 \\ 4/36, & 2 \le m < 3 \\ 9/36, & 3 \le m < 4 \\ 16/36, & 4 \le m < 5 \\ 25/36, & 5 \le m < 6 \\ 1, & m \ge 6 \end{cases}$$



(c) $\mathbb{E}m = 1 \times \frac{1}{36} + 2 \times \frac{3}{36} + 3 \times \frac{5}{36} + 4 \times \frac{7}{36} + 5 \times \frac{9}{36} + 6 \times \frac{11}{36}$

$= \frac{161}{36}$

≈ 4.47

$\mathbb{E}m^2 = 1^2 \times \frac{1}{36} + 2^2 \times \frac{3}{36} + 3^2 \times \frac{5}{36} + 4^2 \times \frac{7}{36} + 5^2 \times \frac{9}{36} + 6^2 \times \frac{11}{36}$

$= \frac{791}{36}$

≈ 21.97

$\Rightarrow$ By the short-cut formula : $Var(m) = \mathbb{E}m^2 - (\mathbb{E}m)^2$

$= 21.97 - 4.47^2$

≈ 1.99 $\square$.

4. **10 points** (Memoryless property of geometric distribution) From Lecture ~~10~~ 9, we have known that exponential distribution has so-called "memoryless property", where if $X \sim \text{Exp}(\lambda)$ with parameter $\lambda > 0$, for any positive numbers x and a,

$$\mathbb{P}(X > x + a | X > a) = \mathbb{P}(X > x). \tag{1}$$

Show that the geometric distribution has the same property too.

Hint: $X \sim \text{Geometric}(p)$, where $0 < p < 1$. Just need to show that ~~(2)~~ (1). holds.

Proof: Suppose $X \sim$ Geometric (p). Then:

$$\mathbb{P}(X > x+a | X > a) = \frac{\mathbb{P}(X > x+a, X > a)}{\mathbb{P}(X > a)}$$

$$= \frac{\mathbb{P}(X > x+a)}{\mathbb{P}(X > a)}$$

$$= \frac{(1-p)^{x+a}}{(1-p)^{a}}$$

$$= (1-p)^{x}$$

$$= \mathbb{P}(X > x). \quad \square$$

Homework 3

100 points

S1201, Summer 2022
Due July 30, 23:59PM EST

5. **$2 \times 6 = 12$ points** A bulk-buying company handles packets in three different sizes: 27, 125 and 512 cubic feet, called type 1, type 2 and type 3 respectively. Let X_1, X_2, X_3 be the number of type 1, 2 and 3 packets shipped during a given week. We know that: $\mathbb{E}X_1 = 200$, $\mathbb{E}X_2 = 250$, $\mathbb{E}X_2 = 100$, $\mathrm{SD}(X_1) = 10$,$\mathrm{SD}(X_1) = 12$, $\mathrm{SD}(X_3) = 8$.

These are NOT the variances !!!

(a) Calculate the expected volume in cubic feet of type 1, type 2 and type 3 packets shipped in one week. (Answer for each type separately)

(b) Calculate the variance of the volume of type 1, type 2, and type 3 packets shipped in one week. (Answer for each type separately)

Solutions: (a) Denote volumes (in cubic feet) of types 1/2/3 packets shipped in one week as V_1, V_2, V_3 :

Then : $V_1 = 27X_1$, $V_2 = 125X_2$, $V_3 = 512X_3$

By linearity of the expectation:

$$\mathbb{E}V_1 = \mathbb{E}(27X_1) = 27\mathbb{E}X_1 = 5400$$

$$\mathbb{E}V_2 = \mathbb{E}(125X_2) = 125\mathbb{E}X_2 = 31250$$

$$\mathbb{E}V_3 = \mathbb{E}(512X_3) = 512\mathbb{E}X_3 = 51200$$

(all units are feet^3).

(b) By "linearity" of the expectation:

$$\mathrm{Var}(V_1) = \mathrm{Var}(27X_1) = 27^2\,\mathrm{Var}(X_1) = 72900$$

$$\mathrm{Var}(V_2) = \mathrm{Var}(125X_2) = 125^2\,\mathrm{Var}(X_2) = 2250000$$

$$\mathrm{Var}(V_3) = \mathrm{Var}(512X_3) = 512^2\,\mathrm{Var}(X_3) = 16777216.$$ □

6. **4 + 2 + 5 = 11 points** Based on past experience, we believes that 60% of road test taker in Ithaca fail their first test. There are 100 people who recently took their first test.

   (a) What are the mean and standard deviation of the **proportion** of these 100 first-time test takers who will pass the exam?
   Hint: Proportion should be the number of people passing the exam divided by 100.

   (b) What assumptions underlie your model? List one of them. Do you think it is satisfied here or not? Any reasonable answers count!

   (c) What's the probability that over 70% of these 100 people fail the exam? Use the normal approximation if possible (check the condition!).
   Hint: Remember to use the continuity correction.

Solutions: (a) sample proportion $\hat{p} = \frac{1}{100} \sum_{i=1}^{n} \mathbb{1}(\text{ith person fails})$.

And $100\hat{p} \sim \text{Bin}(n=100, \ p=0.6)$

$\Rightarrow \mathbb{E}(100\hat{p}) = np = 60, \ \text{Var}(100\hat{p}) = np(1-p) = 24$

$\Rightarrow \mathbb{E}\hat{p} = 0.6, \ \text{Var}(\hat{p}) = 0.0024 \Rightarrow \mathbb{E}(1-\hat{p}) = 0.4. \ \text{Var}(1-\hat{p}) = 0.0024.$

(b) ① The test results of these test takers are independent.

② These people are representative of the population.

For ①: Yes. I think so.

For ②: We don't know.

(c) Check: $n = 100 > 30, \quad n \times \min\{p, 1-p\} = 40 > 10.$ ✓

$100\hat{p} \sim \text{Bin}(n=100, \ p=0.6)$

$\Rightarrow$ By normal approximation: $\frac{n\hat{p} - np}{\sqrt{np(1-p)}} = \frac{100(\hat{p}-p)}{4.899} \overset{d}{\approx} N(0,1).$

$\Rightarrow \mathbb{P}(\hat{p} > 0.7) = \mathbb{P}(100\hat{p} > 70) = \mathbb{P}(100\hat{p} \geq 71)$

$= \mathbb{P}(100\hat{p} \geq 69.5)$

$= \mathbb{P}\left(\underbrace{\frac{100(\hat{p}-p)}{4.899}}_{\sim N(0,1)} \geq \frac{69.5 - 100 \times 0.6}{4.899}\right)$

$\approx \mathbb{P}(Z \geq 1.94)$

$= 1 - \Phi(1.94) \approx 1 - 0.9738 = 0.0262$ $\square$.

7. **4×4 = 16 points** Statistics from Cornell's Northeast Regional Climate Center indicate that Ithaca, New York, gets an average of 35.4" of rain each year, with a standard deviation of 4.2". Assume that a Normal model applies.

   (a) During what percentage of years does Ithaca get more than 40" of rain?
   Hint: Find the corresponding percentage of value 40 in the distribution.

   (b) Less than how much rain falls in the driest 20% of all years?
   Hint: Find the 20% quantile of the distribution.

   (c) A Cornell University student is in Ithaca for 4 years. Let Y represent the sample mean amount of rain for those 4 years. Describe the distribution of Y.

   (d) What's the probability that those 4 years average less than 30" of rain?

Solutions: (a) Rainfall per year $X \sim N(35.4, 4.2^2)$.

$$\Rightarrow \mathbb{P}(X > 40) = \mathbb{P}\Big(\underbrace{\frac{X-35.4}{4.2}}_{\sim N(0,1)} > \frac{40-35.4}{4.2}\Big)$$

$$= \mathbb{P}(Z > 1.10)$$

$$= 1 - \Phi(1.10)$$

$$= 1 - 0.8413 = 0.1587.$$

(b) We want to find x that satisfies $\mathbb{P}(X \leq x) = 0.2$

$$\Rightarrow 0.2 = \mathbb{P}(X \leq x) = \mathbb{P}\Big(\underbrace{\frac{X-35.4}{4.2}}_{\sim N(0,1)} \leq \frac{x-35.4}{4.2}\Big)$$

$$= \Phi\Big(\frac{x-35.4}{4.2}\Big)$$

By z-table: $\frac{x-35.4}{2} \approx -0.84$

$\Rightarrow x = 35.4 - 2 \times 0.84 = 33.72$

(c) $Y = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$, $Y_i \overset{iid}{\sim} N(35.4, 4.2^2)$

$\Rightarrow Y$ is normally distributed.

And since $\mathbb{E}Y = \mathbb{E}\big(\frac{1}{4}\sum_{i=1}^{4} Y_i\big) = 35.4$.

$\mathrm{Var}(Y) = \mathrm{Var}\big(\frac{1}{4}\sum_{i=1}^{4} Y_i\big) = \frac{1}{4} \times 4.2^2 = 4.41$

$\Rightarrow Y \sim N(35.4, 4.41)$.

(d) $\mathbb{P}(Y < 30) = \mathbb{P}\left(\frac{Y - 35.4}{\sqrt{4.41}} < \frac{30 - 35.4}{\sqrt{4.41}} \right)$

$= \mathbb{P}(Z < -2.57)$

$= \Phi(-2.57)$

$\simeq 0.0051$ $\square$.

8. **$2 \times 4 = 8$ points** Let X be a discrete random variable which takes values in the set $\{0, 1, 2\}$. Suppose the distribution function of X is the following:

$$\mathbb{P}(X=0) = \frac{16}{25}, \quad \mathbb{P}(X=1) = \frac{4k}{25}, \quad \mathbb{P}(X=2) = \frac{1}{25}.$$

(a) Find the value of k.

(b) Find $\mathbb{E}X$.

(c) Find $\mathrm{SD}(X)$.

(d) Find $\mathbb{E}[X(1-X^2)]$.

Hint: For (d), consider to use the formula for function of expected values of r.v.'s (in Lecture 8, actually you also proved it in Q2.(a) of this HW!)

Solutions: (a) Since it's a valid pmf: $\frac{16}{25} + \frac{4k}{25} + \frac{1}{25} = 1$

$\Rightarrow \quad k=2.$

(b) $\mathbb{E}X = 0 \times \frac{16}{25} + 1 \times \frac{8}{25} + 2 \times \frac{1}{25} = 0.4$

(c) $\mathbb{E}X^2 = 0^2 \times \frac{16}{25} + 1^2 \times \frac{8}{25} + 2^2 \times \frac{1}{25} = 0.48$

$\Rightarrow \mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0.48 - 0.4^2 = 0.32.$

$\Rightarrow \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = 0.5657.$

(d) Function $g(x) = x(1-x^2)$

$\Rightarrow \mathbb{E}[g(X)] = g(0) \times \frac{16}{25} + g(1) \times \frac{8}{25} + g(2) \times \frac{1}{25}$

$= 0 \times \frac{16}{25} + 0 \times \frac{8}{25} + 2 \times (1-2^2) \times \frac{1}{25}$

$= -0.24. \quad \square$

9. **$3 \times 4 = 12$ points** Suppose that 30% of all students who have to buy a textbook for a particular course want a new copy, whereas the other 70% want a used copy. Consider randomly selecting 40 purchasers.

   (a) What are the mean value and standard deviation of the number of students who want a new copy of the book?

   (b) What is the probability that the number of students who want new copies is more than two standard deviations away from the mean value? (Can ignore the continuity correction)

   Hint: It's ok to use the 3σ-rule to give a quick answer!

   (c) The bookstore has 25 new copies and 25 used copies in stock. If 40 people come in to purchase this text, what is the probability that all of them will get the type of book they want from current stock?

   Hint: Define X = the number who want a new copy. For what values of X will all 40 people get what they want?

Solutions: Let X = the number of students who want a new copy.

(a) $X \sim \text{Bin}(n=40, p=0.3)$

$\Rightarrow \mathbb{E}X = np = 12$, $SD(X) = \sqrt{Var(X)} = \sqrt{np(1-p)} \approx 2.898$

(b) Since $n = 40 > 30$, $n \times \min\{p, 1-p\} = 12 > 10$,

we can approximate X by a normal distribution.

Then by 3σ-rule, the probability $\approx \Phi(2) - \Phi(-2)$

≈ 0.95.

(c) When $15 \le X \le 25$, all people get what they want

(Think about it, why?)

$\sim N(0,1)$.

$$\Rightarrow \mathbb{P}(15 \le X \le 25) = \mathbb{P}\left(\frac{15-12-0.5}{2.898} \le \underbrace{\frac{X-12}{2.898}} \le \frac{25-12+0.5}{2.898}\right)$$

$$\approx \mathbb{P}(0.86 \le Z \le 4.66)$$

$$= \Phi(4.66) - \Phi(0.86)$$

$$\approx 1 - 0.8051$$

$$= 0.1949. \quad \square.$$

10. **$2 \times 3 = 6$ points** Let W be an exponential random variable with expected value 2.

    (a) Find the value of $\mathbb{E}(5W)$.

    (b) Calculate $\mathbb{P}(W > 5)$.

Solutions: (a) $\mathbb{E}(5W) = 5 \cdot \mathbb{E}W = 5 \times 2 = 10$.

(b) Suppose $W \sim Exp(\lambda)$: Then $\mathbb{E}W = \lambda^{-1} = 2$

$\Rightarrow \lambda = \frac{1}{2}$

$\Rightarrow \mathbb{P}(W > 5) = e^{-5\lambda} = e^{-5/2} \approx 0.0821$. $\square$.

11. **(Extra problem, not required) 2+2+4 = 8 bonus points** You roll a die, winning nothing if the number of spots is odd, \$7 for a 2 or a 4, and \$16 for a 6.

(a) Find the expected value and standard deviation of your prospective winnings.

(b) You play twice. Find the mean and standard deviation of your total winnings.

(c) You play 40 times. What's the probability that you win at least \$250?

Hint: For (c), rolling a die is random and the outcomes are mutually independent, so the Central Limit Theorem guarantees that the distribution of total winnings in 40 times is approximately normal. Remember to use the continuity correction.

Solutions: (a) X = my winnings (in \$)

Then: $\mathbb{P}(X=0) = \frac{1}{2}$, $\mathbb{P}(X=7) = \frac{1}{3}$, $\mathbb{P}(X=16) = \frac{1}{6}$.

$$\Rightarrow \mathbb{E}X = 0 \times \frac{1}{2} + 7 \times \frac{1}{3} + 16 \times \frac{1}{6} = 5.$$

$$\mathbb{E}X^2 = 0^2 \times \frac{1}{2} + 7^2 \times \frac{1}{3} + 16^2 \times \frac{1}{6} = 59$$

$$\Rightarrow \mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 59 - 25 = 34.$$

$$SD(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{34} \approx 5.831$$

(b) $X_1, X_2 \overset{iid}{\sim} X$ defined in (a).

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 = 10$$

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 68$$

$$\Rightarrow SD(X_1+X_2) = \sqrt{\mathrm{Var}(X_1+X_2)} = \sqrt{68} \approx 8.246$$

(c) Y = total winnings in 40 rounds

X_i = winning in i-th round. $\overset{iid}{\sim} X$

$$\Rightarrow Y = \sum_{i=1}^{40} X_i.$$

By CLT: $Y \overset{d}{\approx} N(40\mathbb{E}X, 40\mathrm{Var}(X))$

$= N(200, 1360)$

$$\Rightarrow \quad P(Y \geq 250) \approx P(Y \geq 250 - 0.5) \quad \text{(continuity correction)}$$

$$= P\Big(\underbrace{\frac{Y-200}{\sqrt{1300}}}_{\overset{d}{\approx} N(0,1).\ \text{by standardization}} \geq \frac{250-0.5-200}{\sqrt{1300}}\Big)$$

$$\approx P(Z \geq 1.34)$$

$$= 1 - \Phi(1.34)$$

$$= 1 - 0.9099$$

$$= 0.0901. \qquad \square.$$