Section 4

Review

- **Random Variable (rv)**: A numeric function $X : \Omega \to \mathbb{R}$ of the outcome.
- **Range/Support:** The support/range of a random variable X, denoted Ω_X, is the set of all possible values that X can take on.
- **Discrete Random Variable (drv)**: A random variable taking on a countable (either finite or countably infinite) number of possible values.
- **Probability Mass Function (pmf) for a discrete random variable $\mathbf{X}$**: a function $p_X : \Omega_X \to [0, 1]$ with $p_X(x) = \mathbb{P}(X = x)$ that maps possible values of a discrete random variable to the probability of that value happening, such that $\sum_x p_X(x) = 1$.
- **Cumulative Distribution Function (CDF) for a random variable $\mathbf{X}$**: a function $F_X : R \to R$ with $F_X(x) = \mathbb{P}(X \leqslant x)$
- **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 $X_1, \ldots, X_n$,

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

- **Variance**: Let X be a random variable and $\mu = \mathbb{E}[X]$. The variance of X is defined to be $\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2]$. Notice that since this is an expectation of a non-negative random variable ($(X - \mu)^2$), variance is always non-negative. With some algebra, we can simplify this to $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- **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 $\sigma = \sqrt{\mathrm{Var}(X)}$.
- **Property of Variance**: Let $a, b \in \mathbb{R}$ and let X be a random variable. Then, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
- **Independence**: Random variables X and Y are independent iff

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

  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 $X_1, \ldots, X_n$ 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, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. 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, $\mathrm{Var}(aX + bY + c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$.

Task 1 – Content Review

a) True or false: the range of a random variable X is the set of probabilities corresponding to the possible values X can take on.

b) What is the relationship between standard deviation and variance of a random variable X?

- ◯ $\sigma = (\text{Var}(X))^2$
- ◯ $\sigma = \text{Var}(X^2)$
- ◯ $\text{Var}(X) = \sigma^2$

c) Let X be the random variable representing the outcome of taking the sum of a 3-dice roll of 6-sided dice. Which function would you use to determine the probability that $X = 7$?

- ◯ CDF (cumulative distribution function)
- ◯ PMF (probability mass function)

d) Let X be the random variable representing the outcome of taking the sum of a 3-dice roll of 6-sided dice. Which function would you use to determine the probability that $X \leqslant 7$?

- ◯ CDF (cumulative distribution function)
- ◯ PMF (probability mass function)

e) A random variable X has the PMF

$$p_X(x) = \begin{cases} 1/4 & x = -1 \\ 1/4 & x = 0 \\ 1/2 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $\mathbb{E}[X]$?

- ◯ -1/4
- ◯ 3/4
- ◯ 1
- ◯ 2

f) A random variable X has the PMF

$$p_X(x) = \begin{cases} 1/4 & x = -1 \\ 1/4 & x = 0 \\ 1/2 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $\text{Var}[X]$?

- ◯ $3/4$
- ◯ 1
- ◯ $((1/4) + 2) - ((\frac{3}{4})^2) = 27/16$
- ◯ $((1/4) + 2) + ((\frac{3}{4})^2) = 45/16$

Task 2 – Identify that range!

Identify the support/range Ω_X of the random variable X, if X is...

a) The sum of two rolls of a six-sided die.

b) The number of lottery tickets I buy until I win it.

c) The number of heads in n flips of a coin with $0 < \mathbb{P}(\text{head}) < 1$.

d) The number of heads in n flips of a coin with $\mathbb{P}(\text{head}) = 1$.

Linearity of Expectation Problems
The next few problems are expectation and linearity of expectation problems. When finding the expected value of a random variable, first think about if the range is small enough so we can come up with the PMF and use the definition of expectation. Also, think about if there is a random variable from the zoo this random variable follows. If neither is possible, we will most likely want to use linearity Here's a general template for that!

1. **Decompose**. Write the random variable X as a sum of random variables: $X = X_1 + X_2 + ... + X_n$. Often, these X_i's are indicator random variables, especially if we're dealing with some kind of count.
2. **Apply LoE**. Apply LoE to $\mathbb{E}[X]$: $\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + ... + \mathbb{E}[X_n] = \sum_{i=1}^{n} \mathbb{E}[X_i]$.
3. **Conquer.** Compute each of $\mathbb{E}[X_i]$ and the plug it in to get the final answer.

Task 3 – Hungry Washing Machine

You have 10 pairs of socks (so 20 socks in total), with each pair being a different color. You put them in the washing machine, but the washing machine eats 4 of the socks chosen at random. Every subset of 4 socks is equally probable to be the subset that gets eaten. Let X be the number of complete pairs of socks that you have left.

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

b) Find $F_X(k)$, the CDF for X.

c) Find $\mathbb{E}[X]$ from the definition of expectation.

d) Find $\mathbb{E}[X]$ using linearity of expectation.

e) Which way was easier? Doing both (a) and (b), or just (c)?

Task 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 probability mass function of X?

b) What is the cumulative distribution function of X, partitioning the intervals on each possible value of X in its range?

c) Find $\mathbb{E}[X]$ directly from the definition of expectation.

d) Find $\mathbb{E}[X]$ again, but this time using linearity of expectation.

Task 5 – Practice

a) Let X be a random variable with $p_X(k) = ck$ for $k \in \{1, \ldots, 5\} = \Omega_X$, and 0 otherwise. Find the value of c that makes X follow a valid probability distribution and compute its mean and variance ($\mathbb{E}[X]$ and $\text{Var}(X)$).

b) Let X be *any* random variable with mean $\mathbb{E}[X] = \mu$ and variance $\text{Var}(X) = \sigma^2$. Find the mean and variance of $Z = \dfrac{X - \mu}{\sigma}$. (When you're done, you'll see why we call this a "standardized" version of X!)

c) Let X, Y be independent random variables. Find the mean and variance of $X - 3Y - 5$ in terms of $\mathbb{E}[X], \mathbb{E}[Y], \text{Var}(X)$, and $\text{Var}(Y)$.

d) Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables each with mean μ and variance σ^2. The sample mean is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find the mean and variance of $\bar{X}$. If you use the independence assumption anywhere, **explicitly label** at which step(s) it is necessary for your equalities to be true.

Task 6 – Symmetric Difference

For two sets A and B, define the **symmetric difference** Δ to be the set

$$A \Delta B = (A \cap B^C) \cup (B \cap A^C) = (A \cup B) \cap (A^C \cup B^C),$$

i.e., the set containing elements that are in exactly one of A and B. For example, if $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$, then $A \Delta B = \{1, 4\}$, since 1 is in A and not in B, and 4 is in B and not in A. $2, 3$ are in A and B, so they are not included in the symmetric difference.

Suppose A and B are random, independent (possibly empty) subsets of $\{1, 2, \ldots, n\}$, where each subset is equally likely to be chosen as A or B. Let X be the random variable that is the size of $A \Delta B$ (in the example above, X would be 2). What is $\mathbb{E}[X]$?

Task 7 – Hat Check

At a reception, n people give their hats to a hat-check person. When they leave, the hat-check person gives each of them a hat chosen at random from the hats that remain. What is the expected number of people who get their own hats back? (Notice that the hats returned to two people are not independent events: if a certain hat is returned to one person, it cannot also be returned to the other person.)

Task 8 – Balls in 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 $\mathbb{E}[X]$.

Task 9 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

a) Find $p_X(k)$, the probability mass function for X.

b) Compute $\mathbb{E}[X]$ from the definition.

c) Compute $\mathbb{E}[X]$ again, but using linearity of expectation.

Task 10 – Expectations, Independence, and Variance

a) Let U be a random variable which is uniform over the set $[n] = \{1, 2, \ldots, n\}$, i.e, $\mathbb{P}(U = i) = \frac{1}{n}$ for all $i \in [n]$. Compute $\mathbb{E}\left[U^2\right]$ and Var (U).

Hint: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ and $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

b) Let Y_1 and Y_2 be the independent outcomes of two fair 6-sided dice rolls, and let $Z = Y_1 + Y_2$. Then, compute $\mathbb{E}\left[Z^2\right]$ and Var (Z).

Hint: Try to use an indirect solution using linearity and independence, without the need of explicitly giving the distribution of Z^2.

Task 11 – Pond fishing

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

a) how many of the next 10 fish I catch are blue, if I catch and release

b) how many fish I had to catch until my first green fish, if I catch and release

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

d) whether or not my next fish is blue

e) how many of the next 10 fish I catch are blue, if I do not release the fish back to the pond after each catch

f) how many fish I have to catch until I catch three red fish, if I catch and release