CSE 312 Section 5

More Random Variable and Discrete Zoo

Administrivia

Announcements & Reminders

- HW3
  - Grades released on gradescope – check your submission to read comments
  - Regrade requests open ~24 hours after grades are released and close after a week

- HW4
  - Due yesterday, 10/22 @ 11:59pm
  - Late deadline Saturday 10/25 @ 11:59pm

- HW5
  - Released
  - Due Wednesday 10/29 @ 11:59pm
  - Late deadline Saturday 11/2 @ 11:59pm

Review & Questions

Any lingering questions from this last week?

Each week in section, we’ll be reviewing the main concepts from this week and putting them into action by going through some practice problems together. But before we get into that review, we’ll try to start off each section with some time for you to ask questions. Was anything particularly confusing this week? Is there anything we can clarify before we dive into the review? This is your chance to clear things up!

Review of Main Concepts

- **Expectation (expected value, mean, or average):** The expectation of a discrete random variable is defined to be

$$\mathbb{E}[X] = \sum_x x\, p_X(x) = \sum_x x\, \mathbb{P}(X = x)$$

The expectation of a function of a discrete random variable $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_x g(x)\, p_X(x)$$

- **Linearity of Expectation:** Let X and Y be random variables, and $a, b, c \in \mathbb{R}$. Then, $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$. Also, for any random variables $X_1, \dots, X_n$,

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

Review of Main Concepts

- **Variance:** Let X be a random variable and $\mu = E[X]$. The variance of X is defined to be

$$Var(X) = \mathbb{E}[(X-\mu)^2]$$

  Notice that since this is an expectation of a non-negative random variable $((X-\mu)^2)$, variance is always nonnegative. With some algebra, we can simplify this to

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- **Standard Deviation:** Let X be a random variable. We define the standard deviation of X to be the square root of the variance, and denote it $\sigma = \sqrt{Var(X)}$
- **Property of Variance:** Let $a, b \in \mathbb{R}$ and let X be a random variable. Then,

$$Var(aX+b) = a^2 Var(X)$$

Review of Main Concepts

- **Independence:** Random variables X and Y are independent iff

$$\forall x \forall y, \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$$

  In this case, we have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (the converse is not necessarily true).

- **i.i.d. (independent and identically distributed):** Random variables $X_1, \dots, X_n$ are i.i.d. (or iid) iff they are independent and have the same probability mass function.

- **Variance of Independent Variables:** If X is independent of Y,

$$Var(X + Y) = Var(X) + Var(Y)$$

  This depends on independence, whereas linearity of expectation always holds. Note that this combined with the above shows that $\forall a, b, c \in \mathbb{R}$ and if X is independent of Y,

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$$

Review Questions

Question #1

$$Var(A + B) = Var(A) + Var(B)$$

- True
- False

Question #1

$$Var(A + B) = Var(A) + Var(B)$$

- True
- False

Question #2

What is $Var(3A + 4)$?

- $3Var(A) + 4$
- $3Var(A)$
- $9Var(A)$
- $Var(A)$

Question #2

What is $Var(3A + 4)$?

- $3Var(A) + 4$
- $3Var(A)$
- $9Var(A)$
- $Var(A)$

Question #3

$$E[A + B] = E[A] + E[B]$$

- True
- False

Question #3

$$E[A + B] = E[A] + E[B]$$

- True
- False

Question #4

What is $E[3A + 4]$?

- $3E[A] + 4$
- $3E[A]$
- $9E[A]$
- $E[A]$

Question #4

What is $E[3A + 4]$?

- $3E[A] + 4$
- $3E[A]$
- $9E[A]$
- $E[A]$

Problem 2 - Pond fishing

Part a

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many of the next 10 fish I catch are blue, if I catch and release

- $$\text{Bin}\left(10, \frac{B}{N}\right)$$
- $$\text{Ber}\left(\frac{B}{N}\right)$$
- $$\text{Bin}\left(1, \frac{B}{N}\right)$$

Part a

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many of the next 10 fish I catch are blue, if I catch and release

- $$\text{Bin}\left(10, \frac{B}{N}\right)$$

- $$\text{Ber}\left(\frac{B}{N}\right)$$

- $$\text{Bin}\left(1, \frac{B}{N}\right)$$

Part b

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many fish I had to catch until my first green fish, if I catch and release

- $$\text{Ber}\left(\frac{G}{N}\right)$$
- $$\text{Bin}\left(1, \frac{G}{N}\right)$$
- $$\text{Geo}\left(\frac{G}{N}\right)$$

Part b

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many fish I had to catch until my first green fish, if I catch and release

- $$\text{Ber}\left(\frac{G}{N}\right)$$
- $$\text{Bin}\left(1, \frac{G}{N}\right)$$
- $$\text{Geo}\left(\frac{G}{N}\right)$$

Part c

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many red fish I catch in the next five minutes, if I catch on average r red fish per minute

- $$\text{Poi}(5R)$$
- $$\text{Bin}\left(5, \frac{R}{N}\right)$$
- $$\text{Poi}(5r)$$

Part c

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

How many red fish I catch in the next five minutes, if I catch on average r red fish per minute

- $$\mathsf{Poi}(5R)$$
- $$\mathsf{Bin}\left(5, \frac{R}{N}\right)$$
- $$\mathsf{Poi}(5r)$$

Part d

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

Whether or not my next fish is blue

- $\text{Poi}(5B)$
- $\text{Bin}\left(1, \frac{R}{N}\right)$
- $\text{Ber}\left(\frac{B}{N}\right)$

Part d

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

Whether or not my next fish is blue

- $\text{Poi}(5B)$
- $\text{Bin}\left(1, \frac{R}{N}\right)$
- $\text{Ber}\left(\frac{B}{N}\right)$

Problem 3 - Balls and Bins

Problem 3 - Balls and Bins

Let X be the number of bins that remain empty when m balls are distributed into n bins randomly and independently. For each ball, each bin has an equal probability of being chosen. (Notice that two bins being empty are not independent events: if one bin is empty, that decreases the probability that the second bin will also be empty. This is particularly obvious when $n = 2$ and $m > 0$.) Find $\mathbb{E}[X]$.

Work on this problem with the people around you, and then we'll go over it together!

Problem 3 - Balls and Bins

Let X be the number of bins that remain empty when m balls are distributed into n bins randomly and independently. For each ball, each bin has an equal probability of being chosen. (Notice that two bins being empty are not independent events: if one bin is empty, that decreases the probability that the second bin will also be empty. This is particularly obvious when $n = 2$ and $m > 0$.) Find $\mathbb{E}[X]$.

Let X_i be 1 if bin i is empty, and 0 otherwise.

$$X = \sum_{i=1}^{n} X_i$$

$$\mathbb{E}[X_i] = 1 * \mathbb{P}(X_i = 1) + 0 * \mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = \left(\frac{n-1}{n}\right)^m$$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n * \left(\frac{n-1}{n}\right)^m$$

Problem 4 - 3 Sided Die

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the PMF of X?

(b) Find $E[X]$.

(c) Find $Var(X)$.

Work on this problem with the people around you, and then we'll go over it together!

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the PMF of X?

First let us define the range of X. A three sided-die can take on values {1,2,3}. Since X is the sum of two rolls, the range of X is $\Omega_X = \{2, 3, 4, 5, 6\}$.

We must define two random variables R_1, R_2 with R_1 being the roll of the first die, and R_2 being the roll of the second die. Then, $X = R_1 + R_2$. Note that $\Omega_{R1} = \Omega_{R2} = \{1,2,3\}$$. With that in mind we can find the PMF of X:

Work on this problem with the people around you, and then we'll go over it together!

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the PMF of X?

This gives us the following:

$$\begin{aligned}
p_X(k) = \mathbb{P}(X = k) &= \sum_{i \in \Omega_{R1}} \mathbb{P}(R_1 = i, R_2 = k - i) \\
&= \sum_{i \in \Omega_{R1}} \mathbb{P}(R_1 = i) \cdot \mathbb{P}(R_2 = k - i) \qquad \text{(By independence of the rolls)} \\
&= \sum_{i \in \Omega_{R1}} \frac{1}{3} \cdot p_{R2}(k - i) \\
&= \frac{1}{3} \left(p_{R2}(k - 1) + p_{R2}(k - 2) + p_{R2}(k - 3)\right)
\end{aligned}$$

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(a) What is the PMF of X?

At this point, we can evaluate the pmf of X for each value in the range of X, noting that $p_{R2}(k - i) = 0$ if $k - i \notin \Omega_{R2}$, $1/3$ otherwise. We get:

$$p_X(k) = \begin{cases} 1/9 & k = 2 \\ 2/9 & k = 3 \\ 3/9 & k = 4 \\ 2/9 & k = 5 \\ 1/9 & k = 6 \\ 0 & \text{otherwise} \end{cases}$$

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(b) Find $E[X]$.

There are two ways to find the expected value of X. We could apply the *definition of expectation* using the PMF found in part (a). This gives us

$$\mathbb{E}[X] = \sum_{k=2}^{6} k p_X(k) = 2 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} + 4 \cdot \frac{3}{9} + 5 \cdot \frac{2}{9} + 6 \cdot \frac{1}{9} = \boxed{4}$$

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(b) Find $E[X]$.

Alternatively, we can use *linearity of expectation* here. Let R_1 be the roll of the first die, and R_2 the roll of the second. Then, $X = R_1 + R_2$.
By linearity of expectation, we get:

$$\mathbb{E}[X] = \mathbb{E}[R_1 + R_2] = \mathbb{E}[R_1] + \mathbb{E}[R_2]$$

We compute:

$$\mathbb{E}[R_1] = \sum_{i \in \Omega_{R1}} i \cdot \mathbb{P}(R_1 = i) = \sum_{i \in \Omega_{R1}} i \cdot \frac{1}{3} = \frac{1}{3}(1 + 2 + 3) = 2$$

Similarly, $E[R_2] = 2$, since the rolls are independent.

Plugging into our expression for the expectation of X gives us:

$$E[X] = 2 + 2 = \boxed{4}$$

Problem 4 - 3 Sided Die

Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)

(c) Find $\mathrm{Var}[X]$.

We know from the definition of variance that

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

We can compute the $\mathbb{E}[X^2]$ term as follows:

$$\mathbb{E}[X^2] = \sum_{x=2}^{6} x^2 p_X(x) = \frac{2^2 \cdot 1 + 3^2 \cdot 2 + 4^2 \cdot 3 + 5^2 \cdot 2 + 6^2 \cdot 1}{9} = \frac{52}{3}$$

Plugging this into our variance equation gives us

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{52}{3} - 4^2 = \boxed{\frac{4}{3}}$$

Problem 10 - Poisson Practice

Problem 10 – Poisson Practice

Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

a) What is the probability of getting exactly 5 days of snow in a year?

a) According to the Poisson model, what is the probability of getting 367 days of snow?

Work on this problem with the people around you, and then we'll go over it together!

Problem 10 – Poisson Practice

Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

a) What is the probability of getting exactly 5 days of snow in a year?

Problem 10 – Poisson Practice

- Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

a) What is the probability of getting exactly 5 days of snow in a year?

$$\text{Let } X \sim \mathrm{Poi}(3) \text{ Then } p_X(5) = \frac{3^5 e^{-3}}{5!} \approx 0.1008$$

Problem 10 – Poisson Practice

Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

b) According to the Poisson model, what is the probability of getting 367 days of snow?

Problem 10 – Poisson Practice

- Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

b) According to the Poisson model, what is the probability of getting 367 days of snow?

Let $X \sim \text{Poi}(3)$ Then $p_X(367) = \frac{3^{367} e^{-3}}{367!} \approx 1.08 \times 10^{-610}$

Problem 10 – Poisson Practice

- Seattle averages 3 days with snowfall per year. Suppose the number of days with snowfall follows a Poisson distribution.

b) According to the Poisson model, what is the probability of getting 367 days of snow?

Let $X \sim \text{Poi}(3)$ Then $p_X(367) = \frac{3^{367} e^{-3}}{367!} \approx 1.08 \times 10^{-610}$

That's a very small estimate, but of course the true probability is 0. Recall that using a Poisson distribution is a modeling assumption, it may produce nonzero probabilities for events that are practically impossible.

Problem 7 – True or False?

Problem 7 – True or False?

- a) For any random variable X, we have $E[X^2] \geq E[X]^2$

Problem 7 – True or False?

a) For any random variable X, we have $E[X^2] \geq E[X]^2$

True

Since $0 \leq \mathrm{Var}(X) = E[(X - E[X])^2]$, since the squaring necessitates the result is non-negative.

Problem 7 – True or False?

- b) Let X, Y be random variables. Then, X and Y are independent if and only if $E[XY] = E[X]E[Y]$

Problem 7 – True or False?

b) Let X, Y be random variables. Then, X and Y are independent if and only if $E[XY] = E[X]E[Y]$

False

The forward implication is true, but the reverse is not. For example, if X is the discrete uniform random variable on the set $\{-1, 0, 1\}$ such that $P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$, and $Y = X^2$, we have $E[X] = 0$, so $E[X]E[Y] = 0$. However, since $X = X^3$, $E[XY] = E[XX^2] = E[X^3] = E[X] = 0$, we have that $E[X]E[Y] = 0 = E[XY]$. However, X and Y are not independent; indeed, $P(Y = 0|X = 0) = 1 \neq \frac{1}{3} = P(Y = 0)$

Problem 7 – True or False?

c) Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ be independent. Then, $X + Y \sim \text{Binomial}(n + m, p)$

Problem 7 – True or False?

e) Let $X \sim \mathrm{Binomial}(n, p)$ and $Y \sim \mathrm{Binomial}(m, p)$ be independent. Then, $X + Y \sim \mathrm{Binomial}(n + m, p)$

True

X is the sum of n independent Bernoulli trials, and Y is the sum of m. So $X + Y$ is the sum of $n + m$ independent Bernoulli trials, so $X + Y \sim \mathrm{Binomial}(n + m, p)$.

Problem 7 – True or False?

- d) Let $X_1, \dots, X_{n+1}$ be independent $\mathrm{Bernoulli}(p)$ random variables. Then, $\mathrm{E}[\sum_{i=1}^{n} X_i X_{i+1}] = np^2$.

Problem 7 – True or False?

d) Let $X_1, \dots, X_{n+1}$ be independent $\mathrm{Bernoulli}(p)$ random variables. Then, $\mathrm{E}[\sum_{i=1}^{n} X_i X_{i+1}] = np^2$.

True

Notice that $X_i X_{i+1}$ is also Bernoulli (only takes on 0 and 1), but is 1 if and only if both are 1, so $X_i X_{i+1} \sim \mathrm{Bernoulli}(p^2)$. The statement holds by linearity, since $E[X_i X_{i+1}] = p^2$.

Problem 7 – True or False?

- e) Let $X_1, \ldots, X_{n+1}$ be independent $\text{Bernoulli}(p)$ random variables. Then, $\text{Y} = \sum_{i=1}^{n} X_i X_{i+1} \sim \text{Binomial}(n, p^2)$.

Problem 7 – True or False?

e) Let $X_1, \dots, X_{n+1}$ be independent $\mathrm{Bernoulli}(p)$ random variables. Then, $\mathrm{Y} = \sum_{i=1}^{n} X_i X_{i+1} \sim \mathrm{Binomial}(n, p^2)$.

False

They are all Bernoulli p^2 as determined in the previous part, but they are not independent. $P(X_1X_2 = 1|X_2X_3 = 1) = P(X_1 = 1) = p \neq p^2 = P(X_1X_2 = 1)$.

Problem 7 – True or False?

f) If $X \sim \text{Bernoulli}(p)$, then $nX \sim \text{Binomial}(n, p)$.

Problem 7 – True or False?

f) If $X \sim \text{Bernoulli}(p)$, then $nX \sim \text{Binomial}(n, p)$.

False

The range of X is $\{0, 1\}$, so the range of nX is $\{0, n\}$. nX cannot be $\text{Bin}(n, p)$, otherwise its range would be $\{0, 1, \dots, n\}$.

Problem 7 – True or False?

- g) If $X \sim \text{Binomial}(n, p)$, then $\frac{X}{n} \sim \text{Bernoulli}(p)$.

Problem 7 – True or False?

g) If $X \sim \text{Binomial}(n, p)$, then $\frac{X}{n} \sim \text{Bernoulli}(p)$.

False

Again, the range of X is $\{0, 1, \ldots, n\}$, so the range of $\frac{X}{n}$ is $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$. Hence it cannot be $\text{Ber}(p)$, otherwise its range would be $\{0, 1\}$.

Problem 7 – True or False?

h) For any two independent random variables. X, Y, we have
$\mathrm{Var}(X - Y) = \mathrm{Var}(X) - \mathrm{Var}(\mathrm{Y})$.

Problem 7 – True or False?

h) For any two independent random variables. X, Y, we have $\mathrm{Var}(X - Y) = \mathrm{Var}(X) - \mathrm{Var}(\mathrm{Y})$.

False

$$\mathrm{Var}(X - Y) = \mathrm{Var}\big(X + (-Y)\big) = \mathrm{Var}(X) + (-1)^2\mathrm{Var}(Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

That’s All, Folks!

Thanks for coming to section this week!
Any questions?