MATH170E - Discussion Supplements for Winter 23

Contents are motivated from [1], and lecture notes from Jun Yin scribed by Thomas Blakey. [1]

Contents

1 Discussion 1

We will review some basic topics from Math 61 - Discrete Structures that I think would be useful for this class.

Basic Set Notations:

- Sets can be described using $\{x_1, \cdots, x_n\}$ (rooster notation) or $\{x \in U : P(x)\}$ (set builder notation). The second notation reads "the set of all elements in the set U such that $P(x)$ is true". Furthermore, we denote the empty set, the set containing nothing, as $\{\} = \emptyset$.

  Example: The set of natural number $\mathbb{N} = \{1, 2, 3 \cdots\}$, the set of integers $\mathbb{Z} = \{0, 1, -1, 2, -2, \cdots\}$, the set of positive integers $\mathbb{Z}^+ = \{0, 1, 2, 3, \cdots\} = \{x \in \mathbb{Z} : x \geq 0\}$.

- We write $x \in A$ for "x is an element of the set A", and $x \notin A$ for "x is not an element of the set A".

  Example: $1 \in \mathbb{Z}, -1 \notin \mathbb{N}$.

- We write $A \subset B$ for "A is a subset of B". This means that each element in A must be in B. Similarly, we write $A \subsetneq B$ for "A is not a subset of B".

  Example: $\{1, 2\} \subset \{1, 2, 3\}$, but $\{1, 2, 3\} \subsetneq \{1, 2\}$

- A set can contain sets and tuples (you can think of tuples as n dimensional coordinates).

  Example: $\{1, (2, 3)\}, \{\{\}, 2\}, \{\emptyset, \{\emptyset\}\}$. Note that $2 \notin \{1, (2, 3)\}$, and $\{\{\}\} = \{\emptyset\}$ is not empty (since it contains an element, the empty set)!

- Repeated entries in a set are not included (ie we are only interested in "what is inside", and not "how many of each is inside".)

  Example: $\{1, 2, 2\} = \{1, 2\}$.

- We let $|A|$ denote the number of elements in the set A.

  Example: $|\{1, 2, 2\}| = |\{1, 2\}| = 2$, $|\{\{\}\}| = |\{\emptyset\}| = 1$.

- We write $A \cup B$ for "the union of sets A and B". This means that as long as the element appears in either A or B, it will be in the set $A \cup B$.
  We write $A \cap B$ for "the intersection of sets A and B". For an element to be in $A \cap B$, it must be in both sets A and B.
  We write $A \setminus B$ for "the complement of B in A". It contains all elements in A that are not in B.
  We write $B^c = U \setminus B$ for "the complement of the set B", usually understood in the context of a larger ambient set U. It contains all elements that are in the ambient set U, but not in the set of interest B.

  Example: $\{1, 2\} \cup \{3\} = \{1, 2, 3\}$, $\{1, 2\} \cap \{(1, 2)\} = \emptyset$ (the tuple $(1, 2)$ is not exactly the same as each of the individual elements 1 and 2).
  Furthermore, $0 \notin \mathbb{R} \setminus \mathbb{Z}$ is true.

Principles of Counting.

- Multiplication Principle.
  Suppose that a procedure E_1 has n_1 outcomes and for each of these possible outcomes, a procedure E_2 has n_2 outcomes. Then, the composite experiment E_1E_2 that consists of performing first E_1 and then E_2 has n_1n_2 possible outcomes.

  Example: To make an egg sandwich, you were considering if you should have your eggs either fried, scrambled, or boiled. Furthermore, there were three choices of bread available - white, wholemeal, and sourdough.
  By the multiplication principle, if we demand that the two pieces of bread used must be of the same type, then the number of different egg sandwiches that you could make is $3 \times 3 = 9$.
  On the other hand, if the order of the bread is important and both the "upper" and the "lower" bread can be different, then the number of different egg sandwiches that you can make is $3 \times 3 \times 3 = 27$.
  Food for thought: What would this value be if the order of the bread is not important but the "upper" and "lower" bread slices could be different?

- Permutations vs Combinations. The difference lies in the requirement of the "order" in which you pick/form arrangements.

  Example 1. The number of ways to pick 3 marbles out of a bag of 7, in which each marble is unique, is given by $7 \times 6 \times 5$ (by the multiplication principle), if the order in which you pick is important. In other words, picking a red marble before a blue is considered to be different as compared to picking a blue marble before a red. We denote this by $7P3$ (read as "7 permute 3") $= 7 \times 6 \times 5$
  If the order is not important, you would have to account for the fact that there are "repeated" cases. For instance, if you picked a red, blue, and green marble (which we denote as R, G, and B), then the following ordered arrangements are equivalent if they are now unordered - RGB, RBG, BRG, BGR, GRB, GBR. Hence, to "unorder" them, we pay the price of dividing the number of ordered arrangements by 6, since each of the 6 ordered arrangments corresponds to an unordered arrangement. It is worth noting that we obtain the number 6 by counting the number of permutations of 3 distinct objects (given by $3 \times 2 \times 1$).
  Henceforth, the number of ways to pick 3 marbles out of the bag in which the order is not important is given by $\frac{7\times6\times5}{3\times2\times1}$. This is denoted by $7C3$ (read as "7 choose 3") $= \frac{7\times6\times5}{3\times2\times1}$.

  This motivates the following definitions:

  - $n! := 1 \times 2 \times \cdots \times n$,
  - $nPr := n \times (n-1) \times \cdots (n-r+1) = \frac{n!}{(n-r)!}$; the number of ordered arrangements,
  - $\binom{n}{r} = nCr := \frac{nPr}{r!} = \frac{n!}{(n-r)!r!}$; the number of unordered arrangements.

  Note that it is easier to think of permutations as a two-step procedure - choose and rearrange. The "choosing" step is equivalent to looking at unordered arrangements, and the "rearrange" step arranges the objects of interest, inflating the number of arrangements by $r!$ (where r is the size of each arrangement). Hence, $nPr = nCr \times r!$.

  Example 2. The number of ways to select a president, a vice president, a secretary, and a treasurer in a club of ten people is given by $10P4 = 10 \cdot 9 \cdot 8 \cdot 7 = \frac{10!}{6!} = 5040$. In this case, we "permute" since the order is important, ie. here, we can think of the order as each position in the executive committee, which is important. Alternatively, you can first choose 4 people out of the 10, and permute them according to the individual executive committee positions which are distinguishable, in which $10P4 = 10C4 \times 4!$.

2 Discussion 2

Basic Set Theory II:
Algebra of Sets. For any sets A, B, C, we have the following properties:

- **Commutative Laws.**
  $A \cup B = B \cup A$,
  $A \cap B = B \cap A$.
- **Associative Laws.**
  $(A \cup B) \cup C = A \cup (B \cup C)$,
  $(A \cap B) \cap C = A \cap (B \cap C)$.
- **Distributive Laws.**
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- **De Morgan's Laws.**
  $(A \cup B)^c = A^c \cap B^c$,
  $(A \cap B)^c = A^c \cup B^c$.

You are free to use this to proof any properties of sets mathematically by quoting the appropriate properties/laws.

Remark: Another way to "prove" that two sets are equal, say $A = B$, is to show that $A \subset B$ and $B \subset A$ (recall that $\subset$ means "subset", ie $\subseteq$ that you see in high school). We shall not go over that - if you are interested, this is covered in Math 61 or you could ask me about that in office hours.

To visualize properties of sets, it is usually easy to draw the corresponding Venn Diagram.

Properties of Probability:

- The collection of all possible outcomes of a random experiment is denoted by S, and is called the **sample space**.
- An **event** A refers to a subset of the sample space, ie $A \subset S$.
- For events $A_1, \cdots, A_k$, we say that they are **mutually exclusive** (ie disjoint) if for any $i \neq j$ and $i, j \in \{1, 2, \cdots, k\}$, we have $A_i \cap A_j = \emptyset$.
- For events $A_1, \cdots, A_k$, we say that they are **exhaustive** if $A_1 \cup \cdots \cup A_k = S$.

For example, if we denote that possible outcomes of a die roll to be the sample space, then $S = \{1, 2, 3, 4, 5, 6\} = [[1, 6]]$. (Note that $[[1, 6]]$ is the notation used in class, representing integers from 1 to 6.) A possible event could be rolling an even number, corresponding to $A_1 = \{2, 4, 6\}$, which one can see that $A \subset S$. Furthermore, if we denote A_2 as the event of rolling a odd number, then $A_2 = \{1, 3, 5\}$. One can then check that A_1 and A_2 are both mutually exclusive (since $A_1 \cap A_2 = \emptyset$) and exhaustive (since $A_1 \cup A_2 = \{1, 2, 3, 4, 5, 6\} = S$).

In an abstract manner, we define **probability** as follows:

Definition 3. **Probability** is a real-valued set function $\mathbb{P}$ that assigns, to each event A in the sample space S, a number $\mathbb{P}(A)$, called the probability of the event A, such that the following properties are satisfied:

1. $\mathbb{P}(A) \geq 0$,
2. $\mathbb{P}(S) = 1$, and
3. if $A_1, A_2, \cdots$ are events, and $A_i \cap A_j = \emptyset$ for $i \neq j$ (ie mutually exclusive), then
$$\mathbb{P}(A_1 \cup A_2 \cup \cdots A_k) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_k)$$
for each positive integer k, and
$$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$$
for an infinite, but countable, number of events.

By viewing probability as a function that takes in a set and spits out a real number, we have the following properties:

Theorem 4. The following properties of $\mathbb{P}$ holds:

1. For each event A, we have $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$,
2. $\mathbb{P}(\emptyset) = 0$.
3. For arbitrary events A, B, if $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$,
4. For each event A, we have $\mathbb{P}(A) \leq 1$.
5. For arbitrary events A, B, we have
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$
6. (Principle of Inclusion-Exclusion for 3 events.) For arbitrary events A, B, and C, we have
$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

For a proof of theorem 4, refer to the textbook.

Example 5. Let A and B be two events of a sample space S such that $\mathbb{P}(A) = 0.5$ and $\mathbb{P}(B) = 0.6$.

(i) Prove that $\mathbb{P}(A \cap B) \geq 0.1$.

(ii) Suppose that $\mathbb{P}(A \cap B) \leq 0.3$, prove that $\mathbb{P}(A \cup B) \geq 0.8$.

Suggested Solution:

(i) By property 5 of Theorem 4, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \tag{1}$$

Rearranging the equation, we get

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B).$$

Since $A \cap B$ is an event, we have that $\mathbb{P}(A \cap B) \leq 1$ holds. (See property 4 of theorem 4.) This implies that

$$\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \leq 1.$$

Substitute the values of $\mathbb{P}(A)$ and $\mathbb{P}(B)$ in, we have

$$\begin{aligned} 1.1 - \mathbb{P}(A \cup B) &\leq 1 \\ \mathbb{P}(A \cup B) &\geq 0.1. \end{aligned}$$

(ii) On the other hand, from (1) and substituting the values of $\mathbb{P}(A)$ and $\mathbb{P}(B)$, we have

$$\mathbb{P}(A \cup B) = 1.1 - \mathbb{P}(A \cap B).$$

Since $\mathbb{P}(A \cap B) \leq 0.3$, then we have $-\mathbb{P}(A \cap B) \geq -0.3$ and hence

$$\mathbb{P}(A \cup B) = 1.1 - \mathbb{P}(A \cap B) \geq 1.1 - 0.3 = 0.8.$$

Principles of Counting II.
Here, we recap some of the concepts covered in Discussion 1:

- Multiplication Principle.
  Suppose that a procedure E_1 has n_1 outcomes and for each of these possible outcomes, a procedure E_2 has n_2 outcomes. Then, the composite experiment E_1E_2 that consists of performing first E_1 and then E_2 has n_1n_2 possible outcomes.

- Permutations vs Combinations. The difference lies in the requirement of the "order" in which you pick/form arrangements. We begin with certain notations:

  - $n! := 1 \times 2 \times \cdots \times n$,
  - $nPr := n \times (n-1) \times \cdots (n-r+1) = \frac{n!}{(n-r)!}$; the number of ordered arrangements,
  - $\binom{n}{r} = nCr := \frac{nPr}{r!} = \frac{n!}{(n-r)!r!}$; the number of unordered arrangements.

  Note that it is easier to think of permutations as a two-step procedure - choose and rearrange. The "choosing" step is equivalent to looking at unordered arrangements, and the "rearrange" step arranges the objects of interest, inflating the number of arrangements by $r!$ (where r is the size of each arrangement). Hence, $nPr = nCr \times r!$.

We now continue with additional concepts which I think would be relevant for this class.

- Distinguishability of objects is equivalent to arrangements of letters.

  Suppose we would like to arrange 3 objects in a row. Then observe that:

  - If the objects are distinguishable; ie 3 humans, then the number of ways to arrange them is given by $3!$.
  - If the objects are indistinguishable; ie 3 pennies, then the number of ways to arrange them is 1. (Since they look the same, it doesnt matter how you would "rearrange" them.)

  Thus, it always makes sense to ask yourself if the objects of interest are distinguishable or not. However, the problem arises if some of the objects are distinguishable and some are not.

  The number of ways to arrange two pennies and a quarter is given by 3. The possible arrangements are as follows (P := Penny, Q := Quarter): QPP, PQP, and PPQ; thus, 3 ways.

  In fact, the same argument holds if we replace pennies by blue marbles and the quarter by red marbles (assuming the marbles are indistinguishable except for their colors). Furthermore, the same strategy holds; set objects of type one to be a "letter", set objects of type two to be another "letter", and the different arrangementes are represented by letters! In other words, the number of ways to arrange two blue marbles and one red marble is given by 3, corresponding to RBB, BRB, and BBR; where R := Red marble, B:= Blue marble.

- Number of arrangements of letters. This is best understood by examples:

  - Number of ways to arrange the letters of the word "MATH" is given by $4!$, since M, A, T, and H are all distinguishable from one another.
  - Number of ways to arrange the letters of the word "ALLY" is given by $\frac{4!}{2!}$. The strategy is as follows. First, we assume that all 4 letters are distinguishable, corresponding to $4!$. Next, since the two Ls' are actually indistinguishable, we divide by the number of permutations of the two Ls', given by $2!$. This can be visualized as follows:

    1. If the Ls' are distinct, then we have A, L_1, L_2, and Y. Here, the subscript represents distinguishability. Then, we obtain that the number of ways to arrange the words are $4! = 24$. The arrangements are:
       AL_1L_2Y, AL_1YL_2, AL_2L_1Y, AL_2YL_1, AYL_1L_2, AYL_2L_1,
       YL_1L_2A, YL_1AL_2, YL_2L_1A, YL_2AL_1, YAL_1L_2, YAL_2L_1,
       L_1AL_2Y, L_1AYL_2, L_1L_2AY, L_1L_2YA, L_1YAL_2, L_1YL_2A,
       L_2AL_1Y, L_2AYL_1, L_2L_1AY, L_2L_1YA, L_2YAL_1, L_2YL_1A.

2. However, we note that the Ls' are in fact the same! In reality, we see that AL_1YL_2 is the same as AL_2YL_1. It helps to match pairs that are exactly the same, and hence we can observe that there are a total of 12 pairs, each pair corresponding a different arrangement if we swapped the positions of L_1 and L_2. In other words, out of the 24 arrangements, we see that for each arrangement with L_1 coming first, we have a corresponding arrangement with L_2 coming first such that they are equivalent if the Ls' are indistinguishable. This then justifies dividing by $2!$, the number of ways to rearrange the Ls'.

– The number of ways to arrange $n_1 + n_2 + \cdots n_k$ letters for which we have k distinguishable letters, with n_i copies of that letter for each of these k letters, is given by

$$\binom{n_1 + \cdots + n_k}{n_1, n_2, \cdots, n_k} = \frac{(n_1 + n_2 + \cdots + n_k)!}{n_1! n_2! \cdots n_k!}.$$

Try to convince yourself with the above logic that this is true. Observe that the denominator accounts for "overcounting" if we take a simple factorial over the number of letters, and this is as explained in the previous bullet point above. The above notation is from the textbook, and it is called a "multinomial coefficient".

For this concept, practice makes perfect. Hence, we shall dive deep into the following examples.

Example 6. A bridge hand is found by taking 13 cards at random and without replacement from a deck of 52 playing cards. Find the number of possible hands containing 5 spades, 4 hearts, 3 diamonds and 1 club.

Suggested Solution: For each of these suits, there are 13 different cards (of different face values). Hence, this is equivalent to choosing 5 of out of the 13 different spades, 4 out of the 13 different hearts, 3 out of the 13 different diamonds, and 1 out of the 13 different clubs. For each choice of spades (/hearts/diamonds/club), this does not change the number of possible hands for the remaining suits. Hence, by multiplication principle, the number of possible hands is

$$\binom{13}{5} \times \binom{13}{4} \times \binom{13}{3} \times \binom{13}{1} = 3421322190.$$

Example 7. A poker hand is found by taking 5 cards at random and without replacement from a deck of 52 playing cards. Find the number of possible hands consisting of only two distinct suits, with 3 cards for the first suit and 2 for the second.

Suggested Solution: Out of the 4 suits, we would like to choose 2 of them. Note here that the order matters, as we associate the order with the number of cards in the suit (ie first suit picked corresponds to 3 cards; second suit picked corresponds to 2 cards). Then, for the first suit, we will have to pick 3 out of the 13 cards of that suit available in the deck. Similarly, we have to pick 2 out of the 13 cards of the second suit from the deck. By the multiplication principle, we thus have

$$4P2 \times \binom{13}{3} \times \binom{13}{2} = \binom{4}{2} \times 2! \times \binom{13}{3} \times \binom{13}{2} = 267696.$$

Example 8. Consider the scenario in Example 7, but instead, we are taking 5 cards at random **with replacement** from a deck of 52 playing cards. Find the number of possible hands consisting of only two distinct suits, with 3 cards for the first suit and 2 for the second.

Suggested Solution: The solution to this is similar to that in the previous example. However, instead of "choosing" cards for each suit, we note that we are allowed to pick the "same" card from the deck. Thus, the number of ways to pick 3 cards of the same suit is given by 13^3. Similarly, the number of ways to pick 2 cards of another suit is given by 13^2. Hence, by the multiplication rule, we have

$$4P2 \times 13^5 = 4455516.$$

Example 9. You have 8 distinct pieces of food in your pantry. You want to choose two for breakfast, three for lunch, and three for dinner. How many ways can you do so?

Suggested Solution: The way to imagine this problem is to have eight slots corresponding to food item 1 to 8. Then, we attempt to assign 2 B's, 3 L's, and 3 D's (corresponding to the two items to be picked for breakfast, three items to be picked for lunch, and three items to be picked for dinner). Equivalently, the question is asking for the number of ways to arrange the letters BBLLLDDD. Using the concept of multinomial coefficients, we have

$$\frac{8!}{2!3!3!} = 560.$$

Example 10. How many ways are there to select a president, a vice president, and two secretaries in a club of ten people?

Suggested Solution: We can model this as a two-step process. The first step is to choose 4 candidates out of the 10. In this step, the order does not matter; we instead delay the assignment of candidates to their positions in the second step. Since these candidates are distinguishable (as they are humans), we now consider four slots corresponding to candidates 1 to 4. Then, we attempt to assign 1 P, 1 V, and 2 Ss' to the 4 slots. Equivalently, this is then the number of ways to arrange the letters PVSS. By the multiplication principle, we thus have

$$\binom{10}{4} \times \frac{4!}{2!1!1!} = 2520.$$

3 Discussion 3

Probabilities for Equally Likely Outcome Experiments.
Consider a sample space $S = \{e_1, \cdots, e_n\}$ for n possible outcomes of an experiment. If each of the outcomes is **equally likely**, then it means that the associated probability set function defined on S is such that

$$\mathbb{P}(\{e_i\}) = \frac{1}{n}.$$

If the number of outcomes for a given event A is $|A|$ (which is basically the size of the set or the number of elements in the set), then we have under the assumption of equally likely outcomes that

$$\mathbb{P}(A) = \frac{|A|}{|S|}.$$

Example 11. A poker hand is found by taking 5 cards at random and without replacement from a deck of 52 playing cards. Assuming that each hand is equally likely, find the probability of obtaining a hand consisting of only two distinct suits, with 3 cards for the first suit and 2 for the second.

Suggested Solution: Let us denote the set A to be the event as mentioned in the problem. Out of the 4 suits, we would like to choose 2 of them. Note here that the order matters, as we associate the order with the number of cards in the suit (ie first suit picked corresponds to 3 cards; second suit picked corresponds to 2 cards). Then, for the first suit, we will have to pick 3 out of the 13 cards of that suit available in the deck. Similarly, we have to pick 2 out of the 13 cards of the second suit from the deck. By the multiplication principle, we thus have

$$|A| = 4P2 \times \binom{13}{3} \times \binom{13}{2} = \binom{4}{2} \times 2! \times \binom{13}{3} \times \binom{13}{2} = 267696.$$

On the other hand, if we consider the sample space to be all possible 5 card hands (in which the order does not matter), then we have

$$|S| = \binom{52}{5} = 2598960.$$

Hence, we have

$$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{429}{4165} \approx 0.103.$$

Example 12. Consider an experiment of tossing a fair coin 170 times. Find the probability of getting at least two heads. Leave your answer in a fraction.

Suggested Solution: The corresponding sample space is described by $S = \{H, T\}^{170}$. By writing it in this way, we are assuming that the order matters (since $(T, H, T, T, \cdots) \neq (H, T, T, T, \cdots)$). Furthermore, let A denote the event that we have at least two heads (H). Hence, we have

$$|S| = 2^{170}$$

since for each component, there are two ways to pick the value at that component (either H or T). On the other hand, note that there are many possible ways in which we obtain at least two heads out of 170 coin tosses. Instead, it is useful to think of complements. Observe that A^c refers to the even that we have at most one head, which we denote as B_0 and B_1 for the events corresponding to exactly zero and one head respectively. Since $A^c = B_0 \cup B_1$, which are disjoint, the corresponding number of ways to obtain that is given by

$$\begin{aligned}
|B_0| &= 1, \\
|B_1| &= 170, \\
|A^c| &= |B_0| + |B_1| = 1 + 170 = 171.
\end{aligned}$$

since there is only one way to obtain all tails, while there are 170 ways to obtain all but one tail, corresponding to the different possible arrangements that the single "H" can be out of the 170 available. Hence, we have

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{171}{2^{170}} = \frac{2^{170} - 171}{2^{170}}.$$

Conditional Probabilities

Definition 13. The **conditional probability** of an event A occurring given that event B has occurred, is defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

provided that $\mathbb{P}(B) > 0$.

One way to think about this definition is to think about the sample space shrinking to just B (since we already know at this point that event B has occurred). On the other hand, with respect to the original sample space S, if we want to know the probability that A has occurred and that we already knew that B occurred, the numerator would be described by the event $A \cap B$.

Here are some properties of conditional probabilities:

Theorem 14. The following properties of $\mathbb{P}$ holds. For any given event A and B such that $\mathbb{P}(B) > 0$, we have

1. $\mathbb{P}(A|B) \geq 0$,
2. $\mathbb{P}(B|B) = 1$, and
3. if $A_1, \cdots, A_k$ are mutually exclusive events, then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots A_K|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) + \cdots + \mathbb{P}(A_k|B)$$

   for each positive integer k, and

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) + \cdots$$

   for an infinite but countable number of events.
4. Properties 1 - 6 in the theorem for Discussion Supplement 2 holds by replacing $\mathbb{P}(\cdot)$ with $\mathbb{P}(\cdot|B)$.

For a proof of theorem 14, refer to the textbook.
Consequently, we have the following property:

(Multiplication Rule.) The probability that two events, A and B, both occur is given by

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

if $\mathbb{P}(A) > 0$ or by

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B)$$

provided that $\mathbb{P}(B) > 0$.

We will cover more stuff on conditional probabilities in the next discussion (or whenever Bayes' Theorem is covered in lectures). Here are some examples of the application of conditional probabilities. Examples 16 and 17 illustrate that we can view conditional probability problems with the use of a probability tree.

Example 15. Consider an experiment of tossing a fair coin 170 times. Find the probability of getting at most two heads, given that you have at least two heads.

Suggested Solution: Similar to Example 12, we denote the corresponding sample space by $S = \{H, T\}^{170}$. By writing it in this way, we are assuming that the order matters (since $(T, H, T, T, \cdots) \neq (H, T, T, T, \cdots)$). Furthermore, let A be the event that we have at most two heads, and B denote the event that we have at least two heads (H).
Thus, we have

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Note that the event $A \cap B$ refers to the number of ways of getting exactly 2 heads, with each outcome equally likely. Hence, we have

$$|A \cap B| = \frac{170!}{168!2!} = \binom{170}{2} = \frac{170 \times 169}{2} = 14365.$$

Note that this corresponds to the number of ways to arrange 2H and 168T in a row, or choose two positions out of the 170 where we would place the heads in. The third equality is obtained purely by the definition of $\binom{170}{2}$. From our computation in Example 12, we have

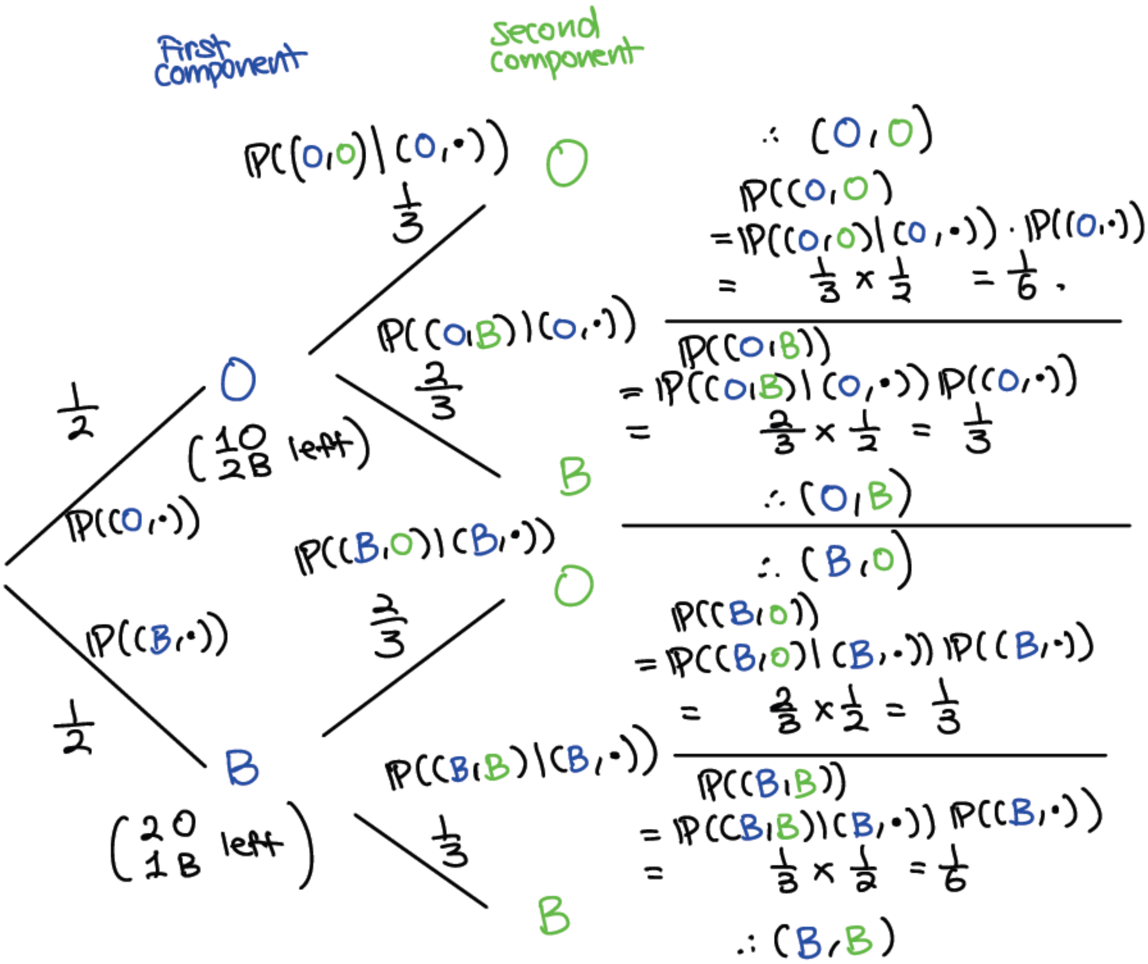
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{14365}{2^{170}}}{\frac{2^{170}-171}{2^{170}}} = \frac{14365}{2^{170} - 171}.$$

Example 16. An urn contains four colored balls; two orange and two blue. Two balls are selected at random without replacement, and you are told that at least one of them is orange. What is the probability that both balls are orange?

Suggested Solution: The sample space S is described by $S = \{(x, y) \in \{B, O\}^2\}$. However, note that each outcome $(O, O), (O, B), (B, O)$, and (B, B) are not equally likely (since the balls are selected without replacement). The required probability is given by

$$\begin{aligned}
\mathbb{P}(\{(O,O)\}| \geq 1 \times O) &= \mathbb{P}(\{(O,O)\}|\{(O,O),(B,O),(O,B)\}) \\
&= \frac{\mathbb{P}(\{(O,O)\} \cap \{(O,O),(B,O),(O,B)\})}{\mathbb{P}(\{(O,O),(B,O),(O,B)\})} \\
&= \frac{\mathbb{P}(\{(O,O)\})}{\mathbb{P}(\{(O,O),(B,O),(O,B)\})} \\
&= \frac{\mathbb{P}(\{(O,O)\})}{\mathbb{P}(\{(O,O)\}) + \mathbb{P}(\{(B,O)\}) + \mathbb{P}(\{(O,B)\})},
\end{aligned}$$

in which the last inequality follows from the fact that obtaining a different tuple is mutually exclusive (we can't draw both oranges and an orange with a blue at the same time).
To compute each probability, we refer to the probability tree below. In the tree below and subsequent computations, we shall omit $\{\}$ for brevity.



Hence, we have

$$\begin{aligned}\text{Required Probability} &= \frac{\mathbb{P}((O,O))}{\mathbb{P}((O,O))+\mathbb{P}((B,O))+\mathbb{P}((O,B))}\\ &= \frac{\mathbb{P}((O,O)|(O,\cdot))\mathbb{P}((O,\cdot))}{\mathbb{P}((O,O)|(O,\cdot))\mathbb{P}((O,\cdot))+\mathbb{P}((B,O)|(B,\cdot))\mathbb{P}((B,\cdot))+\mathbb{P}((O,B)|(O,\cdot))\mathbb{P}((O,\cdot))}\\ &= \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}+\frac{1}{3}} = \frac{1}{5}.\end{aligned}$$

Example 17. An urn contains 8 red and 7 blue balls. A second urn contains an unknown number of red balls and 9 blue balls. A ball is drawn from each urn at random, and the probability of getting two balls of the same color is $\frac{151}{300}$. How many red balls are in the second urn?

Suggested Solution. Let x be the number of red balls in the second urn. Consider the sample space as $S = \{R, B\}^2$, with the first component denoting the color of ball drawn from the first urn, and the second component for that of the second urn. Consequently, the probability of getting two balls of the same color is the event $A = \{(R, R), (B, B)\}$. The required probability (using the multiplication rule) is given by

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{(R,R)\}) + \mathbb{P}(\{(B,B)\}) \\
&= \mathbb{P}(\{(R,R)\}|\{(R,\cdot)\})\mathbb{P}(\{(R,\cdot)\}) + \mathbb{P}(\{(B,B)\}|\{(B,\cdot)\})\mathbb{P}(\{(B,\cdot)\}) \\
&= \mathbb{P}(R \text{ from second urn})\mathbb{P}(R \text{ from first urn}) + \mathbb{P}(B \text{ from second urn})\mathbb{P}(B \text{ from second urn}) \\
&= \left(\frac{x}{x+9}\right)\left(\frac{8}{7+8}\right) + \left(\frac{9}{x+9}\right)\left(\frac{7}{7+8}\right) \\
&= \frac{151}{300} \text{ as given in the question.}
\end{aligned}$$

Thus, to obtain x, we have to solve

$$\left(\frac{x}{x+9}\right)\left(\frac{8}{7+8}\right) + \left(\frac{9}{x+9}\right)\left(\frac{7}{7+8}\right) = \frac{151}{300}$$

in which one can obtain that

$$x = 11.$$

4 Discussion 4

Independence.

Definition 18. Events A and B are **independent** if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Otherwise, A and B are called **dependent** events.

On the other hand, one can show that this is equivalent to having

- $\mathbb{P}(A|B) = \mathbb{P}(A)$ and
- $\mathbb{P}(B|A) = \mathbb{P}(B)$.

The following records some properties of independent events:

Theorem 19. If A and B are independent events, then the following pairs of events are also independent:

- A and B^c,
- A^c and B, and
- A^c and B^c.

The proof of this theorem is in the textbook; though you are not allowed to quote this for HW 3 Q 1(a) (re-prove it on your own)!
For more than two events, we have the following notion of independence:

Definition 20. Events A, B, and C are **mutually independent** if and only if the following two conditions hold:

(a) A, B, and C are **pairwise independent**, that is, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, and $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, **and**

(b) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Note that such a definition can easily extend to more than three events. Here are some exercises on independence.

Example 21. Prove that if A, B, and C are mutually independent, then $A \cup B$ is independent of C.

Suggested Solution: It suffices to show that $\mathbb{P}((A \cup B) \cap C) = \mathbb{P}(A \cup B)\mathbb{P}(C)$ under mutual independence. Indeed, observe that

$$\begin{aligned}
\mathbb{P}((A \cup B) \cap C) &\overset{(1)}{=} \mathbb{P}((A \cap C) \cup (B \cap C)) \\
&\overset{(2)}{=} \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C) \\
&\overset{\text{ind}}{=} \mathbb{P}(A)\mathbb{P}(C) + \mathbb{P}(B)\mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \\
&= \mathbb{P}(C)(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)) \\
&\overset{\text{ind}}{=} \mathbb{P}(C)(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&\overset{(2)}{=} \mathbb{P}(C)\mathbb{P}(A \cup B) \\
&= \mathbb{P}(A \cup B)\mathbb{P}(C).
\end{aligned}$$

Here, (1) refers to the distributive law of intersections and unions, and (2) refers to the formula $\mathbb{P}(X \cup Y) = \mathbb{P}(X) + \mathbb{P}(Y) - \mathbb{P}(X \cap Y)$ with the appropriate choices of X and Y. Furthermore, "ind" here represents independence.
An alternative method here is to use a Venn diagram. From there, one can deduce that

$$\mathbb{P}((A \cup B) \cap C) = \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C),$$

and the same set of arguments then follows.

Example 22. Suppose that events $A, B,$ and C are mutually independent, and that $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = x$ for some $x \in [0, 1]$. Furthermore, suppose that we have that $\mathbb{P}(A \cup B \cup C) = 1$. Prove that we must have $x = 1$ necessarily.

Hint: Use the fact that $x^3 - 3x^2 + 3x - 1 = (x - 1)^3$.

Suggested Solution: Using the inclusion-exclusion principle, we have

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \\ &\overset{\text{ind}}{=} \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(B)\mathbb{P}(C) + \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \\ &= x + x + x - x^2 - x^2 - x^2 + x^3 \\ 1 &= 3x - 3x^2 + x^3 \\ x^3 - 3x^2 + 3x - 1 &= 0. \end{aligned}$$

Using the hint, note that the only solution to $x^3 - 3x^3 + 3x - 1 = (x - 1)^3 = 0$ is given by $x = 1$ (unless you include complex numbers, but probabilities cannot be complex!).

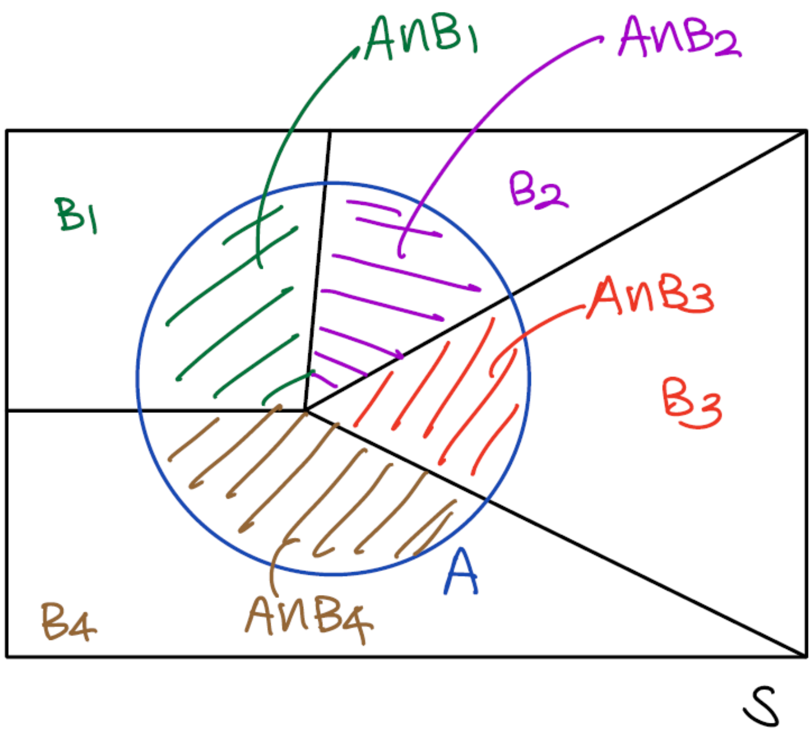
Bayes' Theorem.
Let $B_1, B_2, \cdots, B_m$ constitute a partition of the sample space S. That is, $S = B_1 \cup B_2 \cup \cdots \cup B_m$ and $B_i \cap B_j =$ for $i \neq j$. (In other words, the sets are mutually exclusive and exhaustive.) Suppose that $\mathbb{P}(B_i) \geq 0$ for each $i = 1, 2, \cdots, m$. For a given event A, we can write A as a union of m mutually exclusive events, namely,

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_m \cap A).$$

This then implies that

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i=1}^{m} \mathbb{P}(B_i \cap A) \\ &= \sum_{i=1}^{m} \mathbb{P}(B_i)\mathbb{P}(A|B_i), \end{aligned}$$

where the last step is obtained by the definition of conditional probability. This is also known as the **law of total probability**. A pictorial representation is given below.



If $\mathbb{P}(A) > 0$, then we have

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)}$$

for $k = 1, 2, \cdots, m$. Replacing $\mathbb{P}(A)$ with the law of total probability, we then have the **Bayes' theorem**, given by

$$\underbrace{\mathbb{P}(B_k|A)}_{\text{"posterior" probability}} = \frac{\overbrace{\mathbb{P}(B_k)}^{}_{\text{"prior" probability}} \quad \mathbb{P}(A|B_k)}{\underbrace{\sum_{i=1}^{m} \mathbb{P}(B_i)\mathbb{P}(A|B_i)}_{\text{"normalization" factor}}}$$

Some remarks for the formula above:

- We can think of $\mathbb{P}(B_k)$ as the prior probability (ie prior to knowing that A has happened, what is the probability of B_k), and $\mathbb{P}(B_k|A)$ as the posterior probability (ie posterior to/after knowing that A has happened, what is the probability of B_k happening).

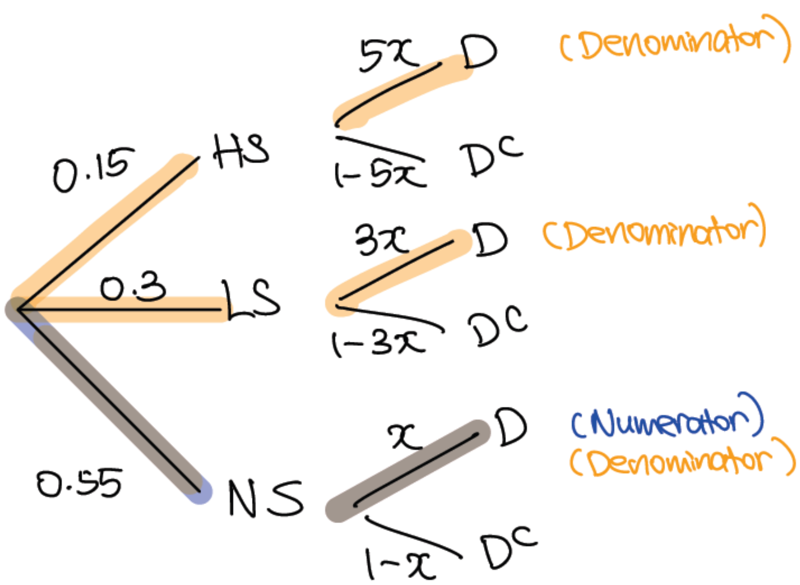
- Another useful way to use this formula is that one can use $\mathbb{P}(A|B_k)$ (and of course $\mathbb{P}(A|B_i)$ for the other $i \neq k$) to compute $\mathbb{P}(B_k|A)$. The formula is best understood (in my opinion) as a way to do as above, but not with the mathematical notation, but rather, a probability tree. This will be elaborated in the following examples.

Example 23. (Exercise 1.5.11.) At the beginning of a certain study of a group of persons, 15% were classified as heavy smokers, 30% were classified as light smokers, and 55% were classified as non-smokers. In a five-year study, it was determined that the death rates of heavy and light smokers were five and three times that of non-smokers, respectively. A randomly selected participant died over the five-year period. Calculate the probability that the participant was a non-smoker.

Denote HS as the set of heavy smokers, LS as the set of light smokers, NS as the set of non-smokers, and D be the set of dead participants. The required probability is $\mathbb{P}(NS|D)$. By Bayes' theorem, we have

$$\mathbb{P}(NS|D) = \frac{\mathbb{P}(NS \cap D)}{\mathbb{P}(D)}.$$

To compute this, we appeal to a probability tree with the following information: $\mathbb{P}(HS) = 0.15, \mathbb{P}(LS) = 0.3, \mathbb{P}(NS) = 0.55$. Furthermore, since the death rate of a non-smoker is unknown, let us denote it by x. This implies that $\mathbb{P}(D|NS) = x, \mathbb{P}(D|LS) = 3x$, and $\mathbb{P}(D|HS) = 5x$. Now, we see that we are given a conditional probability of $D|$something, and are asked to compute the conditional probability of something$|D$, which is why we appeal to Bayes' theorem. The branch corresponding to $NS - D$ goes to the numerator, while by the law of total probability, all branches that end up in D will contribute to the probability that the participant is dead and thus in the denominator.



Hence, we have

$$\begin{aligned}\mathbb{P}(NS|D) &= \frac{\mathbb{P}(D|NS)\mathbb{P}(NS)}{\mathbb{P}(D|NS)\mathbb{P}(NS) + \mathbb{P}(D|LS)\mathbb{P}(LS) + \mathbb{P}(D|HS)\mathbb{P}(HS)} \\ &= \frac{0.55\cancel{x}}{0.55\cancel{x} + 0.3(3\cancel{x}) + 0.15(5\cancel{x})} \\ &= \frac{1}{4}.\end{aligned}$$

Example 24. (Example 1.5.3.) A Pap smear is a screening procedure used to detect cervical cancer. Let $T-$ denote a negative test, $T+$ denote a positive test, $C+$ denote that a woman has cervical cancer, and $C-$ denote that a woman does not have cervical cancer. For women with this cancer, there are about 16% *false negatives*, that is

$$\mathbb{P}(T-|C+) = 0.16$$

For women without cancer, there are about 10% *false positives*, that is,

$$\mathbb{P}(T+|C-) = 0.1$$

In the United States, there are about 8 women in $100,000$ who have this cancer; that is

$$\mathbb{P}(C+) = 0.00008.$$

Compute the probability that a woman has cervical cancer, given that the Pap smear test indicates positive.

Suggested Solution: By Bayes Theorem, we have

$$\begin{aligned}
\mathbb{P}(C+|T+) &= \frac{\mathbb{P}(C+\cap T+)}{\mathbb{P}(T+)} \\
&= \frac{\mathbb{P}(T+|C+)\mathbb{P}(C+)}{\mathbb{P}(T+|C+)\mathbb{P}(C+) + \mathbb{P}(T+|C-)\mathbb{P}(C-)} \\
&= \frac{(0.00008)(1-0.16)}{(0.00008)(1-0.16) + (1-0.00008)(0.10)} \\
&= \frac{(0.00008)(0.84)}{(0.00008)(0.84) + (0.99992)(0.10)} \\
&= \frac{672}{672 + 999920} \\
&= 0.000672.
\end{aligned}$$

One could also use a probability tree to help visualize this problem.

The physical interpretation of the result is as follows: there is a 0.067$\%$ chance of actually having cervical cancer given that the test is positive. The reason for the ineffectiveness is that the percentage of women having cervical cancer is extremely small while coupled with the relatively large error rates of the procedure; mainly 0.16 and 0.1.

5 Discussion 5

Random Variables and Discrete Random Variables.
Let us first formulate the definition of a random variable as follows:

Definition 25. Given a random experiment with a sample space S, a function X that assigns one and only one real number $X(s) = x$ for each element $s \in S$ is called a **random variable**. The **space** of X is the set of real numbers $\{x : X(s) = x, s \in S\}$.

Mathematically, we think of $X : S \to \mathbb{R}$.

Examples:

- Let X denote the random variable corresponding to the outcomes of a fair six-sided dice roll. Then, $S = [[1,6]]$, and $X(1) = 1$, $X(2) = 2$, $X(3) = 3, \cdots, X(6) = 6$.
- Let Y denote the random variable corresponding to the sum of outcomes of two fair six-sided dice rolls. Then, $S = [[1,6]]^2$, and $X((i,j)) = i + j$ for $i, j \in [[1,6]]$. For instance, $X((2,3)) = 5$ (corresponding to outcomes of 2 and 3 from the first and second dice roll respectively).

More often than not, a random variable is associated with a probability (as defined in Discussion Supplement 2), in the sense that for each event $A \subset S$, we can compute $\mathbb{P}(X \in A)$. In such a case, we say that the random variable X has a **distribution** associated with it. For instance, with Y given in the example above, if $A = \{1,3\}$, then $\mathbb{P}(T \in \{1,3\})$ refers to the probability that the output of the random variable is going to be 1 or 3. Since Y represents the sum of outcomes of two fair six-sided dices, $Y = 1$ is impossible. For $Y = 3$, this is only obtainable by $(1,2)$ or $(2,1)$ in the sample space, each happening with probability $\frac{1}{36}$. Consequently, we say that $\mathbb{P}(Y \in A) = \mathbb{P}(Y = 1, Y = 3) = \frac{1}{18}$, and we can then do so for any event $A \subset S = [[1,6]]^2$.

Furthermore, if S is finite or countably infinite, we say that X is a **discrete random variable**. The associated function f that assigns a probability for each $x \in \mathbb{R}$ is known as a **probability mass function**, or commonly denoted by pmf. We can think of $f(x) = \mathbb{P}(X = x)$ for $x \in \mathbb{R}$. Furthermore, since $f(x) = \mathbb{P}(X = x)$ is a probability, we can think of f as a function $f : \mathbb{R} \to [0,1]$, with the property that

$$\sum_{x \in \mathbb{R}} f(x) = 1.$$

Here, we note that since S is finite (or countably infinite), and X is a function, then $X : S \to \mathbb{R}$ says that $f(x) \neq 0$ for at most finitely (or countably infinitely) many $x \in \mathbb{R}$ and thus the above sum makes sense. In addition, denote the **range** of X to be the set $\{x \in \mathbb{R} : f(x) \neq 0\}$. From the above remark, this set must be finite (or countably infinite).

Last but not least, note that this is often defined as $p_X(x) := f(x) = \mathbb{P}(X = x)$.
Next, we shall introduce the definition of expectation below:

Definition 26. If p_X is the pmf of the discrete random variable X, then we have

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

where it is defined whenever the sum exists. In general, if g is a function that takes in a discrete random variable and outputs a discrete random variable, then we have

$$\mathbb{E}(g(X)) = \sum_x g(x) p_X(x).$$

Note that in the definition above, it is intuitively understood that we sum over all possible $x \in \mathbb{R}$, and it suffices to restrict ourselves to x such that $p_X(x) \neq 0$ (ie summing over the range of the distribution). The intuitive ways of thinking about expectation are that

“Expectation is the sum of possibility $\times$ probability” or

“Expectation is the weighted sum of possibility, weighted by its probability”.

Some properties of expectations include:

Proposition 27. Let a, b be constants, X, Y be random variables, and f, g be functions that take in a random variable and output a random variable. Then, we have

$$\mathbb{E}(af(X) + bg(Y)) = a\mathbb{E}(f(X)) + b\mathbb{E}(g(Y)).$$

In other words, expectation is linear.

Example 28. (Exercise 2.1.10, modified) A fair four-sided dice has two faces numbered 0 and two faces numbered 1. Another fair four-sided dice has its faces numbered $0, 1, 4,$ and 5. The two dices are rolled. Let X and Y be the respective outcomes of the roll. Let $W = X + Y$ be the corresponding random variable.

Determine the probability mass function of W and compute $\mathbb{E}(W)$. Show that $\mathbb{E}(W) = \mathbb{E}(X) + \mathbb{E}(Y)$, verifying that expectation is linear.

Suggested Solution: To do so, we consider the following table:

$W = X + Y$	$Y = 0$	$Y = 1$	$Y = 4$	$Y = 5$
$X = 0$	0	1	4	5
$X = 1$	1	2	5	6

Note that since each combination is equally likely, each element in the table above has a probability of $\frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$ chance of happening. Since 1 and 5 appear twice, we have $\mathbb{P}(W = 1) = \mathbb{P}(W = 5) = 2 \times \frac{1}{8} = \frac{1}{4}$. On the other hand, since $0, 2, 4,$ and 6 appear only once, we have $\mathbb{P}(W = 0) = \mathbb{P}(W = 2) = \mathbb{P}(W = 4) = \mathbb{P}(W = 6) = \frac{1}{8}$. Henceforth, the pmf is denoted by

$$p_W(w) = f(w) = \mathbb{P}(W = w) = \begin{cases} \frac{1}{4} & \text{if } w = 1, 5, \text{ and} \\ \frac{1}{8} & \text{if } w = 0, 2, 4, 6, \text{ and} \\ 0 & \text{otherwise .} \end{cases}$$

(Note that it is usually assumed to be zero outside of the inputs that you have specified, or you could include that explicitly if you want to.)
By definition, the expectation can be computed as follows:

$$\begin{aligned} \mathbb{E}(W) &= \sum_w w p_W(w) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{1}{4} + 2 \times \frac{1}{8} + 4 \times \frac{1}{8} + 5 \times \frac{1}{4} + 6 \times \frac{1}{8} \\ &= \boxed{3}. \end{aligned}$$

On the other hand, one can repeat a similar computation with the definition of expectation to show that

$$\begin{aligned} \mathbb{E}(X) &= 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2} \\ \mathbb{E}(Y) &= 0 \times \frac{1}{4} + 1 \times \frac{1}{4} + 4 \times \frac{1}{4} + 5 \times \frac{1}{4} = \frac{5}{2}. \end{aligned}$$

Hence, we have

$$\mathbb{E}(W) = 3 = \frac{1}{2} + \frac{5}{2} = \mathbb{E}(X) + \mathbb{E}(Y).$$

This is expected since the average outcomes of the sum of two dices should be the same as the sum of the average outcomes of the individual dices.

Example 29. Consider an urn containing a red ball and n indistinguishable blue balls, where n is a positive integer ($n \geq 1$). Let X be the random variable denoting the number of blue balls obtained if we take 2 balls out of the urn.

(i) Write down the probability mass function of X as a function of n.

(ii) Compute $\mathbb{E}(X^2)$.

Suggested Solutions:

(i) If $n = 1$, we have 1 red and 1 blue ball in the urn. If we were to pick 2 balls out of it, then the only possible outcome is picking the only red and blue balls. Hence, we have $p_X(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$

For an arbitrary $n > 1$, we have 1 red balls and n blue balls. Note that X can only be either $0, 1$, or 2, since we are only picking out two balls from the urn. In fact, $X \neq 0$ since we only have a red ball, and $X = 0$ implies that both balls that we have picked are red, which is impossible. Consequently, we see that $\mathbb{P}(X = 1) = \frac{1}{n+1} \times \frac{n}{n} + \frac{n}{n+1} \times \frac{1}{n} = \frac{2}{n+1}$ (corresponding to picking red then blue, or blue then red). Similarly (or using the complement method), we thus have $\mathbb{P}(X = 2) = \frac{n-1}{n+1}$. Thus, the probability mass function is given by

$$p_X(x) = \begin{cases} \frac{2}{n+1} & \text{if } x = 1, \\ \frac{n-1}{n+1} & \text{if } x = 2 \end{cases}$$

for $n \geq 2$.

(ii) By definition, we thus have

$$\mathbb{E}(X^2) = \sum_x x^2 p_X(x) = 1^2 \left(\frac{2}{n+1}\right) + 2^2 \left(\frac{n-1}{n+1}\right) = \boxed{\frac{4n-2}{n+1}}.$$

Common Types of Discrete Random Variables.

Below, we summarize the common discrete random variables covered in class up till now.

Distribution	Notation	Parameters	Range	pmf: $p_X(x)$	$\mathbb{E}(X)$	Equivalent Model
Bernoulli	Be(p)	$p \in [0,1]$	$\{0,1\}$	$\begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$	p	1 Coin Flip, Heads w.p p.
Binomial	B(n,p)	$n \in \mathbb{N}$ $p \in [0,1]$	$[[0,n]]$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	n Coin Flips, Heads w.p p.
Geometric	Geom(p)	$p \in (0,1]$	$\mathbb{N} = \{1,2,\cdots\}$	$p(1-p)^{x-1}$	$1/p$	#Coin Flips till first Head, Heads w.p p.
Poisson	Poisson(λ)	$\lambda \geq 0$	$\mathbb{Z}_{\geq 0} = \{0,1,\cdots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	Independent calls in a day, with average of λ calls.

Some algebra of discrete random variables:

- If $X_1, X_2, \cdots, X_n \sim$ Be(p) and independent, then $X_1 + X_2 + \cdots + X_n \sim$ B(n,p).

  "n independent coin tosses are made up of n individual independent coin tosses".

- If $X_1 \sim$ Poisson(λ_1) and $X_2 \sim$ Poisson(λ_2) and are independent, then $X_1 + X_2 \sim$ Poisson$(\lambda_1 + \lambda_2)$.

  "If I receive an average of λ_1 calls in a day, and an average λ_2 calls in another day, then I would receive an average of $\lambda_1 + \lambda_2$ calls in two days."

Example 30. (Exercise 2.4.13.) It is claimed that for a particular lottery, $\frac{1}{10}$ of the 50 million tickets will win a prize. What is the probability of winning at least one prize if you purchase ten tickets?

Suggested Solution: A large number of lottery tickets suggests that it is reasonable to assume that the probability of winning a ticket is $\frac{1}{10}$ and is approximately independent (ie if you have observed that you won a prize from one of the tickets, the probability of winning a prize in the subsequent ticket is still approximately $\frac{1}{10}$).

In that case, let X be the random variable denoting the number of winning tickets out of 10. Then, we must have $X \sim \text{B}(10, 0.1)$ (approximately).

The required probability is $\mathbb{P}(X \geq 1)$. This is then equivalent to

$$\begin{aligned} \mathbb{P}(X \geq 1) &= 1 - \mathbb{P}(X = 0) \\ &= 1 - \binom{10}{0} 0.1^0 (1 - 0.1)^{10} \\ &= 1 - 0.9^{10} \approx 0.651. \end{aligned}$$

Example 31. (Example 2.7.4.) In a large city, telephone calls to 911 come on average of two every 3 minutes. Compute an approximation to the probability of five or more calls arriving in a 9-minute period.

Suggested Solution: Let X be the random variable denoting the number of calls arriving in a 9-minute interval period. If we can write $X = X_1 + X_2 + X_3$ where each X_i are random variables representing the number of calls arriving in a 3-minute interval with $X_i \sim \text{Poisson}(2)$, then $X \sim \text{Poisson}(6)$. In other words, the average number of calls in a 9-minute interval is 6.

Consequently, the required probability is $\mathbb{P}(X \geq 5)$. We compute this via

$$\begin{aligned} \mathbb{P}(X \geq 5) &= 1 - \mathbb{P}(X \leq 4) \\ &= 1 - \sum_{x=0}^{4} \frac{e^{-6} x^6}{x!} \\ &\approx 0.715. \end{aligned}$$

Note that in exams, unless you are able to do an infinite sum on your calculator (I don't even know if calculators are allowed), then one should compute the right-tail probabilities (ie $\mathbb{P}(X \geq \text{something})$) using the complement method. This holds for any distribution with a countably infinite range.

Example 32. If X follows a Poisson distribution and $1.5\,\mathbb{P}(X=1)=\mathbb{P}(X=3)$, compute $\mathbb{P}(X=4)$.

Suggested Solution: Let $X \sim \text{Poisson}(\lambda)$ for some $\lambda > 0$ to be determined. Using the given information and the pmf of X, we have

$$\begin{aligned}
1.5\,\mathbb{P}(X=1) &= \mathbb{P}(X=3) \\
1.5\frac{e^{-\lambda}\lambda^1}{1!} &= \frac{e^{-\lambda}\lambda^3}{3!} \\
1.5\lambda &= \frac{\lambda^3}{6} \\
\lambda^3 - 9\lambda &= 0 \\
\lambda(\lambda^2-9) &= 0 \\
\lambda(\lambda-3)(\lambda+3) &= 0
\end{aligned}$$

This implies that λ is either $0, -3$, or 3. However, since $\lambda > 0$, then $\lambda = 3$ necessarily. Henceforth, we plug this into the pmf of X at 4 to obtain

$$\mathbb{P}(X=4) = \frac{e^{-\lambda}\lambda^4}{4!} = \frac{e^{-3}3^4}{4!} = \frac{27}{8}e^{-4}.$$

Example 33. Let X be a random variable representing the number of heads out of 5 independent coin flips, in which the probability of getting a head on each of the flips is p. Find the value of $p \in (0,1)$ that maximizes the probability of obtaining exactly two heads.

Suggested Solution: Note that the required probability is given by

$$\mathbb{P}(X=2) = \binom{5}{2} p^2(1-p)^3 = 10p^2(1-p)^3.$$

Let $g(p) = 10p^2(1-p)^3$ for $p \in (0,1)$. Maximizing $g(p)$ reduces to a Calculus problem. We first compute

$$g'(p) = 20p(1-p)^3 - 30p^2(1-p)^2.$$

To obtain critical points, we solve

$$\begin{aligned}
20p(1-p)^3 - 30p^2(1-p)^2 &= 0 \\
20p(1-p)^3 &= 30p^2(1-p)^2 \\
20(1-p) &= 30p \\
50p &= 20 \\
p &= \boxed{\frac{2}{5}}.
\end{aligned}$$

Hence, $p = 0.4$ is the only critical point on $(0,1)$.
To check that this value of p indeed maximizes the probability, we compute

$$g''(p) = -20(-1 + 9p - 18p^2 + 10p^3)$$

and see that

$$g''(0.4) = -7.2 < 0,$$

indicating that $p = 0.4$ is a maximum point for g.

Remark: If we are optimizing over the number of coin flips rather than the probability, we are doing so over discrete values. The strategy to do so is to find the range of values of n such that $\mathbb{P}(2 \text{ heads out of } n \text{ flips}) \geq \mathbb{P}(2 \text{ heads out of } n+1 \text{ flips})$ and similarly for $\mathbb{P}(2 \text{ heads out of } n \text{ flips}) \geq \mathbb{P}(2 \text{ heads out of } n-1 \text{ flips})$. An example of this appears in Homework 4, Question 4.

6 Discussion 6

Common Types of Discrete Random Variables.

Below, we summarize the common discrete random variables covered in class up till now.

Distribution	Notation	Parameters	Range	pmf: $p_X(x)$	$\mathbb{E}(X)$	Equivalent Model
Bernoulli	Be(p)	$p \in (0,1]$	$\{0,1\}$	$\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$	p	1 Coin Flip, Heads w.p p.
Binomial	B(n,p)	$n \in \mathbb{N}$ $p \in [0,1]$	$[[0,n]]$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	n Coin Flips, Heads w.p p.
Geometric	Geom(p)	$p \in (0,1]$	$\mathbb{N} = \{1,2,\cdots\}$	$p(1-p)^{x-1}$	$1/p$	#Coin Flips till first Head, Heads w.p p.
Poisson	Poisson(λ)	$\lambda \geq 0$	$\mathbb{Z}_{\geq 0} = \{0,1,\cdots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	Independent calls in a day, with average of λ calls.

Some algebra of discrete random variables:

- If $X_1, X_2, \cdots, X_n \sim$ Be(p) and independent, then $X_1 + X_2 + \cdots + X_n \sim$ B(n,p).

  "n independent coin tosses are made up of n individual independent coin tosses".

- If $X_1 \sim$ Poisson(λ_1) and $X_2 \sim$ Poisson(λ_2) and are independent, then $X_1 + X_2 \sim$ Poisson$(\lambda_1 + \lambda_2)$.

  "If I receive an average of λ_1 calls in a day, and an average λ_2 calls in another day, then I would receive an average of $\lambda_1 + \lambda_2$ calls in two days."

Note that $\mathbb{E}(X)$ can be derived by using the formula

$$\mathbb{E}(X) = \sum_x x p_X(x).$$

Another useful fact is as follows. If $X \sim$ Geom(p), then for a natural number x, we have $\mathbb{P}(X > x) = (1-p)^x$ since obtaining the first success anytime after the x-th attempt is equivalent to failing the first x attempts. We will prove the equivalent in Example 34.

Refer to either the lecture notes or the textbook for the derivation of expectations of the various discrete random variables above. However, we shall include some useful formulas below:

- Geometric Series. If we have $a + ar + ar^2 + \cdots + ar^n$, then the sum is given by $a\left(\frac{r^{n+1}-1}{r-1}\right)$ for $r \neq 1$. Mathematically, for $r \neq 1$, if we cancel the common factor of a on both sides, we have

  $$\sum_{x=0}^{n} r^x = \left(\frac{1-r^{n+1}}{1-r}\right).$$

  If $r \in (0,1)$. we can take the limit as $n \to \infty$ on both sides and noting that $\lim_{n\to\infty} r^{n+1} = 0$, we thus obtain

  $$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r}$$

  This is used to show that the pmf of a geometric distribution sums to 1.

- Binomial Formula.

  $$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}.$$

  This is used to show that the pmf of a binomial distribution sums to 1.

- Taylor Series for Exponential Functions.
$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x.$$
This is used to show that the pmf of a Poisson distribution sums to 1.

- Arithmetico-Geometric Series (used to evaluate expectation of geometric-like distributions). For $r \in (0, 1)$, we had
$$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r}.$$
Differentiating with respect to r on both sides (while noting that for the $x = 0$ term in the infinite sum, it is a constant 1, and thus taking derivative makes it disappear, and hence the sum now starts from 1), we have
$$\sum_{x=1}^{\infty} x r^{x-1} = \frac{1}{(1-r)^2}.$$
For instance, we have that $1 + 2r + 3r^2 + 4r^3 + \cdots = \frac{1}{(1-r)^2}$.

Example 34. (Exercise 2.6.1, Modified.) An excellent free-throw shooter attempts several free throws until she misses. Let $p \in (0,1)$ be her probability of missing a free throw. The game stops if she misses a free throw.

(i) What is the probability of having the first miss on the 13th attempt?

(ii) What is the probability of having the first miss on the 13th attempt or later?

(iii) What is the expected number of misses that she will make?

Leave your answers in terms of p.

Suggested Solution: Let X be the random variable denoting that the Xth attempt is the first miss committed by the excellent free-throw shooter. From the above information, we know that $X \sim$ Geom(p).

(i) The required probability is $\mathbb{P}(X = 13) = (1-p)^{12}p$.

(ii) The required probability is $\mathbb{P}(X \geq 13) = 1 - \mathbb{P}(X < 13) = 1 - \mathbb{P}(X \leq 12) = 1 - \sum_{x=1}^{12}(1-p)^{x-1}p$. Alternatively, this constitutes having to not miss in the first twelve attempts with probability $1-p$. Equivalently, we thus have that the required probability is $(1-p)^{12}$.

We shall prove that the two expressions are equivalent. This requires the use of a geometric series formula as follows. $\sum_{x=0}^{n} r^x = \frac{1-r^{n+1}}{1-r}$. Observe that

$$\sum_{x=1}^{12}(1-p)^{x-1} = \sum_{x=0}^{11}(1-p)^x = \frac{1-(1-p)^{12}}{1-(1-p)} = \frac{1-(1-p)^{12}}{p}$$

Hence, we have

$$p\sum_{x=1}^{12}(1-p)^{x-1} = 1-(1-p)^{12}$$

and thus

$$1 - p\sum_{x=1}^{12}(1-p)^{x-1} = (1-p)^{12},$$

proving that they are indeed equivalent.

(iii) Since $X \sim$ Geom(p), we have that $\mathbb{E}(X) = \frac{1}{p}$.

Example 35. (BOGO Sort.) You randomly reorder $1, 2, \cdots, n$ for $n \geq 2$ repeatedly until you get the original reordering back. What is the expected number of times you have to reorder the numbers until this happens?

Remark: This is related to the average time complexity of BOGO sort.

Let X denote the random variable corresponding to getting the order of $1, 2, \cdots, n$ right on the Xth attempt. The probability of getting the order right on each attempt is given by $\frac{1}{n!}$ (since the numerator corresponds to one element in the sample space corresponding to permutations of the numbers $1, 2, \cdots, n$, while the denominator corresponds to the number of elements in this sample space). Henceforth, we have $X \sim \text{Geom}\left(\frac{1}{n!}\right)$. Consequently, we have

$$\mathbb{E}(X) = \frac{1}{\frac{1}{n!}} = n!.$$

Remark. If you are familiar with algorithms in computer science, you will know that the corresponding average time complexity is given by $\mathcal{O}(n \times n!)$, where we have to perform $n!$ random arrangements on average to get the order right, with each re-arrangement taking $\mathcal{O}(n)$ time to do (and $\mathcal{O}(n)$ time to check if the given arrangement is right).

Example 36. (Memory-less Property of Geometric Distribution.) Let X be a geometric distribution. Show that

$$\mathbb{P}(X > k + j | X > k) = \mathbb{P}(X > j)$$

where k, j are non-negative integers.

Remark: This is known as the memory-less property of geometric distribution. In other words, knowing that you have failed for the k times in a geometric distribution and wanting to know the probability of failing for $k + j$ times is the same as wanting to know the probability of failing just j times since the distribution "forgets" that you have failed k times and "resets" the counter from 0 if such a piece of information is given.

Suggested Solution: By the definition of conditional probability, we have

$$\begin{aligned}\mathbb{P}(X > k + j | X > k) &= \frac{\mathbb{P}(X > k + j \cap X > k)}{\mathbb{P}(X > k)} \\ &= \frac{\mathbb{P}(X > k + j)}{\mathbb{P}(X > k)} \\ &= \frac{(1 - p)^{k+j}}{(1 - p)^k} \\ &= (1 - p)^j \\ &= \mathbb{P}(X > j).\end{aligned}$$

Recall that we have argued similarly in Example 34, in which $\mathbb{P}(X > x) = (1 - p)^x$ since obtaining the first success anytime after the x-th attempt is equivalent to failing the first x attempts.

Example 37. A hospital receives 70% of its flu vaccine from Company Good and the remainder from Company Evil. Each shipment contains a large vial of vaccine. From Company Good, 10% of the vials are ineffective. From Company Evil, 80% are ineffective. The hospital tests 10 randomly selected vials from one shipment for their effectiveness.

Compute the probability that this shipment comes from Company Good, given that exactly two of these 10 vials are effective.

Note that given that the shipment comes from Company Good (denoted by G), we model the corresponding random variable of $X|G \sim \mathrm{B}(10, 0.9)$ corresponding to the number of vials of vaccines out of 10 that is effective. Similarly, for the shipment coming from Company Evil (denoted by E), we model the corresponding random variable of $X|E \sim \mathrm{B}(10, 0.2)$ corresponding to the number of vials of vaccines out of 10 that is effective. Rigorously, we denote

$$\begin{aligned} G &= \{\text{Shipment came from Company Good}\}, \\ E &= \{\text{Shipment came from Company Evil}\}, \\ A &= \{\text{Exactly two vials are effective}\}, \end{aligned}$$

By Bayes's Theorem, we have

$$\mathbb{P}(G|A) = \frac{\mathbb{P}(G \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|G)\,\mathbb{P}(G)}{\mathbb{P}(A|G)\,\mathbb{P}(G) + \mathbb{P}(A|E)\,\mathbb{P}(E)} = \frac{\mathbb{P}(X|G=2)\,\mathbb{P}(G)}{\mathbb{P}(X=2|G)\,\mathbb{P}(G) + \mathbb{P}(X|E=2)\,\mathbb{P}(E)}.$$

Since the hospital receives 70% of its flu vaccine from Company Good, then we have that $\mathbb{P}(G) = 0.7$. Henceforth, we have $\mathbb{P}(E) = 0.3$.

To calculate $\mathbb{P}(X|G=2)$, we know that $X|G$ follows $\mathrm{B}(10, 0.9)$ if we knew that the vials came from G. Hence,

$$\mathbb{P}(X|G=2) = \binom{10}{2}(0.9)^2(0.1)^8.$$

To calculate $\mathbb{P}(X|E=2)$, we know that X follows $\mathrm{B}(10, 0.2)$ if we knew that the vials came from E. Hence,

$$\mathbb{P}(X|E=2) = \binom{10}{2}(0.2)^2(0.8)^8.$$

Plugging this into the conditional probability expression above, we have

$$\begin{aligned} \text{“}\mathbb{P}(G|X=2)\text{”} = \mathbb{P}(G|A) &= \frac{\mathbb{P}(X|G=2)\,\mathbb{P}(G)}{\mathbb{P}(X|G=2)\,\mathbb{P}(G) + \mathbb{P}(X|E=2)\,\mathbb{P}(E)} \\ &= \frac{\binom{10}{2}(0.9)^2(0.1)^8 \times 0.7}{\binom{10}{2}(0.9)^2(0.1)^8 \times 0.7 + \binom{10}{2}(0.2)^2(0.8)^8 \times 0.3} \end{aligned}$$

Example 38. (Exercise 2.3.16.) Let X equal the number of flips of a fair coin that are required to observe the same face on two consecutive flips.

(i) Compute the pmf $p_X(x)$ of X.

(ii) Show that this is indeed a pmf, that is, $\sum_x p_X(x) = 1$.

(iii) Compute $\mathbb{E}(X)$.

Suggested Solutions:

(i) It helps to work with a small number of flips first. For instance, if $X = 2$, this implies that we have observed the same face in two consecutive flips. This then implies that we must have HH or TT, out of four possible combinations obtained from two flips. This implies that $\mathbb{P}(X = 2) = \frac{2}{4} = \frac{1}{2}$.

For $X = 3$, this implies that you only observe the two consecutive flips in the last two flips (else the value of X would be different). The only possibility is given by _HH or _TT (since we did not specify what the two consecutive flips should be; so it could be H or T). Furthermore, if this must be obtained in the last two flips, the blanks must have "alternating" faces; ie we must have THH or HTT. This is out of 8 possible combinations. Thus, we have $\mathbb{P}(X = 3) = \frac{2}{8} = \frac{1}{4}$.

Similarly, for $X = 4$, we must have __HH or __TT. The blanks must have alternatingly different faces, and consequently must be HTHH or THTT (ie we had THHH, then $X = 3$ instead since we observe two heads at the third flip). Using a similar idea, we thus have $\mathbb{P}(X = 4) = \frac{2}{2^4} = \frac{1}{2^3}$.
Correspondingly, we can then deduce that

$$p_X(x) = \frac{2}{2^x} = \frac{1}{2^{x-1}} \quad \text{for } x = 2, 3, 4 \cdots .$$

(ii) The question is equivalent to proving that

$$\sum_{x=2}^{\infty} \frac{1}{2^{x-1}} = 1.$$

First, recall from the geometric formula that

$$\sum_{n=1}^{N} r^n = \frac{1 - r^{N+1}}{1 - r}.$$

If $r \in (0, 1)$, then we can take limits as $N \to \infty$ on both sides to obtain

$$\sum_{n=0}^{\infty} r^n = \lim_{N\to\infty} \sum_{n=0}^{N} r^n = \frac{1 - \lim_{N\to\infty} r^{N+1}}{1 - r} = \frac{1}{1 - r}.$$

Substitute $r = \frac{1}{2}$, we obtain

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{1 - \frac{1}{2}} = 2.$$

Replacing n by $n - 1$, we thus have

$$\sum_{n=1}^{\infty} \frac{1}{2^{n-1}} = 2$$

Taking out the $n = 1$ term, we have

$$1 + \sum_{n=2}^{\infty} \frac{1}{2^{n-1}} = \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} = 2$$

$$\sum_{n=2}^{\infty} \frac{1}{2^{n-1}} = 1.$$

This is exactly what we need to prove, but replacing n with x.

(iii) By the formula of expectation, we have

$$\mathbb{E}(X) = \sum_{x=2}^{\infty} x p_X(x) = \sum_{x=2}^{\infty} \frac{x}{2^{x-1}}.$$

This sum is not easy to evaluate, but we shall mimic the method that the professor used in lectures. Using the arithmetico-geometric series formula, we have

$$\sum_{x=1}^{\infty} x r^{x-1} = \frac{1}{(1-r)^2}.$$

Substitute $r = \frac{1}{2}$ above, we then have

$$\sum_{x=1}^{\infty} \frac{x}{2^{x-1}} = \frac{1}{\left(1-\frac{1}{2}\right)^2} = 4.$$

Since the required sum starts from $x = 2$, we take it out the $x = 1$ term from the series to obtain

$$\begin{aligned} \frac{1}{2^{1-1}} + \sum_{x=2}^{\infty} \frac{x}{2^{x-1}} = \sum_{x=1}^{\infty} \frac{x}{2^{x-1}} &= 4 \\ \sum_{x=2}^{\infty} \frac{x}{2^{x-1}} &= 4 - 1 = \boxed{3}. \end{aligned}$$

7 Discussion 7

Mathematical Properties of Expectation and Variance.

Recall that if X is a discrete random variable with pmf $p_X(x)$, then we have

$$\mathbb{E}(X) = \sum_x x p_X(x).$$

Here are some properties related to expectations:

- For any two random variables X, Y, we have

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- For any random variable X and real number $\alpha \in \mathbb{R}$, we have

$$\mathbb{E}(\alpha X) = \alpha\, \mathbb{E}(X).$$

- For any real number c (ie constant), we have

$$\mathbb{E}(c) = c.$$

- If f is a function that takes in a random variable and outputs a random variable, then

$$\mathbb{E}(f(X)) = \sum_x f(x) p_X(x).$$

  In particular,

$$\mathbb{E}(X^2) = \sum_x x^2 p_X(x).$$

Note that if we are instead interested in the pmf of $Y = f(X)$, one has to first determine the support of $Y = f(X)$ and compute that probability of $Y = y$ at each of these values. Refer to Example 39 for an example of this.

Next, we define the **variance** as follows. For a random variable X, we denote the **variance** of X as

$$\text{Var}(X) = \mathbb{E}(X-\mu)^2, \quad \text{where } \mu = \mathbb{E}(X).$$

One important property (that you will derive in HW 6 Question 3) is that

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mu^2 \tag{2}$$

Some useful properties of variance include:

- For any real number α,

$$\text{Var}(\alpha X) = \alpha^2 \text{Var}(X).$$

- For any real number c,

$$\text{Var}(X+c) = \text{Var}(X).$$

We shall prove one of these in Example 40 below.

Some algebra of discrete random variables:

- If $X_1, X_2, \cdots, X_n \sim \text{Be}(p)$ and independent, then $X_1 + X_2 + \cdots + X_n \sim \text{B}(n, p)$.

  "n independent coin tosses are made up of n individual independent coin tosses".

- If $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ and are independent, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

  "If I receive an average of λ_1 calls in a day, and an average λ_2 calls in another day, then I would receive an average of $\lambda_1 + \lambda_2$ calls in two days."

Refer to either the lecture notes or the textbook for the derivation of expectations of the various discrete random variables above. However, we shall include some useful formulas below:

- Geometric Series. If we have $a + ar + ar^2 + \cdots + ar^n$, then the sum is given by $a\left(\frac{r^{n+1}-1}{r-1}\right)$ for $r \neq 1$. Mathematically, for $r \neq 1$, if we cancel the common factor of a on both sides, we have

  $$\sum_{x=0}^{n} r^x = \left(\frac{1 - r^{n+1}}{1 - r}\right).$$

  If $r \in (0, 1)$. we can take the limit as $n \to \infty$ on both sides and noting that $\lim_{n\to\infty} r^{n+1} = 0$, we thus obtain

  $$\sum_{x=0}^{\infty} r^x = \frac{1}{1 - r}$$

  This is used to show that the pmf of a geometric distribution sums to 1.

- Binomial Formula.

  $$(a + b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}.$$

  This is used to show that the pmf of a binomial distribution sums to 1.

- Taylor Series for Exponential Functions.

  $$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x.$$

  This is used to show that the pmf of a Poisson distribution sums to 1.

- Arithmetico-Geometric Series (used to evaluate expectation of geometric-like distributions). For $r \in (0, 1)$, we had

  $$\sum_{x=0}^{\infty} r^x = \frac{1}{1 - r}.$$

  Differentiating with respect to r on both sides (while noting that for the $x = 0$ term in the infinite sum, it is a constant 1, and thus taking derivative makes it disappear, and hence the sum now starts from 1), we have

  $$\sum_{x=1}^{\infty} x r^{x-1} = \frac{1}{(1 - r)^2}.$$

  For instance, we have that $1 + 2r + 3r^2 + 4r^3 + \cdots = \frac{1}{(1-r)^2}$.

Some examples of these were included in Discussion Supplement 6, which you might find to be useful in HW 6 for this week.

Example 39. Let X be the random variable denoting the outcome of a fair $6-$sided dice.

(i) Compute $\mathbb{E}(X)$ directly by definition.

(ii) Compute $\mathbb{E}(X^2)$ directly by definition.

(iii) Using your answer in (i) and (ii), compute $\mathbb{E}((X-3.5)^2)$.

(iv) Let $Y=(X-3.5)^2$. Compute the pmf of Y, ie $p_Y(y)$.

(v) Use (iv) to compute $\mathbb{E}(Y)$ (which is basically $\mathbb{E}((X-3.5)^2)$) and show that this agrees with (iii).

(vi) What is the value of Var(X)?

Suggested Solution: Recall that the pmf of X is given by

$$p_X(x)=\begin{cases}\frac{1}{6} & \text{if } x=1,2,3,4,5,6\\ 0 & \text{otherwise}\end{cases}.$$

(i) $\mathbb{E}(X)=\sum_x xp_X(x)=\sum_{x=1}^6 x\left(\frac{1}{6}\right)=\frac{1}{6}\sum_{x=1}^6 x=\frac{1}{6}(1+2+3+4+5+6)=3.5.$

(ii) $\mathbb{E}(X^2)=\sum_x x^2p_X(x)=\sum_{x=1}^6 x^2\left(\frac{1}{6}\right)=\frac{1}{6}\sum_{x=1}^6 x^2=\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)=\frac{91}{6}.$

(iii) Note that we have

$$\begin{aligned}\mathbb{E}((X-3.5)^2)&=\mathbb{E}(X^2-7X+3.5^2)\\ &=\mathbb{E}(X^2)-7\,\mathbb{E}(X)+3.5^2\\ &=\frac{91}{6}-7\times 3.5+3.5^2\\ &=\frac{35}{12}.\end{aligned}$$

(iv) First, observe that if $x=1$, implies $y=2.5^2$; $x=2$ implies $y=1.5^2$; $x=3$ implies $y=0.5^2$, $x=4$ implies $y=0.5^2$, $x=5$ implies $y=1.5^2$ and $x=6$ implies $y=2.5^2$.
Thus, the support of $(X-3.5)^2$ is only contains 3 elements, $0.5^2,1.5^2$, and 2.5^2, with contribution from $x=3,4$, $x=2,5$, and $x=1,6$ respectively, each with probability $\frac{1}{6}$ for each of the 6 faces of the dice. Consequently, we have

$$p_Y(y)=\begin{cases}\frac{1}{3} & \text{if } x=0.5^2,1.5^2,2.5^2\\ 0 & \text{otherwise}\end{cases}.$$

(v) By definition,

$$\begin{aligned}\mathbb{E}(Y)&=\sum_y yp_Y(y)\\ &=\frac{1}{3}(0.5^2+1.5^2+2.5^2)\\ &=\frac{35}{12}.\end{aligned}$$

(vi) If you have not already realized this, since $\mathbb{E}(X)=3.5$, then
Var$(X)=\mathbb{E}((X-\mu)^2)=\mathbb{E}((X-\mathbb{E}(X))^2)=\mathbb{E}((X-3.5)^2)=\frac{35}{12}$.

Example 40. Let X be a random variable and c be a real number. Prove that

$$\text{Var}(X+c) = \text{Var}(X).$$

Suggested Solution: Using property (2) for variance, we have

$$\begin{aligned}
\text{Var}(X+c) &= \mathbb{E}((X+c)^2) - (\mathbb{E}(X+c))^2 \\
&= \mathbb{E}(X^2 + 2cX + c^2) - (\mathbb{E}(X) + c)^2 \\
&= \mathbb{E}(X^2) + \mathbb{E}(2cX) + c^2 - (\mathbb{E}(X))^2 - 2c\,\mathbb{E}(X) - c^2 \\
&= \mathbb{E}(X^2) + 2c\,\mathbb{E}(X) + c^2 - (\mathbb{E}(X))^2 - 2c\,\mathbb{E}(X) - c^2 \\
&= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= \text{Var}(X).
\end{aligned}$$

Here, we have used all the various properties of expectation and applied property (2) in the last step.

Conditional Expectations.

Consider A as an event (of some sample space S) and X as a discrete random variable. In some sense, we can think of $X|A$ as a new random variable updated by the fact that event A has occurred. The conditional distribution has a pmf denoted by $p_{X|A}(x)$. Consequently, we define the conditional expectation as

$$\mathbb{E}(X|A) = \sum_x x p_{X|A}(x).$$

To compute the full expectation, we can use the **Total Expectation Theorem** that goes as follows. Let A_k be a partition of the sample space S. Then, we have

$$\mathbb{E}(X) = \sum_k \mathbb{E}(X|A_k)\,\mathbb{P}(A_k).$$

Note that one can draw parallel with the law of total probability, which is given by

$$\mathbb{P}(B) = \sum_k \mathbb{P}(B|A_k)\,\mathbb{P}(A_k)$$

for any event B (of a possibly different sample space). We shall look at a couple of examples on conditional expectations and using the total expectation theorem below.

Example 41. A hospital receives 70% of its flu vaccine from Company Good and the remainder from Company Evil. Each shipment contains a large vial of vaccine. From Company Good, 10% of the vials are ineffective. From Company Evil, 80% are ineffective. The hospital tests 10 randomly selected vials from one shipment for their effectiveness.

(i) Compute the expected number of vials that are effective, given that it comes from Company Good.

(ii) Compute the expected number of vials that are effective.

Suggested Solution: Rigorously, we denote

$$\begin{aligned} G &= \{\text{Shipment came from Company Good}\}, \\ E &= \{\text{Shipment came from Company Evil}\}, \end{aligned}$$

and X be the number of effective vials.

(i) Given G, we know that $X|G \sim \mathrm{B}(10, 0.9)$ since we are testing 10 vials from one shipment, and assume that the probability that each vial is effective is 0.9 (since 10% of the vials are ineffective). Thus, using the fact that for $Y \sim \mathrm{B}(n,p)$ has $\mathbb{E}(Y) = np$, we have

$$\mathbb{E}(X|G) = 10 \times 0.9 = 9.$$

(ii) Similarly, we know that $X|E \sim \mathrm{B}(10, 0.2)$, and thus $\mathbb{E}(X|E) = 10 \times 0.2 = 2$. Using the total expectation theorem and using the fact that $\mathbb{P}(G) = 0.7$ and $\mathbb{P}(E) = 0.3$ (as concluded from the context of the question), we have

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X|G)\,\mathbb{P}(G) + \mathbb{E}(X|E)\,\mathbb{P}(E) \\ &= 9 \times 0.7 + 2 \times 0.3 \\ &= \boxed{6.9}. \end{aligned}$$

Example 42. First, flip a total of 10 fair coins, and denote the number of heads to be X. Then, we roll a $X-$sided fair dice and record the result as Y. Compute $\mathbb{E}(Y)$.

Hint: Recall that for an arithmetic series, we have $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

Suggested Solution:

Note that this is a two-step process, in the sense that Y depends on X. Thus, it makes sense to first compute $\mathbb{E}(Y|X = x)$. (We can think of the event being conditioned on is the event that $X = x$ for a given x.) Indeed, if $X = x$, this implies that we are flipping a x-sided dice and asking for its expectation. Henceforth, each side has an equal probability of $1/x$ happening. By the definition of expectation, we have

$$\begin{aligned}
\mathbb{E}(Y|X = x) &= \sum_{k} k \times p_{Y|X=x}(k) \\
&= \sum_{k=1}^{x} k\left(\frac{1}{x}\right) \\
&= \frac{1}{x}\sum_{k=1}^{x} k \\
&= \frac{1}{x}\frac{x(x+1)}{2} \\
&= \frac{x+1}{2}.
\end{aligned}$$

Thus, note that for $X = x$, we can only take $x = 0, \cdots, 10$ (corresponding to the number of possible heads from flips of 10 fair coins) and hence these events form a partition. Furthermore, the probability that the event $X = x$ occurs can be computed by using the fact that $X \sim \text{B}(10, 0.5)$.

Consequently, we have by the total expectation theorem,

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{x=0}^{10} \mathbb{E}(Y|X = x)\,\mathbb{P}(X = x) \\
&= \sum_{x=0}^{10} \frac{x+1}{2}\binom{10}{x}\left(\frac{1}{2}\right)^{10} \\
&= \frac{1}{2}\sum_{x=0}^{10}\binom{10}{x}\left(\frac{1}{2}\right)^{10} + \frac{1}{2}\sum_{x=0}^{10} x\binom{10}{x}\left(\frac{1}{2}\right)^{10} \\
&= \frac{1}{2} + \frac{1}{2}\left(10 \times \frac{1}{2}\right) \\
&= \boxed{3}.
\end{aligned}$$

Here, we have used the fact that (recall $X \sim \text{B}(10, 0.5)$)

- $\sum_{x=0}^{10}\binom{10}{x}\left(\frac{1}{2}\right)^{10}$ is trying to sum over the pmf of X, which we know must give 1, and
- $\sum_{x=0}^{10} x\binom{10}{x}\left(\frac{1}{2}\right)^{10}$ is trying to compute $\mathbb{E}(X)$, in which we know must give $n \times p = 10 \times \frac{1}{2}$.

8 Discussion 8

Two Random Variables and their Properties.

We begin with a formal definition:

Definition 43. Let X and Y be two discrete random variables defined on a discrete sample space. Let S be the corresponding two-dimensional sample space of X and Y. The probability that $X = x$ and $Y = y$ is denoted by $f(x, y) = p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$. Here, $p_{X,Y}(\cdot)$ is known as the **joint probability mass function** (joint pmf) of X and Y. Furthermore, it satisfies the following properties:

(i) $0 \leq p_{X,Y}(x, y) \leq 1$,

(ii) $\sum_{x,y} p_{X,Y}(x, y) = 1$,

(iii) $\mathbb{P}((X, Y) \in A) = \sum_{(x,y)\in A} p_{X,Y}(x, y)$, where A is a subset of the sample space S.

In particular, for property (iii), an example would be $\mathbb{P}(X = 1, Y = 2 \text{ or } X = 2, Y = 1) = p_{X,Y}(1, 2)+p_{X,Y}(2, 1)$. Some additional properties include:

(i) Starting from the joint pmf, we can obtain the individual pmf by **marginalizing** over the other random variable. In other words, $p_X(x) = \sum_y p_{X,Y}(x, y)$, obtained at each x by summing over all y, is also known as the marginal pmf of X.
Similarly, the marginal pmf of Y is given by $p_Y(y) = \sum_x p_{X,Y}(x, y)$.

(ii) We say that X and Y are **independent** (ie $X \perp Y$) if and only if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for each $(x, y) \in S$.
Otherwise, they are dependent.
(Note: Indeed, recall that $p_{X,Y}(x, y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\})\,\mathbb{P}(\{Y = y\}) = p_X(x)p_Y(y)$, where "independence" here is equivalent to $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ in the usual sense.)

(iii) If $X \perp Y$, then $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y))$ for any functions f and g.
Example: $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$ if $X \perp Y$.

Some additional algebraic properties include:

(iv) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ (always hold regardless if X and Y are independent).
Other equivalent properties of $\mathbb{E}(\cdot)$ for one random variable holds.

(v) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (only if $X \perp Y$).

(vi) $\mathbb{E}(X + Y|B) = \mathbb{E}(X|B) + \mathbb{E}(Y|B)$ for any event B.
Other equivalent properties of $\mathbb{E}(\cdot)$ for one random variable holds, but we replace $\mathbb{E}(\cdot)$ with $\mathbb{E}(\cdot|B)$.

Remark: You might have seen $\mathbb{E}(X + Y|X)$ or some equivalent forms of this. This is somewhat of an abuse of notation since you are conditioning on a random variable. What it means is that for a given $X = x$, we then compute the expectation of $X + Y$ (by treating $X = x$ as given, or in other words, a constant). This yields $x + \mathbb{E}(Y)$. Now, allowing x to vary over the support of the random variable X, we have $X + \mathbb{E}(Y)$ given X, or $X + \mathbb{E}(Y|X)$ where $|X$ here is to record the fact that this is computed on the assumption that X is given. We might possibly cover this in the next discussion.

i.e, $\mathbb{E}(X + Y|X) = X + \mathbb{E}(Y|X)$.
Note that the above concepts generalize to more than 2 random variables.

We shall look at a couple of examples below to see how these concepts come into play.

Example 44. Consider two random variables X and Y (which are possibly not independent) with joint probability mass function below:

$X \setminus Y$	1	2
1	$1/8$	$3/8$
2	$1/4$	$1/4$

(i) Compute the marginal pmf of X and Y.

(ii) Compute $\mathbb{E}(X)$ and $\mathbb{E}(Y)$.

(iii) Compute $\mathbb{E}(XY)$.

(iv) Prove that X and Y are not independent.

Suggested Solution:

(i) By definition, for each x, we sum over all y of the joint pmf $p_{X,Y}(x,y)$. For example, we have

$$p_X(1) = \sum_y p_{X,Y}(1,y) = p_{X,Y}(1,1) + p_{X,Y}(1,2) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}.$$

Similarly, we have

$$p_X(2) = \sum_y p_{X,Y}(2,y) = p_{X,Y}(2,1) + p_{X,Y}(2,2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Hence, we have

$$p_X(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we have

$$p_Y(1) = \sum_x p_{X,Y}(x,1) = p_{X,Y}(1,1) + p_{X,Y}(2,1) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}.$$

For $p_Y(2)$, we can either do the same, or use the fact that the sum over pmf gives 1 to obtain $p_Y(2) = 1 - \frac{3}{8} = \frac{5}{8}$. Hence, we have

$$p_Y(y) = \begin{cases} \frac{3}{8} & \text{if } y = 1 \\ \frac{5}{8} & \text{if } y = 2 \\ 0 & \text{otherwise.} \end{cases}$$

(ii) We can use the definition of expectation as per usual with the pmf computed from (i). Thus, we have

$$\mathbb{E}(X) = \sum_x x p_X(x) = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = \frac{3}{2}.$$

Similarly,

$$\mathbb{E}(Y) = \sum_y y p_X(y) = 1 \times \frac{3}{8} + 2 \times \frac{5}{8} = \frac{13}{8}.$$

(iii) By definition of expectation of a function of two random variables, we have $\mathbb{E}(XY) = \sum_{x,y} xy p_{X,Y}(x,y)$. Thus

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{x,y} xy p_{X,Y}(x,y) \\ &= 1 \cdot 1 \cdot p_{X,Y}(1,1) + 1 \cdot 2 \cdot p_{X,Y}(1,2) + 2 \cdot 1 \cdot p_{X,Y}(2,1) + 2 \cdot 2 \cdot p_{X,Y}(2,2) \\ &= 1 \cdot 1 \cdot \frac{1}{8} + 1 \cdot 2 \cdot \frac{3}{8} + 2 \cdot 1 \cdot \frac{1}{4} + 2 \cdot 2 \cdot \frac{1}{4} \\ &= \frac{19}{8}.\end{aligned}$$

(iv) Suppose that X and Y are independent, then we must have $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$. However, we see that $\mathbb{E}(X)\,\mathbb{E}(Y) = \frac{3}{2} \times \frac{13}{8} = \frac{39}{16} \neq \frac{19}{8} = \mathbb{E}(XY)$, hence a contradiction. Thus, X and Y must be dependent.

Example 45. Suppose X, Y, and Z are mutually independent and they all follow Poisson(1). Compute

$$\mathbb{E}(4X - Y|X + Y + Z = 12).$$

Suggested Solution: By properties of expectation, we have

$$\mathbb{E}(4X|X + Y + Z = 12) = 4\,\mathbb{E}(X|X + Y + Z = 12) - \mathbb{E}(Y|X + Y + Z = 12).$$

To compute $\mathbb{E}(X|X + Y + Z = 12)$ and $\mathbb{E}(Y|X + Y + Z = 12)$, we argue that since X, Y, and Z all follows the same distribution, then we must have the same value for $\mathbb{E}(Y|X + Y + Z = 12)$ and $\mathbb{E}(Z|X + Y + Z = 12)$. Let $\alpha = \mathbb{E}(X|X + Y + Z = 12) = \mathbb{E}(Y|X + Y + Z = 12) = \mathbb{E}(Z|X + Y + Z = 12)$. By properties of expectation, we have

$$\mathbb{E}(X+Y+Z|X+Y+Z = 12) = \mathbb{E}(X|X+Y+Z = 12)+\mathbb{E}(Y|X+Y+Z = 12)+\mathbb{E}(Z|X+Y+Z = 12) = 3\alpha.$$

On the other hand, since $X + Y + Z$ is given to be 12, then $X + Y + Z = 12$ is a constant. Thus,

$$\mathbb{E}(X + Y + Z|X + Y + Z = 12) = 12.$$

Combining both, we have

$$3\alpha = 12 \implies \alpha = 4.$$

Thus, we have

$$\mathbb{E}(4X|X + Y + Z = 12) = 4\,\mathbb{E}(X|X + Y + Z = 12) - \mathbb{E}(Y|X + Y + Z = 12) = 4\alpha - \alpha = 3\alpha = 12.$$

Example 46. Suppose that $X \sim \mathrm{B}\left(2, \frac{1}{2}\right)$, while Y is the outcome of a fair 4-sided dice roll. Furthermore, suppose that X and Y are independent. Let $Z = \max\{X, Y\}$.

(i) Compute $\mathbb{E}(X/Y)$.

(ii) Prove that $\{Z \le z\} = \{X \le z\} \cap \{Y \le z\}$ for any real number z.

(iii) The cumulative distribution function (cdf) of a random variable Z, denoted by $F_Z(z)$, is given by $\mathbb{P}(Z \le z)$. (That is, $F_Z(z) = \mathbb{P}(Z \le z)$.) Determine the cdf of Z.

(iv) Using your answer in (ii), determine the pmf of Z, $p_Z(z)$.

Suggested Solution:

(i) Since X and Y are independent, we have $\mathbb{E}(X/Y) = \mathbb{E}(X)\,\mathbb{E}(1/Y)$. Since $X \sim \mathrm{B}\left(2, \frac{1}{2}\right)$, then $\mathbb{E}(X) = 2 \times \frac{1}{2} = 1$. On the other hand, to compute $\mathbb{E}(1/Y)$, we will have to appeal to the definition. Since the pmf of Y is given by $p_Y(y) = \frac{1}{4}$ if $y = 1, 2, 3$, or 4 and 0 otherwise, we have

$$\mathbb{E}\left(\frac{1}{Y}\right) = \sum_y \frac{1}{y} p_Y(y) = \frac{1}{1} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{25}{48}.$$

Hence, we have

$$\mathbb{E}(X/Y) = \mathbb{E}(X)\,\mathbb{E}(1/Y) = 1 \times \frac{25}{48} = \frac{25}{48}.$$

Remark: In general, $\mathbb{E}(1/Y) \ne \frac{1}{\mathbb{E}(Y)}$!!

(ii) To prove that two sets A, B are equal, we will want to show that $A \subset B$ and $B \subset A$. In other words, we would like to show that if we take an element in A, then it must be in B and vice versa.

$\{Z \le z\} \subset \{X \le z\} \cap \{Y \le z\}$. This reads " if $Z \le z$, then $X \le z$ and $Y \le z$."
Indeed, if $Z \le z$, then $\max\{X, Y\} \le z$. As the max of two values is still smaller than z, then we have that both must be smaller than z, that is, $X \le z$ and $Y \le z$.

$\{Z \le z\} \supset \{X \le z\} \cap \{Y \le z\}$. This reads " if $X \le z$ and $Y \le z$, then $Z \le z$."
Indeed, if both $X \le z$ and $Y \le z$, then the maximum of these two must still be $\le z$, that is, $\max\{X, Y\} \le z$. This in turn implies that $Z \le z$ since $Z = \max\{X, Y\}$.

(iii) Using (ii), we know that

$$\mathbb{P}(Z \le z) = \mathbb{P}(\{Z \le z\}) = \mathbb{P}(\{X \le z\} \cap \{Y \le z\}) = \mathbb{P}(X \le z)\,\mathbb{P}(Y \le z).$$

In the last equality, we have used the fact that X and Y are independent. Now, we just have to iterate through different values of z.

- $\mathbb{P}(Z \le z)$ for $z \le 0$ is equal to zero, since a $4-$sided dice cannot output any value less than or equals to 0. Thus, $\mathbb{P}(Z \le z) = \mathbb{P}(X \le z)\,\mathbb{P}(Y \le z) = \mathbb{P}(X \le z) \times 0 = 0$.
- $\mathbb{P}(Z \le 1) = \mathbb{P}(X \le 1)\,\mathbb{P}(Y \le 1) = (1 - \mathbb{P}(X = 2))\frac{1}{4} = \left(1 - \frac{1}{4}\right) \times \frac{1}{4} = \frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$. Here, we have used the fact that $\mathbb{P}(X \le 1) = 1 - \mathbb{P}(X = 2)$ since X only takes values $0, 1$, and 2. Furthermore, for $X \sim \mathrm{B}\left(2, \frac{1}{2}\right)$, it is easy to compute $\mathbb{P}(X = 2)$ since it is equivalent to getting two tails for two fair coin flips, which is with probability $\frac{1}{4}$.
- For $z \ge 2$, we have $\mathbb{P}(X \le z) = 1$ (since X only takes values $0, 1$, and 2), so it only remains to compute $\mathbb{P}(Y \le z)$. Thus, we have
  $\mathbb{P}(Z \le 2) = \mathbb{P}(Y \le 2) = \frac{2}{4}$. (Since this includes the faces 1 and 2 out of the 4 possible faces.)
  $\mathbb{P}(Z \le 3) = \mathbb{P}(Y \le 3) = \frac{3}{4}$. (Since this includes the faces 1 and 2 out of the 4 possible faces.)
  $\mathbb{P}(Z \le 4) = \mathbb{P}(Y \le 4) = \frac{4}{4} = 1$. (Since this includes the faces 1 and 2 out of the 4 possible faces.)
- In fact, for $z \ge 4$, we will always get $\mathbb{P}(Z \le z) = 1$ since $X \le z$ and $Y \le z$ is always satisfied as X takes values $0, 1$, and 2, while Y takes values $1, 2, 3$, and 4 only.
- Last but not least, we note that if z is not an integer, say $z = 1.5$, then $\mathbb{P}(Z \le 1.5) = \mathbb{P}(Z \le 1)$ since both X and Y only takes discrete values. In fact, $\mathbb{P}(Z \le 1.99999) = \mathbb{P}(Z \le 1)$ but $\mathbb{P}(Z \le 2) = \mathbb{P}(Z \le 2)$. The jump in the cdf if occurs, only occurs at the integers.

- In fact, the jump from $\mathbb{P}(Z \leq 1.9999)$ (which is equals to $\mathbb{P}(Z \leq 1)$) to $\mathbb{P}(Z \leq 2)$ is due to the inclusion of the probability $\mathbb{P}(Z = 2)$. Thus, to compute $\mathbb{P}(Z = 2)$, this is equivalent to taking $\mathbb{P}(Z \leq 2) - \mathbb{P}(Z \leq 1)$ or $\mathbb{P}(Z \leq 2) - \mathbb{P}(Z < 2)$.

Thus, we have

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 1 \\ \frac{3}{16} & \text{if } 1 \leq z < 2 \\ \frac{2}{4} & \text{if } 2 \leq z < 3 \\ \frac{3}{4} & \text{if } 3 \leq z < 4 \\ 1 & \text{if } z \geq 4 \end{cases}$$

(iv) From the final remark in the previous part, we can then obtain

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z \leq 1) - \mathbb{P}(Z < 1) = \frac{3}{16} - 0 = \frac{3}{16}.$$

Similarly, we have

$$\mathbb{P}(Z = 2) = \mathbb{P}(Z \leq 2) - \mathbb{P}(Z < 2) = \frac{2}{4} - \frac{3}{16} = \frac{5}{16}.$$

$$\mathbb{P}(Z = 3) = \mathbb{P}(Z \leq 3) - \mathbb{P}(Z < 3) = \frac{3}{4} - \frac{2}{4} = \frac{1}{4}.$$

$$\mathbb{P}(Z = 4) = \mathbb{P}(Z \leq 4) - \mathbb{P}(Z < 4) = 1 - \frac{3}{4} = \frac{1}{4}.$$

Hence, the pmf is given by

$$p_Z(z) = \begin{cases} \frac{3}{16} & \text{if } z = 1 \\ \frac{5}{16} & \text{if } z = 2 \\ \frac{1}{4} & \text{if } z = 3 \\ \frac{1}{4} & \text{if } z = 4 \\ 0 & \text{otherwise.} \end{cases}$$

Note that in general, if the support of Z is on discrete integers (ie $\mathbb{P}(Z = k) = 0$ if k is not an integer), then the above formula can be generalized to

$$\mathbb{P}(Z = k) = \mathbb{P}(Z \leq k) - \mathbb{P}(Z \leq k - 1).$$

This might be useful for HW 7.

Example 47. Let $Y = X_1 + X_2 + X_3$, where X_1, X_2, and X_3 are mutually independent and all follows Be(p) for some $p \in (0,1)$.

(i) Compute $\mathbb{E}(X_1^2)$ by definition.

(ii) Compute $\mathbb{E}(Y)$.

(iii) Compute $\mathbb{E}(Y^2)$.
Hint: Use the fact that $(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$.

(iv) Compute Var(Y).

(v) Suppose $Z \sim$ B$(3, 0.6)$. Hence or otherwise, compute Var(Z).

Suggested Solution:

(i) Recall that the pmf of Be(p) is given by

$$p_{X_1}(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

By definition, we have

$$\mathbb{E}(X_1^2) = \sum_k k^2 p_{X_1}(k) = 1^2 \cdot p + 0^2 \cdot (1-p) = p.$$

Note that by the exact same argument, we have $\mathbb{E}(X_2^2) = \mathbb{E}(X_3^2) = p$.

(ii) By linearity of expectation, we have

$$\mathbb{E}(Y) = \mathbb{E}(X_1 + X_2 + X_3) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = p + p + p = 3p.$$

Here, we have used the fact that if $X \sim$ Be(p), then $\mathbb{E}(X) = p$.

(iii) Note that $Y^2 = (X_1 + X_2 + X_3)^2 = X_1^2 + X_2^2 + X_3^2 + 2X_1X_2 + 2X_1X_3 + 2X_2X_3$. Hence, we have

$$\begin{aligned} \mathbb{E}(Y^2) &= \mathbb{E}(X_1^2 + X_2^2 + X_3^2 + 2X_1X_2 + 2X_1X_3 + 2X_2X_3) \\ &= \mathbb{E}(X_1^2) + \mathbb{E}(X_2^2) + \mathbb{E}(X_3^2) + \mathbb{E}(2X_1X_2) + \mathbb{E}(2X_1X_3) + \mathbb{E}(2X_2X_3) \\ &= \mathbb{E}(X_1^2) + \mathbb{E}(X_2^2) + \mathbb{E}(X_3^2) + 2\,\mathbb{E}(X_1)\,\mathbb{E}(X_2) + 2\,\mathbb{E}(X_1)\,\mathbb{E}(X_3) + 2\,\mathbb{E}(X_2)\,\mathbb{E}(X_3) \\ &= p + p + p + 2p^2 + 2p^2 + 2p^2 \\ &= 3p + 6p^2 \end{aligned}$$

Here, we have used the fact that since X_1, X_2, and X_3 are independent, then $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\,\mathbb{E}(X_2)$, $\mathbb{E}(X_1X_3) = \mathbb{E}(X_1)\,\mathbb{E}(X_3)$, and $\mathbb{E}(X_2X_3) = \mathbb{E}(X_2)\,\mathbb{E}(X_3)$.

(iv) Using the fact that Var$(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$, we have

$$\text{Var}(Y) = 3p + 6p^2 - (3p)^2 = 3p - 3p^2 = 3p(1-p).$$

(v) Recall that if X_i are Be(p), then the sum of n copies of independent X_i gives B(n,p). Hence, we have that if we set $p = 0.6$, then $Y = X_1 + X_2 + X_3$ is basically just B$(3, 0.6)$ and is the same as Z, and we have computed its variance in (iv). Thus, we have

$$\text{Var}(Y) = 3(0.6)(1 - 0.6) = 0.72.$$

Remark: Using a similar trick, we can compute that if $Z \sim$ B(n,p), then Var$(Z) = np(1-p)$. This is covered in the following example.
Furthermore, we know from Homework 6 Problem 1 that if $W \sim$ Poisson(λ), then $\mathbb{E}(W^2) = \lambda + \lambda^2$, so Var$(W) = \mathbb{E}(W^2) - (\mathbb{E}(W)^2) = \lambda + \lambda^2 - (\lambda)^2 = \lambda$. These two variances are good to know for the purpose of the final exam, and I will include them in a table of distributions towards the final discussion/ office hour before finals.

Example 48. Let $Y = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \cdots X_n$ are mutually independent and all follows Be(p) for some $p \in (0,1)$. Then $Y \sim$ B(n,p) for some positive integer n.

Prove that
$$\text{Var}(Y) = np(1-p).$$

Suggested Solution: Recall from Example 47, we have that $\mathbb{E}(X_i) = \mathbb{E}(X_i^2) = p$. Furthermore, recall that Var$(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$. Hence, we have to compute $\mathbb{E}(Y)$ and $\mathbb{E}(Y^2)$. For the latter, by writing $Y = \sum_{i=1}^n X_i$ and using various properties of expectations, we have

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n p = np.$$

Next, to compute $\mathbb{E}(Y^2)$, we use the following formula:

$$\left(\sum_{i=1}^n X_i\right)^2 = \sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j$$

which can be verified using mathematical induction.
(Refer to Example 47 for when $n = 3$, this is given in the hint.)
Thus, we have that

$$\mathbb{E}(Y^2) = \mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j).$$

Since X_i and X_j are independent, we have $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j) = p^2$ since $\mathbb{E}(X_k) = p$ for each k. On the other hand, we know that $\mathbb{E}(X_i^2) = p$. This yields

$$\mathbb{E}(Y^2) = \sum_{i=1}^n p + \sum_{i \neq j} p^2 = np + (n^2 - n)p^2.$$

Here, we have used the fact that $\sum_{i \neq j}$ contains $n^2 - n$ terms, since $i = j$ for $i = 1$ to n contains n terms, and the unrestricted sum $\sum_{i,j}$ contains n^2 terms since the summation over each index is for n times.

By the definition of variance, we thus have

$$\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = np + (n^2 - n)p^2 - (np)^2 = np - np^2 = np(1-p).$$

9 Discussion 9

Conditional Probability Mass Function.
Given two discrete random variables X and Y, the **conditional probability mass function** of X given that $Y = y$, is defined by

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

provided that $p_Y(y) > 0$.
Similarly, the conditional pmf of Y given $X = x$, is defined by

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

provided that $p_X(x) > 0$.
Note that these are variants of the definition of conditional probability, which says that

$$p_{X|Y}(x,y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

and similarly for $p_{Y|X}$.
We shall include an example of computing the conditional pmf below.

Example 49. Let X be the outcome of a fair $6-$sided dice and $Y \sim \text{Be}\left(\frac{X}{6}\right)$.

(i) Determine the conditional pmf $p_{Y|X}(y|x)$.

(ii) Determine the joint pmf $p_{X,Y}(x,y)$.

(iii) Determine the marginal pmf of $p_Y(y)$. Give the name of this distribution with its corresponding parameter.

(iv) Determine the conditional pmf $p_{X|Y}(x|y)$.

Suggested Solution:

(i) Recall that $p_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$. If $X = x$, then $x \in [[1,6]]$ since it is the outcome of a dice roll. Note that Y, as a Bernoulli random variable, can only take values of 0 and 1. If $X = x$, then $Y|X = x \sim \text{Be}\left(\frac{x}{6}\right)$. so $\mathbb{P}(Y = 1|X = x) = \frac{x}{6}$ and $\mathbb{P}(Y = 0|X = x) = 1 - \frac{x}{6}$. This yields

$$p_{Y|X}(y|x) = \begin{cases} \frac{x}{6} & \text{if } y = 1, \\ 1 - \frac{x}{6} & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

(ii) By the definition of conditional probability, we have that $p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{Y = y\}|\{X = x\})\,\mathbb{P}(\{X = x\}) = p_{Y|X}(y|x)p_X(x)$. Since $p_X(x) = \frac{1}{6}$ for $x \in [[1,6]]$, we have

$$p_{X,Y}(x,y) = \begin{cases} \frac{x}{36} & \text{if } y = 1, \text{ and } x \in [[1,6]], \\ \frac{6-x}{36} & \text{if } y = 0, \text{ and } x \in [[1,6]], \\ 0 & \text{otherwise.} \end{cases}$$

(iii) To compute the marginal pmf, recall that $p_Y(y) = \sum_x p_{X,Y}(x,y)$, and in this case, for each $y \in \{0,1\}$, x can take integer values from 1 to 6. Hence, we have

$$p_Y(1) = \sum_{x=1}^{6} p_X(x,1) = \sum_{x=1}^{6} \frac{x}{36} = \frac{1}{36}(1+2+3+4+5+6) = \frac{21}{36}.$$

On the other hand, we can either compute $p_Y(0)$ similarly or recall the fact that for a pmf, the sum of probabilities is 1, and thus $p_Y(0) = \frac{15}{36}$ necessarily (since $Y = y$ can only take non-zero probabilities for $y \in \{0,1\}$). Since Y can only take values 0 and 1 and has a probability of success = $\frac{21}{36}$, then $Y \sim \text{Be}\left(\frac{21}{36}\right)$.

(iv) By the definition of conditional probability, we have that

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Substituting the relevant formulas from the previous parts, we have

$$p_{X|Y}(x|y) = \begin{cases} \frac{x/36}{21/36} = \frac{x}{21} & \text{if } y = 1, \text{ and } x \in [[1,6]], \\ \frac{(6-x)/36}{15/36} = \frac{6-x}{15} & \text{if } y = 0, \text{ and } x \in [[1,6]], \\ 0 & \text{otherwise.} \end{cases}$$

First Step Analysis.

In general, when we are faced with the task of computing the probability of an event A or the expectation of a random variable X, the trick is to pick a wise choice of partition and apply either the law of total probability or the total expectation theorem. Recall that for a partition $\{B_k\}$ of a sample space S, we then have

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A|B_k)\,\mathbb{P}(B_k).$$

and

$$\mathbb{E}(X) = \sum_k \mathbb{E}(X|B_k)\,\mathbb{P}(B_k).$$

A common choice of partition is to condition on what happens after the first step (and hence the name of this method - first step analysis). First, we describe all possible outcomes of the first step with B_k, which would partition the sample space. Then, we apply the relevant equation accordingly. In some cases, we would usually obtain some function of $\mathbb{P}(A)$ (if we are computing $\mathbb{P}(A)$) or $\mathbb{E}(X)$ (if we are computing $\mathbb{E}(X)$) on the right-hand side of the equation. The problem then reduces to an algebraic one, and we can then solve the algebraic equation to compute $\mathbb{P}(A)$ and/or $\mathbb{E}(X)$.

On the other hand, if $\mathbb{E}(X)$ depends on some given integer n, which we denote as $\mathbb{E}(X_n)$, there are cases in which instead of being able to have $\mathbb{E}(X_n)$, we have $\mathbb{E}(X_{n-1})$ or other combinations of previous iterations instead. This then reduces to "solving a recurrence relation".

Example 50. Two $6-$sided fair dices are rolled together repeatedly. Find the probability that a total of 12 comes up before a total of 11.

Suggested Solution:
Let

$$\begin{aligned} A &= \{\text{Total of } 12 \text{ before a total of } 11\} \\ N &= \{\text{Total of } \le 10 \text{ on the first roll}\}, \\ F &= \{\text{Total of } 11 \text{ on the first roll}\}, \\ P &= \{\text{Total of } 12 \text{ on the first roll}\}. \end{aligned}$$

We can see that N, F, and P partitions the sample space of $[[2, 12]]$. By the law of total probability, we have

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|N)\,\mathbb{P}(N) + \mathbb{P}(A|F)\,\mathbb{P}(F) + \mathbb{P}(A|P)\,\mathbb{P}(P) \\ &= \mathbb{P}(A|N)\frac{33}{36} + \mathbb{P}(A|F)\frac{2}{36} + \mathbb{P}(A|P)\frac{1}{36} \\ &= \mathbb{P}(A)\frac{33}{36} + 0 \cdot \frac{2}{36} + 1 \cdot \frac{1}{36} \\ &= \mathbb{P}(A)\frac{33}{36} + \frac{1}{36}, \end{aligned}$$

where we have computed the relevant probabilities of N, F, and P accordingly. Furthermore, given that we have N, then this is neither the required event nor the event that counts as a failure. Hence, this is as if we start the entire event again, and thus $\mathbb{P}(A|N) = \mathbb{P}(A)$. On the other hand, if F happens, then we have failed, and thus the probability of obtaining a total of 12 before a total of 11 will never happen, and thus $\mathbb{P}(A|F) = 0$. Furthermore, if we had a total of 12 on our first roll, then the event is attained, and we thus have $\mathbb{P}(A|P) = 1$. Solving the above equation, we have

$$\begin{aligned} \frac{3}{36}\,\mathbb{P}(A) &= \frac{1}{36}, \\ \mathbb{P}(A) &= \frac{1}{3}. \end{aligned}$$

Example 51. Let $X \sim \text{Geom}(p)$. Using the method of first-step analysis, compute

(i) $\mathbb{E}(X)$,

(ii) $\mathbb{E}(X^2)$, and hence

(iii) $\text{Var}(X)$.

Suggested Solution:

(i) Let B be the event representing failing on the first attempt. Then, we see that B and B^c partition the sample space. By the total expectation theorem, we have

$$\mathbb{E}(X) = \mathbb{E}(X|B)\,\mathbb{P}(B) + \mathbb{E}(X|B^c)\,\mathbb{P}(B^c).$$

Note that $\mathbb{P}(B) = p$ and $\mathbb{P}(B^c) = 1 - p$ since the probability of passing an attempt is described by p in the parameter of the geometric distribution.

Given that we have succeeded in the first attempt, then the expected number of attempts required to get the first success is then given by 1, ie $\mathbb{E}(X|B) = 1$. On the other hand, given that we have failed in the first attempt, then the expected number of attempts required to get the first success is given by $\mathbb{E}(X+1) = 1 + \mathbb{E}(X)$ (here 1 refers to the first attempt that we have failed, and $\mathbb{E}(X)$ here means that we are resetting our counter back to right before we start the sequence of attempts).

Thus implies that

$$\begin{aligned} \mathbb{E}(X) &= p + (\mathbb{E}(X) + 1)(1-p) \\ p\,\mathbb{E}(X) &= 1 \\ \mathbb{E}(X) &= \frac{1}{p}. \end{aligned}$$

(ii) Using a similar idea, we replace X by X^2 to obtain

$$\mathbb{E}(X^2) = \mathbb{E}(X^2|B)\,\mathbb{P}(B) + \mathbb{E}(X^2|B^c)\,\mathbb{P}(B^c).$$

Furthermore, the expected number of squared attempts is at $1^2 = 1$ given that we have succeeded on our first attempt. On the other hand, $\mathbb{E}(X^2|B^c) = \mathbb{E}((X+1)^2)$. (Recall that $+1$ here is to include the failed first attempt.) Then, we have $\mathbb{E}(X^2|B^c) = \mathbb{E}((X+1)^2) = \mathbb{E}(X^2+2X+1) = \mathbb{E}(X^2)+2\,\mathbb{E}(X)+1 = \mathbb{E}(X^2)+\frac{2}{p}+1$ using our answer in (i). This then implies that

$$\mathbb{E}(X^2) = p + \left(\mathbb{E}(X^2) + \frac{2}{p} + 1\right)(1-p)$$

and hence

$$\mathbb{E}(X^2) = \frac{1}{p} + \frac{2(1-p)}{p^2}.$$

(iii) Last but not least, we use the usual formula for variance to obtain

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Example 52. (Gambler Ruin.) Consider a coin-flipping game with two players where each player has a 50% chance of winning with each flip of the coin. After each flip of the coin, the loser transfers one penny to the winner. The game ends when one player has all the pennies.

Let $f(n)$ be the probability that player A wins if player A has n pennies (assuming that the total number of pennies is at $m > n$).

Explain why $f(n) = \frac{1}{2}f(n-1) + \frac{1}{2}f(n+1)$.

Suggested Solution:

Let W be the event that player A wins, and B be the event that player A wins the first coin flip. Observe that B and B^c partition the sample space. By the law of total probability, we thus have

$$\mathbb{P}(W) = \mathbb{P}(W|B)\,\mathbb{P}(B) + \mathbb{P}(W|B^c)\,\mathbb{P}(B^c).$$

Since the coin is fair, we have that $\mathbb{P}(B) = \mathbb{P}(B^c) = \frac{1}{2}$. Furthermore, we have that $\mathbb{P}(W|B) = f(n+1)$, since upon winning the first coin flip, player A now has $n+1$ pennies, and this reduces to asking for the probability of player A winning given that the player has $n+1$ pennies. Similarly, $\mathbb{P}(W|B^c) = f(n-1)$, since the player would have lost a penny given that the player lost the first coin flip. Plugging all the relevant numbers in, we have

$$f(n) = \frac{1}{2}f(n+1) + \frac{1}{2}f(n-1),$$

as required.

Continuous Random Variable - An Introduction.

Recall that if X is a discrete random variable, then $\mathbb{P}(X = x) = p_X(x)$ is described by the pmf, and for an event A, we have

$$\mathbb{P}(X \in A) = \sum_{k \in A} p_X(k).$$

(That is, we sum over all probabilities that contribute to the event A.) For instance, if X takes integer values,

$$\mathbb{P}(X \leq 3) = \sum_{k=-\infty}^{3} p_X(k). \tag{3}$$

If X is a continuous random variable, then X takes values on the real line. Since X varies continuously on the real line, it is hard to pinpoint the probability that X is exactly equal to a number. say 3, since it could take values 2.999999 or 3.00000101 and they are all not included in the probability that $X = 3$. We instead use the concept of "density", in intuitively says how much probability per unit length on the real line is associated with the random variable X of interest. In other words, for a small Δx around 3, we say that

$$\mathbb{P}(x \in (3 - \Delta x/2, 3 + \Delta x/2)) \approx f_X(3)\Delta x$$

where $f_X(x)$ now records the probability "density" here (ie probability per unit length), and thus is known as the **probability density function** (pdf) associated with the random variable X.

With such a concept, we can generalize the formula in (3) to the continuous case as follows:

$$F_X(k) = \mathbb{P}(X \leq k) = \int_{-\infty}^{k} f_X(x)\mathrm{d}x. \tag{4}$$

for any $k \in \mathbb{R}$. The function on the left-hand side of (4) is also known as the **cumulative distribution function** (cdf).

With that, here are some basic properties associated with a continuous random variable X:

- As a pdf, the "sum" over all probabilities must be equal to one. This implies that

$$\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1.$$

- The above point implies that

$$\lim_{k \to \infty} F_X(k) = 1$$

  since it must contain all possible values that "X can take".

- From (4) and the fundamental theorem of calculus, we can differentiate both sides of (4) and obtain

$$\frac{\mathrm{d}}{\mathrm{d}k} F_X(k) = f_X(k). \tag{5}$$

The most basic example of a continuous random variable that you will encounter is the uniform random variable, denoted by $X \sim \text{Unif}[a, b]$ for some $a < b$. For a uniform random variable supported on (a, b), this implies that

$$f_X(x) = \begin{cases} C & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

where the pdf is constant in its support. To determine the value of C, we use the fact that the integral over $\mathbb{R}$ of the pdf gives one, and thus

$$1 = \int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \int_a^b f_X(x)\mathrm{d}x = C(b - a).$$

This then implies that

$$C = \frac{1}{b - a},$$

and thus

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

To obtain the cdf, we appeal to (4). This will be illustrated in the following example.

Example 53. Suppose that $X \sim$ Unif$[1,7]$. Determine the cdf of X, $F_X(x)$.

Suggested Solution: From the notes above, we have that

$$f_X(x) = \begin{cases} \frac{1}{6} & \text{for } x \in [1,7] \\ 0 & \text{otherwise.} \end{cases}$$

Applying (4), we have that

$$F_X(k) = \int_{-\infty}^{k} p_X(x)\mathrm{d}x.$$

Since $f_X(x) = 0$ for $x < 1$, then for $k < 1$, we have

$$F_X(k) = \int_{-\infty}^{k} f_X(x)\mathrm{d}x = \int_{-\infty}^{k} 0\mathrm{d}x = 0.$$

Hence, for $k \geq 1$, it suffices for the integral to start from 1. This implies that we have

$$F_X(k) = \int_{1}^{k} f_X(x)\mathrm{d}x.$$

In ot For $k \in [1,7]$, we have that

$$F_X(k) = \int_{1}^{k} f_X(x)\mathrm{d}x = \int_{1}^{k} \frac{1}{6}\mathrm{d}x = \frac{1}{6}(k-1).$$

For $k > 7$, as the pdf for $x \geq 7$ is zero, we have that

$$F_X(k) = \int_{1}^{k} f_X(x)\mathrm{d}x = \int_{1}^{7} \frac{1}{6}\mathrm{d}x = 1.$$

Overall, this implies that

$$F_X(k) = \begin{cases} 0 & \text{if } k < 1 \\ \frac{1}{6}(k-1) & \text{if } 1 \leq k \leq 7, \\ 1 & \text{if } k > 7. \end{cases}$$

10 Discussion 10

Continuous Random Variable - An Introduction.
Recall that if X is a discrete random variable, then $\mathbb{P}(X = x) = p_X(x)$ is described by the pmf, and for an event A, we have

$$\mathbb{P}(X \in A) = \sum_{k \in A} p_X(k).$$

(That is, we sum over all probabilities that contribute to the event A.) For instance, if X takes integer values,

$$\mathbb{P}(X \leq 3) = \sum_{k=-\infty}^{3} p_X(k). \tag{6}$$

If X is a continuous random variable, then X takes values on the real line. Since X varies continuously on the real line, it is hard to pinpoint the probability that X is exactly equal to a number. say 3, since it could take values 2.999999 or 3.00000101 and they are all not included in the probability that $X = 3$. We instead use the concept of "density", in intuitively says how much probability per unit length on the real line is associated with the random variable X of interest. In other words, for a small Δx around 3, we say that

$$\mathbb{P}(x \in (3 - \Delta x/2, 3 + \Delta x/2)) \approx f_X(3)\Delta x$$

where $f_X(x)$ now records the probability "density" here (ie probability per unit length), and thus is known as the **probability density function** (pdf) associated with the random variable X.
With such a concept, we can generalize the formula in (3) to the continuous case as follows:

$$F_X(k) = \mathbb{P}(X \leq k) = \int_{-\infty}^{k} f_X(x)\mathrm{d}x. \tag{7}$$

for any $k \in \mathbb{R}$. The function on the left-hand side of (7) is also known as the **cumulative distribution function** (cdf).
With that, here are some basic properties associated with a continuous random variable X:

- As a pdf, the "sum" over all probabilities must be equal to one. This implies that
$$\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1.$$
- The above point implies that
$$\lim_{k\to\infty} F_X(k) = 1$$
since it must contain all possible values that "X can take".
- From (7) and the fundamental theorem of calculus, if $F_X(k)$ is differentiable (which is for a continuous random variable, but not necessarily for a discrete random variable), we can differentiate both sides of (7) and obtain
$$\frac{\mathrm{d}}{\mathrm{d}k}F_X(k) = f_X(k). \tag{8}$$
- $\mathbb{P}(X = k) = 0$, since the probability of a continuous random variable taking an exact value is 0. Note that this is **not the same as** $p_X(k)$, which represents the probability **density**.
- Furthermore, $\mathbb{E}(g(X)) = \int g(x)f_X(x)\mathrm{d}x$, where $f_X(x)$ is the pdf of X.

Common continuous random variables:

Distribution	Notation	Parameters	Range	pdf: $f_X(x)$	cdf: $F_X(x)$	$\mathbb{E}(X)$	Var(X)
Uniform	Unif$[a,b]$	$a, b \in \mathbb{R}$	$[a,b]$	$\begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$	$\begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a,b] \\ 1 & \text{if } x > b. \end{cases}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Normal/ Gaussian	$N(\mu, \sigma^2)$	$\mu \in \mathbb{R}$, $\sigma \geq 0$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ \forall x$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ	σ^2
Exponential	Exp(λ)	$\lambda > 0$	$[0, \infty)$	$\begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$	$\begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Note that $\Phi(k) = F_Z(k)$ where $Z \sim N(0,1)$. In other words, it is the cdf of a standard normal distribution (ie of distribution $N(0,1)$).

Other properties: Let $\alpha \in \mathbb{R}$. Then,

- Recall that $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ and $\mathbb{E}(\alpha X) = \alpha\,\mathbb{E}(X)$.
- Var$(\alpha X) = \alpha^2$Var(X).
- If X and Y are independent, Var$(X+Y) =$ Var$(X) +$ Var(Y).
- If X and Y are normal and **independent**, then so it $\alpha X + \beta Y$ for any $\alpha, \beta \in \mathbb{R}$.
  In fact, if $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then
  $$\alpha X + \beta Y \sim N(\alpha\mu_x, \alpha^2\sigma_X^2) + N(\beta\mu_Y, \beta^2\sigma_Y^2) \sim N(\alpha\mu_X + \beta\mu_Y, \alpha^2\sigma_X^2 + \beta^2\sigma_Y^2).$$
- Note that the pdf of a normal distribution is symmetric about its mean (see the pdf $f_X(x)$ from the table above). Refer to the picture drawn in the lecture for a pictorial representation of this.

Hints/Strategy:

- To determine the pdf $p_X(k)$ of a random variable, always start from the cdf $F_X(k)$ since it always exists, and it is just a usual probability question;
  $$F_X(k) = \mathbb{P}(X \leq k).$$
  Then, use the fundamental theorem of calculus to obtain
  $$p_X(k) = \frac{\mathrm{d}}{\mathrm{d}k}F_X(k).$$
- To obtain probabilities for a Gaussian random variable, note that you are only allowed to express your answer in terms of $\Phi(z)$, that is, if $Z \sim N(0,1)$, then $\Phi(z) = \mathbb{P}(Z \leq z)$. Thus, we try to "normalize" the Gaussian random variable as follows. If $X \sim N(\mu, \sigma^2)$, then $X - \mu \sim N(0, \sigma^2)$, and thus $\frac{X-\mu}{\sigma} \sim N(0,1)$. An explicit example of this will be covered below.

Example 54. Suppose that $X \sim \text{Unif}[1,3]$. Let $Y = X^2$.

(i) Compute the pdf of Y, $f_Y(y)$.

(ii) Compute $\mathbb{E}(Y)$.

Suggested Solution:

(i) Step 1: Compute the cdf of Y.

Let y be given. Note that if $y < 0$, then $\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = 0$ (since a non-negative value cannot be less than or equals to a negative value). For $y \geq 0$, we have

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}).$$

If $\sqrt{y} < 1$ and hence $y < 1$, then $\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \sqrt{y}) = 0$.
If $\sqrt{y} > 3$ and hence $y > 9$, then $\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \sqrt{y}) = 1$ (since the full pdf is only supported from 1 to 3, so after 3, we get the full probability). If $1 \leq \sqrt{y} \leq 3$, then we have

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \sqrt{y}) = \frac{\sqrt{y}-1}{3-1} = \frac{\sqrt{y}-1}{2}.$$

Hence, we have

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 1, \\ \frac{\sqrt{y}-1}{2} & \text{if } 1 \leq y \leq 9, \\ 1 & \text{if } y > 9. \end{cases}$$

Step 2: Use the fundamental theorem of calculus to compute pdf of Y.

Note that from the cdf above, it is differentiable everywhere except at $y = 1$ and $y = 9$ (which honestly does not matter since modifying the value of the pdf at these points does not change the cdf).

Then, we have, since $f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y)$, then,

$$f_Y(y) = \begin{cases} 0 & \text{if } y < 1 \text{ or } y > 9, \\ \frac{1}{4\sqrt{y}} & \text{if } 1 \leq y \leq 9. \end{cases}$$

(ii) $\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y = \int_1^9 \frac{1}{4}\sqrt{y} \mathrm{d}y = \frac{1}{4}\frac{2}{3}\left(9^{\frac{3}{2}} - 1^{\frac{3}{2}}\right) = \frac{26}{6} = \frac{13}{3}$.

Example 55. Suppose that $X \sim N(1,2)$, $Y \sim N(2,4)$, and are independent.

(i) Compute the distribution for
$$2X + Y - 1.$$

(ii) Compute the probability $\mathbb{P}(4 < 2X + Y < 16)$ in terms of Φ, the cdf of $N(0,1)$.

Suggested Solution:

(i) We know that since X and Y are independent and are normal, then so is $2X+Y$. In fact, we can compute its distribution as follows.

- $\mathbb{E}(2X+Y) = 2\,\mathbb{E}(X) + \mathbb{E}(Y) = 2 \cdot 1 + 2 = 4.$
- $\text{Var}(2X+Y) = 4\text{Var}(X) + \text{Var}(Y) = 4 \cdot 2 + 4 = 12.$

Hence, $2X + Y \sim N(4, 12)$.

Next, note that since -1 only shifts the mean of the normal distribution, we should expect that
$$2X + Y - 1 \sim N(3, 12).$$

Rigorously, we can think of $1 = N(1,0)$ (ie a normal random variable that is independent and has 0 variance), and use the properties of normal distributions. Thus, we have

- $\mathbb{E}(2X+Y-1) = \mathbb{E}(2X+Y) - \mathbb{E}(1) = 4 - 1 = 3..$
- $\text{Var}(2X+Y-1) = \text{Var}(2X+Y) = 12.$

Hence, $2X + Y - 1 \sim N(3, 12)$.

(ii) The trick is to manipulate the expression above into the cdf of $Z \sim N(0,1)$. Indeed, observe that

$$\begin{aligned}
\mathbb{P}(4 < 2X + Y < 16) &= \mathbb{P}(4 - 4 < 2X + Y - 4 < 16 - 4) \\
&= \mathbb{P}(0 < 2X + Y - 4 < 12) \\
&= \mathbb{P}\left(\frac{0}{\sqrt{12}} < \frac{2X+Y-4}{\sqrt{12}} < \frac{12}{\sqrt{12}}\right) \\
&= \mathbb{P}\left(0 < Z < \sqrt{12}\right) \\
&= \mathbb{P}(Z < \sqrt{12}) - \mathbb{P}(Z < 0) \\
&= \Phi(\sqrt{12}) - \Phi(0) \\
&= \Phi(\sqrt{12}) - \frac{1}{2}.
\end{aligned}$$

Indeed, we see that in the intermediate steps,

- $2X + Y - 4 \sim N(0, 12)$,
- $\frac{2X+Y-4}{\sqrt{12}} \sim N(0,1)$, and
  (Recall that $\text{Var}\left(\frac{W}{\sqrt{12}}\right) = \left(\frac{1}{\sqrt{12}}\right)^2 \text{Var}(W) = \frac{1}{12}\text{Var}(W) = \frac{1}{12}(12) = 1$ for $W \sim N(0,12)$.)
- The normal distribution is symmetric about its mean, hence, $\Phi(0) = 0.5$. (the area on the left and the right of $z = 0$ is the same, and the sum of areas must be 1, so each side has an area of $\frac{1}{2}$).

Example 56. Suppose $Z \sim N(0,1)$, and denote its cdf by Φ and its pdf by ϕ.

(i) Prove that for any $x \in \mathbb{R}$,
$$\phi'(x) = -x\phi(x).$$

(ii) Hence, show that $\phi'(1) + \phi'(-1) = 0$.

Suggested Solution:

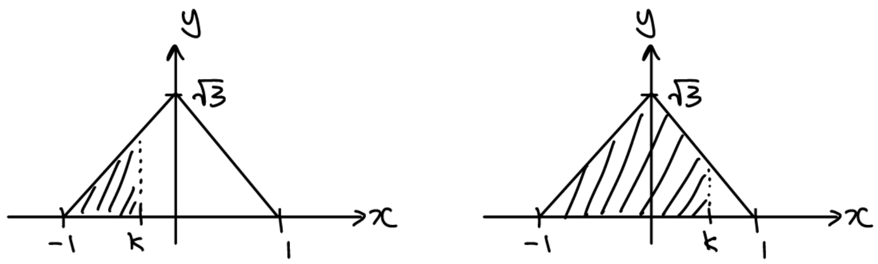
(i) This can be done by direct computation:
$$\phi'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\phi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = -x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = -x\phi(x).$$
Here, we have used the pdf of a normal distribution in the table above, substituting $\mu = 0$ and $\sigma = 1$.

(ii) Using the above formula, we have
$$\phi'(1) + \phi'(-1) = -1\phi(1) + 1\phi(-1) = -\phi(1) + 1\phi(1) = 0.$$
Here, we use the fact that the pdf is symmetric about its mean, and thus $\phi(1) = \phi(-1)$. (In fact, you could have also obtained this by substituting the correct values into the pdf.)

Example 57. Consider an equilateral triangle with coordinates $(-1,0),(1,0)$, and $(0,\sqrt{3})$. Randomly select a point in this triangle, and let X denote the $x-$coordinate of this point. Compute the pdf of X.

Suggested Solution:

Step 1: Compute the cdf of X. Consider $\mathbb{P}(X \leq k)$. Note that this refers to all points in the triangle with x-coordinates less than or equals to k. Hence, this is indicated by the area of the shaded region (see attached image below).



If $-1 \leq k < 0$, then $\mathbb{P}(X \leq k) = \frac{\frac{(1+k)^2}{1^2}\frac{1}{2\sqrt{3}}}{\sqrt{3}} = \frac{1}{2}(1+k)^2$. (Here, we are using the fact that the shaded triangle and the left half triangle with one of the sides lying on the y-axis are similar, so the area of the shaded region to the area of the big left half triangle (which is $1/(2\sqrt{3})$) equals to the square of the ratio of the lengths. Furthermore, for the full probability to be 1, we need to divide by the total area of the triangle, which is at $1/\sqrt{3}$.)

If $k = 0$, then $\mathbb{P}(X \leq k) = \frac{1}{2}$ by symmetry (it is just one half of the full triangle).

If $0 < k \leq 1$, then $\mathbb{P}(X \leq k) = 1 - \frac{(1-k)^2}{1^2}\frac{1}{2}$ (by a similar argument, and using the fact that the area of the shaded region is the area of the whole triangle minus the area of the unshaded triangle).

Hence, we have

$$F_X(k) = \begin{cases} 0 & \text{if } k < -1, \\ \frac{1}{2}(1+k)^2 & \text{if } -1 \leq k \leq 0, \\ 1 - \frac{1}{2}(1-k)^2 & \text{if } 0 < k \leq 1, \\ 1 & \text{if } k > 1. \end{cases}$$

Step 2: Use the fundamental theorem of calculus. From $f_X(k) = \frac{\mathrm{d}}{\mathrm{d}k}F_X(k)$, we have

$$f_X(k) = \begin{cases} 1+k & \text{if } -1 \leq k \leq 0, \\ 1-k & \text{if } 0 < k \leq 1, \\ 0 & \text{otherwise }. \end{cases}$$

11 Finals Revision

Discrete Random Variables:

Distribution	Notation	Parameters	Range	pmf: $p_X(x)$	$\mathbb{E}(X)$	Var(X)	Equivalent Model
Bernoulli	Be(p)	$p \in [0,1]$	$\{0,1\}$	$\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$	p	$p(1-p)$	1 Coin Flip, Heads w.p p.
Binomial	B(n,p)	$n \in \mathbb{N}$ $p \in [0,1]$	$[[0,n]]$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	n Coin Flips, Heads w.p p.
Geometric	Geom(p)	$p \in (0,1]$	$\mathbb{N} = \{1,2,\cdots\}$	$p(1-p)^{x-1}$	$1/p$	$\frac{1-p}{p^2}$	#Coin Flips till first Head, Heads w.p p.
Poisson	Poisson(λ)	$\lambda \geq 0$	$\mathbb{Z}_{\geq 0} = \{0,1,\cdots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	Independent calls in a day, with average of λ calls.

Continuous Random Variables:

Distribution	Notation	Parameters	Range	pdf: $f_X(x)$	cdf: $F_X(x)$	$\mathbb{E}(X)$	Var(X)
Uniform	Unif$[a,b]$	$a,b \in \mathbb{R}$	$[a,b]$	$\begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$	$\begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a,b] \\ 1 & \text{if } x > b. \end{cases}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Normal/ Gaussian	$N(\mu,\sigma^2)$	$\mu \in \mathbb{R}$ $\sigma \geq 0$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ \forall x$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ	σ^2
Exponential	Exp(λ)	$\lambda > 0$	$[0,\infty)$	$\begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$	$\begin{cases} 0 & \text{if } x < 0 \\ 1-e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Note that $\Psi(k) = F_Z(k)$ where $Z \sim N(0,1)$. In other words, it is the cdf of a standard normal distribution (ie of distribution $N(0,1)$).

MATH170E - Selected Revision Problems for Finals.

Note: The following are selected modified problems from the previous homework problems and discussion supplements (usually problems that I did not manage to go through in class for that week). These are solely done for practice purposes and are **not necessarily reflective** of the nature of questions in the final exam.

Exercise 1. Consider an experiment of tossing a fair coin 13 times. Compute the probability of getting 12 heads given that there are at least two heads, simplifying your answer as much as possible.

Exercise 2. Let A and B be two events and assume that they are independent. Suppose that $\mathbb{P}(A) = 0.2, \mathbb{P}(B) = 0.4$, and $\mathbb{P}(A \cup B) = 0.52$. Prove that A and B^c are independent.

Exercise 3. At the beginning of a certain study of a group of persons, 15% were classified as heavy smokers, 30% were classified as light smokers, and 55% were classified as non-smokers. In a five-year study, it was determined that the death rates of heavy and light smokers were five and three times that of non-smokers, respectively. A randomly selected participant died over the five-year period. Calculate the probability that the participant was a light smoker.

Exercise 4. It is claimed that for a particular lottery, $\frac{1}{10}$ of the 50 million tickets will win a prize.

(i) Let X be the number of winning tickets out of the 10 that you have purchased. Explain briefly why we can approximate X by $X \sim B(10, 0.1)$.

(ii) Compute the probability of winning at least two prize if you purchase ten tickets.

Exercise 5. Let X be a random variable representing the number of heads out of 5 independent coin flips, in which the probability of getting a head on each of the flips is p. Find the value of $p \in (0, 1)$ that maximizes the probability of obtaining exactly two heads.

Exercise 6. John is looking to apply for a credit card. He notices that the credit card company offers three different sign-on bonuses online each day he visits the website. Let us denote the bonuses by B_1, B_2, and B_3. The probability of obtaining the bonus B_1 is twice that for B_2, and the probability of obtaining the bonus for B_2 is twice of that of B_3.

(i) Compute the probability of obtaining the bonus B_3 on a randomly given day.

(ii) Suppose that John revisits the website each day starting from 13 March 2023 (inclusive of 13 March). What is the probability that John sees the bonus B_3 for the first time on 20 March 2023 or later?
(That is the date of our finals! Don't be late for it!)

Exercise 7. Let X equal the number of flips of a fair coin that are required to observe the same face on two consecutive flips.

(i) Compute the pmf $p_X(x)$ of X.

(ii) Prove that this is indeed a pmf, that is, $\sum_x p_X(x) = 1$.
Hint: Use the fact that the sum of the pmf of a geometric random variable is 1.

Exercise 8. Let $X \sim \text{Be}(0.8)$, and $Y = 4X - 2$, and $Z = Y^5$.

(i) Compute the pmf of Z.

(ii) Compute $\mathbb{E}(Z)$.

(iii) Compute $\text{Var}(Z)$.

Exercise 9. First, flip a total of 10 fair coins, and denote the number of heads to be X. Then, we roll a $X-$sided fair dice and record the result as Y. Compute $\mathbb{E}(Y)$.

Hint: Recall that for an arithmetic series, we have $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.

Exercise 10. Suppose that $X \sim \text{B}\left(2, \frac{1}{2}\right)$, while Y is the outcome of a fair 4-sided dice roll. Furthermore, suppose that X and Y are independent.

(i) Compute $\mathbb{P}(X + Y < 6)$.

(ii) Compute $\mathbb{E}(X/Y)$.

(iii) Compute $\frac{\mathbb{E}(X)}{\mathbb{E}(Y)}$.

(iv) Explain why the answer in (ii) and (iii) are not the same, even though X and Y are independent.

(v) Compute $\mathbb{E}(XY|Y = 1)$.

Exercise 11. Let X_1, X_2 each follows Poisson(3) and are independent. Let $Y = 2X_1$. Compute

(i) $\mathbb{E}(X_1 + X_2)$,

(ii) $\text{Var}(X_1 + X_2)$,

(iii) $\mathbb{E}(Y)$,

(iv) $\text{Var}(Y)$.

(v) Is Y a Poisson random variable? Explain your answer.

Remark: This exercise shows that the sum of two Poisson distributions could be Poisson, but $2\times$ a Poisson distribution is not necessarily a Poisson distribution.

Exercise 12. Consider two random variables X and Y (which are possibly not independent) with joint probability mass function below:

$X \setminus Y$	1	2
1	$1/8$	$3/8$
2	$1/4$	$1/4$

Compute $\mathbb{E}(X|Y = 1)$.

Exercise 13. Let X be the outcome of a fair $6-$sided dice and $Y \sim \text{Be}\left(\frac{X}{6}\right)$.

(i) Determine the conditional pmf $p_{Y|X}(y|x)$.

(ii) Determine the conditional pmf $p_{X|Y}(x|y)$.

Exercise 14. Two $6-$sided fair dices are rolled together repeatedly. Find the probability that a total of 12 comes up before a total of 11.

Exercise 15. (Gambler Ruin.) Consider a coin-flipping game with two players where each player has a 50% chance of winning with each flip of the coin. After each flip of the coin, the loser transfers one penny to the winner. The game ends when one player has all the pennies. Suppose that player A starts off with 2 pennies, and player B starts off with 1 penny.

Let $f(n)$ be the probability that player A wins if player A has n pennies for $n \in [[0, 3]]$.

(i) Explain why $f(n) = \frac{1}{2}f(n-1) + \frac{1}{2}f(n+1)$ for $n = 1, 2$.

(ii) Explain why $f(0) = 0$ and $f(3) = 1$.

(iii) Using the above parts, find the probability that A wins.

Exercise 16. Let $X \sim N(0, 1)$ and $Y = X^3$. Denote the pdf of X as ϕ and the cdf of X as Φ.

(i) Compute the cdf of Y in terms of ϕ and Φ.

(ii) Compute the pdf of Y in terms of ϕ and Φ.

Exercise 17. Suppose that $X \sim N(1, 2)$, $Y \sim N(2, 7)$, and are independent. Let the cdf of the standard normal random variable $N(0, 1)$ be given by Φ.

Express $\mathbb{P}(\{2 < X + Y < 5\} \cup \{21 < 2(X + Y) + 1 < 33\})$ in terms of Φ.

Exercise 18. Let $X \sim$ Exp(2), and let $Y = \lambda X$ for some $\lambda > 0$. Compute the pdf of Y. What distribution does Y follows?

Exercise 19. Let $X \sim$ Unif$[2, 4]$. Without explicitly referencing the formula for $\mathbb{E}(X)$ or Var(X) (ie please compute these from scratch), compute Var(X).

Answers and Partial Solutions:

- Answer to Exercise 1: $\frac{13}{2^{13}-1-13}$.
- Partial Solution to Exercise 2: Compute $\mathbb{P}(A \cap B)$ and show that it is equals to $\mathbb{P}(A)\,\mathbb{P}(B)$. Use the fact that if $A \perp B$, then $A \perp B^c$.
- Answer to Exercise 3: $\frac{9}{22}$. Hint: See Discussion Supplement 4.
- Since there are a lot of tickets available, we assume that winning/not winning a prize will not change the probability of winning by much. Hence, we can then assume that the probability of winning a prize in each ticket is independent (approximately) of one another.
  Answer to Exercise 4 (ii): $1 - 0.9^{10} - 0.9^{9}$.
- Hint for Exercise 5: See the last example in Discussion Supplement 5.
- Answer to Exercise 6: (i) $\frac{1}{7}$ (recall that the sum of pmf gives 1, ie either of the bonus must appear on each day), (ii) $\left(1 - \frac{1}{7}\right)^7$.
- Hint for Exercise 7: See the last example in Discussion Supplement 6.
  Answer for Exercise 7 (i): $p_X(x) = \frac{1}{2^{x-1}}$ for $x = 2, 3, 4, \cdots$.
  (ii) $\sum_{x=2}^{\infty} \frac{1}{2^{x-1}} = \sum_{x=1}^{\infty} \frac{1}{2^x} = 1$. (Replace x by $x + 1$.)
- Answers for Exercise 8 (i): $p_Z(z) = \begin{cases} 0.8 & \text{if } z = 32 \\ 0.2 & \text{if } z = -32. \end{cases}$ (ii): 19.2. (iii): 655.36.
- Hint for Exercise 9: See the last example in Discussion Supplement 7. Answer for Exercise 9: 3.
- Answers for Exercise 10 (i): $1 - \frac{1}{4} \times \frac{1}{4} = \frac{15}{16}$. (ii): $\frac{25}{48}$. (iii) $\frac{5}{2}$. (iv) $\mathbb{E}(X/Y) = \mathbb{E}(X) \cdot \mathbb{E}(1/Y)$. However, it is not necessarily true that $\mathbb{E}(1/Y) = 1/\,\mathbb{E}(Y)$!! (v): $\mathbb{E}(XY|Y = 1) = \mathbb{E}(X \cdot 1|Y = 1) = \mathbb{E}(X|Y = 1) = \mathbb{E}(X)$ $(X \perp Y)$ and hence is equals to $\mathbb{E}(X) = 1$.
- Answers for Exercise 11 (i): 6, (ii): 6, (iii): 6, (iv): 12. (v): Since $\mathbb{E}(Y) \neq \mathrm{Var}(Y)$, then Y cannot be a Poisson random variable (since all Poisson distributions have the same expectation and variance).
- Answer for Exercise 12 $\frac{5}{3}$.
- Answer for Exercise 13 (i) $p_{Y|X}(y|x) = \begin{cases} \frac{x}{6} & \text{if } y = 1, \\ 1 - \frac{x}{6} & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$
  (ii) $p_{X|Y}(x|y) = \begin{cases} \frac{x/36}{21/36} = \frac{x}{21} & \text{if } y = 1, \text{ and } x \in [[1, 6]], \\ \frac{(6-x)/36}{15/36} = \frac{6-x}{15} & \text{if } y = 0, \text{ and } x \in [[1, 6]], \\ 0 & \text{otherwise.} \end{cases}$
  (Refer to the first example in Discussion Supplement 9.)
- Hint for Exercise 14: Refer to the second example in Discussion Supplement 9.
  Answer for Exercise 14: $\frac{1}{3}$.
- Hint for Exercise 15: Refer to the fourth example in Discussion Supplement 9.
  Answer for Exercise 15 (iii): $\frac{2}{3}$.
- Answer for Exercise 16 (i): $F_Y(y) = \Phi\left(y^{\frac{1}{3}}\right)$ (ii): $f_Y(y) = \frac{1}{3} y^{-\frac{2}{3}} \phi\left(y^{\frac{1}{3}}\right)$.
- Answer for Exercise 17: $\Phi\left(\frac{13}{3}\right) + \Phi\left(\frac{2}{3}\right) - \Phi\left(\frac{7}{3}\right) - \Phi\left(-\frac{1}{3}\right)$.
- Answer for Exercise 18: $f_Y(y) = \begin{cases} e^{-(2/\lambda)x} & \text{if } y \geq 0, \\ 0 & \text{if } y < 0. \end{cases}$ $Y \sim \mathrm{Exp}(2/\lambda)$.
- Answer for Exercise 19: $\frac{1}{3}$.

References

[1] Dale Zimmerman Robert Hogg, Elliot Tanis. *Probability and Statistical Inference*. Pearson, 10th edition, 2023.