Lecture Notes for MATH 380

1. ALGEBRA OF SETS

Definition. A **random phenomenon** is a phenomenon that has an unpredicted outcome on every try but the overall distribution of all possible outcomes is well-defined. For example, flipping a coin is a random phenomenon because we don't know what comes up on the next flip, heads or tails, but we know that if it is a fair coin, then heads should come up roughly 50% of the time.

Definition. The **sample space** is the set of all possible outcomes of a random phenomenon. It is denoted by S.

Examples. (1) Flipping a coin once. The sample space $S = \{H, T\}$.
(2) Flipping a coin twice. The sample space $S = \{HH, HT, TH, TT\}$.
(3) Flipping a coin until one head or three tails appear. The sample space is $S = \{H, TH, TTH, TTT\}$.
(4) Rolling a die. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$.
(5) Rolling a die twice (or rolling two dice). The sample space is $S = \{(1,1), (1,2), \dots, (6,5), (6,6)\}$.

Definition. An **event** is a subset of the sample space. Events are denoted by capital Latin letters from the beginning of the alphabet. For example, A, B, C, D, A_1, A_2, etc.

Example. A coin is flipped twice. Define event A as *at least one head appears*. We can write $A = \{HH, HT, TH\}$.

Example. Two dice are rolled. List all outcomes in the event that the sum of two rolls is equal to 8. We write $B = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$.

Definition. An **complement** of event A, denoted A^c or $\bar{A}$, is the collection of all outcomes that are in the sample space S but not in A.

Example. A coin is flipped twice. The complement of A=*at least one head appears* is $\bar{A}$=*no heads appear*$= \{TT\}$.
Definition. An **intersection** of two events A and B, written $A \cap B$, is the set of all outcomes that are in both A and B.
Definition. A **union** of events A and B, denoted by $A \cup B$, is the set of all outcomes that are either in A or in B or in both.

Note. A good way to remember the notation is to notice that $\cup$ resembles the letter U (union).

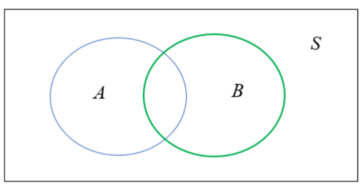
Example. Consider a standard pocket deck of cards, with 52 cards, 4 suits (hearts and diamonds are red suits, and club and spades are black suits), 13 cards in each suit

(2 through 10, Jack, Queen, King, and Ace). Suppose one card is randomly drawn from this deck of cards. The events $A =$ *an ace is drawn* $= \{A\heartsuit, A\diamondsuit, A\clubsuit, A\spadesuit\}$, and $B =$ *a black card is drawn* $= \{2\clubsuit, \dots, A\clubsuit, 2\spadesuit, \dots, A\spadesuit\}$. The intersection $A \cap B = \{A\clubsuit, A\spadesuit\}$, and the union

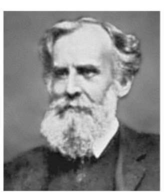
$$A \cup B = \{2\clubsuit, \dots, K\clubsuit, 2\spadesuit, \dots, K\spadesuit, 2\clubsuit, \dots, A\clubsuit, 2\spadesuit, \dots, A\spadesuit\}.$$

Note. To list all outcomes in an intersection of $A \cap B$, go through the list of the outcomes in A and keep only those that are also in B. To list all outcomes in the union $A \cup B$, list all outcomes in A and then go through the list of outcomes in B and list only those that were not already listed, to avoid listing the intersection twice.

Definition. A **Venn diagram** is a visual tool that helps depict events relative to each other. The sample space is drawn as a rectangle, inside which events are drawn as circles. The most general Venn diagram for two sets looks like this:

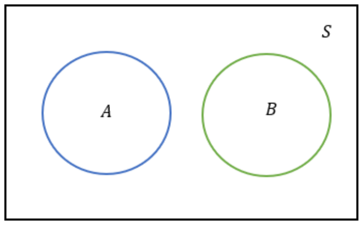


Historical Note. John Venn (1834 –1923) was a British logician and philosopher, who introduced the diagram in 1880.



Definition. A **null** event (or an **empty** event) is and event that contains no outcomes. The notation is $\emptyset$.

Definition. Two events are **disjoint** (or **mutually exclusive**) if their intersection is empty, that is, $A \cap B = \emptyset$. On a Venn diagram, disjoint events are depicted as non-overlapping circles.

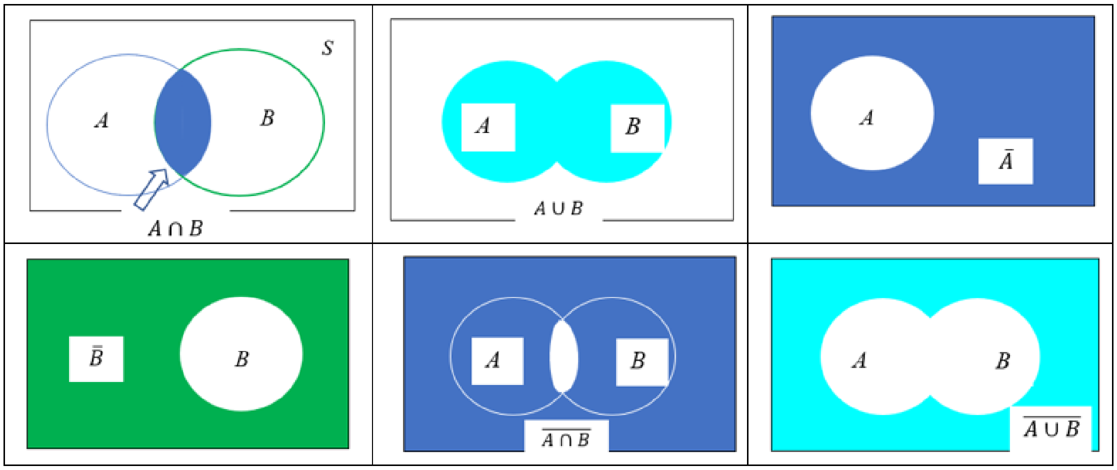


Exercise 1.1. $A \cap \emptyset = \emptyset$, $A \cup A^c = S$, $A \cup \emptyset = A$, $A \cap S = A$, $A \cap A^c = \emptyset$, $A \cup S = S$.

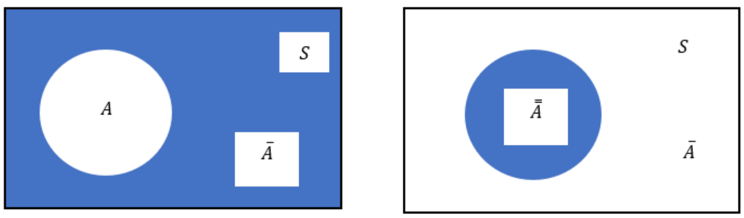
Exercise 1.2. An engineering firm is hired to determine if certain waterways in Virginia are safe for fishing. Samples are taken from three rivers, and the rivers are classified by letters F for "safe for fishing" or N for "not safe for fishing."
(a) Describe in words an event $A = \{FFF, NFF, FFN, NFN\}$. Answer: $A =$the second river is safe for fishing.
(b) Find $A \cap B$, if $B =$ at least one river is safe for fishing. Answer: $A \cap B = A = \{FFF, NFF, FFN, NFN\}$.

Exercise 1.3. Which of the following pairs of events are mutually exclusive?
(a) A golfer scoring the lowest 18-hole round in a 72-hole tournament and losing the tournament. Answer: Not mutually exclusive, two events can happen at the same time.
(b) A poker player getting a flush (all cards in the same suit) and 3 of a kind on the same 5-card hand. Answer: Mutually exclusive, cannot happen at the same time.
(c) A mother giving birth to a baby girl and a set of twin daughters on the same day. Answer: Not mutually exclusive, can happen at the same time.
(d) A chess player losing the last game and winning the match. Answer: Not mutually exclusive, can happen at the same time.

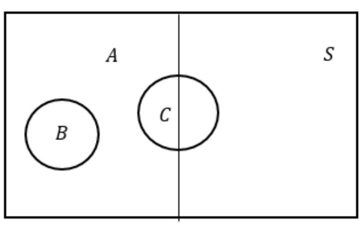
Exercise 1.4. Draw a Venn diagram for two general events A and B and show where on the diagram are $A \cap B$, $A \cup B$, $\bar{A}$, $\bar{B}$, $\overline{A \cap B}$, and $\overline{A \cup B}$.



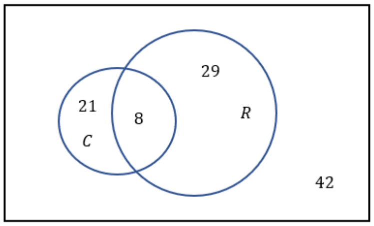
Exercise 1.5. Use Venn diagram to prove $(A^c)^c = A$.



Exercise 1.6. Draw a card at random from a deck of cards. Define events $A =$ the card is red, $B = \{J\diamondsuit,\ Q\heartsuit,\ K\diamondsuit\}$, $C =$ the card is an ace. Draw a Venn diagram for these three events.



Exercise 1.7. One hundred horses were made to listen to classical and rock music. Twenty-nine horses exhibited some head movements when classical music was played, 37 when rock music was played, and 8 when both were played. How many horses exhibited head movements to



(a) at least one type of music? $|C \cup R| = 21 + 8 + 29 = 58$.

(b) to classical music only? $|C \cap \bar{R}| = 21$.

(c) to both types of music? $|C \cap R| = 8$.

(d) only one type of music? $|(C \cap \bar{R}) \cup (\bar{C} \cap R)| = 21 + 29 = 50$.

(e) neither music? $|\overline{C \cup R}| = 42$.

Exercise 1.8. Which of the following statements are correct?

(a) $2 \in \{1, 2, 3\}$. Answer: Correct, 2 is an element of the set.
(b) $2 \subset \{1, 2, 3\}$. Answer: Wrong, an element is not a subset of a set.
(c) $\{2\} \in \{1, 2, 3\}$. Answer: Wrong, a set is not an element of a set.
(d) $\{2\} \subset \{1, 2, 3\}$. Answer: Correct, a set is a subset of a bigger set.

Exercise 1.9. Three students are selected at random from a chemistry class and classified as male or female.
(a) List the elements of the sample space S using the letter M for "male" and F for "female." Answer: $S = \{MMM, MMF, MFM, FMM, FFM, FMF, MFF, FFF\}$.
(b) Define a second sample space S_1 where the elements represent the number of females selected. Answer: $S_1 = \{0, 1, 2, 3\}$.

Exercise 1.10. A fair coin is flipped until two tails or three heads appear. Write down the sample space. Answer: $S = \{TT, THT, THHT, THHH, HTT, HHTT, HTHT, HHTH, HHH, HTHH\}$.

Exercise 1.11. In each of the following cases, describe a sample space S for the indicated random phenomenon.

(a) A patient with a usually fatal form of cancer is given a new treatment. The response variable is the length of time that the patient lives after treatment. Answer: $S = \{T \geq 0\}$.
(b) A student enrolls in a probability course and at the end of the semester receives a letter grade. Answer: $S = \{A, B, C, D, F\}$.

(c) A basketball player shoots two free throws. Answer: $S = \{HH, HM, MH, MM\}$ where H=hit, M=miss.

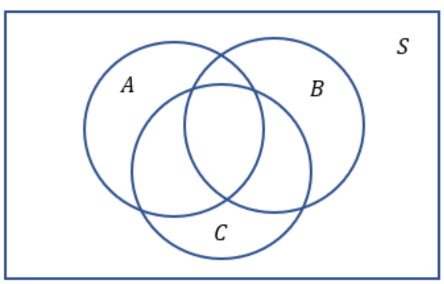
(d) A year after knee surgery, a patient is asked to rate the amount of pain in the knee. A seven-point scale is used, with 1 corresponding to no pain and 7 corresponding to extreme discomfort. Answer: $S = \{1, 2, 3, 4, 5, 6, 7\}$.

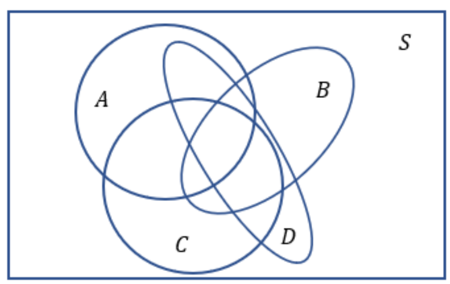
(e) Choose a student in your class at random. Ask how many hours that student spent studying during the past 24 hours. Answer: $S = \{0, 1, 2, \dots, 24\}$.

(f) In a test of a new package design, a carton of a dozen eggs is dropped from a height of 1 foot. The number of broken eggs is counted. Answer: $S = \{0, 1, \dots, 12\}$.

(g) A nutrition researcher feeds a new diet to a young male white rat. The response variable is the weight (in grams) that the rat gains in 8 weeks. Answer: $S = \{0, 1, 2, \dots\}$.

Exercise 1.12. Draw a general Venn diagram for three events. Can you draw one for four events?





2. Probability: Axiomatic Definition

Overly simplifying, we can say that a **probability** is a numerical measure of how likely it is that a specific event will occur. It is a number between 0 and 1, inclusively. We can think of a probability as:

• **Long-term proportion**. For example, $\mathbb{P}(H) = 0.5$ means that when we flip a coin many times, roughly half will be heads.
• **Fraction**. For example, there are 5 red and 3 green balls, we choose one at random, and the probability of choosing a red ball is $5/(5 + 3) = 5/8$.
• **Personal belief**. For example, you might be estimating that the probability that you will pass a final exam in this course is 0.8.

Now we give a formal axiomatic definition of probability.

Definition. Let $A_1, A_2, \ldots,$ be an infinite set of disjoint events, that is, $A_i \cap A_j = \emptyset$ for any $i \neq j$. A **probability** $\mathbb{P}$ of an event A is defined by the following two axioms: (1) $0 \leq \mathbb{P}(A) \leq 1$, $\mathbb{P}(\emptyset) = 0$, and $\mathbb{P}(S) = 1$, and (2) $\mathbb{P}\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Example. Suppose we have a finite set of events $A_1, A_2, \ldots, A_n$. We need to show that $\mathbb{P}\big(\bigcup_{i=1}^{n} A_i\big) = \sum_{i=1}^{n} \mathbb{P}(A_i)$. To this end, we can supplement our finite set with infinitely many empty events and use the second axiom in the definition of probability. We write $\mathbb{P}\big(\bigcup_{i=1}^{n} A_i\big) = \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup \emptyset \cup \emptyset \cup \ldots) = \{\text{by the axiom}\} = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \cdots = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n)$, since the probability of an empty set is equal to zero.

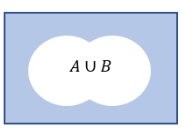
Complement Rule. We will show that $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$. Indeed, we can notice that A and $\bar{A}$ constitute a finite partition of S, so using the result of the previous exercise, we can write $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = \mathbb{P}(A \cup \bar{A}) = \mathbb{P}(S)$, so $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

Additive Rule. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
Proof: Let's assume that probabilities are equivalent to areas on a Venn diagram. Then the area of $A \cup B$ is the same as the area of A plus the area of B, but the area of the intersection $A \cap B$ is counted twice, so we have to subtract it off ones.

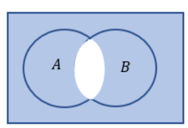
Note. Given any three of the four quantities in the additive rule, we can find the fourth.

De Morgan's Laws. For any events A and B, the following two rules (called **De Morgan's Laws**) always hold: $\mathbb{P}(\overline{A \cup B}) = \mathbb{P}(\bar{A} \cap \bar{B})$, and $\mathbb{P}(\overline{A \cap B}) = \mathbb{P}(\bar{A} \cup \bar{B})$.
Proof: The quickest way to prove the laws is through Venn diagrams.

Here we depict the left-hand sides of De Morgan's Laws.

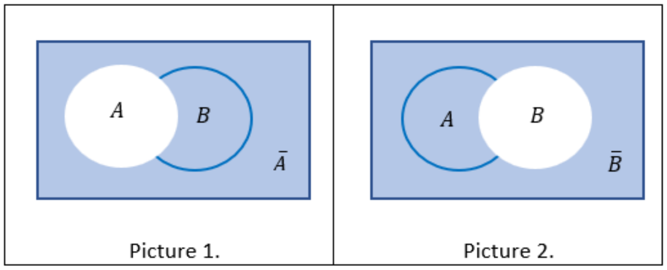


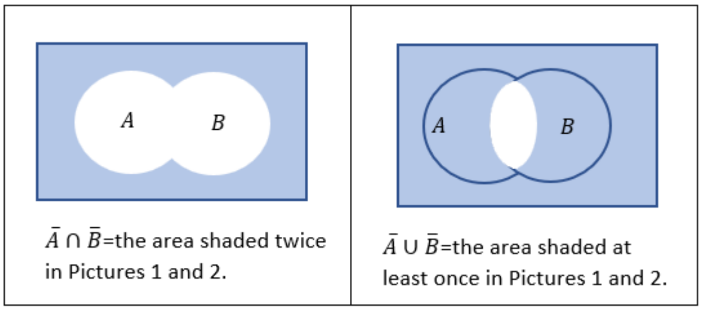
$\overline{A \cup B}$



$\overline{A \cap B}$

And here we draw the Venn diagrams for the right-hand sides.





We can see that the areas are the same for both sides of the respective identities.

Note. The laws become very intuitive if described in simple words. In the first law, the left-hand side says that neither event happens, and the right-hand side states the same by saying that A doesn't happen and B doesn't happen. Likewise, in the second law, the left-hand side says that A and B don't happen at the same time, whereas the right-hand side restates it by saying that either A doesn't happen or B doesn't happen.

Historical Note. Augustus De Morgan (1806-1871) was a British mathematician and logician.

Note. The way to remember De Morgan's Laws is as follows: the bar on top over the union of A and B distributes over each event, and the union symbol $\cup$ is turned upside down to become an intersection $\cap$. Likewise, in the second law, the bar above the intersection of A and B goes over each event individually and the intersection symbol $\cap$ turns over to become the union $\cup$.

Exercise 2.1. Match one of the probabilities that follow with each statement given. The list of probabilities is: 0, 0.01, 0.3, 0.54, 0.99, and 1.

(a) This event is impossible. It can never occur. Answer: 0.
(b) This event is certain; it will occur on every trial of the random phenomenon. Answer: 1.
(c) This event is very unlikely, but it will occur once in a while in a long sequence of trials. Answer: 0.01.
(d) This event is very probable but still fails to occur once in a while. Answer: 0.99.
(e) This event will occur more often than not. Answer: 0.54.

Exercise 2.2. A major credit card company is interested in the main use of credit cards by people in various income brackets. The data in the table below are obtained from 2,000 randomly selected cardholders.

Annual Income	Use of the card: Clothes	Food	Travel	Other
over $100,000	50	10	100	40
$50,000 to $100,000	70	30	300	100
$25,000 to $49,000	400	400	100	100
below $25,000	50	200	30	20

(a) What is the probability that a randomly chosen cardholder has an annual income between $25,000 and $49,000? Answer: $(400+400+100+100)/2000 = 1/2$.
(b) What is the probability that he uses his credit card for buying food and he earns below $25,000 each year? Answer: $200/2000 = 0.1$.
(c) What is the probability that his annual income earns over $100,000 or that he uses his card for travel? Answer: $(50+10+100+40+300+100+30)/2000 = 630/2000 = 0.315$.

Exercise 2.3. $\mathbb{P}(A) = 1/4$, $\mathbb{P}(B) = 1/2$, $\mathbb{P}(A \cup B) = 5/8$. Write using proper mathematical notion and compute:
(a) $\mathbb{P}(A \text{ or } B)$. Answer: $\mathbb{P}(A \cup B) = 5/8$.
(b) $\mathbb{P}(A \text{ and } B)$. Answer: By the additive rule, $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = 1/4 + 1/2 - 5/8 = 2/8 + 4/8 - 5/8 = 1/8$.
(c) $\mathbb{P}(\text{only } A \text{ happens})$. Answer: $\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = 1/4 - 1/8 = 1/8$.
(d) $\mathbb{P}(\text{exactly one event happens})$. Answer: $\mathbb{P}\big((A \cap \bar{B}) \cup (\bar{A} \cap B)\big) = \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B) = \{\text{by the additive rule}\} = \mathbb{P}(A \cup B) - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)) = 5/8 - 1/8 = 4/8 = 1/2$.
(e) $\mathbb{P}(A \text{ or } B \text{ but not both})$. Answer: Same event as in part (d).
(f) $\mathbb{P}(A \text{ and } B \text{ do not happen simultaneously})$. Answer: $\mathbb{P}(\overline{A \cap B}) = 1 - \mathbb{P}(A \cap B) = 1 - 1/8 = 7/8$.
(g) $\mathbb{P}(\text{neither event happens})$. Answer: $\mathbb{P}(\overline{A \cup B}) = 1 - \mathbb{P}(A \cup B) = 1 - 5/8 = 3/8$.

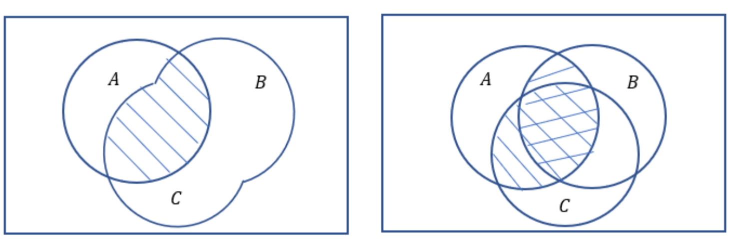
Exercise 2.4. Write down the additive rule for three events. Answer: Think of the probability of an event as an area on a Venn diagram. Consider the area of the union of three events $A \cup B \cup C$. If we add the areas of A, B, and C, we double-count the intersections $A \cap B$, $A \cap C$, and $B \cap C$, so we need to subtract them once. Focusing now on the area of the intersection of all three events $A \cap B \cap C$, we see that we added it three times and then subtracted it three times, so we need to add it back on once. Therefore, the additive rule for three events looks like this:

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

Note. The additive rule can be extended to any number of events. We need to add probabilities of odd-folds (1, 3, 5, etc.), and subtract probabilities of all even-folds (2, 4, 6, etc.).

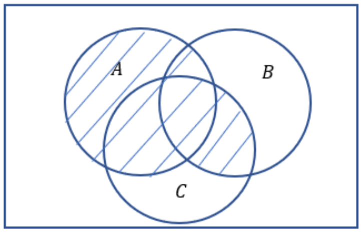
Exercise 2.5. It is given that $\mathbb{P}(A) = 0.55$ and $\mathbb{P}(B) = 0.72$.
(a) Can A and B be disjoint? Answer: no, A and B cannot be disjoint. We prove this by contradiction. Suppose they are disjoint, then $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ and by the additive rule, we must have $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.55 + 0.72 - 0 = 1.27$ which is larger than 1. It is impossible, so A and B cannot be disjoint.
(b) What is the smallest possible value of the probability of their intersection? Answer: The probability of the intersection would be the smallest if the events fill up the entire sample space, that is, $\mathbb{P}(A \cup B) = \mathbb{P}(S) = 1$, and thus, by the additive rule, $\min \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = 0.55 + 0.72 - 1 = 0.27$.
(c) What is the largest possible value of the probability of their intersection? Answer: The probability of the intersection would be the largest if one of the events is entirely contained within the other. Since the probability of A is smaller than the probability of B, A should be the one contained within B. So, $\max \mathbb{P}(A \cap B) = \mathbb{P}(A) = 0.55$.

Exercise 2.6. Show that the distributive law holds for unions and intersections.
(a) Show that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

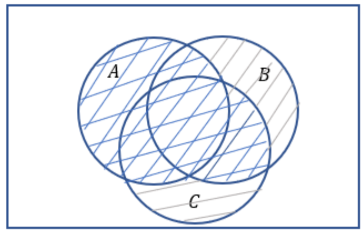


$A \cap (B \cup C)$ $\quad$ $(A \cap B) \cup (A \cap C)$

(b) Show that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.



$A \cup (B \cap C)$



$(A \cup B) \cap (A \cup C)$

Exercise 2.7. Express the following probabilities in terms of the probabilities of A, B, and the intersection $A \cap B$.
(a) $\mathbb{P}(A^c \cup B^c)$. Answer: By De Morgan's Law, $\mathbb{P}(A^c \cup B^c) = \mathbb{P}\big((A \cap B)^c\big) = 1 - \mathbb{P}(A \cap B)$.
(b) $\mathbb{P}(A^c \cap B^c)$. Answer: By De Morgan's Law, $\mathbb{P}(A^c \cap B^c) = \mathbb{P}\big((A \cup B)^c\big) = 1 - \mathbb{P}(A \cup B) = \{\text{by the additive rule}\} = 1 - \big(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\big) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B)$.
(c) $\mathbb{P}(A^c \cap (A \cup B))$. Answer: Put in simple words, we want to compute the probability that A didn't happen but at the same time the union of A and B happened. It means that only B happened. Formally, we write $\mathbb{P}(A^c \cap (A \cup B)) = \mathbb{P}\big((A^c \cap A) \cup (A^c \cap B)\big) = \mathbb{P}(\emptyset \cup (A^c \cap B)\big) = \mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

3. Conditional Probability

Definition. The **conditional** probability of event A given that an event B has happened, denoted by $\mathbb{P}(A|B)$, is defined by the formula

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

Note. $\mathbb{P}(A|B)$ is pronounced as "the conditional probability of A given B. By definition, the conditional probability is the probability of both events over the probability of the given event.

Note. The above definition is very intuitive. Think of a Venn diagram for two events A and B, and assume that areas are probabilities. Since we know that B has happened, we confine our attention to the circle that corresponds to event B. How can A also happen? Only if we are in the $A \cap B$, the intersection of A and B. What fraction of the area of circle B is the area of $A \cap B$? This is by definition the conditional probability of A given B.

Example. Sixty percent of students in a high school are females. Of those students who take a math class, 40% are females. Find the probability that a student is taking a math class if we know that the student is female.

Let F be the event that the student is female, and M that the student is taking a math class.
We need to find the conditional probability of M given F. We write

$$\mathbb{P}(M|F) = \frac{\mathbb{P}(M \cap F)}{\mathbb{P}(F)} = \frac{0.4}{0.6} = 2/3 = 0.67.$$

Remark. Depending on the setting, sometimes it is easier to compute conditional probabilities by the method of **reduction of the sample space**.

Example. A sack contains small and large marbles of two colors, red and blue. The frequency distribution is given in the table below.

	Red	Blue
Small	8	6
Large	10	14

Suppose we want to compute the conditional probability of drawing a large marble given that it is blue. We can reduce the sample space to the blue marbles only. There are a total of 6+14=20, of which 10 are large. Therefore, the conditional probability that the chosen marble is large given that it is blue is 14/20=0.7.

For comparison, we can compute the conditional distribution using the formula in the definition. We obtain

$$\mathbb{P}(\text{large}|\text{blue}) = \frac{\mathbb{P}(\text{large and blue})}{\mathbb{P}(\text{blue})} = \frac{14/(8+6+10+14)}{(6+14)/(8+6+10+14)}$$
$$= \frac{14/38}{20/38} = 14/20 = 0.7.$$

Notice that when using the formula in the definition, we divide the top and bottom by the grand total, which can be canceled. The reduction of the sample space method allows us to avoid computing the grand total.

Further, we can rewrite the formula in the definition as

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

This expression is often referred to as **unconditional** probability. It allows to compute probabilities **by conditioning**.

Example. In a city, 30% of the population are senior citizens. Eighty percent of senior citizens are subscribed to HBO. Forty-five percent of non-senior citizens are subscribed to HBO. We want to compute the probability that a randomly chosen person in this city is subscribed to HBO. We condition on whether the person is a senior citizen or not. We write

$$\begin{aligned}\mathbb{P}(\text{subscribed}) &= \mathbb{P}(\text{subscribed and senior}) + \mathbb{P}(\text{subscribed and not senior}) \\ &= \mathbb{P}(\text{subscribed} \mid \text{senior})\mathbb{P}(\text{senior}) + \mathbb{P}(\text{subscribed} \mid \text{not senior})\mathbb{P}(\text{not senior}) \\ &= (0.8)(0.3) + (0.45)(1-0.3) = 0.555.\end{aligned}$$

Exercise 3.1. Draw a marble from a box containing 3 green, 1 white, and 5 black marbles. If the drawn marble is not white, find the probability that it is green. Do the calculations in two ways:
(a) by definition of conditional probability. Answer:

$$\mathbb{P}(\text{green} \mid \text{not white}) = \frac{\mathbb{P}(\text{green and not white})}{\mathbb{P}(\text{not white})} = \frac{\mathbb{P}(\text{green})}{\mathbb{P}(\text{not white})}$$

$$= \frac{\mathbb{P}(\text{green}}{1-\mathbb{P}(\text{white})} = \frac{3/(3+1+5)}{1-1/(3+1+5)} = \frac{3/9}{8/9} = 3/8.$$

(b) by reduction of the sample space. Answer: We reduce the sample space to 8 non-white marbles, of which 3 are green. Therefore, $\mathbb{P}(\text{green} \mid \text{not white}) = 3/8$.

Exercise 3.2. A random sample of 200 adults are classified by gender and their level of education attained.

Education	**Male**	**Female**
Elementary	38	45
Secondary	28	50
College	22	17

If a person is picked at random from this group, find the probability that

(a) the person is a male, given that the person has a secondary education. Answer: We reduce the sample space to $28 + 50 = 78$ adults with secondary education, of which 28 are males. So, $\mathbb{P}(\text{male} \mid \text{secondary education}) = 28/78 = 14/39 = 0.359$.

(b) the person does not have a college degree, given that the person is a female. Answer: We reduce the sample space to $45+50+17 = 112$ females, of which $45+50 = 95$ don't have a college degree. Thus, $\mathbb{P}(\text{does not have a college degree} \mid \text{female}) = 95/117 = 0.812$.

Exercise 3.3. The probability that a doctor correctly diagnoses a particular illness is 0.7. Given that the doctor makes an incorrect diagnosis, the probability that the patient files a lawsuit is 0.9. What is the probability that the doctor makes an incorrect diagnosis and the patient files a lawsuit? Answer: We are given that $\mathbb{P}(\text{correct diagnosis}) = 0.7$, and $\mathbb{P}(\text{lawsuite} \mid \text{wrong diagnosis}) = 0.9$. we need to find the probability of intersection $\mathbb{P}(\text{wrong diagnosis and lawsuit})$. By conditioning on a wrong diagnosis, we write

$$\begin{aligned}\mathbb{P}(\text{wrong diagnosis and lawsuit}) &= \mathbb{P}(\text{lawsuit} \mid \text{wrong diagnosis})\,\mathbb{P}(\text{wrong diagnosis}) \\ &= (0.9)(1-0.7) = 0.27.\end{aligned}$$

Exercise 3.4. The probability that a vehicle entering the Luray Caverns has a Canadian license plate is 0.12, the probability that it is a camper is 0.28, and the probability that it is a camper with a Canadian license plate is 0.09. What is the probability that

(a) a camper entering the Luray Caverns has a Canadian license plate? Answer:

$$\mathbb{P}(\text{Canadian plate} \mid \text{camper}) = \frac{\mathbb{P}(\text{camper with Canadian plate})}{\mathbb{P}(\text{camper})} = \frac{0.09}{0.28} = 0.321.$$

(b) a vehicle with a Canadian license plate entering the Luray Caverns is a camper? Answer:

$$\mathbb{P}(\text{camper} \mid \text{Canadian plate}) = \frac{\mathbb{P}(\text{camper with Canadian plate})}{\mathbb{P}(\text{Canadian plate})} = \frac{0.09}{0.12} = 0.75.$$

(c) a vehicle entering the Luray Caverns does not have a Canadian plate or is not a camper? Answer: $\mathbb{P}\big(\overline{\text{Canadian plate}} \text{ or } \overline{\text{camper}}\big) = \{\text{by De Morgan's Law}\} = \mathbb{P}\big(\overline{\text{Canadian plate and camper}}\big) = 1 - \mathbb{P}(\text{Canadian plate and camper}) = 1 - 0.09 = 0.91$.

Exercise 3.5. There are r red marbles, and b blue marbles in a box. We choose one at a time **without replacement** (that is, we don't return the marble into the box after drawing it). Show that the probability of choosing a red marble on the second draw is $r/(r+b)$, the same as on the first draw. Answer: Conditioning on the color of the first marble, we write

$$\begin{aligned}\mathbb{P}(\text{second red}) &= \mathbb{P}(\text{second red and first red}) + \mathbb{P}(\text{second red and first blue}) \\ &= \mathbb{P}(\text{second red} \mid \text{first red})\mathbb{P}(\text{firstred}) + \mathbb{P}(\text{second red} \mid \text{first blue})\mathbb{P}(\text{first blue})\end{aligned}$$

$$= \frac{r-1}{r-1+b} \cdot \frac{r}{r+b} + \frac{r}{r+b-1} \cdot \frac{b}{r+b} = \frac{r}{(r+b-1)(r+b)}(r-1+b) = \frac{r}{r+b}.$$

4. Independence

Definition. Events A and B are said to be **independent** the knowledge of whether B occurred doesn't change the probability of A. Put mathematically, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Remark. This is a very intuitive definition of independence. It basically means that A and B have nothing to do with each other.

Another definition of independence is often used in practice because it doesn't require computing the conditional probability. The second definition is not intuitive at all, but the two definitions are equivalent.

Second Definition. Events A and B are said to be **independent** if the probability of their intersection is equal to the product of individual probabilities, that is if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Next, we should that the two given definitions are equivalent. Suppose the first definition holds. Then, conditioning on B, we can write

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) \overset{1st\ def}{=} \mathbb{P}(A)\mathbb{P}(B),$$

which means that the second definition holds. Now assume that the second definition holds. Then we have

$$\mathbb{P}(A \mid B) \overset{def}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \overset{2nd\ def}{=} \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A),$$

implying that the first definition holds.

Remark. The statements in the above definitions work in both directions. That is, A and B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ (or $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$). It means that if we need to show that two events are independent, we need to prove that either probability identity holds. If the identities don't hold, then the events are **not independent** (or **dependent**). On the other hand, if we are told that the two events are independent, then we can use the identities for computations.

Example. A single card is drawn at random from a standard deck of 52 cards. Are the events $A = \{\text{an ace is drawn}\}$ and $B = \{\text{a black card is drawn}\}$ independent?
Solution: (1) Let's use the first definition of independence first. We need to check whether $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Assume that B has happened. We can reduce the sample space to the 26 black cards, or which 2 are aces, thus we compute $\mathbb{P}(A \mid B) = 2/26 = 1/13$. On the other hand, there are 4 aces among the 52 cards, so $\mathbb{P}(A) = 4/52 = 1/13$. Since the two probabilities are equal, the identity holds and so the two events are independent.
(2) Now we use the second definition to show independence. We have $\mathbb{P}(A \cap B) =$

$2/52 = 1/26$ since there are two black aces among the 52 cards. Individual probabilities are calculated as $\mathbb{P}(A) = 4/52 = 1/13$ and $\mathbb{P}(B) = 26/52 = 1/2$. Since $1/26 = (1/13)(1/2)$, the identity holds and so the events are independent.

Example. A single card is drawn at random from a standard deck of 52 cards. Are the events $J = \{$ a Jack is drawn$\}$ and $F = \{$ a face card is drawn$\}$ independent? Note that a **face card** is any Jack, Queen, or King.
Solution: (1) Suppose F is the given event. We can reduce the sample space to the 12 face cards, of which 4 are Jacks. Hence, $\mathbb{P}(J \mid F) = 4/12 = 1/3$. The original probability of drawing a Jack is 4/52=1/13. Since the two probabilities are not equal, the events are not independent.
(2) Alternatively, we can use the second definition of independence and show that the second identity doesn't hold. To this end, we note that the event J is completely included in F, and so their intersection is J. Thus, $\mathbb{P}(J \cap F) = \mathbb{P}(J) \neq \mathbb{P}(J)\mathbb{P}(F)$. And so, the events are not independent.

Remark. The advantage of the second definition is that it can be easily generalized to more than two events. However, the statement becomes one-directional in this case. For example, if events A, B, and C are independent, then $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. More generally, if events $A_1, \dots, A_n$ are independent, then $\mathbb{P}\big(\cap_{i=1}^n A_i\big) = \prod_{i=1}^n \mathbb{P}(A_i)$. The converse is not always true. To show that three or more events are independent, we would need to show that all pairs, three-folds, etc. are independent. For instance, to show that A, B, and C are independent, we need to demonstrate that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Example. A building has three smoke detectors which act independently of each other. Each smoke detector fails to detect smoke with a probability of 0.01. What is the probability that smoke will be detected in this building?

Solution: Smoke will be detected in the building if at least one smoke detector detects smoke. The complement of this event is that all three smoke detectors fail to detect smoke. Since smoke detectors are independent, the probability that all three fail can be found as the product of probabilities that each of the detectors fails: $\mathbb{P}($ 1st fails, 2nd fails, 3rd fails$) = \mathbb{P}($1st fails$)\mathbb{P}($2nd fails$)\mathbb{P}($3rd fails$) = (0.01)^3 = 0.000001$. The probability that a fire is detected is then $1 - 0.000001 = 0.999999$.

Example. A coin is flipped two times. The sample space is $S = \{HH, HT, TH, TT\}$. Suppose $\mathbb{P}(H) = \mathbb{P}(T) = 0.5$, meaning the coin is **fair**. To assign probabilities to the outcome, we assume that the flips are independent, so we have $\mathbb{P}(HH) = \mathbb{P}(H)\mathbb{P}(H) = (0.5)(0.5) = 0.25$, $\mathbb{P}(HT) = \mathbb{P}(H)\mathbb{P}(T) = (0.5)(0.5) = 0.25$, $\mathbb{P}(TH) = \mathbb{P}(T)\mathbb{P}(H) = (0.5)(0.5) = 0.25$, and $\mathbb{P}(TT) = \mathbb{P}(T)\mathbb{P}(T) = (0.5)(0.5) = 0.25$. Note that these four probabilities must sum up to one. More generally, let's assume it is a **biased** coin with $\mathbb{P}(H) = 0.8$ and $\mathbb{P}(T) = 0.2$. Using independence of flips, we assign probabilities as follows: $\mathbb{P}(HH) = \mathbb{P}(H)\mathbb{P}(H) = (0.8)(0.8) = 0.64$, $\mathbb{P}(HT) =$

$\mathbb{P}(H)\mathbb{P}(T) = (0.8)(0.2) = 0.16$, $\mathbb{P}(TH) = \mathbb{P}(T)\mathbb{P}(H) = (0.2)(0.8) = 0.16$, and $\mathbb{P}(TT) = \mathbb{P}(T)\mathbb{P}(T) = (0.2)(0.2) = 0.04$. Again, these probabilities do sum up to one.

Proposition. Suppose A and B are independent events. Then (1) A and $\bar{B}$ are independent, and (2) $\bar{A}$ and $\bar{B}$ are independent.

Proof: (1) $\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \{$ by independence of A and $B\} = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\bar{B})$, thus A and $\bar{B}$ are independent.

(2) $\mathbb{P}(\bar{A} \cap \bar{B}) = \{$ by De Morgan's Law$\} = \mathbb{P}(\overline{A \cup B}) = 1 - \mathbb{P}(A \cup B) = \{$by the additive rule$\} = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)) = \{$by independence of A and $B\} = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = 1 - \mathbb{P}(A) - \mathbb{P}(B)(1 - \mathbb{P}(A)) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(\bar{A})\mathbb{P}(\bar{B})$, and thus $\bar{A}$ and $\bar{B}$ are independent.

Exercise 4.1. A string of Christmas lights contains 20 lights. The lights are wired in series so that if any light fails, the whole string will go dark. Each light has a probability of 0.02 of failing during a 3-year period. The lights fail independently of each other.

(a) What is the probability that a string of lights will remain bright for 3 years? Answer: $\mathbb{P}(\text{all 20 lights work}) = \{\text{by independence}\} = \big(\mathbb{P}(\text{light works})\big)^{20} = \big(1 - 0.02\big)^{20} = (0.98)^{20} = 0.6676$.
(b) What is the minimum number of lights needed for the string to be equally likely to go dark or remain bright? Answer: We want to find n such that $\mathbb{P}(\text{remain bright}) = (0.98)^n = 0.5$. Hence, $n = \lceil \ln(0.5)/\ln(0.98) \rceil = \lceil 34.3096 \rceil = 35$.

Exercise 4.2. Here is a two-way table of the composition of the 99th Congress (elected in 1986) by party and seniority. The entries in the body of the table should be the probabilities that a randomly chosen member of Congress has both the stated seniority and party affiliation. Only the two marginal distributions of party alone and seniority alone are given. If party and seniority were independent, what would be the probabilities in the body of the table?

Seniority	Democrat	Republican	Total
<2years			0.1
2-9 years			0.6
$\geq$ 10 years			0.3
Total	0.6	0.4	

Answer: To populate the table, we need to multiply the respective marginal probabilities for each cell. We get

Seniority	Democrat	Republican	Total
<2years	(0.1)(0.6)=0.06	(0.1)(0.4)=0.04	0.1
2-9 years	(0.6)(0.6)=0.36	(0.6)(0.4)=0.24	0.6
$\geq$ 10 years	(0.3)(0.6)=0.18	(0.3)(0.4)=0.12	0.3
Total	0.6	0.4	

Note that all the probabilities in the table add up to 1, as it should be. Indeed, $0.06 + 0.04 + 0.36 + 0.24 + 0.18 + 0.12 = 1$.

Exercise 4.3. A general can plan a campaign to fight one major battle or three small battles. He believes that he has a probability of 0.6 of winning the large battle and a probability of 0.8 of winning each of the small battles. Victories or defeats in the small battles are independent. The general must win either the large battle or all three small battles to win the campaign. Which strategy should he choose?

Answer: Using the independence of small battles, the probability to win all three of them equal to the product of individual probabilities, that is, $(0.8)^3 = 0.512$, which is smaller than 0.6, so we recommend the general fight the major battle.

Exercise 4.4. A manufacturer of flu vaccine is concerned about the quality of its flu serum. Batches of serum are processed by three different departments having rejection rates of 0.1, 0.08, and 0.12, respectively. The inspections by the three departments are sequential and independent. What is the probability that a batch of serum survives the first two departments but is rejected by the third department? Answer: $(0.9)(0.92)(0.12) = 0.09936$.

Exercise 4.5. Can mutually exclusive events be independent? Explain your answer. Answer: No, mutually exclusive events cannot be independent since if one event happens, the other event cannot happen. Thus, whatever the original probability of the event $\mathbb{P}(A)$ was, if the other event B happens, the conditional probability $\mathbb{P}(A|B)$ turns to zero.

Exercise 4.6. John will get a D in French class with a probability of 0.4. His cousin, independently, will get a D in biology class with a probability of 0.7. Compute the probability that

(a) both will get a D. Answer: $(0.4)(0.7) = 0.28$.

(b) neither will get a D. $(1 - 0.4)(1 - 0.7) = (0.6)(0.3) = 0.18$.

(c) either will get a D. Answer: $1 - \mathbb{P}(\text{neither will get a D}) = 1 - 0.18 = 0.82$.

Exercise 4.7. Refer to Exercise 1.10. A fair coin is flipped until two tails or three heads appear. The sample space is $S = \{TT, THT, THHT, THHH, HTT, HHTT, HTHT, HHTH, HHH, HTHH\}$. Assuming that the flips are independent and the coin is fair, assign probabilities to the outcomes and check that the probabilities sum up to one.

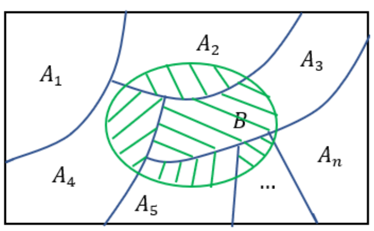
Answer: The one outcome of length two is assigned probability $(1/2)^2 = 1/4$, the three outcomes of length three are assigned probability $(1/2)^3 = 1/8$, and the six outcomes of length four are assigned probability $(1/2)^4 = 1/16$. The sum of the probabilities is $(1)(1/4) + (3)(1/8) + (6)(1/16) = 1/4 + 3/8 + 6/16 = 4/16 + 6/16 + 6/16 = 1$.

5. Bayes' Rule

Theorem (Bayes' Rule). Suppose $A_1, A_2, \ldots, A_n$ is a partition of the sample space S. That is, these sets don't overlap and fill up the entire sample space. Let B denote an event. Then for any fixed $i, i = 1, \ldots, n$, the following identity is true:

$$\mathbb{P}(A_i \mid B) \;=\; \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.$$

Proof. As shown in the Venn diagram below, B can be written as the union of non-overlapping intersections $B \cap A_1, \ldots, B \cap A_n$. Some of these intersections could be empty. Thus, we have $\mathbb{P}(B) = \sum_{j=1}^{n} \mathbb{P}(B \cap A_j)$.



Conditioning on A_j, we obtain $\mathbb{P}(B \cap A_j) = \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)$. Putting it all together, and applying the definition of conditional probability, we obtain

$$\mathbb{P}(A_i \mid B) \;=\; \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \;=\; \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}. \quad \square$$

Historical Note. The Reverand Thomas Bayes (circa 1701-1761) was an English statistician, philosopher, and Presbyterian minister.

Remark. Bayes' Rule has a very significant practical application. The probabilities $\mathbb{P}(A_j), j = 1, \ldots, n$ constitute the initial knowledge about the partition $A_j, j =$

$1, \ldots, n$. It can be referred to as **prior probabilities**. Once the event B occurs, investigators use Bayes' formula to update their knowledge about the probabilities of the partitioning events (called **posterior probabilities**). In fact, this formula is so important that it gave rise to an entirely new branch of Statistics – Bayesian Statistics, in which probabilities are recalculated as more data become available.

Example. A surgery is effective in 90% of patients with early stages of myeloma, and in 50% of patient with advanced stages. Among myeloma patients in a large hospital, 85% have an early stage of the disease. A myeloma patient had an effective surgery. What is the probability that the patient has an early (advanced) stage of the disease? Answer: By the Bayes' Rule,

$$\mathbb{P}(\text{early}|\text{effective}) = \frac{\mathbb{P}(\text{effective}|\text{early})\mathbb{P}(\text{early})}{\mathbb{P}(\text{effective}|\text{early})\mathbb{P}(\text{early}) + \mathbb{P}(\text{effective}|\text{advanced})\mathbb{P}(\text{advanced})}$$
$$= \frac{(0.9)(0.85)}{(0.9)(0.85) + (0.5)(0.15)} = 0.910714.$$

The probability that the patient has an advanced stage of the disease can be found by the complement rule,

$$\mathbb{P}(\text{advanced}|\text{effective}) = 1 - 0.910714 = 0.089286.$$

Note that the prior probability that a patient has an early stage of myeloma is 0.85, and since the surgery is likely to be effective, the posterior probability of the early stage is even higher (0.91). On the other hand, it is much less likely that the surgery is effective in advanced myeloma patients, the prior probability of the advanced stage of 0.15 becomes even less (0.08) after we learn that the surgery was effective.

Exercise 5.1. A large industrial firm uses three local motels to provide overnight accommodations for its clients. From experience, it is known that 20% of the clients are assigned rooms at the Ramada Inn, 50% at the Sheraton, and 30% at the Lakeview Motor Lodge. If the plumbing is faulty in 5% of the rooms at the Ramada Inn, 4% of the rooms at the Sheraton, and 8% of the rooms at the Lakeview Motor Lodge, what is the probability that a person with a room having faulty plumbing was assigned accommodations at the Ramada Inn? Sheraton? Lakeview Motor Lodge?
Answer: Let's R stand for Ramada Inn, S for Sheraton, L for Lakeview Motor Lodge, and F for faulty plumbing. We are given that the prior probabilities are $\mathbb{P}(R) = 0.2$, $\mathbb{P}(S) = 0.5$, and $\mathbb{P}(L) = 0.3$. Also, the conditional probabilities are specified as $\mathbb{P}(F|R) = 0.05$, $\mathbb{P}(F|S) = 0.04$, and $\mathbb{P}(F|L) = 0.08$. Once faulty plumbing is observed, the posterior probabilities are computed as follows.

$$\mathbb{P}(R|F) = \frac{\mathbb{P}(F|R)\mathbb{P}(R)}{\mathbb{P}(F|R)\mathbb{P}(R) + \mathbb{P}(F|S)\mathbb{P}(S) + \mathbb{P}(F|L)\mathbb{P}(L)}$$
$$= \frac{(0.05)(0.2)}{(0.05)(0.2) + (0.04)(0.5) + (0.08)(0.3)} = \frac{0.01}{0.01 + 0.02 + 0.024} = \frac{0.01}{0.054} = 0.1852,$$

$$\mathbb{P}(S|F) = \frac{\mathbb{P}(F|S)\mathbb{P}(S)}{\mathbb{P}(F|R)\mathbb{P}(R) + \mathbb{P}(F|S)\mathbb{P}(S) + \mathbb{P}(F|L)\mathbb{P}(L)}$$
$$= \frac{(0.04)(0.5)}{(0.05)(0.2) + (0.04)(0.5) + (0.08)(0.3)} = \frac{0.02}{0.054} = 0.3704,$$

and

$$\mathbb{P}(L|F) = \frac{\mathbb{P}(F|L)\mathbb{P}(L)}{\mathbb{P}(F|R)\mathbb{P}(R) + \mathbb{P}(F|S)\mathbb{P}(S) + \mathbb{P}(F|L)\mathbb{P}(L)}$$
$$= \frac{(0.08)(0.3)}{(0.05)(0.2) + (0.04)(0.5) + (0.08)(0.3)} = \frac{0.024}{0.054} = 0.4444.$$

Comparing prior and posterior probabilities, for Ramada Inn there is a slight decrease (0.2 vs. 0.1852), for Sheraton, there is a larger decrease (0.5 vs 0.3704, and for Lakeview Motor Lodge, there is a large increase (0.3 vs. 0.4444). This is not surprising as Lakeview Motor Lodge has a higher chance of faulty plumbing. Note also that the prior probabilities as well as posterior probabilities add up to one, as it must be.

Exercise 5.2. An insurance company believes that people can be divided into two classes: accident-prone and non-accident-prone. An accident-prone person will have an accident within a year with a probability of 0.4, whereas a non-accident-prone person, with a probability of 0.2.

(a) If 30% of the population is accident-prone, what is the probability that a new policyholder will have an accident within a year? Answer: Let A denote an accident proneness, N stand for non-accident proneness, and E stand for event (accident). From the setting of the problem, we obtain that $\mathbb{P}(A) = 0.3$, $\mathbb{P}(N) = 1 - 0.3 = 0.7$, $\mathbb{P}(E|A) = 0.4$, and $\mathbb{P}(E|N) = 0.2$. We compute

$$\mathbb{P}(E) = \mathbb{P}(E|A)\mathbb{P}(A) + \mathbb{P}(E|N)\mathbb{P}(N) = (0.3)(0.4) + (0.7)(0.2) = 0.12 + 0.14 = 0.26.$$

(b) Suppose that the new policyholder has an accident. What is the probability that he is accident-prone? non-accident-prone? Answer: We compute the posterior probabilities using the Bayes' Rule. We write

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(E|A)\mathbb{P}(A)}{\mathbb{P}(E)} = \frac{0.12}{0.26} = 0.4615,$$

and

$$\mathbb{P}(N|E) = \frac{\mathbb{P}(E|N)\mathbb{P}(N)}{\mathbb{P}(E)} = \frac{0.14}{0.26} = 0.5385.$$

The probability that a policyholder is accident-prone increased (0.3 vs. 0.4615), whereas the probability of a non-accident-prone policyholder decreased (0.7 vs. 0.5385) since accident-prone people are more likely to be involved in an accident.

Exercise 5.3. Pacemakers were implanted into one hundred cardiac patients. Two types of pacemakers were used. Fifty patients received single-chamber pacemakers, while the others received dual-chamber type. Previous clinical trials revealed that 25% of single-chamber pacemakers have instances of false alarm, whereas only 15% of dual-chamber pacemakers cause false alarms.
(a) Suppose a patient's pacemaker caused a false alarm. What is the probability that the patient has a single-chamber pacemaker? A dual-chamber one?
Answer: Let S denote a single-chamber pacemaker, D stand for dual-chamber one, and F stand for false alarm. We are given that $\mathbb{P}(S) = 0.5 = \mathbb{P}(D), \mathbb{P}(F|S) = 0.25$, and $\mathbb{P}(F|D) = 0.15$. We compute

$$\mathbb{P}(S|F) = \frac{\mathbb{P}(F|S)\mathbb{P}(S)}{\mathbb{P}(F|S)\mathbb{P}(S) + \mathbb{P}(F|D)\mathbb{P}(D)}$$

$$= \frac{(0.25)(0.5)}{(.25)(0.5) + (0.15)(0.5)} = \frac{0.125}{0.125 + 0.075} = \frac{0.125}{0.2} = 0.625,$$

and $\mathbb{P}(D|F) = 1 - 0.625 = 0.375$.

Note that the probability of a single-chamber pacemaker increased from 0.5 to 0.625, and the probability of dual-chamber pacemaker respectively decreased. This change is expected since single-chamber pacemakers are more likely to cause false alarms.
(b) Suppose another false alarm has occurred in the same patient. Update the posterior probabilities.
Answer: Assuming that 0.625 and 0.375 are now our respective prior probabilities, and using the Bayes' formula again, we compute

$$\mathbb{P}(S|2nd\ F) = \frac{(0.25)(0.625)}{(0.25)(0.625) + (0.15)(0.375)} = \frac{0.15625}{0.2125} = 0.7353,$$

and $\mathbb{P}(D|2nd\ F) = 1 - 0.7353 = 0.2647$. Note that the probability of a single-chamber pacemaker increased even more, whereas the other probability dropped.

6. Combinatorics

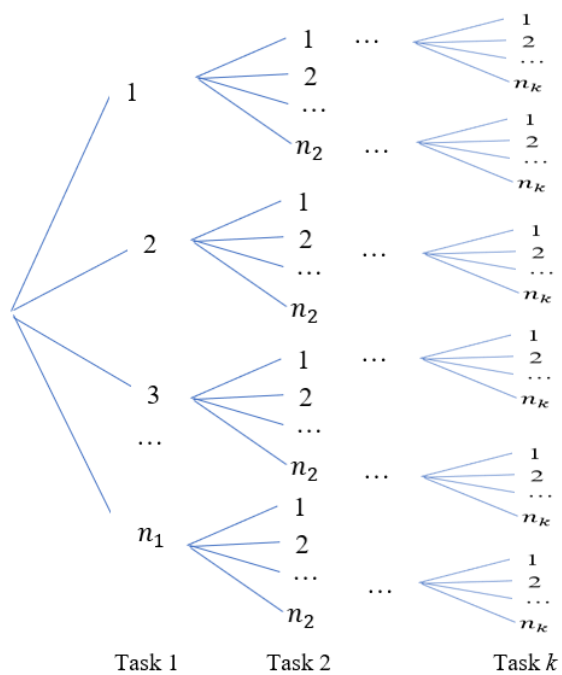
Counting Principle

Counting Principle (or Counting Rule). If a job requires completing k tasks, and the ith task can be completed in n_i ways, where $i = 1, \dots, k$, then the job can be done in $(n_1)(n_2)\cdots(n_k)$ ways.

Example. Bob is buying a new car. There are two body styles: sedan or hatchback; five colors: black, red, green, blue, or navy; and three models: standard, sports, or luxury. How many total choices does Bob have? There are three tasks: to choose body style (2 choices), to choose color (5 choices), and to choose model (3 choices).

By the counting rule, the total number of choices is (2)(5)(3)=30.

A good illustration of the counting principle is given by a **dendrogram** (also called **tree diagram**). It depicts a branch for every choice. First, there are n_1 branches that reflect the choices for task 1. Then, for every fixed choice in task 1, there are n_2 branches (choices) for task 2, etc. Finally, the total number of choices is the number of terminal branches (or leaves) on that dendrogram, which can be found by multiplying the number of choices for every task. Schematically, a dendrogram looks like this:



Exercise 6.1. A nursery rhyme starts as follows:

As I was going to St. Ives
I met a man with seven wives.
Each wife had seven sacks.
Each sack had seven cats.
Each cat had seven kittens.

How many kittens did the traveler meet? Answer: 7 wives, 7 sacks, 7 cats, 7 kittens, for a total of $7^4 = 2401$ kittens.

Exercise 6.2. Eloise is buying an ice-cream. There are six flavors of ice cream and three kinds of cones. How many different two-scoop ice-creams can she order? Answer: There are six choices for the first scoop, six choices for the second scoop, and three choices for the cone, for a total of $(6)(6)(3) = 108$.

Exercise 6.3. The Indiana license plate looks like this 1ABC234. How many different license plates of this design are possible? Answer: There are 10 choices for each of the four digits and 26 choices for each of the three letters, so assuming there are no restrictions on what the digits and letters could be, a total of $(10)^4(26)^3 = 175,760,000$ license plates of this design are possible.

Exercise 6.4. How many three-digit numbers can be formed from the digits 1,4,5,7, and 9, if digits can be used more than once? How many of them will be even numbers? Answer: There are five choices for each of the three positions, so $(5)(5)(5) = 125$ three-digit numbers can be formed. For it to be an even number, it must end in 4, so there are $(5)(5)$=25 even numbers possible.

Exercise 6.5. One must choose a four-digit PIN number. Each digit can be chosen from 0 to 9, and digits cannot be repeated. How many different possible PIN numbers can be chosen? Answer: There are 10 choices for the first digit, 9 choices for the second digit, 8 choices for the third digit, and 7 choices for the last digit, for a total of $(10)(9)(8)(7) = 5,040$ PIN numbers.

Factorial

Definition. The **factorial** of a number n, denoted by $n!$, is the product of all integers from n down to 1, that is, $n! = (n)(n-1)(n-2)\cdots(2)(1)$. By definition, $0! = 1$.

Example. The following factorials are mostly often used in calculations and should be memorized.

$$0! = 1, \quad 1! = 1, \quad 2! = (2)(1) = 2, \quad 3! = (3)(2)(1) = 6,$$
$$4! = (4)(3)(2)(1) = 24, \quad 5! = (5)(4)(3)(2)(1) = 120, \quad 6! = (6)(5)(4)(3)(2)(1) = 720.$$

Remark. The following extremely useful identity holds: $n! = (n)(n-1)!$. This follows directly from the definition. It can be extended further, if necessary. For example, $n! = (n)(n-1)(n-2)!$.

Exercise 6.6. Compute $\frac{5!}{3!}$. Answer: $\frac{5!}{3!} = \frac{(5)(4)(3!)}{3!} = (5)(4) = 20$.

Exercise 6.7. Compute $\frac{6!}{2!(6-2)!}$. Answer:

$$\frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = \frac{(6)(5)(4!)}{(2)(4!)} = \frac{(6)(5)}{2} = 15.$$

Exercise 6.8. Compute $\frac{20!}{17!(20-17)!}$. Answer:

$$\frac{20!}{17!(20-17)!} = \frac{20!}{17!3!} = \frac{(20)(19)(18)(17!)}{(17!)(6)} = (20)(19)(3) = 1,140.$$

Definition. A **permutation** is an <u>ordered</u> arrangement of objects.

Definition. A **combination** is an <u>unordered</u> arrangement of objects.

Example. If we have three objects named A, B, and C, then there are six permutations ABC, ACB, BAC, BCA, CAB, and CBA (because order matters), and only one combination ABC (because order doesn't matter).

Example. If we have three objects A, B, and C, then there are six two-object permutations AB, BA, AC, CA, BC, and CB (order matters), and only three two-object combinations AB, AC, and BC (order doesn't matter).

Proposition. The number of permutations of k objects chosen from among n objects is

$$n(n-1)(n-2)\cdots(n-k+1) \;=\; \frac{n!}{(n-k)!}.$$

Proof. There are n choices for the first position, $n-1$ choices for the second position, etc. There are $(n-k+1)$ choices for the kth position. By the counting rule, we multiply the number of choices.

Example. In how many ways can k people sit in row? We are looking for the number of ordered arrangements of k people chosen from among k people. The number of permutations for $n = k$ is $\dfrac{k!}{(k-k)!} = \dfrac{k!}{0!} = k!$. It can also be derived by noticing that there are k choices for the first position, $k-1$ choices for the next position, and so on, until the last person who will sit in the last position, for a total of $k!$ choices.

Proposition. The number of combinations of k objects chosen from among n objects is

$$\binom{n}{k} \;=\; \frac{n!}{k!(n-k)!}.$$

The symbol $\dbinom{n}{k}$ is called a **binomial coefficient** and is read "n choose k".

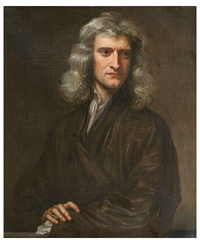
Proof. There are $\dbinom{n}{k}$ combinations (unordered arrangements) of k objects chosen from among n objects, which can be ordered in $k!$ ways. Thus, $\dbinom{n}{k}(k!)$ is equal to the number of permutations $\dfrac{n!}{(n-k)!}$. From here, the result follows.

Remark. $\binom{n}{k}$ is called a binomial coefficient because it appears in **Newton's binomial** ("binomial" means a polynomial with two members), an identity that states:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

$$= \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n.$$

Historical Note. Sir Isaac Newton (1642-1726) was an English mathematician, physicist, astronomer, and philosopher.



Exercise 6.9. Compute $\binom{3}{0}, \binom{3}{1}, \binom{3}{2}\binom{3}{3}, \binom{5}{2}$, and $\binom{5}{3}$. Answer:

$$\binom{3}{0} = \frac{3!}{0!(3-0)!} = \frac{3!}{0!3!} = 1, \quad \binom{3}{1} = \frac{3!}{1!(3-1)!} = \frac{3!}{1!2!} = 3,$$

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3!}{2!1!} = 3, \quad \binom{3}{3} = \frac{3!}{3!(3-3)!} = \frac{3!}{3!0!} = 3,$$

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!} = \frac{(5)(4)(3!)}{2!3!} = \frac{(5)(4)}{2} = 10,$$

and

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{(5)(4)(3!)}{3!2!} = \frac{(5)(4)}{2} = 10.$$

Exercise 6.10. Show that

(a) $\binom{n}{0} = 1$ for any $n \geq 0$. That is, there is only one way to choose none from n objects. It is by doing nothing. Answer:

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{0!\,n!} = 1.$$

(b) $\binom{n}{1} = n$ for any $n \geq 1$. That is, either one of the n objects can be chosen, resulting in the n choices. Answer:

$$\binom{n}{1} = \frac{n!}{1!(n-1)!} = \frac{(n)(n-1)!}{1!(n-1)!} = n.$$

(c) $\binom{n}{k} = \binom{n}{n-k}$ for any $n \geq k$. That is, to choose k objects from n objects is equivalent to choosing $n-k$ objects that will remain untouched. Answer:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!k!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}.$$

Exercise 6.11. Ten cars are in a race. In how many ways can they win the first, second, and third places? Answer: We are interested in the number of ordered arrangements (permutations). There are 10 choices for the first place, 9 choices for the second place, and 8 choices for the third place, for a total of $(10)(9)(8) = 720$ choices.

Exercise 6.12. Ten cars are in a race. Three cars will qualify for the next race. In how many different ways can this happen? Answer: We are interested in the number of unordered arrangements (combinations) of three objects chosen from among 10 objects. The number of ways is $\binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10!}{3!7!} = \frac{(10)(9)(8)}{6} = (10)(3)(4) = 120$.

Exercise 6.13. There are 10 people in a room. Everyone shakes hands with everyone else. How many handshakes take place? Answer: We need to choose two people for a handshake and the order doesn't matter, therefore, there are $\binom{10}{2} = \frac{10!}{2!(10-2)!} = \frac{(10)(9)(8!)}{(2)(8!)} = \frac{(10)(9)}{2} = 45$.

Exercise 6.14. In how many ways can a committee of 4 be chosen from 5 girls and 5 boys if
(a) all are equally eligible? Answer: The number of combinations of 10 choose 4 is $\binom{10}{4} = \frac{10!}{4!(10-4)!} = \frac{(10)(9)(8)(7)(6!)}{(4)(3)(2)(6!)} = (10)(3)(7) = 210$.

(b) the committee must include 2 girls and 2 boys? Answer: The number of combinations of 5 choose 2 is $\binom{5}{2} = \frac{5!}{2!3!} = \frac{(5)(4)}{2} = 10$, so there are 10 ways to choose 2 girls and 10 ways to choose 2 boys. By the counting rule, the total number of ways is $(10)(10) = 100$.

Exercise 6.15. In how many ways can 7 books by William Shakespeare be arranged on a shelf, if

(a) there are no restrictions. Answer: Seven books can be arranged in order in 7!=5,040 ways.

(b) if “Much Ado About Nothing” must be in the middle? Answer: Let "Much Ado About Nothing" be in the middle. We still have 6 books that have to be arranged in order. This can be done in 6!=720 ways.

Exercise 6.16. In order to play basketball, ten children at a playground divide themselves into team A and team B of five each. How many different divisions are possible? Answer: The children need to choose 5 to be on team A, and the others will be on team B. Also, the order doesn't matter, so the total number of divisions is $\binom{10}{5} = \frac{10!}{5!\,5!} = \frac{(10)(9)(8)(7)(6)}{(5)(4)(3)(2)} = (3)(2)(7)(6) = 252.$

Exercise 6.17. From a group of five women and seven men, randomly choose five people. What is the probability that

(a) two women and three men are chosen? Answer: The total number of ways to choose 5 out of 12 people is $\binom{12}{5}$. The number of ways to choose 2 women from among 5 women is $\binom{5}{2}$ and the number of ways to choose 3 men from among 7 men is $\binom{7}{3}$. Thus the probability is

$$\frac{\binom{5}{2}\binom{7}{3}}{\binom{12}{5}} = \frac{\frac{5!}{2!\,3!}\cdot\frac{7!}{3!\,4!}}{\frac{12!}{5!\,7!}} = \frac{\frac{(5)(4)}{2}\cdot\frac{(7)(6)(5)}{6}}{\frac{(12)(11)(10)(9)(8)}{(5)(4)(3)(2)}} = \frac{(10)(35)}{(11)(9)(8)} = 0.4419.$$

(b) women are not chosen? Answer: The number of ways to choose 5 men from among 7 men is $\binom{7}{5}$, thus the probability is

$$\frac{\binom{5}{0}\binom{7}{5}}{\binom{12}{5}} = \frac{\frac{7!}{5!\,2!}}{\frac{12!}{5!\,7!}} = \frac{21}{(11)(9)(8)} = 0.0265.$$

(c) Mr. N. is chosen? Answer: Let Mr. N be chosen. Still, 4 more people should be chosen from among the remaining 11 people. Therefore, the probability is

$$\frac{\binom{11}{4}}{\binom{12}{5}} = \frac{\frac{11!}{4!\,7!}}{\frac{12!}{5!\,7!}} = \frac{\frac{(11)(10)(9)(8)}{(4)(3)(2)}}{(11)(9)(8)} = \frac{(11)(10)(3)}{(11)(9)(8)} = 0.4167.$$

(d) Mr. N. is the only man chosen? Answer: The other 4 people chosen must be women. Thus, the probability is

$$\frac{\binom{5}{4}}{\binom{12}{5}} = \frac{5}{(11)(9)(8)} = 0.0063.$$

Exercise 6.18. A police department in a small city consists of ten officers. The department policy is that five officers patrol the streets, two work full-time at the station, and three are on reserve.

(a) How many different divisions are possible? Answer: The number of ways to choose 5 officers from 10 to patrol the streets is $\binom{10}{5}$. Of the remaining 5 officers, the number of ways to choose 2 to work full-time at the station is $\binom{5}{2}$ and the remaining 3 officers will be on reserve. The total number of ways is $\binom{10}{5}\binom{5}{2} = \frac{(10)(9)(8)(7)(6)}{(5)(4)(3)(2)} \cdot 10 = (3)(2)(7)(6)(10) = 2,520$.

(b) What is the probability that Officer Larson patrols the streets? Answer: Let Officer Larson patrol the streets. To choose 4 more officers to patrol the street from the remaining 9 officers can be done is $\binom{9}{4}$ way. Of the remaining 5 officers, 2 should be chosen to work full-time at the station. Hence, the total number of ways is $\binom{9}{4}\binom{5}{2} = \frac{(9)(8)(7)(6)}{(4)(3)(2)} \cdot 10 = (3)(7)(6)(10) = 1,260$.

(c) What is the probability that Officer Larson works at the station? Answer: Let Officer Larson work at the station. The number of ways to choose 5 officers from among the remaining 9 to patrol the streets is $\binom{9}{5}$, and from the other 4 officers, one will be chosen to work at the station with Officer Larson. The total number of ways is $\binom{9}{5}\binom{4}{1} = \frac{(9)(8)(7)(6)}{(4)(3)(2)} \cdot 4 = (3)(7)(6)(4) = 504$.

(d) What is the probability that Officer Larson is on reserve? Answer: Let Officer Larson be on reserve. Of the remaining 9 officers, 5 should be chosen to patrol the streets and of the other 4 officers, 2 should be chosen to work at the station. Thus the total number of way is $\binom{9}{5}\binom{4}{2} = (3)(7)(6)(6) = 756$.

Note that $1,260 + 504 + 756 = 2,520$, as it should be.

7. DISCRETE RANDOM VARIABLE

Definition. A **random variable** is a variable that assumes certain values on each outcome of a random phenomenon.

Notation. Random variables are typically denoted by a large letter from the second half of the Latin alphabet. For example, X, Y, Z, T, W, X_1, X_2, etc.

Definition. A **discrete** random variable assumes a finite or countably infinite number of values.

Definition. A **probability mass function** (pmf) of a discrete random variable X is $p_X(x) = \mathbb{P}(X = x)$. It has two properties: (1) $0 \leq p_X(x) \leq 1$, and (2) $\sum_x p_X(x) = 1$. To define a pmf, one needs to identify all possible values with the respective probabilities.

Example. A fair coin is flipped two times. Let X be the number of heads that appear. The sample space is $S = \{HH, HT, TH, TT\}$. The values that X assumes are $X(HH) = 2$, $X(HT) = X(TH) = 1$, and $X(TT) = 0$. The pmf is $p(0) = \mathbb{P}(X = 0) = \mathbb{P}(TT) = (1/2)(1/2) = 1/4$, $p(1) = \mathbb{P}(X = 1) = \mathbb{P}(HT) + \mathbb{P}(TH) = (1/2)(1/2) + (1/2)(1/2) = 1/2$, and $\mathbb{P}(X = 2) = \mathbb{P}(HH(= (1/2)(1/2) = 1/4$. Note that each of these probabilities falls between 0 and 1, and the probabilities add up to one.

Example. A fair coin is flipped until a head or three tails appear. We want to find the distribution of the number of flips required. We specify the sample space $S = \{H, TH, TTH, TTT\}$. Let X be the number of flips. The pmf of X is $p(1) = \mathbb{P}(X = 1) = \mathbb{P}(H) = 1/2$, $p(2) = \mathbb{P}(X = 2) = \mathbb{P}(TH) = (1/2)(1/2) = 1/4$, $p(3) = \mathbb{P}(X = 3) = \mathbb{P}(TTH) + \mathbb{P}(TTT) = (1/2)(1/2)(1/2)(2) = 1/4$. Note that the probabilities sum up to one.

Definition. The **expected value** (or **expectation**, or **mean**, or **average**) of a discrete random variable X is computed as $\mathbb{E}X = \displaystyle\sum_x x\, p_X(x)$.

Example. Let X be the number of heads that appear when a fair coin is flipped two times. Above we found the probability function of X: $p(0) = p(2) = 1/4$, and $p(1) = 1/2$. The expected number of heads is $\mathbb{E}X = (0)(1/4) + (1)(1/2) + (2)(1/4) = 1/2 + 1/2 = 1$.

Example. Let X be the number of flips required when a fair coin is flipped until a head or three tails appear. Earlier we found the pmf of X, which is $p(1) = 1/2$, and $p(2) = p(3) = 1/4$. The expected number of flips is $\mathbb{E}X = (1)(1/2) + (2)(1/4) + (3)(1/4) = 2/4 + 2/4 + 3/4 = 7/4 = 1.75$.

Remark. In the physical world, the expectation represents the center of mass. If we consider all the values that X assumes and place at those values ball with weights proportional to the respective probabilities, then the expected value of X represents the values at which a fulcrum should be located to make the system balanced. Indeed, let $\mathbb{E}X$ denote the center of mass. A system is balanced when the sum of the weights multiplied by the distances to the center of mass is equal to zero, that is, $\sum_x p(x)(x - \mathbb{E}X) = 0$. Opening the parentheses, we rewrite this identity as $\sum_x xp(x) = \mathbb{E}X\big(\sum_x p(x)\big)$. Since the probabilities add up to one, we have that necessarily $\mathbb{E}X = \sum_x x\, p(x)$, which is the definition of the expected value.

Proposition (Functional Invariance of the Mean). Let X be a discrete random variable, and let $Y = f(X)$ be another random variable obtained from X by applying function f. The distribution function of Y may be very hard to obtain, but the expected value of Y can be computed as the expected value of $f(X)$, that is, $\mathbb{E}Y = \mathbb{E}(f(X))$.

Definition. The k**th moment** of a discrete random variable X with the pmf $p_X(x)$ is computed as $\mathbb{E}(X^k) = \sum_x x^k\, p_X(x)$.

Definition. The k**th central moment** of a discrete random variable X with the pmf $p_X(x)$ is $\mathbb{E}(X - \mathbb{E}X)^k = \sum_x (x - \mathbb{E}X)^k\, p_X(x)$.

Proposition (Invariance of Mean Under Linear Transformation). Suppose X is a random variable, and a and b are some constants. Then $\mathbb{E}(aX+b) = a\mathbb{E}X+b$.

Proof. The proof is a direct consequence of the linearity of summation. We write $\mathbb{E}(aX+b) \overset{def}{=} \sum_x (ax+b)p_X(x) = a\sum_x xp_X(x) + b\sum_x p_X(x) = a\mathbb{E}X + b$.

Harder to show but also true is

Proposition. For any discrete random variables X and Y, and any constants a and b, $\mathbb{E}(aX+bY) = a\mathbb{E}X + b\mathbb{E}Y$.

Definition. The **variance** of a discrete random variable $\mathbb{V}ar(X) = \mathbb{E}(X-\mathbb{E}X)^2 = \sum_x (x-\mathbb{E}X)^2 p_X(x)$. It represents the sum of squared distances to the mean multiplied by respective probabilities. It is a measure of the spread of the values of X.

Proposition. The **computational formula** for the variance is

$$\mathbb{V}ar(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Proof. $\mathbb{V}ar(X) \overset{def}{=} \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}\big(X^2 - 2(\mathbb{E}X)X + (\mathbb{E}X)^2\big) = \mathbb{E}X^2 - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

Remark. By definition, the expected value of X is the **first moment** of X. The variance of X is the **second central moment**, which according to the computational formula, is the second moment minus the square of the first moment.

Remark. The expected value of X is measured in the same units as X, whereas the variance is measured in these units squared. If, for instance, X is measured in inches, its mean is also measured in inches, but its variance is measured in inches squared. It follows from the fact that the variance is a quadratic form. It makes the variance hard to picture. So, instead, another measure of spread is introduced.

Definition. A **standard deviation** of a discrete random variable X is the square root of the variance, that is $\sqrt{\mathbb{V}ar(X)}$. There is no conventional notation for the standard deviation. Sometimes it is denoted by s_X or σ_X, or *stdev*. It is measured in the same units as X.

Useful Formula. For any real-valued a and b, and any random variable X,

$$\mathbb{V}(aX+b) = a^2\,\mathbb{V}ar(X).$$

Proof. $\mathbb{V}ar(aX+b) = \mathbb{E}\big(aX+b-\mathbb{E}(aX+b)\big)^2 = \mathbb{E}\big(aX+b-a\mathbb{E}X-b\big)^2 = \mathbb{E}\big(a(X-\mathbb{E}X)\big)^2 = a^2\mathbb{E}(X-\mathbb{E}X)^2 = a^2\mathbb{V}ar(X)$.

Exercise 7.1. Which of the following variables have discrete probability distributions?

(a) the number of automobile accidents per year in Virginia. Answer: discrete (0, 1, 2, etc.).

(b) the length of time to play 18 holes of golf. Answer: time flows continuously, so not discrete, unless the scale is specified. For example "How many hours does it take to play 18 holes of golf? Round to the nearest integer."

(c) the amount of milk produced yearly by a particular cow. Answer: Amounts of fluids change continuously, so naturally, it is not a discrete random variable, unless the scale is specified (in tons, for example).

(d) the number of eggs laid each month by a hen. Answer: discrete (0, 1, 2, etc.).

(e) the number of building permits issued each month in a city. Answer: discrete (0, 1, 2, etc.).

(f) the weight of grain produced per acre. Answer: Weights change continuously, so not discrete, unless measured in bushels or on some other scale.

Exercise 7.2. Determine the value of the normalizing constant c that make the function $p(x) = c(x^2 + 4)$ for $x = 0, 1, 2,$ or 3, a true probability mass function. Answer: $p(0) = 4c$, $p(1) = 5c$, $p(2) = 8c$, and $p(3) = 13c$. The sum of probabilities, which should be one, is equal to $30c$, thus, $c = 1/30$.

Exercise 7.3. An entomologist has five bugs. Two of them are nice looking and three are ugly looking. He randomly picks two.

(a) Find the probability distribution of the number of ugly bugs in the sample. Answer: Let X denote the number of ugly-looking bugs in the sample of two. The probability distribution of X is

$$\mathbb{P}(X = 0) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1}{10} = 0.1, \quad \mathbb{P}(X = 1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{6}{10} = 0.6,$$

and

$$\mathbb{P}(X = 2) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} = 0.3.$$

(b) How many ugly-looking bugs should the entomologist expect to see in his sample? Answer:

$$\mathbb{E}X = (0)(0.1) + (1)(0.6) + (2)(0.3) = 1.2.$$

Exercise 7.4. An entomologist has five bugs. Two of them are nice looking and three are ugly looking. He randomly picks three.

(a) Find the probability distribution of the number of ugly bugs in the sample. Answer: Let X denote the number of ugly-looking bugs in the sample of two. The probability distribution of X is

$$\mathbb{P}(X = 1) = \frac{\binom{2}{2}\binom{3}{1}}{\binom{5}{3}} = \frac{3}{10} = 0.3,\ \mathbb{P}(X = 2) = \frac{\binom{2}{1}\binom{3}{2}}{\binom{5}{3}} = \frac{6}{10} = 0.6,$$

and

$$\mathbb{P}(X = 3) = \frac{\binom{2}{0}\binom{3}{3}}{\binom{5}{3}} = \frac{1}{10} = 0.1.$$

(b) How many ugly-looking bugs should the entomologist expect to see in his sample? Answer:

$$\mathbb{E}X = (1)(0.3) + (2)(0.6) + (3)(0.1) = 1.8.$$

Exercise 7.5. Let Y be the random variable with $p(y)$ given in the accompanying table. Find $\mathbb{E}(Y)$, $\mathbb{E}(1/Y)$, $\mathbb{E}(2Y^2 - 4)$, and $\mathbb{V}ar(5 - Y)$.

y	1	2	3	4
p(y)	0.4	0.3	0.2	0.1

Answer: $\mathbb{E}Y = (1)(0.4) + (2)(0.3) + (3)(0.2) + (4)(0.1) = 2$, $\mathbb{E}(1/Y) = (1/1)(0.4) + (1/2)(0.3) + (1/3)(0.2) + (1/4)(0.1) = 77/120 = 0.6417$, $\mathbb{E}(2Y^2 - 4) = 2\mathbb{E}Y^2 - 4 = 2\big((1)^1(0.4) + (2)^2(0.3) + (3)^2(0.2) + (4)^2(0.1)\big) - 4 = 2(5) - 4 = 6$, and $\mathbb{V}ar(5 - Y) = (-1)^2\mathbb{V}ar(Y) = \mathbb{V}ar(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = 5 - (2)^2 = 1$.

Exercise 7.6. You are offered a once-in-a-lifetime opportunity to play the following game. A fair coin is flipped three times. If exactly two heads appear, you will be paid \$10. Otherwise, you have to pay \$7.

(a) Would you be willing to play this game? Answer: Let W be your winnings in this game. The distribution of W is $\mathbb{P}(W = 10) = \mathbb{P}(HHT, HTH, THH) = 3/8$, and $\mathbb{P}(W = -7) = 1 - 3/8 = 5/8$. The expected winnings are $\mathbb{E}W = (10)(3/8) + (-7)(5/8) = (30 - 35)/8 = -5/8 = -0.625$, so on average, you expect to lose 62.5 cents and you should not agree to play this game.

(b) A game is called **fair** if the expected gain is zero for both players. Assume that in the described game, you will be paid \$10 if exactly two heads appear. How much should you pay otherwise to make it a fair game? Answer: Let w denote the amount that you would pay. Then $\mathbb{P}(W = 10) = 3/8$, and $\mathbb{P}(W = w) = 5/8$. We want to expected value of W to be equal to zero, so $\mathbb{E}W = (30 + 5w)/8 = 0$, from where

$w = -\$6$, so you would pay \$6 in a fair game.

Exercise 7.7. A psychic runs the following ad in a magazine: Expecting a baby? Renown psychic will tell you the sex of the unborn child from any photograph of the mother. Cost \$10. Money-back guarantee. This may be a profitable con game. Suppose that the psychic simply replies "girl" to each inquiry. In the worst case, everyone who has a boy will ask for her money back. Find the expected value and standard deviation of the psychic's profit. Answer: Let G denote the psychic's gain. We are given that $\mathbb{P}(G = \$10) = 1/2$, and $\mathbb{P}(G = \$0) = 1/2$. Thus, the expected gain is $\mathbb{E}G = (\$10)(1/2) + (\$0)(1/2) = \$5$. The variance of the gain is $\mathbb{V}ar(G) = (\$10)^2(1/2) + (\$0)^2(1/2) - (\$5)^2 = \$^2 25$, and the standard deviation is $\sqrt{\mathbb{V}ar(G)} = \sqrt{\$^2 25} = \$5$.

Exercise 7.8. The Connecticut State Lottery awards at random, for every 100,000 one-dollar tickets sold,

$$\begin{array}{rrl} 1 & \$5,000 & \text{prize,} \\ 18 & \$200 & \text{prizes,} \\ 120 & \$25 & \text{prizes,} \\ 270 & \$20 & \text{prizes.} \end{array}$$

What is the expected value of the winnings of one ticket in this lottery? Do you want to play? Answer: Denote by W the winning amount of one ticket. The distribution of W is $\mathbb{P}(W = \$5,000) = 1/100,000$, $\mathbb{P}(W = \$200) = 18/100,000$, $\mathbb{P}(W = \$25) = 120/100,000$, $\mathbb{P}(W = \$20) = 270/100,000$, and $\mathbb{P}(W = \$0) = 1-$the other probabilities. The expected winning amount is $\mathbb{E}W = \frac{1}{100,000}\big((\$5,000)(1) + (\$200)(18) + (\$25)(120) + (\$20)(270)\big) = \$17,000/100,000 = \$0.17$ or 17 cents. So, each one-dollar ticket wins on average 17 cents. You should not be playing this lottery.

8. BERNOULLI DISTRIBUTION

Definition. A discrete random variable X assumes a **Bernoulli distribution** if $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. The notation is $X \sim Ber(p)$, which is read as 'X has a Bernoulli distribution with parameter p". Conventionally, the value of 1 is termed a**success**, and 0 is termed a**failure**. The parameter p is termed the **probability of success**.

For Bernoulli distribution, the pmf can be written as $p_X(x) = p^x(1-p)^{1-x},\ x = 0$ or 1. Indeed, when $X = 1$, $p_X(1) = p$, and when $X = 0$, $p_X(0) = 1 - p$. The expected value of X is $\mathbb{E}X = (1)(p) + (0)(1-p) = p$, the probability of success. The variance of X is $\mathbb{V}ar(X) = (1)^2 p + (0)^2(1-p) - p^2 = p - p^2 = p(1-p)$, and the standard deviation is $\sqrt{p(1-p)}$.

Historical Note. Jacob Bernoulli (1655-1705) was one of the many prominent mathematicians in the Swiss Bernoulli family.

Example. A fair coin is flipped once. Let X be the number of heads that appeared, so $X \sim Ber(1/2)$ with the pmf $p_X(x) = (0.5)^x(0.5)^{1-x} = 0.5, \ x = 0$ or 1. The mean is $\mathbb{E}X = 0.5$, the variance is $\mathbb{V}ar(X) = (0.5)(1-0.5) = 0.25$, and the standard deviation is $\sqrt{0.25} = 0.5$.

Example. A biased coin is flipped once. The probability of success is 0.65. The number of heads X has a Bernoulli(0.65) distribution with the pmf $p_X(x) = (0.65)^x(0.35)^{1-x}$, $x = 0$ or 1. The mean is $\mathbb{E}(X) = 0.65$, variance is $\mathbb{V}ar(X) = (0.65)(0.35) = 0.2275$, and standard deviation $\sqrt{0.2275} = 0.4770$.

Exercise 8.1. An infected person will infect a susceptible person with a probability of 0.3, for whom symptoms will appear with a probability of 0.6. What is the distribution of the number of infected individuals who show symptoms after contact with the infected person? What is the expected number of such individuals? What's the standard deviation? Answer: The probability that an individual is infected and shows symptoms is $(0.3)(0.6) = 0.18$, so the number of people who show symptoms is a Bernoulli(0.18) random variable. The expected number of such individuals is $p = 0.18$, and the standard deviation is $\sqrt{p(1-p)} = \sqrt{(0.18)(1-0.18)} = 0.4792$.

Exercise 8.2. On a multiple-choice test with five choices for each question, what is the probability that a student answers a question correctly just by guessing? What are the underlying distribution function, mean, and variance? Answer: We assume that exactly one answer choice is correct, so the probability to pick the correct answer is $p = 1/5 = 0.2$ and the distribution of the number of correct answers is $Bernoulli(0.2)$, with mean $p = 0.2$ and variance $p(1-p) = (0.2)(1-0.2) = 0.16$.

9. BINOMIAL DISTRIBUTION

Definition. A **Bernoulli** trial is a random phenomenon with a binary outcome (success or failure) and a fixed probability of success.

Definition. Consider n independent Bernoulli trials, each with a probability of success p. Let X be the number of successes among these n trials. Then X has a **binomial distribution** with parameters n and p. We write $X \sim B(n, p)$. The probability distribution function can we expressed as $p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}$, $x = 0, 1, \ldots, n$.

Explanation. There are $\begin{pmatrix} n \\ x \end{pmatrix}$ sequences with x successes and $n - x$ failures, each success happening with probability p and each failure occurring with probability $1-p$.

Remark. A Bernoulli distribution is a special case of a binomial distribution for $n = 1$.

Remark. Notice that the pmf of a binomial distribution involves binomial coefficients, which give rise to the name of the distribution. Moreover, Newton's binomial helps to show that the probabilities add up to one. Indeed, we write

$$\sum_{x=0}^{n} \binom{n}{x} p^x(1-p)^{n-x} \;=\; (p+1-p)^n \;=\; 1.$$

Proposition. The expected value of a random variable $X \sim Bi(n, p)$ is $\mathbb{E}X = np$ and its variance is $\mathbb{V}ar(X) = np(1-p)$.

Proof.
$$\begin{aligned}
\mathbb{E}X \;&=\; \sum_{x=0}^{n} x \binom{n}{x} p^x\,(1-p)^{n-x} \;=\; \sum_{x=1}^{n} \frac{(x)(n!)}{x!(n-x)!} p^x\,(1-p)^{n-x} \\
&=\; n\,p \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-1-(x-1))!}\, p^{x-1}(1-p)^{n-1-(x-1)} \\
&=\; \{k = x-1\} \;=\; np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k(1-p)^{n-1-k} \;=\; n\,p,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}X^2 \;&=\; \sum_{x=0}^{n} x^2 \binom{n}{x} p^x\,(1-p)^{n-x} \;=\; \sum_{x=2}^{n} (x)(x-1) \binom{n}{x} p^x\,(1-p)^{n-x} \\
&+ \sum_{x=0}^{n} x \binom{n}{x} p^x\,(1-p)^{n-x} \;=\; \sum_{x=1}^{n} \frac{x(x-1)n!}{x!(n-x)!} p^x(1-p)^{n-x} \;+\; np \\
&=\; n(n-1)\,p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-2-(x-2))!}\, p^{x-2}(1-p)^{n-2-(x-2)} \;+\; np \\
&=\; \{k = x-2\} \;=\; n(n-1)\,p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k(1-p)^{n-2-k} \;=\; n(n-1)\,p^2 + np,
\end{aligned}$$

and so,

$$\mathbb{V}ar(X) = \mathbb{E}X^2 \;-\; (\mathbb{E}X)^2 \;=\; n(n-1)p^2 \;+\; np \;-\; (np)^2 \;=\; np - np^2 = np(1-p).$$

Example. A fair coin is tossed 10 times. Let X be the number of heads. Then $X \sim Bi(10, 0.5)$. The mean number of heads is $\mathbb{E}X = np = (10)(0.5) = 5$, the variance is $\mathbb{V}ar(X) = np(1-p) = (10)(0.5)(1-0.5) = 2.5$, and the standard deviation is $\sqrt{2.5} = 1.58$. The probability that there will be, say, between 4 heads is computed as $\mathbb{P}(X = 4)\binom{10}{4}(0.5)^{10} = \frac{(10)(9)(8)(7)}{(4)(3)(2)}(0.5)^{10} = (10)(3)(7)(0.5)^{10} = 210/1024 = 0.2051$.

Example. A biased coin is tossed 10 times. The probability of a head is 0.65. Let X be the number of heads. Then $X \sim Bi(10, 0.65)$, with mean $\mathbb{E}X = (10)(0.65) = 6.5$, variance $\mathbb{V}ar(X) = (10)(0.65)(1-0.65) = 2.275$, and the standard deviation $\sqrt{2.275} = 1.5083$. The probability $\mathbb{P}(X = 4) = \binom{10}{4}(0.65)^4(0.35)^6 = (210)(0.65)^4(0.35)^6 = 0.0689$.

Exercise 9.1. Which of the following settings is a binomial one? To reemphasize, a binomial setting necessarily has a fixed number of independent Bernoulli trials and the probability of success stays constant from trial to trial.

(a) The gender of the next 50 children born at a local hospital is observed. A random variable of interest is the number of girls. Answer: This is a binomial setting, $n = 50$, p is constant, and trials are assumed independent (no identical twins born).

(b) A couple decides to continue to have children until their first boy is born. A random variable is the number of children they have. Answer: This is not a binomial setting as the number of trials is not fixed.

(c) An auto manufacturer chooses one car from each hour's production for a quality inspection. A random variable is the number of defects in the car's paint. Answer: This is not a binomial setting as the number of trials is not fixed, and the trials are not Bernoulli trials (not 0/1 outcomes).

(d) Joe buys a state lottery ticket every week. A random variable is the number of times in a year that he wins a prize. Answer: This is a binomial setting, assuming that there are a fixed number of weeks in a year, and the probability of winning a prize stays constant from week to week.

Exercise 9.2. Screws produced by a company are defective with probability 0.01 independently of each other. A package of 10 screws is bought.

(a) Find the probability that at most one screw is defective. Answer: Let X be the number of defective screws in the package of 10 screws. Then $X \sim Bi(10, 0.01)$. We compute $\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = (0.99)^{10} + (10)(0.01)(0.99)^9 = 0.9957$.

(b) If the company offers a money-back guarantee that at most one of ten screws is defective, find the proportion of screws the company must replace. Answer: The company will have to replace a package is more than two screws are defective, which happens with probability $1-0.9957 = 0.0043$, so 0.43% of packages (screws) will have to be replaced (43 of 10,000 packages, or 430 of 100,000 screws).

Exercise 9.3. If a family has four children, is it more likely they will have two boys and two girls or three of one sex and one of the other? Assume that the probability of a child being a boy is 1/2 and that the births are independent events. Answer: Denote by X the number of boys. We are given that $X \sim Bi(4, 1/2)$. We compute $\mathbb{P}(X = 2) = \binom{4}{2}(1/2)^4 = 6/16$, whereas $\mathbb{P}(X = 1 \text{ or } X = 3) = \binom{4}{1}(1/2)^4 + \binom{4}{3}(1/2)^4 = 8/16$, so having one boy and three girls or the other way around is more likely than having two boys and two girls.

Exercise 9.4. A travel company uses a tour bus with a capacity of ten passengers but sells twelve tickets. Fortunately for them, one person out of six is a no-show.

(a) What is the probability that everyone who shows up for the tour will be accommodated? Answer: Let X be the number of people who show up for the tour. We know that $X \sim Bi(12, 5/6)$, thus, $\mathbb{P}(X \leq 10) = 1 - \mathbb{P}(X = 11) - \mathbb{P}(X = 12) = 1 - \binom{12}{11}(5/6)^{11}(1/6) - \binom{12}{12}(5/6)^{12} = 0.6187$.

(b) How many people do they expect will come to the bus tour? Answer: $\mathbb{E}X = (12)(5/6) = 10$.

(c) What is the standard deviation of the number of people who will show up? Answer: $\sqrt{(12)(5/6)(1/6)} = 1.2910$.

Exercise 9.5. Consider a multiple-choice quiz with three possible answers for each of the five questions.

(a) What is the probability that a student would answer more than 3 questions correctly just by guessing? Answer: Let X be the number of correctly answered questions. We know that $X \sim Bi(5, 1/3)$. We compute $\mathbb{P}(X > 3) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) = \binom{5}{4}(1/3)^4(2/3) + \binom{5}{5}(1/3)^5 = 11/3^5 = 0.04523$.

(b) How many questions can he expect to answer correctly? Answer: $\mathbb{E}X = (5)(1/3) = 1.67$.

(c) What is the standard deviation of the number of correct answers? Answer: $\sqrt{(5)(1/3)(2/3)} = 1.054$.

Exercise 9.6. The gunner on a small assault boat fires six missiles at an attacking plane. Each has a 10% chance of being on target. If two or more of the shells find their mark, the plane will crash. At the same time, the pilot of the plane fires 2 air-to-surface rockets, each of which has a 25% chance of critically disabling the boat. Would you rather be on the plane or the boat? Answer: Denote by X the number of shells that hit the plane. We have that $X \sim Bi(6, 0.1)$. We compute

$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - (0.9)^6 - (6)(0.1)(0.9)^5 = 0.1143$. Now let Y be the number of rockets that hit the boat. We know that $Y \sim Bi(2, 0.25)$. We compute $\mathbb{P}(Y \leq 1) = 1 - \mathbb{P}(Y = 0) = 1 - (0.75)^{10} = 0.4375$. The probability that the boat sinks is higher than the probability that the plane crashes, so it is better to be on the plane.

10. GEOMETRIC DISTRIBUTION

Definition. Consider a sequence of independent Bernoulli trials, each having the probability of success p. Let X be the total number of trials until the first success is observed. Then X has a **geometric distribution** with parameter p. The pmf of X is $p_X(x) = p(1-p)^{x-1}$, $x = 1, 2$, etc. The notation is $X \sim Geom(p)$.

Remark. To show that these probabilities add up to one, we need to sum up a geometric series, hence the name of the distribution. Recall that the sum of a geometric series is $\sum_{x=k}^{\infty} a^x = \frac{a^k}{1-a}$, $|a| < 1$. We write

$$\sum_{x=1}^{\infty} p(1-p)^{x-1} = \frac{p}{1-p} \sum_{x=1}^{\infty} (1-p)^x = \frac{p}{1-p} \cdot \frac{1-p}{1-(1-p)} = 1.$$

Proposition. The mean of a geometric random variable is $\mathbb{E}(X) = \frac{1}{p}$ and the variance is $\mathbb{V}ar(X) = \frac{1-p}{p^2}$.

Proof. Before we proceed, we need to derive the following two results. For $|a| < 1$, $\sum_{x=1}^{\infty} x\, a^{x-1} = \left(\sum_{x=0}^{\infty} a^x\right)'_a = \left(\frac{1}{1-a}\right)'_a = \frac{1}{(1-a)^2}$. In addition, $\sum_{x=1}^{\infty} x^2\, a^x = a^2 \sum_{x=2}^{\infty} x(x-1)a^{x-2} + \sum_{x=1} x\, a^x = a^2 \left(\sum_{x=2} a^x\right)''_a + a \sum_{x=1} x\, a^{x-1} = a^2 \left(\frac{a^2}{1-a}\right)''_a + \frac{a}{(1-a)^2} = \frac{2a^2}{(1-a)^3} + \frac{a}{(1-a)^2}$. Here we skipped the calculation of the second derivative. It can be done as follows. $\left(\frac{a^2}{1-a}\right)''_a = \left(\frac{(1-a)(2a) - a^2(-1)}{(1-a)^2}\right)'_a = \left(\frac{2a - a^2}{(1-a)^2}\right)'_a = \frac{(1-a)^2(2-2a) - (2a-a^2)2(1-a)(-1)}{(1-a)^4} = \frac{2(1-a)^3 + 2a(2-a)(1-a)}{(1-a)^4}$
$= 2\frac{(1-a)^2 + a(2-a)}{(1-a)^3} = 2\frac{1 - 2a + a^2 + 2a - a^2}{(1-a)^3} = \frac{2}{(1-a)^3}$.

Now applying the first result with $a = 1-p$, we obtain $\mathbb{E}(X) = \sum_{x=1}^{\infty} x\, p(1-p)^{x-1} = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}$. Making use of the second result with $a = 1-p$, we derive the expression for the variance. We get

$$\mathbb{V}ar(X) \;=\; \mathbb{E}X^2 \;-\; (\mathbb{E}X)^2 \;=\; \sum_{x=1}^{\infty} x^2\, p(1-p)^{x-1} \;-\; \Big(\frac{1}{p}\Big)^2$$

$$= \frac{p}{1-p}\sum_{x=1}^{\infty} x^2\,(1-p)^x - \frac{1}{p^2} \;=\; \Big(\frac{p}{1-p}\Big)\Big(\frac{2(1-p)^2}{(1-(1-p))^3} + \frac{1-p}{(1-(1-p))^2}\Big) - \Big(\frac{1}{p^2}\Big)$$

$$= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \;=\; \frac{2(1-p)+p-1}{p^2} \;=\; \frac{1-p}{p^2}.$$

Example. A fair coin is flipped until a head appears. The total number of required flips is a random variable, say, X that has a geometric distribution with parameter $p = 0.5$. The pmf of X is $p_X(x) = (0.5)(1-0.5)^{x-1} = (0.5)^x,\;\; x = 1, 2, ...$ The mean is $\mathbb{E}X = 1/(0.5) = 2$ (meaning that on average, two flips are required to see a head), the variance is $\mathbb{V}ar(X) = (1-0.5)/(0.5)^2 = 2$, and the standard deviation is $\sqrt{2} = 1.4142$. The probability that the first head appears, for example, on the sixth flip is $\mathbb{P}(X = 6) = (0.5)^6 = 0.0156$.

Example. A biased coin is flipped until a head appears. The probability of a head is 0.65. Then X, the total number of flips, has a $Geom(0.65)$ distribution with the pmf $p_X(x) = (0.65)(0.35)^{x-1},\;\; x = 1, 2,$ etc., mean $\mathbb{E}(X) = 1/0.65 = 1.5385$, variance $\mathbb{V}ar(X) = (1-0.65)/(0.65)^2 = 0.8284$, and standard deviation $\sqrt{0.8284} = 0.9102$. The probability that the first head appears on the sixth flip is $\mathbb{P}(X = 6) = (0.65)(0.35)^5 = 0.0034$.

Example. Suppose $X \sim Geom(p)$. To compute the probability that it will take at least k trials to see the first success, we can use the complement rule and write $\mathbb{P}(X \geq k) = 1 - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) - \cdots - \mathbb{P}(X = k-1) = 1 - p - p(1-p) - \cdots - p(1-p)^{k-2} = (1-p) - p(1-p)(1 + (1-p) + \cdots + (1-p)^{k-3}) = (1-p) - p(1-p)\big(\frac{1-(1-p)^{k-2}}{1-(1-p)}\big) = (1-p) - (1-p)(1-(1-p)^{k-2}) = (1-p)^{k-1}$. The same result could be obtained much quicker by noticing that in order for the first head to appear in at least k flips, the first $k-1$ flips must all be failures, that is, $\mathbb{P}(X \geq k) = \mathbb{P}(\text{first } k-1 \text{ flips are all failures}) = (1-p)^{k-1}$.

Remark. Sometimes a geometric distribution is defined as the number of failures before the first success. In this case, the pmf has the form $p_X(x) = p(1-p)^x,\;\; x = 0, 1, ...$ The mean is $\mathbb{E}X = \dfrac{1}{p} - 1 = \dfrac{1-p}{p}$ and the variance is $\mathbb{V}ar(X) = \dfrac{1-p}{p^2}$.

Exercise 10.1. Which of the settings below is a geometric setting? As a reminder, a setting is geometric if a sequence of independent Bernoulli trials with a constant probability of success is observed until the first success happens, at which point the trials stop and the total number of trials is counted.

(a) Four cards are drawn one at a time with replacement from a standard poker deck of cards. A random variable is the number of kings drawn. Answer: No, this is a binomial setting, not geometric. Geometric would be to draw cards with replacement and stop when the first king is drawn and count the total number of cards.

(b) A couple decides to continue to have children until their first boy is born. A random variable is the number of children they have. Answer: This is a geometric setting, assuming a constant probability of having a boy and independent trials (no identical twins).

(c) Cards are drawn without replacement from a standard poker deck of cards. A random variable is the number of draws until the queen of spades shows up. Answer: This is not a geometric setting because drawing is done without replacement and so the probability of success changes from trial to trial. It would be a geometric setting if the drawing were done with replacement.

Exercise 10.2. Joe takes driving tests until he passes. The probability that he passes a test is 0.6, and the tests are independent from each other.

(a) Find the probability that he passes on the second attempt. Answer: He fails first and then passes with probability $(0.4)(0.6) = 0.24$.

(b) Find the probability that he needs at most two attempts. Answer: He passes either on the first or second attempt with a probability of $0.6 + 0.24 = 0.84$.

(c) What is the probability that he needs more than four attempts? Answer: The first four attempts are failures with probability $(0.4)^4 = 0.0256$.

(d) What is the expected number of attempts he needs? Answer: The expected value is $1/(0.6) = 1.67$.

Exercise 10.3. An oil prospector will drill a succession of holes in a given area to find a productive well. The probability that he is successful on a given try is 0.2, independently of other tries.

(a) What is the probability that the third hole drilled is the first to yield a productive well? Answer: Let X be the number of holes drilled until the first productive one. Then $X \sim Geom(0.2)$. We compute $\mathbb{P}(X = 3) = (0.2)(0.8)^2 = 0.128$.

(b) If the prospector can afford to drill at most ten wells, what is the probability that he will fail to find a productive well? Answer: All 10 drills must be failures. This happens with probability $(0.8)^{10} = 0.1074$.

Exercise 10.4. Suppose the probability of a connection during a busy time for international phone calls is 0.3, and attempts are assumed independent.

(a) What is the probability that between 3 and 5 attempts are necessary for a successful call? Answer: Let X be the total number of attempts until the first successful phone call. Then $X \sim Geom(0.3)$. We compute $\mathbb{P}(3 \leq X \leq 5) = \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5) = (0.3)(0.7)^2 + (0.3)(0.7)^3 + (0.3)(0.7)^4 = 0.32193$. A shorter way to compute this probability is to notice that the first success comes after 2 attempts but before 6 attempts, so we need to compute the probability that the first two attempts are failures minus the probability that the first five attempts are failures, that is, $(0.7)^2 - (0.7)^5 = 0.32193$.

(b) How many phone calls are necessary, on average? Answer: $\mathbb{E}X = 1/0.3 = 3.33$.

(c) Compute the standard deviation of the number of attempts necessary. Answer:

$$\sqrt{\frac{1-0.3}{(0.3)^2}} = 2.79.$$

(d) What is the probability that between 5 and 10 attempts are necessary for a successful call? Answer: $(0.3)^4 - (0.3)^10 = 0.0081$.

11. NEGATIVE BINOMIAL DISTRIBUTION

Definition. Consider a sequence of independent Bernoulli trials with a constant probability of success p, and let X be the total number of trails required to see the rth success where r is a fixed number, $r \geq 1$. Then X has a **negative binomial** distribution with parameters r and p. The notation is $X \sim NB(r,p)$. The probability function of X is $p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, r+2, \ldots$. This can be easily seen by noticing that in a sequence of x trials, the last one must be a success, and there are $r-1$ successes somewhere among the first $x-1$ trials. The number of such sequences is $\binom{x-1}{r-1}$. Further, each of r successes happens with probability p and each of $x-r$ failures happens with probability $1-p$.

Remark. A geometric distribution is a special case of negative binomial distribution with $r = 1$.

Proposition. Let $X \sim NB(r,p)$. The mean of X is $\mathbb{E}X = \frac{r}{p}$, and the variance is $\mathbb{V}ar(X) = \frac{r(1-p)}{p^2}$.

Proof. $\mathbb{E}X = \sum_{x=r}^{\infty} x\binom{x-1}{r-1} p^r(1-p)^{x-r} = \sum_{x=r}^{\infty} \frac{x(x-1)!}{(r-1)!(x-r)!} p^r(1-p)^{x-r} = r\sum_{x=r}^{\infty} \frac{x!}{r!(x-r)!} p^r(1-p)^{x-r} = \frac{r}{p}\sum_{x=r}^{\infty} \binom{x}{r} p^{r+1}(1-p)^{x-r} = \{y = x+1,\ \tilde{r} = r+1\} =$

$$\frac{r}{p}\sum_{y=\tilde{r}}^{\infty}\binom{y-1}{\tilde{r}-1}p^{\tilde{r}}(1-p)^{y-\tilde{r}} = \frac{r}{p}.$$

As for the variance, we write $\mathbb{V}ar(X) = \sum_{x=r}^{\infty} x^2\binom{x-1}{r-1}p^r(1-p)^{x-r} - \left(\frac{r}{p}\right)^2 =$

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

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

$$= \{y = x+2,\ \tilde{r} = r+2\} = \frac{r(r+1)}{p^2}\sum_{y=\tilde{r}}^{\infty}\binom{y-1}{\tilde{r}-1}p^{\tilde{r}}(1-p)^{y-\tilde{r}} - \frac{r}{p} - \frac{r^2}{p^2}$$

$$= \frac{r(r+1)}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} = \frac{r^2+r-rp-r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

Remark. The name "negative binomial" comes from the fact that the pmf of a negative binomial distribution can be rewritten to resemble a binomial distribution for the negative number of trials. To see this, we make a substitution $x = r+y$ (here y denotes the number of failures until the rth success). The pmf becomes $\binom{x-1}{r-1}p^r(1-p)^{x-r} = \binom{r+y-1}{r-1}p^r(1-p)^y$. Now we notice that

$$\binom{r+y-1}{r-1} = \frac{(r+y-1)!}{(r-1)!y!} = \frac{(r)(r+1)\cdot\dots\cdot(r+y-1)}{y!}$$

$$= (-1)^y\frac{(-r)(-r-1)\cdot\dots\cdot(-r-y+1)}{y!} = (-1)^y\binom{-r}{y}.$$

Thus, the pmf takes the form $(-1)^y\binom{-r}{y}p^r(1-p)^y$, which resembles a binomial pmf.

Remark. It is not difficult now for us to show that the probabilities add up to one. Indeed, we write $\sum_{x=r}^{\infty}\binom{x-1}{r-1}p^r(1-p)^{x-r} = \sum_{y=0}^{\infty}\binom{r+y-1}{r-1}p^r(1-p)^y =$ $\sum_{y=0}^{-r}(-1)^y\binom{-r}{y}p^r(1-p)^y = p^r\sum_{y=0}^{-r}\binom{-r}{y}(-(1-p))^y = \{\text{by Newton's binomial}\} =$ $p^r\,(1+(-(1-p)))^{-r} = p^r\,p^{-r} = 1.$

Example. Suppose a fair coin is flipped until the fourth head appears, and let X be the total number of flips. Then $X \sim NB(4, 0.5)$ with the probability distribution function $p_X(x) = \mathbb{P}(X = x) = \binom{x-1}{4-1}(0.5)^4(0.5)^{x-4} = \binom{x-1}{3}(0.5)^x,\ x = 4, 5, \dots$. It takes on average, $\mathbb{E}X = 4/0.5 = 8$ flips to see four heads, with the variance $\mathbb{V}ar(X) = (4)(1-0.5)/(0.5)^2 = 8$, standard deviation $\sqrt{8} = 2.8284$. The probability that it takes, say, 6 flips to see the fourth head is $\mathbb{P}(X = 6) = \binom{5}{3}(0.5)^4(0.5)^2 = (10)(0.5)^6 = 0.15625$.

Example. Suppose a biased coin is flipped until the fourth head appears. The probability of a head is 0.65. Let X be the total number of flips. The distribution of X is $NB(4, 0.65)$ and the pmf can be written as $p_X(x) = \binom{x-1}{3}(0.65)^4(0.35)^{x-4}$, $x = 4, 5$, etc. The expected number of required flips is $\mathbb{E}X = 4/0.65 = 6.1538$, variance is $\mathbb{V}ar(X) = (4)(0.35)/(0.65)^2 = 3.3136$, and standard deviation $\sqrt{3.3136} = 1.8203$.

Exercise 11.1. An oil prospector will drill a succession of holes in a given area to find a productive well. The probability that he is successful on a given try is 0.2, independently of other tries.

(a) What is the probability that the fifth hole drilled is the second to yield a productive well? Answer: Let X be the number of drills until the second successful one. Then $X \sim NB(2, 0.2)$. We compute $\mathbb{P}(X = 5) = \binom{5-1}{2-1}(0.2)^2(1-0.2)^3 = \binom{4}{1}(0.2)^2(0.8)^3 = 0.08192$.

(b) What is the expected number of drills needed to find 10 productive wells? Answer: Let X be the number of drills needed to find 10 productive wells. We know that $X \sim NB(10, 0.2)$, with the mean $\mathbb{E}X = 10/0.2 = 50$.

(c) What is the standard deviation of the number of drills needed to find 10 productive wells? Answer: $\sqrt{(10)(1-0.2)/(0.2)^2} = 14.14$.

Exercise 11.2. Suppose the probability of a connection during a busy time for international phone calls is 0.3, and attempts are assumed independent.

(a) What is the probability that from 3 to 5 attempts are necessary in order to place two successful calls? Answer: Let X be the number of attempts until two successful calls are placed. The distribution of X is $NB(2, 0.3)$. We compute $\mathbb{P}(3 \le X \le 5) = \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5) = \binom{2}{1}(0.3)^2(0.7) + \binom{3}{1}(0.3)^2(0.7)^2 + \binom{4}{1}(0.3)^2(0.7)^3 = (0.3)^2(0.7)(2 + (3)(0.7) + (4)(0.7)^2) = 0.3818$.

(b) How many phone calls are required on average to place three successful calls? Answer: $X \sim NB(3, 0.3)$, $\mathbb{E}X = 3/0.3 = 10$.

(c) Compute the standard deviation of the number of attempts necessary to make four successful phone calls. Answer: $X \sim NB(4, 0.3)$, $\sqrt{(4)(1-0.3)/0.3^2} = 5.58$.

Exercise 11.3. A roulette wheel consists of 38 numbers – 1 through 36, 0, and 00. Suppose Rob always bets that the outcome will be one of the numbers 1 through 12.

(a) What is the probability that his second win will occur on the fourth bet? Answer: $X \sim NB(2, 12/38 = 6/19)$, $\mathbb{P}(X = 4) = \binom{3}{1}(6/19)^2(13/19)^2 = 0.14$.

(b) What is the expected number of bets until his second win? Answer: $\mathbb{E}X = 2/(6/19) = 6.33$.

(c) What is the standard deviation of the number of bets until his second win? Answer: $\sqrt{(2)(13/19)/(6/19)^2} = 3.70$.

Exercise 11.4. Find the probability that a person tossing three coins will get either all heads or all tails for the second time on the fifth toss. Answer: $\mathbb{P}(HHH \text{ or } TTT) = 1/8 + 1/8 = 1/4$, $X \sim NB(2, 1/4)$, $\mathbb{P}(X = 5) = \binom{4}{1}(1/4)^2(3/4)^3 = 0.1055$.

Exercise 11.5. In a certain manufacturing process it is known that 1 in every 100 items is defective. What is the probability that the seventeenth item inspected is the third defective item found? Answer: $X \sim NB(3, 0.01)$, $\mathbb{P}(X = 17) = \binom{16}{2}(0.01)^3(0.99)^{14} = \frac{(16)(15)}{2}(0.01)^3(0.99)^{14} = 0.000104$.

12. HYPERGEOMETRIC DISTRIBUTION

Definition. Let N denote the population size, of which K objects are of interest. A random sample of size n is drawn from this population. Note that for a **random sample**, every object in the population is equally likely to be chosen. Let X denote the number of objects of interest in the sample. Then X has a **hypergeometric distribution** with parameters N, K, and n, which is written as $X \sim HG(N, K, n)$. The probability function has the form

$$p_X(x) = \mathbb{P}(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \quad \max(0, n + K - N) \leq x \leq \min(K, n).$$

Example. There are 10 students in a group, 4 females and 6 males. We choose 3 students at random. Let X be the number of females in the sample. Then $X \sim HG(10, 4, 3)$. The probability distribution of X is

$$\mathbb{P}(X = 0) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{(5)(4)(6)}{(10)(9)(8)} = 1/6,$$

$$\mathbb{P}(X = 1) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{(4)(3)(5)(6)}{(10)(9)(8)} = 0.5,$$

$$\mathbb{P}(X=2) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{(6)(6)(6)}{(10)(9)(8)} = 0.3,$$

$$\mathbb{P}(X=3) = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{(4)(6)}{(10)(9)(8)} = 1/30.$$

Note that the probabilities add up to one, that is, $1/6 + 0.5 + 0.3 + 1/30 = 1$.

Example. There are 10 students in a group, 4 females and 6 males. We choose 5 students at random. Let X be the number of females in the sample. Then $X \sim HG(10, 4, 5)$. The probability distribution of X is

$$\mathbb{P}(X=0) = \frac{\binom{4}{0}\binom{6}{5}}{\binom{10}{5}} = \frac{(6)(5)(4)(3)(2)}{(10)(9)(8)(7)(6)} = 1/42,$$

$$\mathbb{P}(X=1) = \frac{\binom{4}{1}\binom{6}{4}}{\binom{10}{5}} = \frac{(4)(3)(5)(5)(4)(3)(2)}{(10)(9)(8)(7)(6)} = 5/21,$$

$$\mathbb{P}(X=2) = \frac{\binom{4}{2}\binom{6}{3}}{\binom{10}{5}} = \frac{(2)(3)(5)(4)(5)(4)(3)(2)}{(10)(9)(8)(7)(6)} = 10/21,$$

$$\mathbb{P}(X=3) = \frac{\binom{4}{3}\binom{6}{2}}{\binom{10}{5}} = \frac{(4)(3)(5)(5)(4)(3)(2)}{(10)(9)(8)(7)(6)} = 5/21,$$

$$\mathbb{P}(X=4) = \frac{\binom{4}{4}\binom{6}{1}}{\binom{10}{5}} = \frac{(6)(5)(4)(3)(2)}{(10)(9)(8)(7)(6)} = 1/42.$$

The sum of these probabilities is $1/42 + 5/21 + 10/21 + 5/21 + 1/42 = 1$, as it should be. Note that it would be impossible for X to be equal to 5 since there are only 4 female students in the group. The values for X range between $\max(0, n+K-N) = \max(0, 5+4-10) = \max(0, -1) = 0$ and $\min(K, n) = \min(4, 5) = 4$.

Example. There are 10 students in a group, 4 females and 6 males. We choose 7 students at random. Let X be the number of females in the sample. Then $X \sim HG(10, 4, 7)$. The probability distribution of X is

$$\mathbb{P}(X=1) = \frac{\binom{4}{1}\binom{6}{6}}{\binom{10}{7}} = \frac{(4)(3)(2)}{(10)(9)(8)} = 1/30,$$

$$\mathbb{P}(X=2) = \frac{\binom{4}{2}\binom{6}{5}}{\binom{10}{7}} = \frac{(6)(6)(3)(2)}{(10)(9)(8)} = 0.3,$$

$$\mathbb{P}(X=3) = \frac{\binom{4}{3}\binom{6}{4}}{\binom{10}{7}} = \frac{(4)(3)(5)(3)(2)}{(10)(9)(8)} = 0.5,$$

$$\mathbb{P}(X=4) = \frac{\binom{4}{4}\binom{6}{3}}{\binom{10}{7}} = \frac{(5)(4)(3)(2)}{(10)(9)(8)} = 1/6.$$

The sum of these probabilities is $1/30 + 0.3 + 0.5 + 1/6 = 1$, as it should be. Note that it would be impossible for X to be equal to 0 since there are only 6 male students in the group and we are choosing 7 for the sample. The values for X range between $\max(0, n+K-N) = \max(0, 7+4-10) = \max(0,1) = 1$ and $\min(K,n) = \min(4,7) = 4$.

Proposition. For $X \sim HG(N,K,n)$, the expected value is $\mathbb{E}X = \frac{K}{N}n$, and the variance is $\mathbb{V}ar(X) = \frac{K}{N}n\left(1-\frac{K}{N}\right)\left(\frac{N-n}{N-1}\right)$.

Proof. To derive the expression for the mean, we first notice that this expression is very intuitive. The true proportion of objects of interest in the sample is K/N, and since the drawn sample is random, the proportion should be preserved in the sample, thus, the expected number of objects of interest in the sample should be K/N times n, the size of the sample. Now we go through the formal proof. For definitiveness, we assume that n is less than K and write sums up to n. We get

$$\mathbb{E}X = \sum_{x=0}^{n} x\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n}\frac{x\cdot\frac{K!}{x!(K-x)!}\cdot\frac{(N-K)!}{(n-x)!(N-K-n+x)!}}{\frac{N!}{n!(N-n)!}}$$

$$= \sum_{x=1}^{n}\frac{K\cdot\frac{(K-1)!}{(x-1)!(K-x)!}\cdot\frac{(N-K)!}{(n-x)!(N-K-n+x)!}}{\frac{N\cdot(N-1)!}{n\cdot(n-1)!(N-n)!}} = \frac{K}{N}n\sum_{x=1}^{n}\frac{\binom{K-1}{x-1}\binom{N-K}{n-x}}{\binom{N-1}{x-1}} = \frac{K}{N}n.$$

The last sum is equal to one since it is the sum of all probabilities in the distribution of a $HG(N-1, K-1, n-1)$ random variable.

Turning now to the derivation of the formula for the variance, we proceed as follows.

$$\mathbb{V}ar(X) = \sum_{x=0}^{n} x^2\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} - \left(\frac{K}{N}n\right)^2 = \sum_{x=2}^{n} x(x-1)\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} + \frac{K}{N}n - \left(\frac{K}{N}n\right)^2$$

$$= \frac{K(K-1)n(n-1)}{N(N-1)}\sum_{x=2}^{n}\frac{\binom{K-2}{x-2}\binom{N-K}{n-x}}{\binom{N-2}{n-2}} + \frac{K}{N}n - \left(\frac{K}{N}n\right)^2$$

(the sum is equal to 1)

$$= \frac{K(K-1)n(n-1)}{N(N-1)} + \frac{K}{N}n - \left(\frac{K}{N}n\right)^2 = \frac{K}{N}n\left(\frac{(K-1)(n-1)}{N-1} + 1 - \frac{K}{N}n\right)$$

$$= \frac{K}{N}n\cdot\frac{(NK-N)(n-1)+(N-1)(N-Kn)}{(N)(N-1)}$$

$$= \frac{K}{N}n\cdot\frac{-NK-Nn+N^2+Kn}{(N)(N-1)} = \frac{K}{N}n\cdot\frac{(N-K)(N-n)}{N(N-1)}$$

$$= \frac{K}{N}n\left(1-\frac{K}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Remark. There is a more intuitive formula for the variance of a random variable X that has a hypergeometric distribution with parameters N, K, and n. Consider the following two-way table

	Of interest	Not of interest	Total
Sampled	x	$n-x$	n
Not sampled	$K-x$	$N-K-(n-x)$	$N-n$
Total	K	$N-K$	N

The variance of X can be expressed as the product of all **marginal totals** divided by the **grand total** squared multiplied by the grand total minus 1, that is, $\mathbb{V}ar(X) = \dfrac{(n)(N-n)(N-K)(K)}{N^2(N-1)}$. It is straightforward to see that this expression is equivalent to the one introduced above. Note also that the expression for the mean of X is the product of the respective marginal totals divided by the grand total, that is, $\mathbb{E}X = \dfrac{n}{K}N$.

Remark. The name "hypergeometric distribution" comes from the fact that the probabilities $\mathbb{P}(X \leq x)$ can be written in terms of **generalized hypergeometric functions**, which definitely lie beyond the scope of this course.

Exercise 12.1. A gardener plants six bulbs selected at random from a box with five tulip bulbs and four daffodil bulbs.

(a) What is the probability that he plants four tulip bulbs and two daffodil bulbs? Answer: Let X be the number of tulip bulbs that the gardener plants. We know that $X \sim HG(9,5,6)$. We compute $\mathbb{P}(X=4) = \dfrac{\binom{5}{4}\binom{4}{2}}{\binom{9}{6}} = \dfrac{(5)(6)(3)(2)}{(9)(8)(7)} = \dfrac{5}{14}$.

(b) What is the expected number of tulip bulbs he plants? Answer: $\mathbb{E}X = (5)(6)/9 = 10/3 = 3.33$.

(c) What is the expected number of daffodil bulbs he plants? Answer: $6 - 3.33 = 2.67$.

Exercise 12.2. What is the probability that a waitress will refuse to serve alcoholic beverages to only two minors if she randomly checks the IDs of five students from among nine students of whom four are not of legal age? Answer: Denote by X the number of minors whose IDs are checked. The distribution of X is $HG(9,4,5)$. We compute $\mathbb{P}(X=2) = \dfrac{\binom{4}{2}\binom{5}{3}}{\binom{9}{5}} = \dfrac{(10)(4)(3)(2)}{(9)(8)(7)(6)} = \dfrac{10}{21}$.

Exercise 12.3. A committee of size three is selected at random from four doctors and two nurses.
(a) Find the probability that two or three doctors are chosen. Answer: Let X be the number of doctors selected for the committee. Then $X \sim HG(6,4,3)$. We obtain

$$\mathbb{P}(X=2)+\mathbb{P}(X=3)=\frac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}}=0.6.$$

(b) What is the expected number of doctors selected for the committee? Answer: $\mathbb{E}X=(4)(3)/6=2$.

(c) What are the variance and standard deviation of the number of doctors selected for the committee? Answer: The variance is

$$\mathbb{V}ar(X)\;=\;\left(\frac{4}{6}\right)(3)\left(1-\frac{4}{6}\right)\left(\frac{6-3}{6-1}\right)\;=\;0.4.$$

The standard deviation is $\sqrt{0.4}=0.63$.

Exercise 12.4. Shipments of fifty items are inspected. The procedure is to take a sample of five and pass the shipment if no more than one is found to be defective. What proportion of 20% defective shipments will be accepted? Answer: Let X be the number of selected items that are defective. The distribution of X is $HG(50,10,5)$. The probability that a shipment is accepted is $\mathbb{P}(X=0)+\mathbb{P}(X=1)=\frac{\binom{10}{0}\binom{40}{5}}{\binom{50}{5}}+\frac{\binom{10}{1}\binom{40}{4}}{\binom{50}{5}}\;=\;\frac{(40)(39)(38)(37)(36)}{(50)(49)(48)(47)(46)}+\frac{(10)(40)(39)(38)(37)(5)}{(50)(49)(48)(47)(46)}\;=\;0.3106+0.4313=0.7419$. So, roughly, 74.2% of shipments will be accepted.

13. POISSON DISTRIBUTION

Definition. A **count** variable is a random variable that counts the number of occurrences of some random phenomenon. It assumes non-negative integer values (0, 1, 2, etc.).

Definition. A count variable X is said to follow a **Poisson** distribution with parameter λ (Greek letter "lambda" - pronounced "lam-da") if the pmf of X is $p_X(x)=\frac{\lambda^x}{x!}\,e^{-\lambda},\ x=0,1,2,$ etc. It is abbreviated $X\sim Poi(\lambda)$.

Historical Note. Poisson distribution was derived empirically by Simeon Poisson (1781-1840), a French mathematician and physicist.

Remark. Poisson distribution models **rare occurrences**, for which 0 or 1 are typical observations, 2 or 3 are seen less often, 4, 5, or 6 even less frequent, and 7 or 8 are unfrequent, and anything above is extremely unfrequent. Some examples of random variables governed by a Poisson distribution are the number of accidents per week or month at a certain intersection; the number of abandoned cars on a certain stretch of highway; the number of customers entering a bank per minute; the number of typographical errors per page in a newspaper or a magazine; number of defects per unit length of fabric; or the number of mice per acre of land.

Definition. Let $X \sim Poi(\lambda)$. The parameter λ is termed the **rate**. It specifies the average number of occurrences per unit interval of time (minute, hour, day, week, year, etc.), or unit piece of material (page of a newsletter, meter of fabric of fixed width, etc.), or unit of area (acre, square mile, etc.).

Remark. It is important to write out explicitly and memorize the probability function for some small values of x. We have

$$p(0) = \frac{\lambda^0}{0!}\,e^{-\lambda} = e^{-\lambda},\ \ p(1) = \frac{\lambda^1}{1!}\,e^{-\lambda} = \lambda\,e^{-\lambda},$$
$$p(2) = \frac{\lambda^2}{2!}\,e^{-\lambda} = \frac{\lambda^2}{2}\,e^{-\lambda},\ \ p(3) = \frac{\lambda^3}{3!}\,e^{-\lambda} = \frac{\lambda^3}{6}\,e^{-\lambda},$$

and

$$p(4) = \frac{\lambda^4}{4!}\,e^{-\lambda} = \frac{\lambda^4}{24}\,e^{-\lambda}.$$

Note that the factorials in the denominator increase very rapidly, making the probabilities very small very quickly.

Remark. To show that all the probabilities add up to one, we need to use Taylor's expansion of the exponential function:

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

From here we derive $\displaystyle\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}\,e^{-\lambda} = e^{\lambda}\cdot e^{-\lambda} = 1$. We can see that $e^{-\lambda}$ is just a normalizing constant in the probability function.

Proposition. Let $X \sim Poi(\lambda)$. Then $\mathbb{E}X = \mathbb{V}ar(X) = \lambda$.

Proof. $\displaystyle\mathbb{E}X = \sum_{x=0}^{\infty} x\,\frac{\lambda^x}{x!}\,e^{-\lambda} = \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}\,e^{-\lambda} = \lambda$, since after the substitution $y = x - 1$, we can see that the sum is equal to one. Focusing now on the variance, we compute

$$\mathbb{V}ar(X) = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} - \lambda^2 = \sum_{x=2}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda} + \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} - \lambda^2$$
$$= \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} e^{-\lambda} + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Proposition. Let λ be the rate of event occurrence per unit of time. If we look at time length of t units, there will be, on average, λt occurrences. So, the number of occurrences per t time units is $Poi(\lambda t)$.

Example. Suppose the number of phone calls X to a customer service department in a credit card company is a Poisson random variable with a rate of 3 per minute. Thus, $X \sim Poi(\lambda)$ where $\lambda = 3/\text{min}$. The probability that there will be at most 2 phone calls in the next minute is $\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = e^{-3} + 3e^{-3} + \frac{3^2}{2}e^{-3} = 13e^{-3} = 0.6472$. The average number of phone calls per minute is $\mathbb{E}X = \lambda = 3$ and the standard deviation is $\sqrt{\lambda} = \sqrt{3} = 1.7321$. Further, the average number of phone calls per 5 minutes, say, is $5\lambda = (5)(3) = 15$. The probability that there will be, say, 10 phone calls in the next 5 minutes is $\frac{(15)^{10}}{10!} e^{-15} = 0.0486$.

Exercise 13.1. Suppose that the number of typos on a single page of a textbook X has a Poisson distribution with parameter $\lambda = 0.5$. In a textbook with 300 pages, around how many pages have at least two errors? Answer: $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - e^{-0.5} - 0.5\, e^{-0.5} = 1 - 1.5\, e^{-0.5} = 0.0902$. Thus, in a 300-page textbook, there are roughly $(300)(0.09) = 27$ pages with at least two typos.

Exercise 13.2. On average, a certain intersection results in three traffic accidents every month.

(a) What is the probability that for any given month, exactly five accidents occur at this intersection? Answer: $X \sim Poi(3)$, $\mathbb{P}(X = 5) = \frac{3^5}{5!} e^{-3} = 0.1008$.

(b) What is the probability that next month there will be fewer than three accidents at this intersection? Answer: $\mathbb{P}(X < 3) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = e^{-3} + 3\,e^{-3} + \frac{3^2}{2} e^{-3} = 8.5\, e^{-3} = 0.4232$.

Exercise 13.3. A certain area of the eastern US is, on average, hit by six hurricanes a year. Find the probability that for a given year this area will be hit by anywhere from six to eight hurricanes. Answer: $X \sim Poi(6)$, $\mathbb{P}(6 \leq X \leq 8) = \mathbb{P}(X = 6) + \mathbb{P}(X = 7) + \mathbb{P}(X = 8) = e^{-6}\left(\frac{6^6}{6!} + \frac{6^7}{7!} + \frac{6^8}{8!}\right) = 0.4016$.

Exercise 13.4. Assume that the average number of cars abandoned weekly on a certain highway is 2.2.

(a) What is the probability that there will be no abandoned cars in the next week? Answer: $X \sim Poi(2.2)$, $\mathbb{P}(X=0) \,=\, e^{-2.2} \,=\, 0.1108$.

(b) What is the standard deviation of the number of abandoned cars in the next week? Answer: $\sqrt{2.2} = 1.4832$.

(c) What is the probability that there be no abandoned cars in the next two weeks? $X \sim Poi((2.2)(2) = 4.4)$, $\mathbb{P}(X=0) \,=\, e^{-4.4} \,=\, 0.0123$.

Exercise 13.5. People enter a casino at a rate of one per every two minutes. Find the probability that exactly four people will enter the casino

(a) between 12 PM and 12:04 PM? Answer: A unit time interval is 2 minutes. We want the information for a 4-minute period, meaning that we are looking at two-unit time intervals. We have that the number of people who enter the casino within a 4-minute period is $X \sim Poi((1)(2) = 2)$, $\mathbb{P}(X=4) \,=\, \frac{2^4}{4!}\, e^{-2} \,=\, 0.0902$.

(b) between 1:00 AM and 1:10 AM? Answer: Let X the number of people who enter the casino within the 10-minute period, then $X \sim Poi((1)(5) = 5)$, and $\mathbb{P}(X=4) \,=\, \frac{5^4}{4!}\, e^{-5} \,=\, 0.1755$.

Exercise 13.6. The earthquakes in California occur at a rate of two per week. Find the probability that at least three earthquakes will occur during

(a) the next two weeks? Answer: $X \sim Poi((2)(2) = 4)$, $\mathbb{P}(X \geq 3) \,=\, 1 \,-\, \mathbb{P}(X = 0) \,-\, \mathbb{P}(X=1) \,-\, \mathbb{P}(X=2) \,=\, 1 \,-\, e^{-4} \,-\, 4\,e^{-4} \,-\, \frac{4^2}{2}\, e^{-4} \,=\, 1 \,-\, 13\, e^{-4} \,=\, 0.7619$.

(b) the next four weeks? Answer: $X \sim ((2)(4) = 8)$, $\mathbb{P}(X \geq 3) \,=\, 1 - e^{-8} \,-\, 8\, e^{-8} \,-\, \frac{8^2}{2}\, e^{-8} \,=\, 1 \,-\, 41\, e^{-8} \,=\, 0.9862$.

14. Chebyshev's Inequality

Theorem. Let X be a random variable with a known mean $\mathbb{E}X = \mu$ (Greek letter "mu" – pronounced "myoo") and a known standard deviation σ (Greek letter "sigma" – pronounced "sig-ma"). For any real-valued $k > 1$, the following inequality is always true:

$$\mathbb{P}(|X - \mu| \,\geq\, k\,\sigma) \,\leq\, \frac{1}{k^2}.$$

Corollary. The inequality involving the complementary probability is also always true. It states:

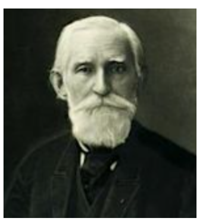
$$\mathbb{P}(|X-\mu| < k\,\sigma) \geq 1 - \frac{1}{k^2}.$$

Proof of the theorem. We write the second central moment (variance) as a sum of two terms under two complementary conditions:

$$\sigma^2 = \mathbb{E}(X-\mu)^2 = \mathbb{E}\Big[(X-\mu)^2\,,\,(X-\mu)^2 \geq k^2\sigma^2\Big] + \mathbb{E}\Big[(X-\mu)^2\,,\,(X-\mu)^2 < k^2\sigma^2\Big]$$
$$\geq \mathbb{E}\Big[(X-\mu)^2\,,\,(X-\mu)^2 \geq k^2\sigma^2\Big] \geq k^2\sigma^2\mathbb{P}\big((X-\mu)^2 \geq k^2\sigma^2\big).$$

From here, $\mathbb{P}\big(|X-\mu| \geq k\sigma\big) = \mathbb{P}\big((X-\mu)^2 \geq k^2\sigma^2\big) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$

Historical Note. This theorem is called **Chebyshev's Inequality** because it was formulated and proved by Pafnuty Lvovich Chebyshev (1821–1894) who was a Russian mathematician and is considered to be the founding father of Russian school of Mathematics.



Example. For $k=2$, we compute $\mathbb{P}(|X-\mu|<2\sigma) \geq 1-\frac{1}{2^2} = \frac{3}{4} = 0.75$. For $k=3$, $\mathbb{P}(|X-\mu|<3\sigma) \geq 1-\frac{1}{3^2} = \frac{8}{9} = 0.89$. It means that for any distribution, at least 75% of observations lie within 2 standard deviations from the mean (in the interval $(\mu-2\sigma,\mu+2\sigma)$), and at least 89% of observations lie within 3 standard deviations from the mean (in the interval $(\mu-3\sigma,\mu+3\sigma)$). Respectively, at most 25% of observations lie more than 2 standard deviations away from the mean in both tails, and at most 11% of observations lie more than 3 standard deviations above or below the mean.

Example. Suppose a random variable X has mean $\mu=3.4$ and standard deviation $\sigma=1.8$. We want to know in what interval around the mean falls at least 60% of observations. By Chebyshev's inequality, we have that $1-\frac{1}{k^2}=0.6$. From here, $k=\sqrt{1/0.4}=\sqrt{2.5}=1.581139$, and thus the interval within which at least 60% of observations fall is $(\mu-k\sigma,\mu+k\sigma) = (3.4-(1.581139)(1.8), 3.4+1.581139)(1.8)) = (0.554, 6.246)$.

Exercise 14.1. Let Y be a random variable with mean 10 and variance 9. Using Chebyshev's inequality, find

(a) a lower bound for $\mathbb{P}(1 < Y < 19)$. Answer: $\sigma = 3$, $\mathbb{P}\big(|Y - 10| < (3)(3)\big) \geq 1 - \frac{1}{3^2} = 0.89$.

(b) the value of α such that $\mathbb{P}(|Y - 10| \geq \alpha) \leq 0.01$. Answer: $\mathbb{P}\big(|Y - 10| \geq \frac{\alpha}{\sigma}\,\sigma\big) = \mathbb{P}\big(|Y - 10| \geq \frac{\alpha}{3}\,(3)\big) \leq \frac{1}{(\alpha/3)^2} = 0.01$. From here, $\alpha = 30$. Thus, at most 1% of observations lie beyond $k = \alpha/3 = 10$ standard deviations from the mean.

Exercise 14.2. The daily production of electric motors at a certain factory averages 120 with a standard deviation of 10.

(a) What can be said about the fraction of days on which the production level falls between 100 and 140? Answer: $\mathbb{P}(100 < X < 140) = \mathbb{P}(|X - 120| < 20) = \mathbb{P}\big(|X - 120| < (2)(10)\big) \geq 1 - \frac{1}{2^2} = 0.75$, so in at least 75% of days production level falls between 100 an 140.

(b) Find the shortest interval certain to contain at least 90% of the daily production levels. Answer: $1 - \frac{1}{k^2} = 0.90$, so $k = \sqrt{10} = 3.162278$. Thus, the shortest interval containing at least 90% of the daily production levels is $\big(120 - (3.162278)(10), 120 + (3.162278)(10)\big) = (88.37, 151.62)$.

Exercise 14.3. Suppose that the distribution of scores on an IQ test has a mean of 100 and a standard deviation of 16. Show that the probability of a student having an IQ of 148 or above, or at 52 and below is at most 1/9. Answer: $\mathbb{P}(X \leq 52 \text{ or } X \geq 148) = \mathbb{P}(|X - 100| \geq 48) = \mathbb{P}\big(|X - 100| \geq (3)(16)\big) \leq \frac{1}{3^2} = \frac{1}{9}$.

Exercise 14.4. Use Chebyshev's inequality to get a lower bound for the number of times a fair coin must be tossed in order for the probability to be at least 0.9 that the ratio of the observed number of heads to the total number of tosses be between 0.4 and 0.6. Answer: Let X be the number of heads, and n be the number of tosses. Then $X \sim Bi(n, 0.5)$. We have $\mu = \mathbb{E}X = (0.5)(n)$, $\sigma^2 = \mathbb{V}ar(X) = (0.5)^2(n)$, and $\sigma = (0.5)\sqrt{n}$. By the Chebyshev's inequality, $\mathbb{P}(0.4 < X/n < 0.6) = \mathbb{P}\big(|X/n - 0.5| < 0.1\big) = \mathbb{P}\big(|X - (0.5)(n)| < (0.1)(n)\big) = \mathbb{P}\big(|X - \mu| < (0.2\sqrt{n})\cdot(0.5)\sqrt{n}\big) = \mathbb{P}\big(|X - \mu| < (0.2\sqrt{n})\sigma\big) \geq 1 - \frac{1}{(0.2\sqrt{n})^2} = 1 - \frac{25}{n} = 0.9$. Hence, $25/n = 1/10$ and so, $n = 250$.

Exercise 14.5. For a certain section of a pine forest, the number of diseased trees per acre Y has a Poisson distribution with mean $\lambda = 10$. The diseased trees are sprayed with an insecticide at a cost of \$3 per tree, plus a fixed overhead cost for equipment rental of \$50. Let C denote the total spraying cost for a randomly selected acre.

(a) Find the expected value and standard deviation for C. Answer: $\mu = \mathbb{E}C = \mathbb{E}(3Y+50) = (3)\mathbb{E}Y + 50 = (3)(10)+50 = 80$, $\mathbb{V}(C) = \mathbb{V}ar(3Y+50) = (9)\mathbb{V}ar(Y) = (9)(10) = 90$, and $\sigma = \sqrt{90} = 9.4868$.

(b) Within what interval would you expect C to lie with probability at least 0.75? Answer: $1 - 1/k^2 = 0.75$, so $k = 2$. Therefore, the interval within which at least 75% of all C's lie is $(\mu - 2\sigma, \mu + 2\sigma) = \big(80 - (2)(9.4868), 80 + (2)(9.4868)\big) = (61.03, 98.97)$.

15. MOMENT GENERATING FUNCTION

Definition. Let X be a random variable. The **moment generating function (mgf)** of X is $M_X(t) = \mathbb{E}e^{tX}$.

Proposition. Let $M_X^{(k)}(t)$ denote the kth derivative of $M_X(t)$. Under this notation, $M_X^{(k)}(0) = \mathbb{E}X^k$, the kth moment of X.

Proof. $M_X'(t) = \big(\mathbb{E}e^{tX}\big)' = \mathbb{E}\big(e^{tX}\big)' = \mathbb{E}(Xe^{tX})$. Here we exchanged expectation and differentiation, which needs justification. We will omit a rigorous proof. Just mention that it is true for all distributions that we consider in this course. Further, we can see that $M_X'(0) = \mathbb{E}(Xe^{(0)(X)}) = \mathbb{E}X$, the first moment of X. Taking the second derivative, we get $M_X''(t) = (M_X'(t))' = \big(\mathbb{E}(Xe^{tX})\big)' = \mathbb{E}(X^2e^{tX})$, and so $M_X''(0) = \mathbb{E}(X^2e^{(0)(X)}) = \mathbb{E}X^2$, the second moment of X. Repeating in the same fashion, it is easy to see that we can derive the k moment as the kth derivative of the mgf computed at zero.

Note. A moment generating function literary generates moments, hence the name.

Example. Let $X \sim Ber(p)$. The mgf of X is $M_X(t) = \mathbb{E}e^{tX} = e^{t(1)}(p) + e^{t(0)}(1-p) = pe^t + 1 - p$. We use the mgf to compute the mean and variance of X. We write $M_X'(t) = (pe^t + 1 - p)' = pe^t$, so the mean is $\mathbb{E}X = M_X'(0) = pe^0 = p$. The second derivative of the mgf is $M_X''(t) = (pe^t)' = pe^t$. Hence, the second moment is $\mathbb{E}X^2 = M_X''(0) = pe^0 = p$. From here, the variance is $\mathbb{V}ar(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1-p)$.

Example. The moment generation function of a hypergeometric distribution doesn't have a closed-form solution. Computation of mgf's for other discrete distributions is left as exercises.

Exercise 15.1. Let $X \sim Bi(n,p)$.
(a) Show that the mgf of X is $M_X(t) = (pe^t + 1 - p)^n$. Answer: $M_X(t) = \mathbb{E}e^{tX} =$

$\sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x} = \big(pe^t+1-p\big)^n$, by Newton's binomial.

(b) From the mgf, obtain the expected value and variance of X. Answer: $M_X'(t) = \Big[(pe^t+1-p)^n\Big]' = n(pe^t+1-p)^{n-1}(pe^t)$. So, $\mathbb{E}X = M_X'(0) = n(pe^0+1-p)^{n-1}(pe^0) = np$. Also, $M_X''(t) = \Big[n(pe^t+1-p)^{n-1}(pe^t)\Big]' = n(n-1)(pe^t+1-p)^{n-2}(pe^t)^2 + n(pe^t+1-p)^{n-1})pe^t$. Hence, the second moment is $\mathbb{E}X^2 = M_X''(0) = n(n-1)(pe^0+1-p)^{n-2}(pe^0)^2 + n(pe^0+1-p)^{n-1}pe^0 = n(n-1)p^2+np$, from where $\mathbb{V}ar(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = n^2p^2 - np^2 + np - (np)^2 = np(1-p)$.

Exercise 15.2. Let $X \sim Geo(p)$.

(a) Show that the mgf of X is $M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$. Answer: $M_X(t) = \sum_{x=1}^{\infty} e^{tx}p(1-p)^{x-1} = \dfrac{p}{1-p}\sum_{x=1}^{\infty}((1-p)e^t)^x = \dfrac{p}{1-p}\cdot\dfrac{(1-p)e^t}{1-(1-p)e^t} = \dfrac{pe^t}{1-(1-p)e^t}$.

(b) Use the mgf to derive the expressions for the mean and variance of X. Answer:

$$M_X'(t) = \frac{(1-(1-p)e^t)(pe^t) - (pe^t)(-(1-p)e^t)}{(1-(1-p)e^t)^2} = \frac{pe^t}{(1-(1-p)e^t)^2},$$

$$M_X''(t) = \Big(\frac{pe^t}{(1-(1-p)e^t)^2}\Big)' = \frac{(1-(1-p)e^t)^2(pe^t) - (pe^t)(2)(1-(1-p)e^t)(-(1-p)e^t)}{(1-(1-p)e^t)^4}$$

$$= \frac{pe^t - 2p(1-p)e^{2t} + p(1-p)^2e^{3t} + 2p(1-p)e^{2t} - 2p(1-p)^2e^{3t}}{(1-(1-p)e^t)^4} = \frac{pe^t - p(1-p)^2e^{3t}}{(1-(1-p)e^t)^4}.$$

Consequently,

$$\mathbb{E}X = M_X'(0) = \frac{pe^0}{(1-(1-p)e^0)^2} = \frac{p}{p^2} = \frac{1}{p},$$

$$\mathbb{E}X^2 = M_X''(0) = \frac{pe^0 - p(1-p)^2e^{(3)(0)}}{(1-(1-p)e^0)^4} = \frac{p - p(1-p)^2}{p^4} = \frac{p - p + 2p^2, -p^3}{p^4} = \frac{2-p}{p^2},$$

and

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

Exercise 15.3. Let $\sim NB(r,p)$.

(a) Prove that the moment generating function of X is $M_X(t) = \Big(\dfrac{pe^t}{1-(1-p)e^t}\Big)^r$. Answer:

$$M_X(t) \;=\; \sum_{x=r}^{\infty} \binom{x-1}{r-1}\, e^{tx}\, p^r (1-p)^{x-r} \;=$$

Let $x = y + r$, so y is the number of failures until the rth success. We write

$$M_X(t) \;=\; \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} e^{t(y+r)} p^r (1-p)^y \;=\; p^r\, e^{tr} \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} ((1-p)e^t)^y$$

$$=\; p^r\, e^{tr} \sum_{y=0}^{-r} (-1)^y \binom{-r}{y} ((1-p)e^t)^y \;=\; p^r\, e^{tr}\, (1-(1-p)e^t)^{-r} \;=\; \Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^r.$$

(b) Obtain the mean and variance of X using the moment generating function. Answer:

$$M_X'(t) \;=\; r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r-1} \Big(\frac{(1-(1-p)e^t)(p\,e^t) \;-\; (p\,e^t)\,(-(1-p)e^t)}{(1-(1-p)e^t)^2}\Big)$$

$$=\; r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r-1} \Big(\frac{p\, e^t}{(1-(1-p)e^t)^2}\Big) \;=\; r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r} \Big(\frac{1}{1-(1-p)e^t}\Big),$$

$$\mathbb{E}X \;=\; M_X'(0) \;=\; r\,\Big(\frac{p\, e^0}{1-(1-p)e^0}\Big)^{r} \Big(\frac{1}{1-(1-p)e^0}\Big) \;=\; \frac{r}{p}.$$

Further,

$$M_X''(t) \;=\; \Big[r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r} \Big(\frac{1}{1-(1-p)e^t}\Big)\Big]'$$

$$=\; r^2\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r} \Big(\frac{1}{(1-(1-p)e^t)^2}\Big) \;+\; r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r} \Big(\frac{-(1-p)e^t}{(1-(1-p)e^t)^2}\Big)$$

$$=\; r\,\Big(\frac{p\, e^t}{1-(1-p)e^t}\Big)^{r} \Big[\frac{r \;-\; (1-p)e^t}{(1-(1-p)e^t)^2}\Big].$$

$$\mathbb{E}X^2 \;=\; M_X''(0) \;=\; r\,\Big(\frac{p\, e^0}{1-(1-p)e^0}\Big)^{r} \Big[\frac{r \;-\; (1-p)e^0}{(1-(1-p)e^0)^2}\Big] \;=\; r\,\frac{r-1+p}{p^2}.$$

The variance is computed as

$$\mathbb{V}ar(X) \;=\; r\,\frac{r-1+p}{p^2} \;-\; \frac{r^2}{p^2} \;=\; \frac{r^2 + r(1-p) - r^2}{p^2} \;=\; \frac{r(1-p)}{p^2}.$$

Exercise 15.4. Let $\sim Poi(\lambda)$.

(a) Show that the mgf of X is $M_X(t) = \exp(\lambda(e^t - 1))$. Answer: $M_X(t) \;=\; \sum_{x=0}^{\infty} e^{tx}\, \frac{\lambda^x}{x!}\, e^{-\lambda} \;=$
$e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda\, e^t)^x}{x!} \;=\; e^{-\lambda}\, e^{\lambda\, e^t} \;=\; \exp(\lambda(e^t - 1)).$

(b) From $M_X(t)$ derive the mean and variance of X. Answer:

$$M'_X(t) = \lambda\, e^t \exp(\lambda(e^t - 1)), \quad M'_X(0) = \mathbb{E}X = \lambda\, e^0 \exp(\lambda(e^0 - 1)) = \lambda,$$
$$M''_X(t) = \Big(\lambda\, e^t \exp(\lambda(e^t - 1))\Big)' = \lambda\, e^t \exp(\lambda(e^t - 1)) + (\lambda\, e^t)^2 \exp(\lambda(e^t - 1)),$$
$$M''_X(0) = \mathbb{E}X^2 = \lambda\, e^0 \exp(\lambda(e^0 - 1)) + (\lambda\, e^0)^2 \exp(\lambda(e^0 - 1)) = \lambda + \lambda^2,$$

and

$$\mathbb{V}ar(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

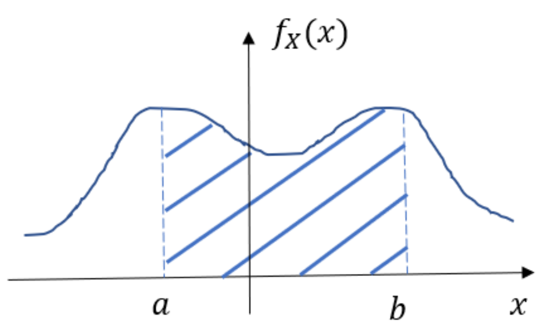
16. CONTINUOUS RANDOM VARIABLE

Definition. A **continuous** random variable assumes values in an interval (it could be an open interval, closed interval, a collection of disjoint intervals, a ray, or an entire real line).

Note. Recall that there are uncountable many points in an interval (called **continuum**), so we cannot assign a non-zero probability to every point in an interval. A proper way to assign probabilities is through a density function.

Definition. A **probability density function (pdf)** is any function $f(x)$ with two properties: $f(x) \geq 0$ for any x, and $\int_{-\infty}^{\infty} f(x)\, dx = 1$. That is, a pdf is always non-zero and the area under the density curve is equal to 1.

Definition. Let X be a continuous random variable with the pdf $f_X(x)$. The **probability** that $a \leq X \leq b$ for some real-valued a and b is computed as the area under the density curve above the interval $[a, b]$. That is, $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$. Note that since the entire area under the curve is equal to one, the probability is well-defined. It will always be a number between 0 and 1. See the illustration below.



To avoid computing an integral every time we need to find a probability, we can define a function that gives us accumulated probability up to a certain point.

Definition. A **cumulative distribution function (cdf)** of a continuous random variable X with the pdf $f_X(x)$ is $F_X(x) = \mathbb{P}(X \leq x) = \displaystyle\int_{-\infty}^{x} f_X(u)\, du.$

Note. We can compute the probability that $a \leq X \leq b$ as the difference of cdfs in the upper point and that in the lower point, That is,

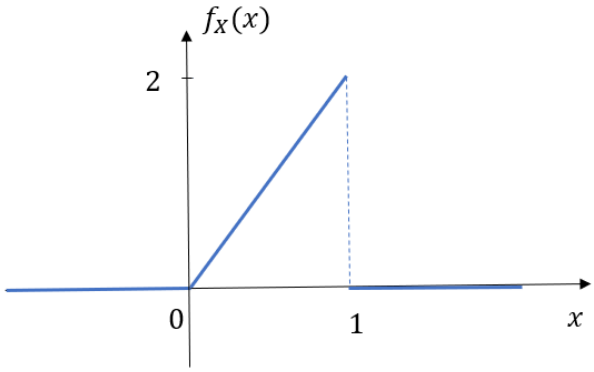
$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx = \int_{-\infty}^b f_X(x)dx - \int_{-\infty}^a f_X(x)dx = F_X(b) - F_X(a).$$

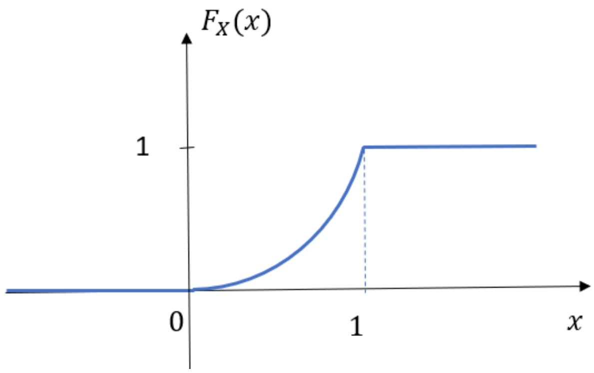
Definition. The **expected value** of a continuous random variable X with pdf $f_X(x)$ is computed as $\mathbb{E}X = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$

Example. Let X have a pdf $f_X(x) = 2x,\ 0 < x < 1$. Note that this density integrates to 1. Indeed, $\int_0^1 2x\, dx = x^2|_0^1 = 1$. The cdf is

$$F_X(x) = \int_{-\infty}^{x} 2u\, du = \begin{cases} 0, & \text{if } x \leq 0, \\ x^2, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

Next, we plot both functions, one underneath the other.





Note that the pdf can have discontinuities whereas the cdf is necessarily a continuous function since it is an integral.

Now we compute the mean, second moment, variance, and standard deviation of X. We write

$$\mathbb{E}X = \int_0^1 x(2x)\,dx = \frac{2}{3}, \;\; \mathbb{E}X^2 = \int_0^1 x^2(2x)\,dx = \frac{1}{2},$$
$$\mathbb{V}ar(X) = \frac{1}{2} - \frac{2^2}{3^2} = \frac{1}{18}, \;\text{ and }\; \sigma = \sqrt{\frac{1}{18}} = 0.2357.$$

Remark. For a continuous random variable $X \sim f_X(x)$, the probability that X is exactly equal to some number is always equal to zero, that is, $\mathbb{P}(X = a) = \int_a^a f_X(x)\,dx = 0$ for any real-valued a. It means, in particular, that when we calculate the probability of X falling within an interval between a and b, it makes no difference whether the end-points of the interval are included or excluded. The probability will still be the same and is computed as follows.

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,dx.$$

Remark. If the cdf $F(x)$ is given, we can find the pdf $f(x)$ by differentiation, that is, $f(x) = F'(x)$.

Example. Let X be a continuous random variable with the cumulative distribution function

$$F(x) = \begin{cases} 0, & x \leq 0, \\ \sqrt{x}, & 0 \leq x \leq 1, \\ 1, & x \geq 1. \end{cases}$$

To find the density function, we compute $f(x) = F'(x) \;=\; \dfrac{1}{2\sqrt{x}},\; 0 \leq x \leq 1.$

Exercise 16.1 $g(x) = ce^{-2x},\ x > 0$. Find c that makes $g(x)$ a pdf. Answer:

$$\int_0^\infty g(x)\,dx \;=\; c\int_0^\infty e^{-2x}\,dx \;=\; c\Big(-\frac{1}{2}e^{-2x}\Big)\Big|_0^\infty \;=\; \frac{c}{2} \;=\; 1, \;\text{ so, }\; c \;=\; 2.$$

Exercise 16.2. The error in the reaction temperature in an experiment is a continuous random variable X having density $f_X(x) \;=\; x^2/3,\;\; -1 < x < 2$.

(a) Show that $f_X(x)$ is a true probability density function. Answer: $f_X(x)$ is everywhere non-negative, and integrates to one:

$$\int_{-1}^2 \frac{x^2}{3}\,dx \;=\; \Big(\frac{x^3}{9}\Big)\Big|_{-1}^2 \;=\; \frac{2^3-(-1)^3}{9} \;=\; \frac{8+1}{9} \;=\; 1.$$

(b) Find the probability that the error is between 0 and 1. Answer:

$$\mathbb{P}(0 < X < 1) \;=\; \int_0^1 \frac{x^2}{3}\,dx \;=\; \frac{1}{9}.$$

(c) Find the average value of the error. Answer:

$$\mathbb{E}X \;=\; \int_{-1}^2 x\cdot\frac{x^2}{3}\,dx \;=\; \int_{-1}^2 \frac{x^3}{3}\,dx \;=\; \Big(\frac{x^4}{12}\Big)\Big|_{-1}^2 \;=\; \frac{2^4-(-1)^4}{12} \;=\; \frac{16-1}{12} \;=\; \frac{15}{12} \;=\; 1.25.$$

(d) Find the variance and standard deviation of the error. Answer:

$$\mathbb{E}X^2 \;=\; \int_{-1}^2 x^2\cdot\frac{x^2}{3}\,dx \;=\; \int_{-1}^2 \frac{x^4}{3}\,dx \;=\; \Big(\frac{x^5}{15}\Big)\Big|_{-1}^2 \;=\; \frac{2^5-(-1)^5}{15} \;=\; \frac{32+1}{15} \;=\; \frac{33}{15} \;=\; 2.2,$$

$\mathbb{V}ar(X) \;=\; \mathbb{E}X^2 - \big(\mathbb{E}X\big)^2 \;=\; 2.2-(1.25)^2 \;=\; 0.6375,$ and $\sigma \;=\; \sqrt{0.6375} \;=\; 0.7984.$

(e) Find the probability that the error is equal to zero. Answer:

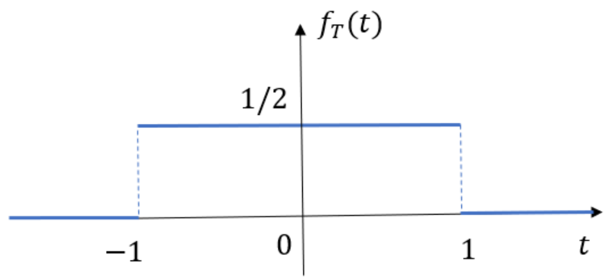
$$\mathbb{P}(X=0) \;=\; \int_0^0 \frac{x^2}{3}\,dx \;=\; 0.$$

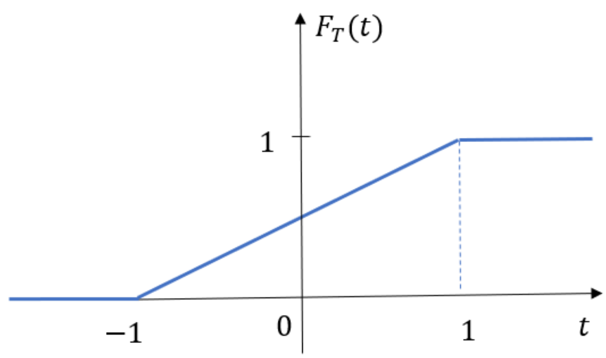
Exercise 16.3. The probability density function of a random variable T is

$$f_T(t) = \begin{cases} 1/2, & \text{if } -1 < t < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Compute the cumulative distribution function of T and make graphs of both functions. Answer: The cdf of T is

$$F_T(t) = \int_{-\infty}^{t} f_T(x)\,dx = \begin{cases} \displaystyle\int_{-\infty}^{t} 0\,dx = 0, & \text{if } t \le -1, \\ \int_{-\infty}^{-1} 0\,dx + \int_{-1}^{t} \frac{1}{2}\,dx = \frac{t+1}{2}, & \text{if } -1 < t < 1, \\ \int_{-\infty}^{t} 0\,dx + \int_{-1}^{1} \frac{1}{2}\,dx + \int_{1}^{t} 0\,dx = 1, & \text{if } t \ge 1. \end{cases}$$





Exercise 16.4. The diameter of a ball bearing produced by a machine is a continuous random variable Y with the cdf

$$F(y) = \begin{cases} 0, & y \le 0, \\ y^{3/2}, & 0 \le y \le 1, \\ 1, & y \ge 1. \end{cases}$$

(a) Find $\mathbb{P}(0.2 < Y < 0.6)$. Answer: $\mathbb{P}(0.2 < Y < 0.6) = F(0.6) - F(0.2) = (0.6)^{3/2} - (0.2)^{3/2} = 0.3753$.

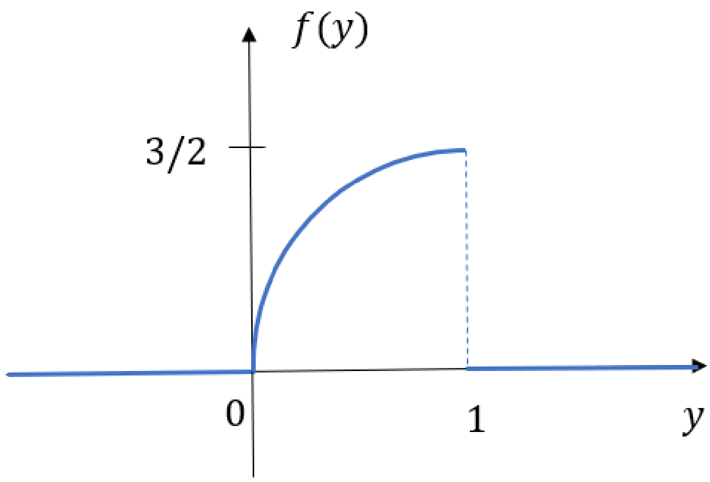
(b) Find $\mathbb{P}(Y > 0.2)$. Answer: $\mathbb{P}(Y > 0.2) = 1 - \mathbb{P}(X \leq 0.2) = 1 - F(0.2) = 1 - (0.2)^{3/2} = 0.9106$.

(c) Find $\mathbb{P}(Y = 0.7)$. Answer: $\mathbb{P}(Y = 0.7) = 0$ because Y is a continuous random variable.

(d) Find $\mathbb{P}(Y = 2)$. Answer: $\mathbb{P}(Y = 2) = 0$ because Y is a continuous random variable.

(e) Compute the pdf of Y and graph it. Answer:

$$f(y) = F'(y) = \frac{3}{2}\sqrt{y}, \quad 0 \leq y \leq 1.$$



Exercise 16.5. Dealer's profit on a new car, in units of \$5,000, is a random variable W having density $f(w) = 2(1-w)$, $0 < w < 1$. Find the average profit per car and the variance of the profit. Answer:

$$\mathbb{E}W = \int_0^1 w \cdot 2(1-w)\,dw = \int_0^1 2w\,dw - \int_0^1 2w^2\,dw = 1 - \frac{2}{3} = \frac{1}{3}.$$

Thus, the average profit per car is $5,000\mathbb{E}W = \$5,000/3 = \$1,6666.67$.

$$\mathbb{V}ar(W) = \int_0^1 w^2 \cdot 2(1-w)\,dw - \left(\frac{1}{3}\right)^2 = \frac{2}{3} - \frac{1}{2} - \frac{1}{9} = \frac{12-9-2}{18} = \frac{1}{18}.$$

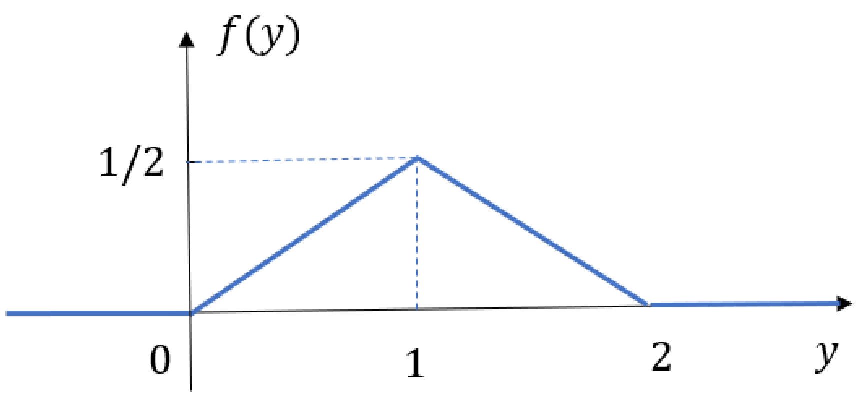
The variance of the profit per car is $(\$5,000)^2/18 = \$^2 1,3888,889.89$.

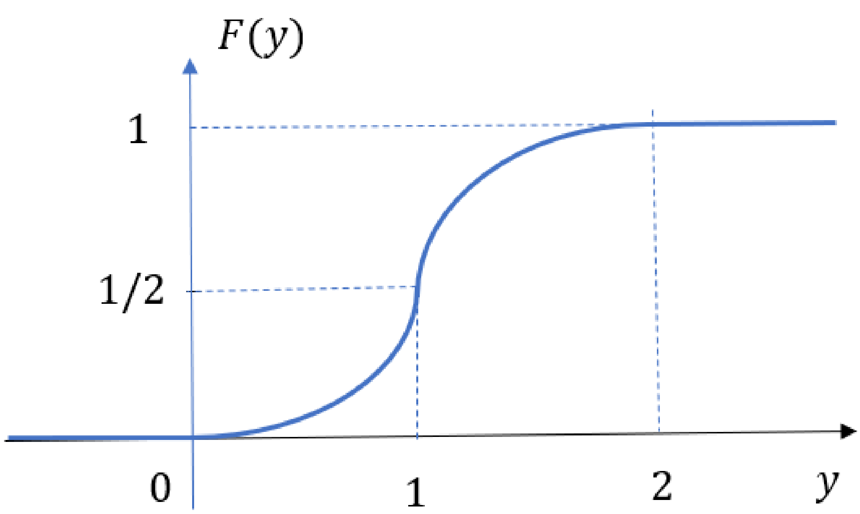
Exercise 16.6. The total number of hours Y, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year has a probability density function

$$f(y) = \begin{cases} 0, & \text{if } y \le 0, \\ y, & \text{if } 0 < y < 1, \\ 2 - y, & \text{if } 1 \le y < 2, \\ 0, & \text{if } y \ge 2. \end{cases}$$

(a) Find the cdf of Y and plot both pdf and cdf. Answer: The cdf is

$$F(y) = \begin{cases} 0, & \text{if } y \le 0, \\ \int_0^y u\,du = \frac{y^2}{2}, & \text{if } 0 < y < 1, \\ \frac{1}{2} + \int_1^y (2-u)\,du = \frac{1}{2} + 2(y-1) - \frac{1}{2}(y^2 - 1) & \\ = -\frac{1}{2}(y-2)^2 + 1, & \text{if } 1 \le y < 2, \\ 1, & \text{if } y \ge 2. \end{cases}$$





(b) Find the expected value of Y. Answer:

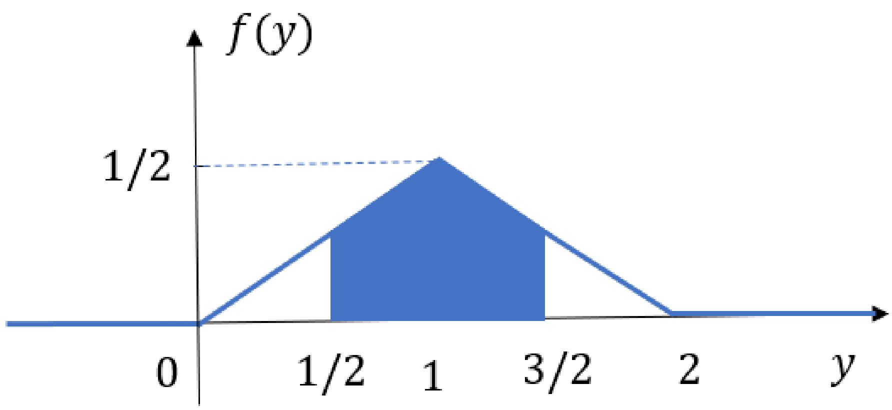
$$\mathbb{E}Y = \int_0^1 y^2\,dy + \int_1^2 y(2-y)\,dy = \frac{1}{3} - 3 - \frac{7}{3} = 1.$$

(c) Find the probability that a family runs a vacuum clean between 50 and 150 hours per year.

$$\mathbb{P}\Big(\frac{1}{2} \le Y \le \frac{3}{2}\Big) = \int_{1/2}^{3/2} f(y)\,dy = \int_{1/2}^{1} y\,dy + \int_1^{3/2} (2-y)\,dy$$
$$= \frac{y^2}{2}\Big|_{1/2}^{1} + \Big(2y - \frac{y^2}{2}\Big)\Big|_1^{3/2} = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}.$$

We can also compute the probability as the area under the density curve. In the figure below, it is the shaded area, which can be computed as one minus the sum of the areas of the two triangles, thus, it is

$$\mathbb{P}\Big(\frac{1}{2} \le Y \le \frac{3}{2}\Big) = 1 - (2)\frac{1}{8} = 1 - \frac{1}{4} = \frac{3}{4}.$$

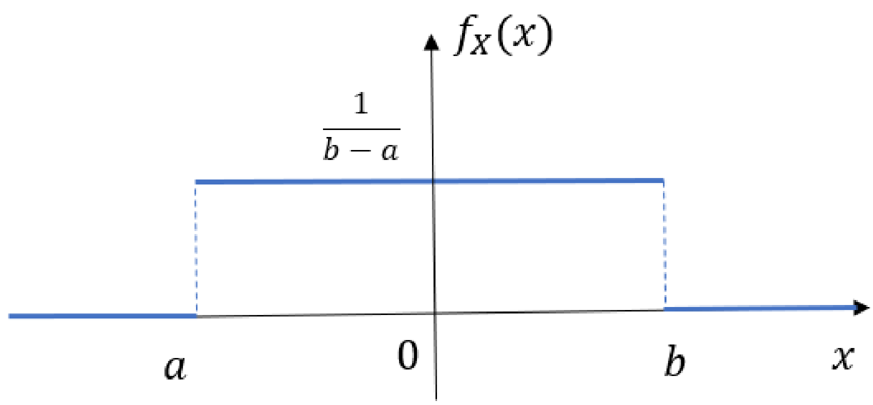


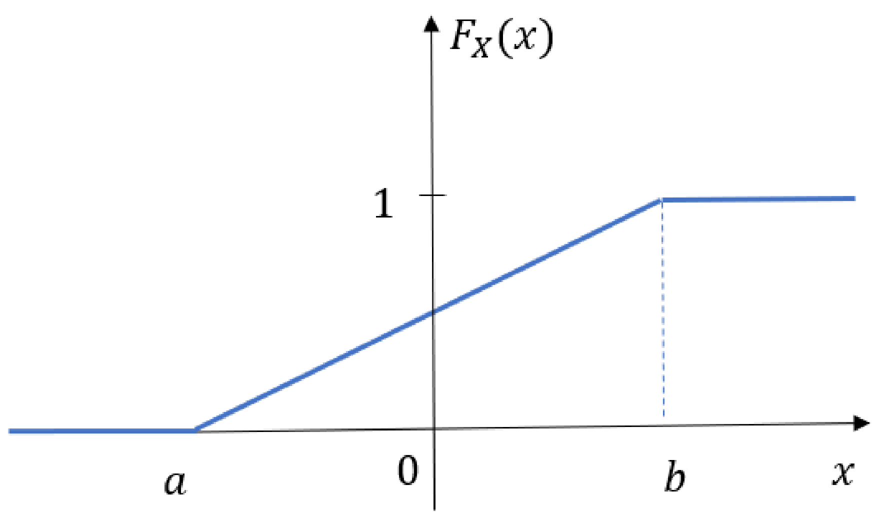
Also, since we already obtained the expression for the cdf, we can use it to compute the probability, $\mathbb{P}\Big(\frac{1}{2} \le Y \le \frac{3}{2}\Big) = F\Big(\frac{3}{2}\Big) - F\Big(\frac{1}{2}\Big) = -\frac{1}{2}\big(\frac{3}{2} - 2\big)^2 + 1 - \frac{1}{2}\cdot\big(\frac{1}{2}\big)^2 = -\frac{1}{8} + 1 - \frac{1}{8} = \frac{3}{4}.$

17. UNIFORM DISTRIBUTION

Definition. A continuous random variable X has a **uniform** distribution on interval $[a,b]$ if its pdf is $f_X(x) = \frac{1}{b-a}$, $a \le x \le b$. It is written as $X \sim Unif(a,b)$. The cdf is found as

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{b-a}\,du = \begin{cases} 0, & \text{if } x \le a, \\ \int_a^x \frac{1}{b-a}\,du = \frac{u}{b-a}\Big|_a^x = \frac{x-a}{b-a}, & \text{if } a \le x \le b, \\ 1, & \text{if } x \ge a. \end{cases}$$





Proposition. Let $X \sim Unif(a,b)$. The mean of X is the middle of the interval, that is, $\mathbb{E}X = \frac{a+b}{2}$. The variance of X is the squared length of the interval divided by 12, that is, $\mathbb{V}ar(X) = \frac{(b-a)^2}{12}$.

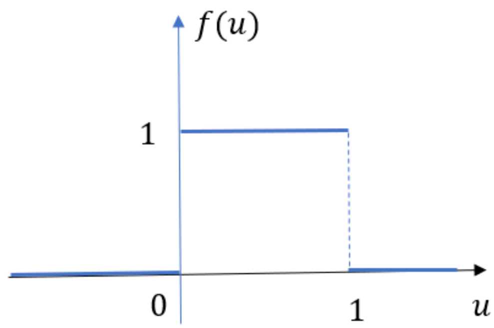
Proof. $\mathbb{E}X = \displaystyle\int_a^b \frac{x}{b-a}\,dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$, and

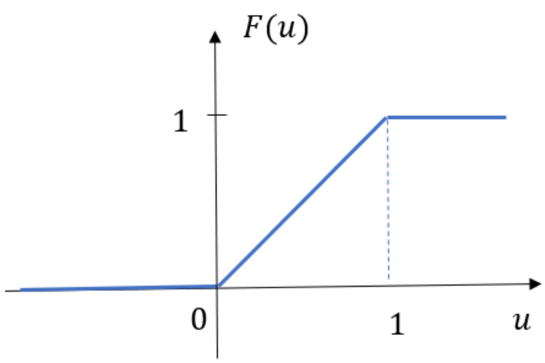
$$\begin{aligned}\mathbb{V}ar(X) &= \int_a^b \frac{x^2}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 = \frac{x^3}{3(b-a)}\Big|_a^b - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{4(a^2+ab+b^2)-3(a^2+2ab+b^2)}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

Example. A **standard uniform** random variable U has a uniform distribution on the unit interval $[0,1]$. For this distribution, $a = 0$ and $b = 1$. The pdf is $f(u) = 1$, if $0 \le u \le 1$, and 0, otherwise. The cdf is

$$F(u) = \begin{cases} 0, & \text{if } u \le 0, \\ u, & \text{if } 0 \le u \le 1, \\ 1, & \text{if } u \ge 1. \end{cases}$$

The expected value of U is $1/2$ and the variance is $1/12$. The pdf and cdf are plotted below.





Note that the notation U is pretty much reserved for a standard uniform random variable. If a random variable is called U, the chances are good that it has a $Unif(0,1)$ distribution.

Remark. The mgf of $X \sim Unif(a,b)$ is $M_X(t) = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{e^{bt} - e^{at}}{t(b-a)}$. The easiest way to derive the mean and variance of X based on the mgf is to apply Taylor's expansion and obtain

$$M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)} = \frac{(1 + bt + \frac{b^2t^2}{2} + \frac{b^3t^3}{6}) - (1 + at + \frac{a^2t^2}{2} + \frac{a^3t^3}{6})}{t(b-a)} + \text{higher order terms}$$

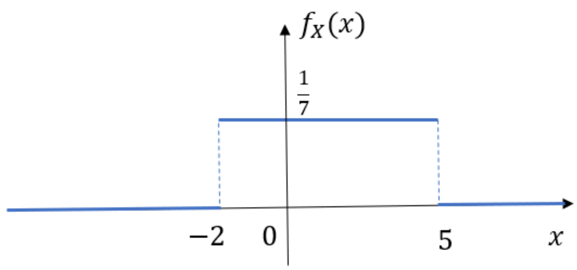
$$= 1 + \frac{a+b}{2} t + \frac{a^2 + ab + b^2}{6} t^2 + \text{higher order terms.}$$

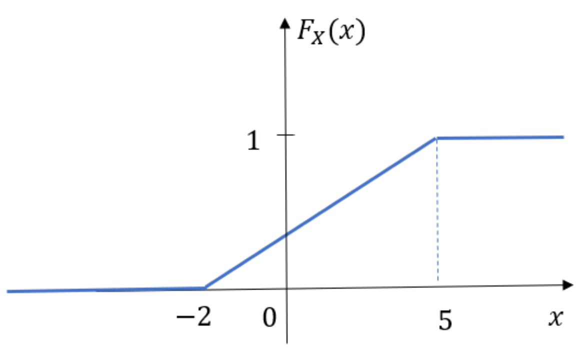
From here, the first moment of X is $\mathbb{E}X = M_X'(0) = \frac{a+b}{2}$, the second moment is $\mathbb{E}X^2 = M_X''(0) = \frac{a^2 + ab + b^2}{3}$, and the variance is $\mathbb{V}ar(X) = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$.

Exercise 17.1. A random variable $X \sim Unif(-2, 5)$.

(a) Write down and plot the pdf and cdf of X. Answer: We are given that $a = -2$ and $b = 5$. The pdf is $f_X(x) = \frac{1}{7}$, if $-2 \le x \le 5$. The cdf is

$$F_X(x) = \begin{cases} 0, & \text{if } x \le -2, \\ \frac{x+2}{7}, & \text{if } -2 \le x \le 5, \\ 1, & \text{if } x \ge 5. \end{cases}$$





(b) Find the mean, variance, and standard deviation of X. Answer: The mean is $\mathbb{E}X = \frac{-2+5}{2} = \frac{3}{2} = 1.5$, the variance is $\mathbb{V}ar(X) = \frac{7^2}{12} = 4.0833$, and the standard deviation is $\sqrt{4.0833} = 2.0207$.

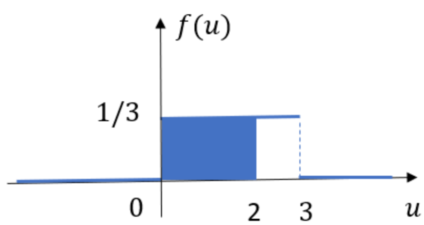
(c) Above what value do the top 10% of all observations lie? Answer: We need to find x_0 so that $\mathbb{P}(X > x_0) = 0.1$ or $F_X(x_0) = 0.9$. We solve $\dfrac{x_0+2}{7} = 0.9$ to get $x_0 = 4.3$.

Exercise 17.2. A 3-foot long stick is cut randomly into two pieces. Let L be the length of the left piece.

(a) Find the pdf and cdf of L. Answer: We have that $L \sim Unif(0,3)$. For this distribution, $a = 0$ and $b = 3$. The pdf is $f_L(x) = \dfrac{1}{3}$, $0 \le x \le 3$. The cdf is

$$F_L(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \dfrac{x}{3}, & \text{if } 0 \le x \le 3, \\ 1, & \text{if } x \ge 3. \end{cases}$$

(b) Find the probability that L is less than two feet long. Do it in three ways: integrating the density, finding the area under the curve geometrically, and using the cdf. Answer: We compute $\mathbb{P}(0 \le L \le 2) = \int_0^2 \frac{1}{3}\,du = \frac{2}{3}$. Alternatively, the probability is equal to the area under the density curve above the interval $[0,2]$ (see the figure below). The area of the shaded rectangle is $\mathbb{P}(0 \le L \le 2) = (2)(1/3) = 2/3$.



The third method of finding the probability is to employ the cdf. We can write $\mathbb{P}(0 \le L \le 2) = F_L(2) - F_L(0) = \frac{2}{3} - 0 = \frac{2}{3}$.

Exercise 17.3. Delta Air Lines quotes a flight time of 2 hours and 5 minutes for its flights from Cincinnati to Tampa. Suppose we believe that actual flight times are uniformly distributed between 2 hours and 2 hours and 20 minutes.

(a) What is the probability that the flight will be more than 10 minutes late? Answer: Let X be the duration of a flight in minutes. Then $X \sim Unif(120, 140)$. If the flight is more than 10 minutes late, its duration is longer than 2 hours and 15 minutes or

135 minutes. We compute $\mathbb{P}(X > 135) = \frac{140-135}{140-120} = \frac{5}{20} = \frac{1}{4} = 0.25$.

(b) Given that the flight is late, find the probability that it is late by more than 10 minutes. Answer:

$$\mathbb{P}(X > 135 \,|\, X > 125) = \frac{\mathbb{P}(X > 135)}{\mathbb{P}(X > 125)} = \frac{(140-135)/(140-120)}{(140-125)/(140-120)} = \frac{5}{15} = \frac{1}{3}.$$

(c) What is the average flight duration? By how many minutes, on average, does a flight arrive late? Answer: $\mathbb{E}X = \frac{120+140}{2} = 130$, which is above 125 by 5 minutes.

(d) What is the standard deviation of flight duration? Answer: $\sigma = \sqrt{\frac{(140-120)^2}{12}} = \frac{10}{\sqrt{3}} = 5.8$ minutes.

Exercise 17.4. A manufacturer of calculators estimates that the lifetime of a calculator is uniformly distributed between 0 and 50 months. If the manufacturer guarantees replacement of calculators that break in the first two months of their lives, about what percent of all calculators are replaced? Answer: Let X be the lifetime of a calculator. We are given that $X \sim Unif(0, 50)$, and we need to find the probability that X is less than 2 months. We compute $\mathbb{P}(X < 2) = \frac{2}{50} = 0.04$ or 4% of calculators are replaced.

Exercise 17.5. A person comes to a bus stop at 7 am. The bus arrives sometime between 7:05 am and 7:20 am.

(a) What is the average waiting time? Answer: Let W be the waiting time. We have $W \sim Unif(5, 20)$. The average of W is $\mathbb{E}W = \frac{5+20}{2} = 12.5$ minutes.

(b) What is the probability that the waiting time is at most 7 minutes? Answer: $\mathbb{P}(W \leq 7) = \frac{7-5}{20-5} = \frac{2}{15} = 0.1333$.

(c) Given that the bus did not arrive for 10 minutes, find the probability that the person has to wait for not more than 5 additional minutes. Answer:

$$\mathbb{P}(W < 15 \,|\, W > 10) = \frac{\mathbb{P}(10 < W < 15)}{\mathbb{P}(W > 10)} = \frac{(15-10)/(20-5)}{(20-10)/(20-5)} = \frac{5}{10} = 0.5.$$

(d) Find the probability that the person waits exactly 5 minutes. Answer: $\mathbb{P}(W = 5) = 0$, since the probability of an exact equality to a value is zero for continuous random variables.

(e) Eighty percent of the time, the waiting time falls below what value? Answer: Let x_0 denote the desired value. We have $\mathbb{P}(X < x_0) = 0.8$. Thus, $\frac{x_0 - 5}{20-5} = 0.8$, from where $x_0 = (0.8)(15) + 5 = 17$.

Exercise 17.6. A balloon has radius $R \sim Uniform(12", 15")$. Find the mean volume of the balloon. Answer: The volume is $V = \frac{4}{3}\pi R^3$. The expected value of the volume is $\mathbb{E}V = \frac{4}{3}\pi \mathbb{E}R^3 = \frac{4}{3}\pi \int_{12}^{15} r^3 \frac{1}{15-12}\, dr = \frac{4}{9}\pi \frac{r^4}{4}\Big|_{r=12}^{r=15} = \frac{\pi}{9}((15)^4 - (12)^4) = 10,433.22$ (inches cubed).

18. EXPONENTIAL DISTRIBUTION

Definition. A continuous random variable X has an **exponential** distribution with parameter β (Greek letter "beta" – pronounced "bey-ta"), written $X \sim Exp(\beta)$ if the pdf of X is $f_X(x) = \frac{1}{\beta} e^{-x/\beta}$, $x > 0, \beta > 0$. The cdf of X is $F_X(x) = \int_0^x \frac{1}{\beta} e^{-u/\beta}\, du = \Big(-e^{-u/\beta}\Big)\Big|_{u=0}^{u=x} = 1 - e^{-x/\beta}$, $x > 0$.

Proposition. Let $X \sim Exp(\beta)$. Then $\mathbb{E}X = \beta$ and $\mathbb{V}ar(X) = \beta^2$.

Proof. The mean of X is $\mathbb{E}X = \int_0^\infty \frac{x}{\beta} e^{-x/\beta}\, dx = \Big(-\frac{x}{\beta}\beta e^{-x/\beta}\Big)\Big|_{x=0}^{x=\infty} + \int_0^\infty e^{-x/\beta}\, dx = \Big(-\beta e^{-x/\beta}\Big)\Big|_{x=0}^{\infty} = \beta$. The variance of X is

$$\begin{aligned}\mathbb{V}ar(X) &= \int_0^\infty \frac{x^2}{\beta} e^{-x/\beta}\, dx - \beta^2 = \Big(-\frac{x^2}{\beta}\beta e^{-x/\beta}\Big)\Big|_{x=0}^{\infty} + \int_0^\infty 2x\, e^{-x/\beta}\, dx - \beta^2 \\ &= 2\int_0^\infty x\, e^{-x/\beta}\, dx - \beta^2 = \Big(-2x\beta e^{-x/\beta}\Big)\Big|_{x=0}^{x=\infty} + 2\beta \int_0^\infty e^{-x/\beta}\, dx - \beta^2 \\ &= 2\beta\Big(-\beta e^{-x/\beta}\Big)\Big|_{x=0}^{x=\infty} - \beta^2 = 2\beta^2 - \beta^2 = \beta^2.\end{aligned}$$

Proposition. Let $X \sim Exp(\beta)$. The mgf of X is $M_X(t) = (1 - \beta t)^{-1}$.

Proof. The moment generating function is computed as

$$\begin{aligned}M_X(t) &= \int_0^\infty e^{tx} \frac{1}{\beta} e^{-x/\beta}\, dx = \frac{1}{\beta}\int_0^\infty e^{-x\left(\frac{1}{\beta}-t\right)}\, dx \\ &= \Big(-\frac{1}{\beta\left(\frac{1}{\beta}-t\right)} e^{-x\left(\frac{1}{\beta}-t\right)}\Big)\Big|_{x=0}^{x=\infty} = \frac{1}{1-\beta t}.\end{aligned}$$

Remark. From the mgf, we can derive the mean and variance of X. We compute

$$M_X'(t) = \Big(\frac{1}{1-\beta t}\Big)' = \frac{\beta}{(1-\beta t)^2}, \quad \mathbb{E}X = M_X'(0) = \frac{\beta}{(1-(\beta)(0))^2} = \beta.$$

Further, $M''_X(t) = \dfrac{2\beta^2}{(1-\beta\,t)^3}$, and so, $\mathbb{E}X^2 = M''_X(0) = \frac{2\beta^2}{1-(\beta)(0))^3} = 2\beta^2$. The variance of X is found as $\mathbb{V}ar(X) = 2\beta^2 - \beta^2 = \beta^2$.

Example. Suppose the time T until remission of disease has an exponential distribution with a mean of $\beta = 6$ weeks. Then $T \sim Exp(6)$ with the pdf $f_T(t) = \dfrac{1}{6}\,e^{-t/6}, \; t > 0$, and the cdf $F_T(t) = 1 - e^{-t/6}$, $t > 0$. The average time remission is $\beta = 6$ weeks, variance is $\beta^2 = 36$ weeks squared, and the standard deviation is $\sqrt{\beta^2} = \beta = 6$ weeks. The probability that the remission occurs within the first, say, 5 weeks is $\mathbb{P}(T \le 5) = F_T(5) = 1 - e^{-5/6} = 0.5654$. The probability of not experiencing remission for 9 weeks is $\mathbb{P}(T > 9) = 1 - F_T(9) = e^{-9/6} = 0.2231$.

Proposition. For $X \sim Exp(\beta)$, the **memoryless property** holds. It states that given that the random variable X is larger than some value, say x, the probability that X is larger than $x + y$, for some real-valued y, depends only on y and not x.

Proof. $\mathbb{P}(X > x + y \,|\, X > x) = \dfrac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > x)} = \dfrac{e^{-(x+y)/\beta}}{e^{-x/\beta}} = e^{-y/\beta}$.

Example. The memoryless property means that, for example, if the waiting time has been x minutes already, the probability that the person has to wait for y additional minutes depends only on y. It doesn't matter that the person has been already waiting for x minutes. The process renews itself every minute, and the remaining waiting time does not depend on the past. The waiting time is memoryless.

Remark. The exponential distribution is the only continuous distribution with the memoryless property. The proof of this fact is beyond the scope of this course. As for discrete distributions, a geometric distribution is the only discrete distribution that possesses the memoryless property. Again, we will not prove the uniqueness here, only the property itself. Let $X \sim Geom(p)$. We write

$$\mathbb{P}(X > n + m \,|\, X > n) = \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m.$$

Remark. Sometimes the pdf of an exponential distribution is written as $f_X(x) = \beta\,e^{-\beta\,x}, \; x > 0, \beta > 0$. In this case $\mathbb{E}X = 1/\beta$ and $\mathbb{V}ar(X) = 1/\beta^2$. To distinguish between the two formulations, it is customary to specify the mean. Typically, the statement would be "X has an exponential distribution with mean so and so". Then in the pdf, we divide by the mean.

Exercise 18.1. The phone calls arriving at a switchboard follow a Poisson distribution with an average of 5 calls per minute. What is the probability that up to 30 seconds will elapse until a call has come? Answer: Let X be the waiting time. The average length of wait is 1/5 of a minute or 12 seconds. Then $X \sim Exp(mean = 12)$. We compute $\mathbb{P}(X < 30) = 1 - e^{-30/12} = 0.9179$.

Exercise 18.2. Treadmills in a gym are occupied for an exponential amount of time with a mean of 17.4 minutes. If a person sees that his favorite treadmill is occupied, what is the probability that he has to wait for at most 10 more minutes to use it? Answer: Let W be the waiting time for this individual. By the memoryless property of an exponential distribution, $W \sim Exp(mean = 17.4)$. We compute $\mathbb{P}(W \leq 10) = 1\text{-}e^{-10/17.4} = 0.4371$.

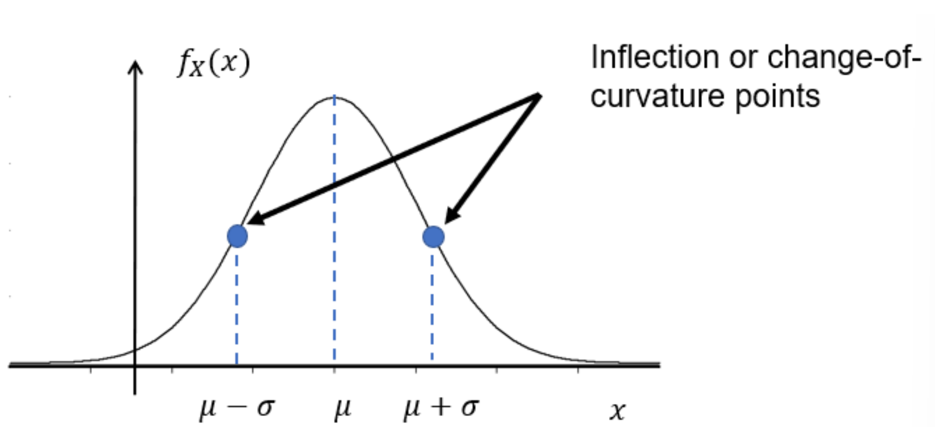
Exercise 18.3. Lifetime T of a component is exponential with mean 5 years. If 6 of these components are installed in different systems, what is the probability that only one of them is still functioning at the end of 8 years? Answer: The probability that a component is functioning is $\mathbb{P}(T > 8) = e^{-8/5} = 0.2019$. Let N be the number of functioning components. Then $N \sim Bi(6, 0.2019)$, and $\mathbb{P}(N = 1) = \binom{6}{1}(0.2019)(1 - 0.2019)^5 = 0.3923$.

Exercise 18.4. The service time T at a bank is an exponentially distributed random variable having a mean of 2 minutes. How many people, on average, should pass until the first person who is served for more than 5 minutes? Answer: $p = \mathbb{P}(T > 5) = e^{-5/2} = 0.0821$. Let X be the number of people until the first success. Then X has a geometric distribution with mean $\mathbb{E}X = \dfrac{1}{p} - 1 = \dfrac{1}{0.0821} - 1 = 11.18$.

19. NORMAL DISTRIBUTION

Definition. A continuous random variable X has a **normal** distribution with mean μ and variance σ^2 if its pdf is $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. It is written $X \sim N(\mu, \sigma^2)$. The cdf of X doesn't have a closed form.

Proposition. The normal density curve is symmetric around the mean μ and is **bell-shaped** with the **inflection points** (also called **change-of-curvature points**) at $\mu - \sigma$ and $\mu + \sigma$ (see the picture).



Proof. To find where the maximum of $f_X(x)$ occurs, we take the first derivative and set is equal to zero. We obtain $f'_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\Big(-\dfrac{x-\mu}{\sigma^2}\Big)\,e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 0$, hence at $x = \mu$, $f_X(x)$ reaches its extremum. To show that it is the maximum, we need to show that the second derivative at this point is negative. We compute

$$f''_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\Big[\Big(-\frac{1}{\sigma^2}\Big) + \Big(\frac{x-\mu}{\sigma^2}\Big)^2\Big]\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Thus, $f''_X(\mu) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\Big[\Big(-\dfrac{1}{\sigma^2}\Big) + \Big(\dfrac{\mu-\mu}{\sigma^2}\Big)^2\Big]\,e^{-\frac{(\mu-\mu)^2}{2\sigma^2}} < 0$, and so, the maximum is reached at $x = \mu$. To find the inflection points, we set the second derivative equal to zero. We have $\dfrac{1}{\sqrt{2\pi\sigma^2}}\Big[\Big(-\dfrac{1}{\sigma^2}\Big) + \Big(\dfrac{x-\mu}{\sigma^2}\Big)^2\Big]\,e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 0$. From here, $\Big(-\dfrac{1}{\sigma^2}\Big) + \Big(\dfrac{x-\mu}{\sigma^2}\Big)^2 = 0$, or $\frac{(x-\mu)^2}{\sigma^2} = 1$, or $x = \mu \pm \sigma$.

Historical Note. A normal distribution was discovered in 1809 by Carl Friedrich Gauss (1777-1855) who was a German mathematician. The normal distribution is sometimes called the Gaussian distribution.

Definition. A normally distributed random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ is said to have a **standard normal** distribution. It is traditionally denoted by Z. Its density function is $f_Z(z) = \dfrac{1}{\sqrt{2\pi}}\,e^{-\frac{z^2}{2}}$, $-\infty < z < \infty$. The cdf of a standard normal distribution is denoted by $\Phi(x)$ (capital Greek letter "phi"- pronounced "fee") and its values are tabulated.

Proposition. Suppose $X \sim N(\mu, \sigma^2)$. Then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

Proof. We start with the cumulative distribution functions and find that

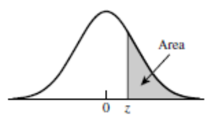
$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}\Big(\frac{X-\mu}{\sigma} \le z\Big) = \mathbb{P}\Big(X \le \mu + \sigma\, z\Big) = F_X(\mu + \sigma\, z).$$

Now we turn to the density functions. We obtain $f_Z(z) = F'_Z(z) = F'_X(\mu + \sigma z) = \sigma f_X(\mu + \sigma z) = \dfrac{\sigma}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu+\sigma z-\mu)^2}{2\sigma^2}} = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \Phi(z)$.

Note. When we subtract the mean and divide by standard deviation, we **standardize** the random variable. We can write $X = \mu + \sigma Z$, so Z tells how many standard deviations away from the mean an observation lies and sometimes is termed the **z-score**.

Example. The life span of a calculator X has a normal distribution with a mean of 54 months and a standard deviation of 8 months. The company guarantees that any calculator that starts malfunctioning within 36 months of the purchase will be replaced. We want to find out what percentage of calculators will be replaced. We write $\mathbb{P}(X < 36) = \mathbb{P}\Big(\dfrac{X-\mu}{\sigma} < \dfrac{36-54}{8}\Big) = \mathbb{P}(Z < -2.25) = \mathbb{P}(Z > 2.25) = 0.0122$, thus, roughly 1.22% of calculators will be replaced.

Table 4 Normal Curve Areas
Standard normal probability in right-hand tail
(for negative values of z, areas are found by symmetry)



	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

From R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

Remark. Even though the direct integration of the normal density is not possible, it still can be shown that it integrates to one. Let $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$. Using a standardizing substitution $z = (x-\mu)/\sigma$, we can rewrite $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$.

Next, we will show that $I^2 = 1$. It will follow that $I = 1$. We have

$$I^2 = \Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\Big)^2 = \Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\Big)\Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy\Big)$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx\, dy.$$

Now, we switch to the polar coordinates $x = r\,cos\,\theta, y = r\,sin\,\theta$, so $x^2 + y^2 = r^2$, and $dxdy = rdrd\theta$ (θ is a Greek letter "theta" - pronounced "they-ta"). We continue

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r\, dr\, d\theta = \int_0^{\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) = 1.$$

Next, we want to show that the parameters μ and σ^2 are indeed the respective mean and variance of a normally distributed random variable.

Proposition. Consider $X \sim N(\mu, \sigma^2)$. The mean of X is $\mathbb{E}X = \mu$ and the variance is $\mathbb{V}ar(X) = \sigma^2$.

Proof. Since $X = \mu + \sigma\, Z$, we have that $\mathbb{E}X = \mu + \sigma\, \mathbb{E}Z$ and $\mathbb{V}ar(X) = \sigma^2\, \mathbb{V}ar(Z)$. Thus, it suffices to show that $\mathbb{E}Z = 0$ and $\mathbb{V}ar(Z) = 1$. We calculate

$$\mathbb{E}Z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\, e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} z\, e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z\, e^{-\frac{z^2}{2}} dz$$
$$= \{y = -z\} = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z\, e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0} (-y)\, e^{-\frac{(-y)^2}{2}} d(-y)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} z\, e^{-\frac{z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} y\, e^{-\frac{y^2}{2}} dy = 0,$$

and

$$\mathbb{V}ar(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2\, e^{-\frac{z^2}{2}} dz = \Big\{u = z\, dv = z\, e^{-\frac{z^2}{2}} dz = e^{-\frac{z^2}{2}} d\Big(\frac{z^2}{2}\Big), du = dz,$$
$$v = -\, e^{-\frac{z^2}{2}}\Big\} = \frac{1}{\sqrt{2\pi}} \Big(-\, z\, e^{-\frac{z^2}{2}}\Big)\Big|_{z=-\infty}^{z=\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = 1.$$

Proposition. Consider $X \sim N(\mu, \sigma^2)$. The moment generating function of X is $M_X(t) = \exp\Big\{\mu\, t + \frac{\sigma^2\, t^2}{2}\Big\}$.

Proof. We derive

$$M_X(t) = \mathbb{E}\big(e^{tX}\big) = \int_{-\infty}^{\infty} e^{tx}\,\frac{1}{2\pi\sigma^2}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} \exp\Big\{-\frac{\big[x^2-2(\mu+t\sigma^2)x+(\mu+\sigma^2 t)^2\big]+\mu^2-(\mu+\sigma^2 t)^2}{2\sigma^2}\Big\}\,dx$$

$$= \exp\Big\{\mu t+\frac{\sigma^2 t^2}{2}\Big\}\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma^2}\exp\Big\{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\Big\}\,dx = \exp\Big\{\mu t+\frac{\sigma^2 t^2}{2}\Big\}.$$

We can use the mgf to obtain the mean and variance of X. We take the first derivative

$$M_X'(t) = (\mu+\sigma^2 t)\,\exp\Big\{\mu t+\frac{\sigma^2 t^2}{2}\Big\},$$

and the second derivative

$$M_X''(t) = \big(\sigma^2+(\mu+\sigma^2 t)^2\big)\,\exp\Big\{\mu t+\frac{\sigma^2 t^2}{2}\Big\}.$$

From here, $\mathbb{E}X = M_X'(0) = \mu$, $\mathbb{E}X^2 = M_X''(0) = \sigma^2+\mu^2$, and $\mathbb{V}ar(X) = \sigma^2+\mu^2-\mu^2 = \sigma^2$.

Exercise 19.1. The diameter of a bearing $X \sim N(3,(0.005)^2)$. The buyer sets specifications on the diameter to be 3.0 ± 0.01 cm. On average, what percentage of bearings will be scrapped? Answer:

$$\mathbb{P}(X<2.99)+\mathbb{P}(X>3.01) = \mathbb{P}\Big(\frac{X-\mu}{\sigma}<\frac{2.99-3}{0.005}\Big)+\mathbb{P}\Big(\frac{X-\mu}{\sigma}>\frac{3.01-3}{0.005}\Big)$$

$$= \mathbb{P}(Z<-2)+\mathbb{P}(Z>2) = (2)\mathbb{P}(Z>2) = (2)(0.0228) = 0.0456.$$

So, about 4.56% of bearings will be scrapped.

Exercise 19.2. The average score for an exam is 74 and the standard deviation is 7. If 12% of the class are given A's, and the grades are assumed to follow a normal distribution, what is the lowest possible A and the highest possible B? (Note: lowest A and highest B are integer numbers). Answer: Let X be a score. It is given that $X \sim N(74,7^2)$. Denote by a the cut-off for an A. It must be that $\mathbb{P}(X>a) = 0.12$. Thus, $\mathbb{P}\Big(\frac{X-\mu}{\sigma}>\frac{a-74}{7}\Big) = \mathbb{P}\Big(Z>\frac{a-74}{7}\Big) = 0.12$, therefore, $\frac{a-74}{7} = 1.175$, and $a = 74+(7)(1.175) = 82.225$. The lowest A is 83 and the highest B is 82.

Exercise 19.3. IQ's of 600 applicants of a college are normally distributed with a mean of 115 and a standard deviation of 13. If the college required an IQ of at least 95, how many of these applicants will be rejected on this basis? Answer: Let X be the IQ of an applicant. We know that $X \sim N(115,(13)^2)$. We calculate $\mathbb{P}(X<95) = \mathbb{P}\Big(Z<\frac{95-115}{13}\Big) = \mathbb{P}(Z<-1.54) = \mathbb{P}(Z>1.54) = 0.0618.$

Hence, about 6.18% of applicants are rejected.

Exercise 19.4. John took an SAT test, for which the scores are $N(490, 75)$, and got 523. His cousin took an ACT test, which scores are $N(18, 6)$, and got 22. Whose score is better? Answer: We standardize both observations and compare their z-scores. John's z-score is $\dfrac{523-490}{75} = 0.44$, meaning that he scored 0.44 standard deviations above the mean. John's cousin's z-score is $\dfrac{22-18}{6} = 0.67$, meaning that John's cousin scored 0.67 standard deviations above the mean. John's cousin scored higher.

Exercise 19.5. The assembly time T of a toy racing car follows a normal distribution with a mean of 55 minutes and a standard deviation of 4 minutes. If a worker starts assembling a toy at 4 pm, what is the probability that she finishes the job by 5 pm? Answer: $\mathbb{P}(T < 60) = \mathbb{P}\Big(Z < \dfrac{60-55}{4}\Big) = \mathbb{P}(Z < 1.25) = 1 - \mathbb{P}(Z > 1.25) = 1 - 0.1056 = 0.8944$.

Exercise 19.6. The stock price S for a company is normally distributed with mean \$300, and the standard deviation \$82.5.

(a) What is the probability that the company will have a stock price of at least \$400? Answer: $\mathbb{P}(S > 400) = \mathbb{P}\Big(Z > \dfrac{400-300}{82.5}\Big) = \mathbb{P}(Z > 1.21) = 0.1131$.

(b) What is the probability that a company will have a stock price no higher than \$150? Answer: $\mathbb{P}(S \leq 150) = \mathbb{P}\Big(Z \leq \dfrac{150-300}{82.5}\Big) = \mathbb{P}(Z \leq -1.82) = \mathbb{P}(Z > 1.82) = 0.0344$.

(c) What is the probability that a company will have a stock price within one standard deviation from the mean? Answer: $\mathbb{P}(-1 < Z < 1) = 1 - (2)\mathbb{P}(Z > 1) = 1 - (2)(0.1587) = 0.6826$.

20. MULTIVARIATE PROBABILITY DISTRIBUTION

Definition. Let X and Y be two discrete random variables. A **joint** (or **bivariate**) **probability distribution function** of X and Y is $p(x, y) = \mathbb{P}(X = x, Y = y)$.

Example. Random variables X and Y have the following joint probability distribution:

		x	
	$p(x,y)$	2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

First, we would like to verify that it is indeed a legitimate joint probability distribution. We check that all the given probabilities add up to 1. Indeed, $0.10+0.15+0.20+0.30+0.10+0.15 = 1$. Next, we see that, for example, $\mathbb{P}(X = 2, Y = 3) = 0.2$, and $\mathbb{P}(X = 4, Y > 1) = 0.30 + 0.15 = 0.45$.

Definition. Let X and Y be two continuous random variables. The **joint cumulative probability function (cdf)** is $F(x,y) = \mathbb{P}(X \le x, Y \le y)$. The **joint probability density function (pdf)** is $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$. Given $f(x,y)$, we can find the joint cdf as $F(x,y) = \displaystyle\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du$.

Example. Two continuous random variables X and Y have a joint density $f(x,y) = \frac{2}{3}(x+2y)$, $0 \le x, y \le 1$. First off, we want to check that it is a true joint density function. We need to verify that it integrates to one. We compute

$$\int_0^1\int_0^1 \frac{2}{3}(x+2y)\,dy\,dx = \int_0^1 \Big(\frac{2xy}{3} + \frac{2y^2}{3}\Big)\Big|_{y=0}^{y=1} dx = \int_0^1 \Big(\frac{2x}{3} + \frac{2}{3}\Big)\,dx$$
$$= \Big(\frac{x^2}{3} + \frac{2x}{3}\Big)\Big|_0^1 = \frac{1}{3} + \frac{2}{3} = 1.$$

Now we can use this joint density to compute, for example,

$$\mathbb{P}\Big(X < \frac{1}{2}, Y > \frac{1}{3}\Big) = \int_0^{1/2}\int_{1/3}^{1} \frac{2}{3}(x+2y)\,dy\,dx = \int_0^{1/2} \Big(\frac{2xy}{3} + \frac{2y^2}{3}\Big)\Big|_{y=1/3}^{y=1} dy$$
$$= \int_0^{1/2} \Big(\frac{2x}{3} + \frac{2}{3} - \frac{2x}{9} - \frac{2}{27}\Big)\,dy = \int_0^{1/2} \Big(\frac{4x}{9} + \frac{16}{27}\Big)\,dx$$
$$= \Big(\frac{2x^2}{9} + \frac{16x}{27}\Big)\Big|_0^{1/2} = \frac{1}{18} + \frac{8}{27} = \frac{19}{54}.$$

Further, we can calculate the joint cdf. It is done as follows.

$$F(x,y) = \int_0^x\int_0^y \frac{2}{3}(u+2v)\,dv\,du = \frac{2}{3}\int_0^x (uv + v^2)\Big|_{v=0}^{v=y} du = \frac{2}{3}\int_0^x (uy + y^2)\,du$$
$$= \frac{2}{3}\Big(\frac{u^2 y}{2} + u y^2\Big)\Big|_{u=0}^{u=x} = \frac{2}{3}\Big(\frac{x^2 y}{2} + xy^2\Big) = \frac{1}{3}xy(x+2y), \ \ 0 \le x, y \le 1.$$

Note that $F(1,1) = 1$, as it should be.

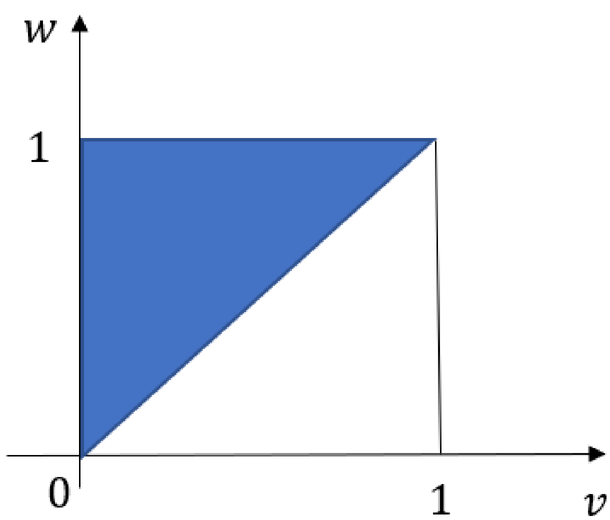
Exercise 20.1. Suppose $p(x,y) = cxy,\ x = 1,2,3,\ y = 1,2,3, x \le y$, is a joint probability distribution of X and Y.

(a) Determine the normalizing constant c. Answer: $\sum_x \sum_y p(x,y) = c\big[(1)(1) + (1)(2) + (1)(3) + (2)(2) + (2)(3) + (3)(3)\big] = 25c = 1$, so $c = 1/25 = 0.04$.

(b) Compute the probability that $X < Y$. Answer: $\mathbb{P}(X < Y) = (0.04)\big[(1)(2) + (1)(3) + (2)(3)\big] = (0.04)(11) = 0.44$.

Exercise 20.2. The joint probability density of V and W is given by $f(v,w) = 2,\ \ 0 < v < w < 1$.

(a) Check that it is indeed a joint density. Answer: We need to show that $f(v,w)$ integrates to 1. The area of integration is the upper triangle that is shaded in this picture.



We can integrate vertically as follows

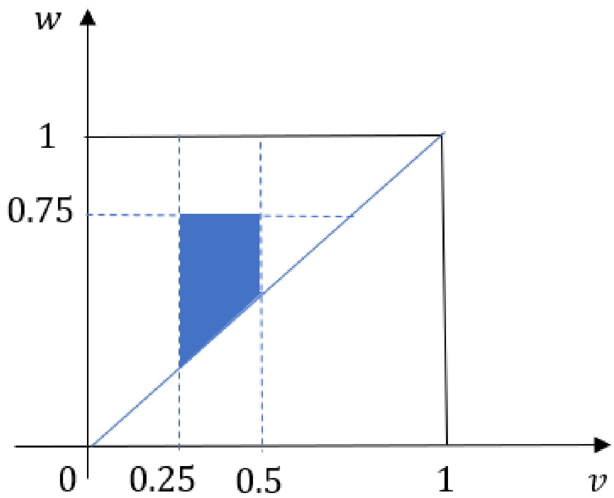
$$\int_0^1 \int_v^1 2\,dw\,dv \;=\; \int_0^1 \big(2w\big)\Big|_{w=v}^{w=1} dv \;=\; \int_0^1 2(1-v)\,dv \;=\; \big(2v - v^2\big)\Big|_0^1 \;=\; 1.$$

Or we can integrate horizontally. We calculate

$$\int_0^1 \int_0^w 2\,dv\,dw \;=\; \int_0^1 \big(2v\big)\Big|_{v=0}^{v=w} dv \;=\; \int_0^1 2w\,dw \;=\; w^2\Big|_0^1 \;=\; 1.$$

Alternatively, we can notice that we can find the double integral just by multiplying the constant density by the area of the shaded triangle to get (2)(1/2)=1.

(b) Compute $\mathbb{P}(0.25 < V < 0.5,\, W \le 0.75)$. Answer: We need to integrate the density over the shaded region (see the picture).



Integration is easier to be done vertically. It requires computing only one integral as opposed to two integrals if we integrate horizontally. We write

$$\mathbb{P}(0.25 < V < 0.5,\, W \le 0.75) \;=\; \int_{0.25}^{0.5}\int_{v}^{0.75} 2\,dw\,dv \;=\; \int_{0.25}^{0.5} 2w\Big|_{w=v}^{w=0.75}\,dv$$

$$= \int_{0.25}^{0.5} 2(0.75-v)\,dv \;=\; (1.5v - v^2)\Big|_{0.25}^{0.5} \;=\; (1.5)(0.5-0.25) \;-\; ((0.5)^2 - (0.25)^2) \;=$$

$$= 0.375 \;-\; 0.1875 \;=\; 0.1875.$$

(c) Find the joint cdf and use it to compute $\mathbb{P}(0.25 < V < 0.5,\, W \le 0.75)$. Answer:

$$F(v,w) \;=\; \int_0^v\int_x^w 2\,dy\,dx \;=\; \int_0^v 2(w-x)\,dx \;=\; (2wx - x^2)\Big|_{x=0}^{x=v} \;=\; v(2w-v), \quad 0 < w < v < 1.$$

From here,

$$\mathbb{P}(0.25 < V < 0.5,\, W \le 0.75) \;=\; F(0.5, 0.75) \;-\; F(0.25, 0.75)$$
$$= (0.5)((2)(0.75) - 0.5) - (0.25)((2)(0.75) - 0.25) \;=\; 0.1875.$$

Exercise 20.3. The bivariate probability distribution of X and Y is

$$p(x,y) = \frac{x+y}{30}, \quad x = 0,1,2,3, \quad y = 0,1,2.$$

(a) Check that the joint probability function sums up to 1. Answer: $\sum_{x=0}^{3}\sum_{y=0}^{2} p(x,y) = \frac{1}{30}\big((0+0)+(0+1)+(0+2)+(1+0)+(1+1)+(1+2)+(2+0)+(2+1)+(2+2)+(3+0)+(3+1)+(3+2)\big) = \frac{30}{30} = 1.$

(b) Find $\mathbb{P}(X \leq 2,\, Y = 1)$. Answer: $\frac{1}{30}\left((0+1)+(1+1)+(2+1)\right) = \frac{6}{30} = 0.2.$

(c) Find $\mathbb{P}(X > 2,\, Y \leq 1)$. Answer: $\frac{1}{30}((3+0)+(3+1)) = \frac{7}{30} = 0.2333.$

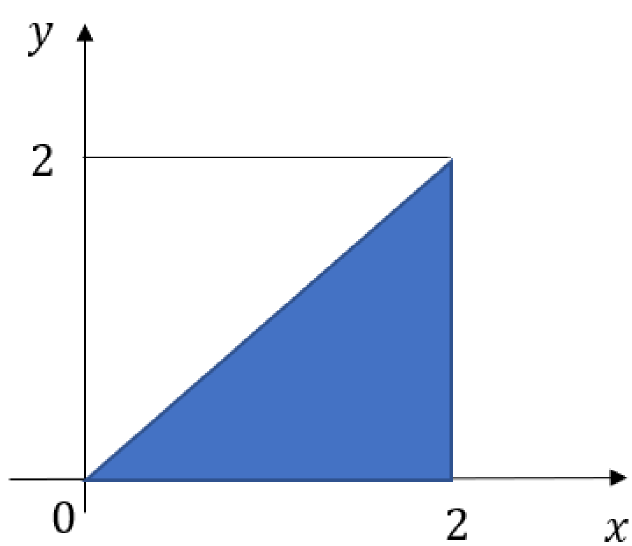
(d) Find $\mathbb{P}(X > Y)$. Answer: $\frac{1}{30}((1+0)+(2+0)+(3+0)+(2+1)+(3+1)+(3+2)) = \frac{18}{30} = 0.6.$

(e) Find $\mathbb{P}(X = Y)$. Answer: $\frac{1}{30}((0+0)+(1+1)+(2+2))\frac{6}{30} = 0.2.$

(f) Find $\mathbb{P}(X + Y = 4)$. Answer: $\frac{1}{30}((2+2)+(3+1)) = \frac{8}{30} = 0.2667.$

Exercise 20.4. A joint density is $f(x,y) = cxy,\;\; 0 < y < x < 2.$

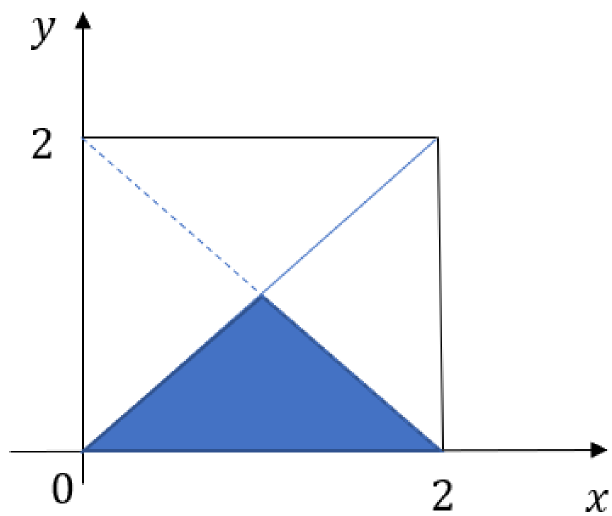
(a) Find the normalizing constant c that makes the total volume under the density graph equal to one. Answer: The joint density is defined over the shaded region as shown in the picture below.



We calculate $\int_0^2 \int_0^x cxy\,dy\,dx \;=\; c\int_0^2 \left(\frac{xy^2}{2}\right)\Big|_{y=0}^{y=x} dx \;=\; c\int_0^2 \frac{x^3}{2}\,dx \;=\; c\left(\frac{x^4}{8}\right)\Big|_0^2 =$

$2c = 1$, so $c = \dfrac{1}{2}$.

(b) Find the probability that $X + Y < 2$. Answer: We need to integrate the density over the shaded region depicted here:



It is more convenient to integrate horizontally. We write

$$\int_0^1 \int_y^{2-y} \frac{1}{2} xy \, dx \, dy = \frac{1}{2} \int_0^1 \left(\frac{x^2 y}{2} \right) \Big|_{x=y}^{x=2-y} dy$$

$$= \frac{1}{4} \int_0^1 \left(y(2-y)^2 - y^3 \right) dy = \int_0^1 (y - y^2) \, dy = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

21. MARGINAL AND CONDITIONAL DISTRIBUTIONS

Definition. Let X and Y be two discrete random variables with the joint probability distribution function $p(x, y)$. The **marginal probability distribution functions** are distribution functions of X alone and Y alone, respectively. They are computed as $p_X(x) = \sum_y p(x, y)$ and $p_Y(y) = \sum_x p(x, y)$.

Note. Marginal distribution functions have this name because, in a two-way table, they would appear as the marginal totals (written on the margins of the table).

Example. In the two-way table example that we have seen earlier, we compute the marginal distributions, which are the row and column totals. We get

		x		
	$p(x,y)$	2	4	$p_Y(y)$
	1	0.10	0.15	0.25
y	3	0.20	0.30	0.50
	5	0.10	0.15	0.25
	$p_X(x)$	0.4	0.6	

We can see that $p_X(2) = 0.4, p_X(4) = 0.6$, $p_Y(1) = 0.25, p_Y(3) = 0.5$, and $p_Y(5) = 0.25$.

Definition. Let $p(x, y)$ be the joint probability distribution function for two discrete random variables X and Y. A **conditional distribution function** of X given Y is

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

Example. In the previous example, we can compute

$$p_{X|Y}(2|1) = \frac{p(2,1)}{p_Y(1)} = \frac{0.1}{0.25} = 0.4, \quad p_{X|Y}(4|1) = \frac{p(4,1)}{p_Y(1)} = \frac{0.15}{0.25} = 0.6,$$

$$p_{X|Y}(2|3) = \frac{p(2,3)}{p_Y(3)} = \frac{0.2}{0.5} = 0.4, \quad p_{X|Y}(4|3) = \frac{p(4,3)}{p_Y(3)} = \frac{0.3}{0.5} = 0.6,$$

and

$$p_{X|Y}(2|5) = \frac{p(2,5)}{p_Y(5)} = \frac{0.1}{0.25} = 0.4, \quad p_{X|Y}(4|5) = \frac{p(4,5)}{p_Y(5)} = \frac{0.15}{0.25} = 0.6.$$

Also, $p_{Y|X}(1|2) = \dfrac{p(2,1)}{p_X(2)} = \dfrac{0.1}{0.4} = 0.25, \quad p_{Y|X}(3|2) = \dfrac{p(2,3)}{p_X(2)} = \dfrac{0.2}{0.4} = 0.5,$

$$p_{Y|X}(5|2) = \frac{p(2,5)}{p_X(2)} = \frac{0.1}{0.4} = 0.25, p_{Y|X}(1|4) = \frac{p(4,1)}{p_X(4)} = \frac{0.15}{0.6} = 0.25,$$

$$p_{Y|X}(3|4) = \frac{p(4,3)}{p_X(4)} = \frac{0.3}{0.6} = 0.5, \text{ and } p_{Y|X}(5|4) = \frac{p(4,5)}{p_X(4)} = \frac{0.15}{0.6} = 0.25.$$

Definition. Let X and Y be two continuous random variables with the joint density function $f(x, y)$. The **marginal density functions** of X and Y are:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \;\text{ and }\; f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

Example. Two continuous random variables X and Y have a joint density $f(x, y) = \frac{2}{3}(x + 2y)$, $0 \le x, y \le 1$. The marginal distribution function of X is

$$f_X(x) = \int_0^1 \frac{2}{3}(x+2y)\,dy = \left(\frac{2xy}{3} + \frac{2\,y^2}{3}\right)\Big|_{y=0}^{y=1} = \frac{2}{3}(x+1), \quad 0 < x < 1.$$

The marginal distribution function of Y is

$$f_Y(y) = \int_0^1 \frac{2}{3}(x+2y)\,dx = \left(\frac{x^2}{3} + \frac{4xy}{3}\right)\Big|_{x=0}^{x=1} = \frac{1}{3}(4y+1), \ \ 0<y<1.$$

Definition. Let $f(x,y)$ be the joint probability density function for two continuous random variables X and Y. A **conditional probability density function** of X given Y is $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$.

Example. In the previous example, we can compute the conditional densities as

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{2}{3}(x+2y)}{\frac{1}{3}(4y+1)} = \frac{2(x+2y)}{4y+1}, \ \ 0<x<1.$$

and

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{2}{3}(x+2y)}{\frac{2}{3}(x+1)} = \frac{x+2y}{x+1}, \ \ 0<y<1.$$

We use the conditional densities to compute, for example,

$$\mathbb{P}(X \le 0.5\,|\,Y=0.25) = \int_0^{0.5} f_{X|Y}(x|0.25)\,dx = \int_0^{0.5} \frac{2(x+2(0.25))}{4(0.25)+1}\,dx$$
$$= \frac{1}{2}\int_0^{0.5}(2x+1)\,dx = \frac{1}{2}(x^2+x)\Big|_0^{0.5} = \frac{1}{2}\left(\frac{1}{4}+\frac{1}{2}\right) = \frac{3}{8},$$

and

$$\mathbb{P}(Y \le 0.5\,|\,X=0.25) = \int_0^{0.5} f_{Y|X}(y|0.25)\,dy = \int_0^{0.5} \frac{0.25+2y}{0.25+1}\,dy$$
$$= \frac{1}{5}\int_0^{0.5}(8y+1)\,dy = \frac{1}{5}(4y^2+y)\Big|_0^{0.5} = \frac{1}{5}\left((4)\frac{1}{4}+\frac{1}{2}\right) = \frac{1}{5}+\frac{1}{10} = \frac{3}{10}.$$

We conclude this example with the following computation:

$$\mathbb{P}\left(X < \frac{1}{2}\,|\,Y > \frac{1}{3}\right) = \int_0^{1/2}\int_{1/3}^1 \frac{2(x+2y)}{4y+1}\,dy\,dx$$
$$= \int_0^{1/2}\int_{1/3}^1 \left(\frac{2x-1}{4y+1}+1\right)dy\,dx = \int_0^{1/2}\left(\frac{2x-1}{4}\ln(4y+1) + y\right)\Big|_{y=1/3}^{y=1}dx$$
$$= \int_0^{1/2}\left[\frac{2x-1}{4}\ln\left(\frac{15}{7}\right)+\frac{2}{3}\right]dx = \left[(x^2-x)\frac{1}{4}\ln\left(\frac{15}{7}\right)+\frac{2x}{3}\right]\Big|_0^{1/2}$$
$$= \frac{1}{3} - \frac{1}{16}\ln\left(\frac{15}{7}\right) = 0.2857.$$

Exercise 21.1. The bivariate probability density of X and Y is

$$f(x,y) = \frac{6-x-y}{8}, \;\; 0 \,<\, x \,<\, 2, \; 2 \,<\, y \,<\, 4.$$

(a) Compute the marginal probability densities of X and Y. Answer: We write

$$f_X(x) \;=\; \int_2^4 \frac{6-x-y}{8}\,dy \;=\; \frac{6y-xy-y^2/2}{8}\Big|_{y=2}^{y=4} \;=\; \frac{3-x}{4}, \;\; 0 \,<\, x \,<\, 2,$$

and

$$f_Y(y) \;=\; \int_0^2 \frac{6-x-y}{8}\,dx \;=\; \frac{6x-x^2/2-xy}{8}\Big|_{x=0}^{x=2} \;=\; \frac{5-y}{4}, \;\; 2 \,<\, y \,<\, 4.$$

(b) What is $\mathbb{P}(X \geq 1 \,|\, Y=3)$? Answer: $f_{X|Y}(x|y) \;=\; \dfrac{f(x,y)}{f_Y(y)} \;=\; \dfrac{\frac{6-x-y}{8}}{\frac{5-y}{4}}$, $f_{X|Y}(x|3) \;=\; \dfrac{\frac{6-x-3}{8}}{\frac{5-3}{4}} \;=\; \dfrac{3-x}{4}$, $\mathbb{P}(X \geq 1 \,|\, Y=3) \;=\; \displaystyle\int_0^1 \frac{3-x}{4}\,dx \;=\; \frac{3x-x^2/2}{4}\Big|_0^1 \;=\; \frac{5}{8}$.

(c) What is $\mathbb{P}(2 < Y < 3 \,|\, X=2)$? Answer: $f_{Y|X}(y|x) \;=\; \dfrac{f(x,y)}{f_X(x)} \;=\; \dfrac{\frac{6-x-y}{8}}{\frac{3-x}{4}}$, $f_{Y|X}(y|2) \;=\; \dfrac{\frac{6-x-3}{8}}{\frac{5-3}{4}} \;=\; \dfrac{4-y}{2}$, $\mathbb{P}(2 < Y < 3 \,|\, X=2) \;=\; \displaystyle\int_2^3 \frac{4-y}{2}\,dy \;=\; \frac{1}{2}(4y-y^2/2)\Big|_2^3 \;=\; \frac{3}{4}$.

Exercise 21.2. The bivariate probability distribution of X and Y is

$$p(x,y) = \frac{x+y}{30}, \;\; x = 0,1,2,3, \;\; y = 0,1,2.$$

(a) Find the marginal pmf $p_X(x)$. Answer:

$$p_X(x) \;=\; \frac{x+0}{30} \;+\; \frac{x+1}{30} \;+\; \frac{x+2}{30} \;=\; \frac{x+1}{10}, \;\; x = 0,1,2,3.$$

(b) Find the marginal pmf $p_Y(y)$. Answer:

$$p_Y(y) \;=\; \frac{0+y}{30} \;+\; \frac{1+y}{30} \;+\; \frac{2+y}{30} \;+\; \frac{3+y}{30} \;=\; \frac{2y+3}{15}, \;\; x = 0,1,2,3.$$

(c) Compute $\mathbb{P}(X > 1 \,|\, Y=1)$. Answer:

$$p_{X|Y}(x|y) \;=\; \frac{p(x,y)}{p_Y(y)} \;=\; \Big(\frac{x+y}{30}\Big)/\Big(\frac{2y+3}{15}\Big) \;=\; \frac{x+y}{4y+6}, \;\; p_{X|Y}(x|1) \;=\; \frac{x+1}{(4)(1)+6} \;=\; \frac{x+1}{10},$$

$$\mathbb{P}(X > 1 \,|\, Y=1) \;=\; p_{X|Y}(2|1) \;+\; p_{X|Y}(3|1) \;=\; \frac{2+1}{10} \;+\; \frac{3+1}{10} \;=\; \frac{7}{10}.$$

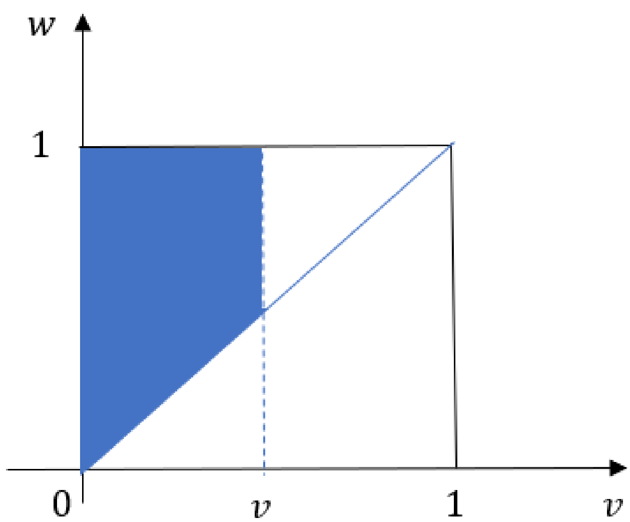
(d) Compute $\mathbb{P}(Y < 2 \,|\, X=2)$. Answer:

$$p_{Y|X}(y|z) \;=\; \frac{p(x,y)}{p_X(x)} \;=\; \Big(\frac{x+y}{30}\Big)/\Big(\frac{x+1}{10}\Big) \;=\; \frac{x+y}{3x+3}, \;\; p_{Y|X}(y|2) \;=\; \frac{2+y}{(3)(2)+3} \;=\; \frac{2+y}{9},$$

$$\mathbb{P}(Y < 2 \,|\, X = 2) \;=\; p_{Y|X}(0|2) \;+\; p_{Y|X}(1|2) \;=\; \frac{2+0}{9} \;+\; \frac{2+1}{9} \;=\; \frac{5}{9}.$$

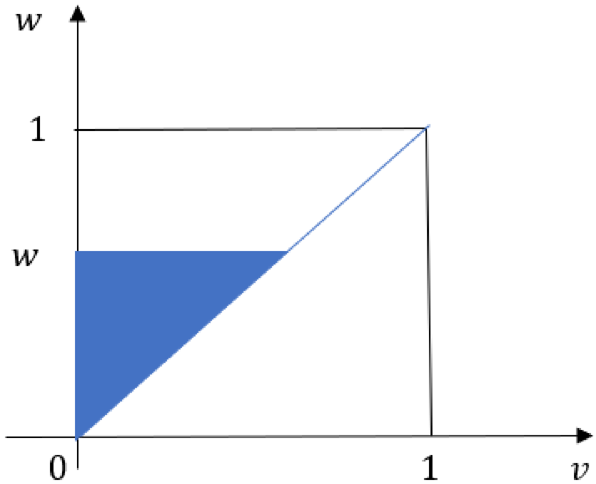
Exercise 21.3. The joint probability density of V and W is given by $f(v,w) = 2,\;\; 0 < v < w < 1.$

(a) Find the marginal density of V. Answer: We fix a value of v, and integrate vertically over the shaded region shown in the picture:



$$f_V(v) \;=\; \int_v^1 2\,dw \;=\; 2(1-v), \;\; 0 \;<\; v \;<\; 1.$$

(b) Find the marginal density of W. Answer: For a fixed w, we integrate horizontally over the shaded region as depicted below:



$$f_W(w) \;=\; \int_0^w 2\,dv \;=\; 2w, \;\; 0 \;<\; w \;<\; 1.$$

(c) Compute $\mathbb{P}(V < 0.5 \,|\, W = 0.75)$. Answer: we calculate

$$\mathbb{P}(V < 0.5 \,|\, W = 0.75) \;=\; \int_0^{0.5} f_{V|W}(v, 0.75)\,dv \;=\; \int_0^{0.5} \frac{f(v, 0.75)}{f_W(0.75)}\,dv$$

$$= \int_0^{0.5} \frac{2}{(2)(0.75)}\,dv \;=\; \left(\frac{4}{3}\right)\left(\frac{1}{2}\right) \;=\; \frac{2}{3}.$$

22. INDEPENDENCE AND COVARIANCE

Definition. Two discrete random variables X and Y are **independent** if and only if $p(x, y) \;=\; p_X(x)\,p_Y(y)$.

Example. Random variables X and Y have the following joint probability distribution:

		x	
	$p(x,y)$	2	4
y	1	0.10	0.15
	3	0.20	0.30
	5	0.10	0.15

We need to find out whether these random variables are independent. First, we compute the marginal distributions:

		x		
	$p(x,y)$	2	4	$p_Y(y)$
	1	0.10	0.35	0.45
y	3	0.20	0.10	0.30
	5	0.10	0.15	0.25
	$p_X(x)$	0.4	0.6	

We can see that $p_X(2) = 0.4, p_X(4) = 0.6$, $p_Y(1) = 0.45, p_Y(3) = 0.30$, and $p_Y(5) = 0.25$. Now, we multiply the marginal distribution functions and compare the product with the joint probability function. If they are equal for all cells in the table, then the random variables are independent. If the equality fails for at least one cell, the variables are not independent. We write $p_X(2)\,p_Y(1) = (0.45)(0.4) = 0.18 \neq 0.10$, hence, the variables are not independent.

Example. It is known that X and Y are independent. Suppose the marginal pmfs are $p_X(2) = 0.4, p_X(4) = 0.6$, $p_Y(1) = 0.45, p_Y(3) = 0.30$, and $p_Y(5) = 0.25$. The joint probability can then be computed as the product of the marginal totals as in the table.

		x		
	$p(x,y)$	2	4	$p_Y(y)$
	1	(0.45)(0.4)=0.18	(0.45)(0.6)=0.27	0.45
y	3	(0.30)(0.4)=0.12	(0.30)(0.6)=0.18	0.30
	5	(0.25)(0.4)=0.10	(0.25)(0.6)=0.15	0.25
	$p_X(x)$	0.4	0.6	

Definition. Two continuous random variables X and Y are **independent** if and only if $f(x,y) = f_X(x)\,f_Y(y)$.

Example. The joint density function is $f(x,y) = 4xy, \;\; 0 < x, y < 1$. The marginal density functions are $f_X(x) = \int_0^1 (4xy)\,dy = 2x\,y^2\Big|_{y=0}^{y=1} = 2x, \;\; 0 < x < 1$, and $f_Y(y) = \int_0^1 (4xy)\,dx = 2y\,x^2\Big|_{x=0}^{x=1} = 2y, \;\; 0 < y < 1$. We can see that the product of the marginal densities is equal to the joint density, $f(x,y) = 4xy = (2x)(2y) = f_X(x)f_Y(y)$, and hence, X and Y are independent.

Example. Suppose $X \sim Exp(mean = 2)$ and $Y \sim Exp(mean = 4)$ are independent. To find, for example, $\mathbb{P}(X < 3, Y > 2)$, we write

$$\mathbb{P}(X < 3, Y > 2) = \int_0^3 \int_2^\infty f(x,y)\,dy\,dx = \int_0^3 \int_2^\infty \left(\frac{1}{2}e^{-x/2}\right)\left(\frac{1}{4}e^{-y/4}\right)dy\,dx$$

$$= \left(\int_0^3 \frac{1}{2}e^{-x/2}dx\right)\left(\int_2^\infty \frac{1}{4}e^{-y/4}\,dy\right) = \left(1 - e^{-3/2}\right)\left(e^{-2/4}\right) = 0.4712.$$

Definition. Let X and Y be two random variables. The **covariance** between X and Y is $\mathbb{C}ov(X,Y) = \mathbb{E}\Big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\Big]$. The computational formula for covariance is $\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$.

Example. Two continuous random variables X and Y have a joint density $f(x,y) = \frac{2}{3}(x+2y)$, $0 \leq x,\, y \leq 1$. The marginal distribution functions of X and Y were computed before. They are $f_X(x) = \dfrac{2}{3}(x+1),\ \ 0 < x < 1$, and $f_Y(y) = \dfrac{1}{3}(4y+1),\ \ 0 < y < 1$. The covariance between X and Y is calculated as

$$\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = \int_0^1\int_0^1 xy\,\frac{2}{3}(x+2y)\,dy\,dx$$

$$-\Big(\int_0^1 x\,\frac{2}{3}(x+1)\,dx\Big)\Big(\int_0^1 y\,\frac{1}{3}(4y+1)\,dy\Big)$$

$$= \int_0^1 \Big[\frac{x^2y^2}{3} + \frac{4xy^3}{9}\Big]\Big|_{y=0}^{y=1} dx - \Big(\int_0^1 \Big[\frac{2x^2}{3} + \frac{2x}{3}\Big]dx\Big)\Big(\int_0^1 \Big[\frac{4y^2}{3} + \frac{y}{3}\Big]dy\Big)$$

$$= \int_0^1 \Big[\frac{x^2}{3} + \frac{4x}{9}\Big]dx - \Big(\Big[\frac{2x^3}{9} + \frac{x^2}{3}\Big]\Big|_0^1\Big)\Big(\Big[\frac{4y^3}{9} + \frac{y^2}{6}\Big]\Big|_0^1\Big)$$

$$= \frac{1}{3} - \Big(\frac{5}{9}\Big)\Big(\frac{11}{18}\Big) = -\frac{3}{162} = -0.01852.$$

Proposition. If X and Y are independent, then $\mathbb{C}ov(X,Y) = 0$, and X and Y are termed **uncorrelated**.

Proof. We use the independence to write

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dy\,x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)\,dy\,x$$

$$= \Big(\int_{-\infty}^{\infty} xf_X(x)\,dx\Big)\Big(\int_{-\infty}^{\infty} yf_Y(y)\,dy\Big) = (\mathbb{E}X)(\mathbb{E}Y).$$

Therefore,

$$\mathbb{C}ov(X) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = (\mathbb{E}X)(\mathbb{E}Y) - (\mathbb{E}X)(\mathbb{E}Y) = 0.$$

Remark. The converse statement to the one in the above proposition is false in general. Two variables can be uncorrelated, but not independent. A simple example is given in the table below. Infinitely many examples can be constructed in the same fashion.

		x		
	$p(x,y)$	0	1	$p_Y(y)$
y	-1	0.3	0.1	0.4
	0	0.2	0.0	0.2
	1	0.3	0.1	0.4
	$p_X(x)$	0.8	0.2	

We can see that $\mathbb{E}(XY) = (0)(-1)(0.3) + (1)(-1)(0.1) + (0)(0)(0.2) + (1)(0)(0.0) + (0)((1)(0.3) + (1)(1)(0.1) = -0.1 + 0.1 = 0$, and $\mathbb{E}Y = (-1)(0.4) + (0)((0.2) + (1)(0.4) = 0$. Therefore, $\mathbb{V}ar(X) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = 0 - 0 = 0$, and so, X and Y are uncorrelated. However, X and Y are not independent since,for example, $\mathbb{P}(X = 0) = 0.8$ and $\mathbb{P}(Y = -1) = 0.4$ but $\mathbb{P}(X = 0, Y = -1) = 0.3 \neq 0.32 = (0.8)(0.4)$.

Proposition (Useful Formulas). The following statements are true.

(1) For any X and Y (independent or not) and any real-valued a and b,

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

(2) If X and Y are independent, then $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

(3) For any X and Y (independent or not) and any real-valued a and b,

$$\mathbb{V}ar(aX + bY) = a^2\mathbb{V}(X) + 2ab\,\mathbb{C}ov(X, Y) + b^2\mathbb{V}ar(Y).$$

(4) If X and Y are uncorrelated, then $\mathbb{V}ar(aX + bY) = a^2\mathbb{V}ar(X) + b^2\mathbb{V}ar(Y)$.

Proof. We will consider only continuous random variables, but the discrete case is proven analogously, with sums in place of integrals.

(1) By linearity of integration,

$$\begin{aligned}
\mathbb{E}(aX + bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by) f(x, y)\, dy\, dx \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x, y)\, dy\, dx + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x, y)\, dy\, dx \\
&= a\int_{-\infty}^{\infty} x\Big[\int_{-\infty}^{\infty} f(x, y)\, dy\Big]\, dx + b\int_{-\infty}^{\infty} y\Big[\int_{-\infty}^{\infty} f(x, y)\, dx\Big]\, dy \\
&= a\int_{-\infty}^{\infty} x\, f_X(x)\, dx + b\int_{-\infty}^{\infty} y\, f_Y(y)\, dy = a\mathbb{E}(X) + b\mathbb{E}(Y).
\end{aligned}$$

(2) By independence,

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x, y)\, dy\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\, dy\, dx$$

$$= \Big(\int_{-\infty}^{\infty} x\, f_X(x)\, dx \Big) \Big(\int_{-\infty}^{\infty} y\, f_Y(y)\, dy \Big) \;=\; (\mathbb{E}X)(\mathbb{E}Y).$$

(3) By definition,

$$\mathbb{V}ar(aX + bY) \;=\; \mathbb{E}\big(aX + bY - \mathbb{E}(aX + bY)\big)^2 \;=\; \mathbb{E}\big(aX \;-\; a\mathbb{E}X \;+\; bY \;-\; b\mathbb{E}Y\big)^2$$
$$= a^2\mathbb{E}(X - \mathbb{E}X)^2 \;+\; 2\,ab\mathbb{E}\big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\big] \;+\; b^2\mathbb{E}(Y - \mathbb{E}Y)^2$$
$$= a^2\mathbb{V}(X) \;+\; 2\,ab\,\mathbb{C}ov(X, Y) \;+\; b^2\mathbb{V}ar(Y).$$

(4) If X and Y are uncorrelated, then $\mathbb{C}ov(X, Y) \;=\; 0$, and so $\mathbb{V}ar(aX + bY) \;=\; a^2\mathbb{V}ar(X) \;+\; b^2\mathbb{V}ar(Y)$.

Exercise 22.1. Random variables X and Y have the following joint probability distribution:

		x	
	$p(x, y)$	0	1
y	-1	0.10	0.25
	0	0	0.30
	1	0.20	0.15

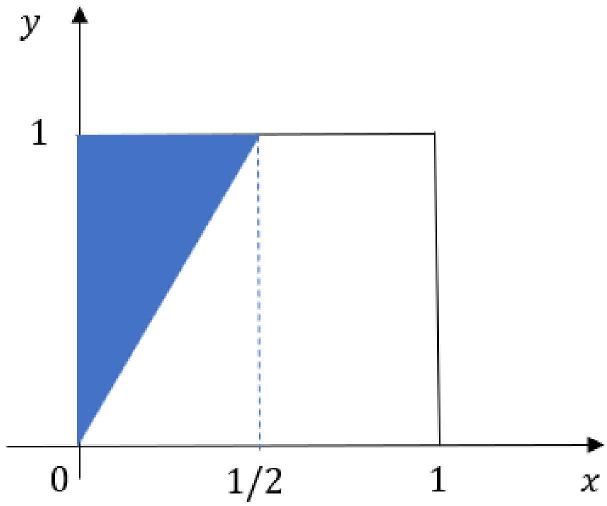
(a) Are X and Y independent? Answer: The marginal distributions are $p_X(0) = 0.3$, $p_X(1) = 0.7$, $p_Y(-1) = 0.35$, $p_Y(0) = 0.3$, and $p_Y(1) = 0.35$. The variables are not independent because, for instance, $p(1, -1) \;=\; 0.25 \;\neq\; 0.245 \;=\; (0.7)(0.35) \;=\; p_X(1)p_Y(-1)$.

(b) Compute the covariance of X and Y. Answer: $\mathbb{C}ov(X{+}Y) \;=\; \mathbb{E}(XY)\,(\mathbb{E}X)(\mathbb{E}Y) \;=\; \big((1)(-1)(0.25){+}(1)(1)(0.15)\big) - \big((1)(0.7)\big)\big((-1)(0.35){+}(1)(0.35)\big) \;=\; -0.1\,(0.7)(0) \;=\; -0.1$.

(c) What is $\mathbb{E}(2X{-}5Y)$? Answer: $\mathbb{E}(2X{-}5Y) \;=\; 2\mathbb{E}X \,-\, 5\mathbb{E}Y \;=\; (2)(0.7){-}(5)(0) \;=\; 1.4$.

(d) What is $\mathbb{V}ar(2X - 5Y)$? Answer: $\mathbb{V}ar(X) = (0.7)(1 - 0.7) = 0.21$, and $\mathbb{V}ar(Y) = (-1)^2(0.35) + (1)^2(0.35) \;=\; 0.7$. We compute $\mathbb{V}ar(2X \,-\, 5Y) \;=\; (2)^2\,\mathbb{V}ar(X) \,-\, (2)(2)(5)\mathbb{C}ov(X, Y) \,+\, (5)^2\,\mathbb{V}ar(Y) \;=\; (4)(0.21) \,-\, (20)(-0.1) \,+\, (5)(0.7) \;=\; 6.34$.

Exercise 22.2. Suppose that random variables X and Y are independent with marginal pdf's $f_X(x) \;=\; 2x$, $0 \le x \le 1$, and $f_Y(y) \;=\; 1$, $0 \le y \le 1$. Calculate $\mathbb{P}(Y > 2X)$. Answer: We need to integrate the joint density $f(x, y) = f_X(x)f_Y(y) = 2x$ over the shaded region shown in the picture:



We integrate, for example, vertically. We write $\mathbb{P}(Y > 2X) = \int_0^{1/2} \int_{2x}^1 2x\,dy\,dx = \int_0^{1/2} 2x(1-2x)\,dx = \left[x^2 - \frac{4}{3}x^3\right]\Big|_0^{1/2} = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$.

Exercise 22.3. Random variables X_1, X_2, and X_3 are independent with marginal densities $f_{X_i}(x_i) = 3x_i^2,\ 0 \le x_i \le 1,\ i = 1, 2, 3$.

(a) Compute $\mathbb{E}(X_1X_2 + X_1X_3 + X_2X_3)$. Answer: Since the three variables are identically distributed, their means are equal, that is $\mathbb{E}X_1 = \mathbb{E}X_2 = \mathbb{E}X_3 = \int_0^1 3x^3\,dx = \frac{3}{4}$, and $\mathbb{E}(X_1X_2 + X_1X_3 + X_2X_3) = (3)(\mathbb{E}X_1)^2 = (3)\left(\frac{3}{4}\right)^2 = \frac{27}{16} = 1.6875$.

(b) Compute $\mathbb{V}ar(X_1 + X_2^2 + X_3^3)$. Answer: Since the variables are independent, their covariances are equal to zero, and so, the variance of the sum is the sum of variances, $\mathbb{V}ar(X_1 + X_2^2 + X_3^3) = \mathbb{V}ar(X1) + \mathbb{V}ar(X_2^2) + \mathbb{V}ar(X_3^3)$. We compute each variance as follows. $\mathbb{V}ar(X1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \int_0^1 3x^4\,dx - (\int_0^1 3x^3\,dx)^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = 0.0375$, $\mathbb{V}ar(X_2^2) = \mathbb{E}X_2^4 - (\mathbb{E}X_2^2)^2 = \int_0^1 3x^6\,dx - (\int_0^1 3x^4\,dx)^2 = \frac{3}{7} - \left(\frac{3}{5}\right)^2 = 0.0686$, and $\mathbb{V}ar(X_2^3) = \mathbb{E}X_3^6 - (\mathbb{E}X_3^3)^2 = \int_0^1 3x^8\,dx - (\int_0^1 3x^5\,dx)^2 = \frac{3}{9} - \left(\frac{3}{6}\right)^2 = 0.0833$. Thus, $\mathbb{V}ar(X1) + \mathbb{V}ar(X_2^2) + \mathbb{V}ar(X_3^3) = 0.1894$.

Exercise 22.4. The bivariate distribution of discrete random variables is

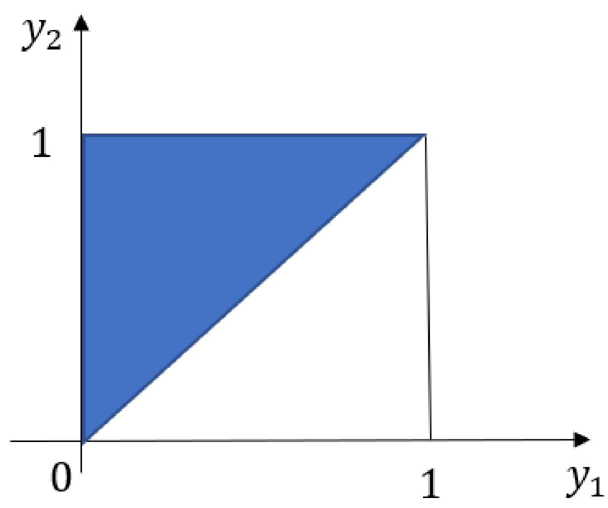
$x \setminus y$	-3	1	2	$p_X(x)$
0	$1/3$	0	$1/2$	$20/24$
1	$1/24$	$1/8$	0	$4/24$
$p_Y(y)$	$9/24$	$3/24$	$12/24$	

(a) Are X and Y independent? Answer: The variables are not independent since, for example, $p_X(0)p_Y(1) = (20/24)(3/24) \neq 0 = p(0,1)$.

(b) Are X and Y uncorrelated? Answer: $\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = \big((1)(-3)(1/24)+(1)(1)(1/8)\big) - \big((1)(4/24)\big)\big((-3)(9/24)+(1)(3/24)+(2)(12/24)\big) = 0 - (1/6)(0) = 0$, therefore, the variables are uncorrelated.

Exercise 22.5. Continuous random variables Y_1 and Y_2 have joint density given by $f(y_1, y_2) = 6y_1,\ 0 \leq y_1 \leq y_2 \leq 1$.

(a) Can Y_1 and Y_2 be independent? Answer: Since the ranges of the variables are interdependent, they cannot be independent. We show it formally by computing the marginal densities. We integrate over the shaded region (see the picture). We get $f_{Y_1}(y_1) = \int_{y_1}^{1} 6y_1\, dy_2 = 6y_1(1-y_1),\ 0 < y_1 < 1$, and $f_{Y_2}(y_2) = \int_0^{y_2} 6y_1\, dy_1 = 3y_2^2,\ 0 < y_2 < 1$.



The product of the marginal densities is $f_{Y_1}(y_1)f_{Y_2}(y_2) = 6y_1(1-y_1)\,3y_2^2$. It is not equal to the joint density $f(y_1, y_2) = 6y_1$, so, the variables are not independent.

(b) Compute $\mathbb{C}ov(Y_1, Y_2)$. Answer:

$$\mathbb{C}ov(Y_1, Y_2) \;=\; \mathbb{E}(XY) \;-\; (\mathbb{E}X)(\mathbb{E}Y)$$

$$= \int_0^1 \int_{y_1}^1 (y_1 y_2)(6y_1)\, dy_2\, dy_1 \;-\; \Big(\int_0^1 \; y_1\, 6y_1(1-y_1)\, dy_1\Big)\Big(\int_0^1 \; y_2\, 3y_2^2\, dy_2\Big)$$

$$= 1 - \frac{3}{5} \;-\; \Big(2 - \frac{3}{2}\Big)\Big(\frac{3}{4}\Big) \;=\; 0.4 \;-\; (0.5)(0.75) \;=\; 0.025.$$

(c) Compute $\mathbb{V}ar(Y_1 - Y_2)$. Answer:

$$\mathbb{V}ar(Y_1) \;=\; \mathbb{E}(Y_1^2) - \big(\mathbb{E}Y_1\big)^2 \;=\; \int_0^1 \; y_1^2\, 6y_1(1-y_1)\, dy_1 \;-\; \Big(\frac{1}{2}\Big)^2 \;=\; \frac{3}{2} - \frac{6}{5} - \frac{1}{4} \;=\; 0.05,$$

and

$$\mathbb{V}ar(Y_2) \;=\; \mathbb{E}(Y_2^2) \;-\; \big(\mathbb{E}Y_2\big)^2 \;=\; \int_0^1 \; y_2^2\, 3y_2^2\, dy_1 \;-\; \Big(\frac{3}{4}\Big)^2 \;=\; \frac{3}{5} \;-\; \frac{9}{16} \;=\; 0.0375.$$

We have $\mathbb{V}ar(Y_1 - Y_2) \;=\; \mathbb{V}ar(Y_1) - 2\mathbb{C}ov(Y_1, Y-2) + \mathbb{V}ar(Y_2) \;=\; 0.05 - (2)(0.025) + 0.0375 \;=\; 0.0375$.

23. FUNCTIONS OF RANDOM VARIABLES

Let X be a continuous random variable, and let $Y = g(X)$ where g is some known function. The goal is to find the cdf and pdf of Y. The method that always works is the **cumulative distribution function method**. Below we consider several examples.

Example. Let $X \sim f_X(x)$, $F_X(x)$, and $Y = aX + b$ for some real-valued a and b. Then the cdf of Y is $F_Y(y) \;=\; \mathbb{P}(Y \;\le\; y) \;=\; \mathbb{P}(aX + b \;\le\; y) \;=\; \mathbb{P}\Big(X \le \dfrac{y-b}{a}\Big) \;=\; F_X\Big(\dfrac{y-b}{a}\Big)$, and the pdf of Y is $f_Y(y) \;=\; F_Y'(y) \;=\; \dfrac{1}{a} f_X\Big(\dfrac{y-b}{a}\Big)$.

Example. Let $X \sim f_X(x)$, $F_X(x), x > 0$, and $Y = \sqrt{X}$. The cdf of Y is

$$F_Y(y) \;=\; \mathbb{P}(Y \;\le\; y) \;=\; \mathbb{P}(\sqrt{X} \;\le\; y) \;=\; \mathbb{P}\Big(X \;\le\; y^2\Big) \;=\; F_X(y^2),$$

and the pdf of Y is $f_Y(y) \;=\; F_Y'(y) \;=\; 2y\, f_X(y^2)$.

Example. Let $X \sim f_X(x)$, $F_X(x)$, and $Y = X^2$. The cdf of Y is

$$F_Y(y) \;=\; \mathbb{P}(Y \;\le\; y) \;=\; \mathbb{P}(X^2 \;\le\; y) \;=\; \mathbb{P}\Big(-\sqrt{y} \;\le\; X \;\le\; \sqrt{y}\Big) \;=\; F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

and the pdf of Y is $f_Y(y) \;=\; F'_Y(y) \;=\; \dfrac{1}{2\sqrt{y}}\, f_X(\sqrt{y}) \;+\; \dfrac{1}{2\sqrt{y}}\, f_X(-\sqrt{y})$.

Exercise 23.1. Suppose $X \sim Exp(mean = \beta)$ and let $Y = X + a$ for some $a > 0$. Show that the pdf of Y is $f_Y(y) \;=\; \dfrac{1}{\beta}\, e^{-(y-a)/\beta},\; y \,>\, a$. It is called a **shifted exponential distribution**. Answer: The cdf of X is $F_X(x) = 1 - e^{-x/\beta}$, $x > 0$. We find the cdf of Y as $F_Y(y) \;=\; \mathbb{P}(Y \;\le\; y) \;=\; \mathbb{P}(X + a \;\le\; y) \;=\; \mathbb{P}(X \;\le\; y - a) \;=\; F_X(y-a) \;=\; 1 - e^{-(y-a)/\beta}, y - a > 0$ or $y > a$, and the pdf of Y is $f_Y(y) \;=\; F'_Y(y) \;=\; \dfrac{1}{\beta}\, e^{-(y-a)/\beta}$, $y > a$.

Exercise 23.2. Consider $X \;\sim\; N(\mu, \sigma^2)$ and let $Y \;=\; e^X$. Prove that $f_Y(y) \;=\; \dfrac{1}{y\sqrt{2\pi\sigma^2}} \exp\Big\{ -\; \dfrac{(\ln(y) \;-\; \mu)^2}{2\sigma^2} \Big\}$. It is called a **log-normal distribution**. Answer: The cdf of Y can be derived as $F_Y(y) \;=\; \mathbb{P}(Y \le y) \;=\; \mathbb{P}\big(e^X \le y\big) \;=\; \mathbb{P}(X \le \ln(y)) \;=\; F_X(\ln(y))$. The density of Y is $f_Y(y) \;=\; F'_Y(y) \;=\; \dfrac{1}{y\sqrt{2\pi\sigma^2}} \exp\Big\{ -\; \dfrac{(\ln(y) \;-\; \mu)^2}{2\sigma^2} \Big\}$.

Exercise 23.3. Let $Z \sim N(0, 1)$ and let $X \,=\, \mu + \sigma\, Z$. Show that $X \,\sim\, N(\mu, \sigma^2)$. Answer: The cdf of X is $F_X(x) \;=\; \mathbb{P}(X \le x) \;=\; \mathbb{P}(\mu + \sigma\, Z \le x) \;=\; \mathbb{P}\Big(Z \le \dfrac{x - \mu}{\sigma}\Big) \;=\; \Phi\Big(\dfrac{x-\mu}{\sigma}\Big)$. The pdf of X is

$$f_X(x) \;=\; F'_X(x) \;=\; \Phi'\Big(\frac{x-\mu}{\sigma}\Big)$$

$$= \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \exp\Big[-\; \frac{1}{2} \Big(\frac{x-\mu}{\sigma}\Big)^2 \Big] \;=\; \frac{1}{\sqrt{2\pi\,\sigma^2}} \exp\Big[-\; \frac{(x-\mu)^2}{2\sigma^2} \Big].$$

Exercise 23.4. Let $X \sim Unif(-\pi/2, \pi/2)$. Take $Y = tan(X)$. Show that Y follows a **Cauchy distribution** with the pdf $f_Y(y) \;=\; \dfrac{1}{\pi(1 + y^2)}$, $-\infty < y < \infty$.

Answer: We start with the cdf of Y. We write $F_Y(y) \;=\; \mathbb{P}(Y \,\le\, y) \;=\; \mathbb{P}(\tan(X) \,\le\, y) \;=\; \mathbb{P}\big(X \le \tan^{-1}(y)\big) \;=\; F_X\big(\tan^{-1}(y)\big),\;\; -\dfrac{\pi}{2} < \tan^{-1}(y) < \dfrac{\pi}{2}$, and thus, the pdf of Y is $f_X(y) \;=\; F'_Y(y) \;=\; F'_X\big(\tan^{-1}(y)\big) \;=\; \big(\tan^{-1}(y)\big)' \,\dfrac{1}{\pi} \;=\; \dfrac{1}{\pi(1+y^2)}, -\infty < y < \infty$.

Exercise 23.5. Consider X and Y independent $Unif(0, 1)$ random variables, and let $W = X + Y$. Show that W has a **triangular distribution** with the pdf given by

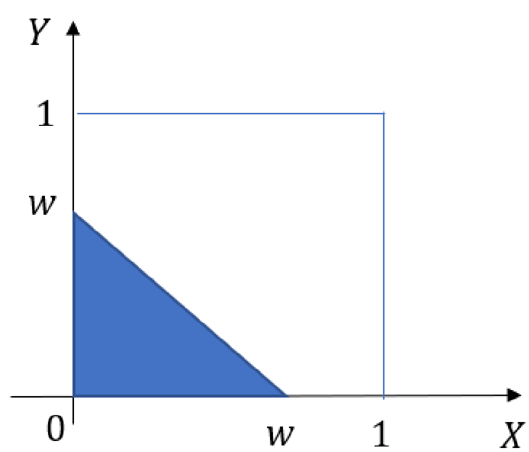
$$f_W(w) = \begin{cases} \left(\dfrac{w^2}{2}\right)' = w, & \text{if } 0 < w < 1, \\ \left(1 - \dfrac{(2-w)^2}{2}\right)' = 2 - w, & \text{if } 1 < w < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Plot the graph of this function. Answer: Starting with the cdf, we write

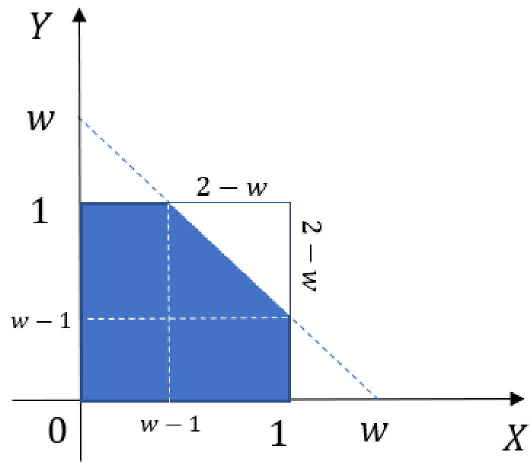
$$F_W(w) = \mathbb{P}(W \le w) = \mathbb{P}(X + Y \le w).$$

There are two cases: $0 < w < 1$ and $1 < w < 2$.

Case 1. $0 < w < 1$. We integrate over the shaded region (see the picture). We get $\mathbb{P}(X + Y \le w) = \int_0^w \int_0^{w-x} dy\, dx = \int_0^w (w - x)\, dx = w^2 - \frac{w^2}{2} = \frac{w^2}{2}$, the area of the shaded triangle since the density is equal to 1.



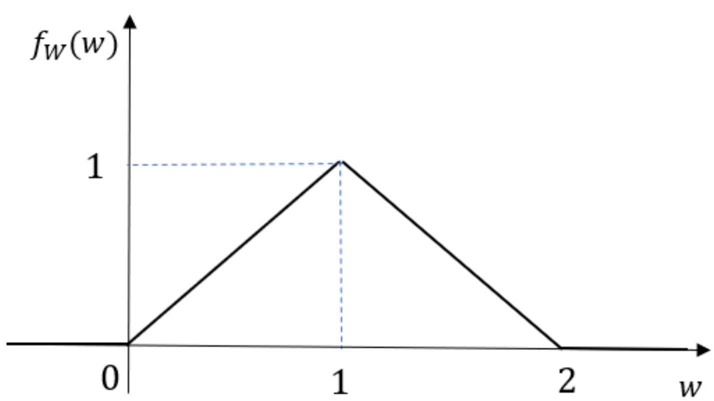
Case 2. $1 < w < 2$. The integration should be done over the shaded region as in the picture, but since we are dealing with a unit density, we just need to find the area of the shaded region, and that can be found as one minus the area of the unshaded triangle, that is, $\mathbb{P}(X + Y \le w) = 1 - \frac{(2-w)^2}{2}$.



The density function combines both cases. We write

$$f_W(w) \;=\; F'_W(w) \;=\; \begin{cases} \left(\dfrac{w^2}{2}\right)' \;=\; w, & \text{if } 0 < w < 1, \\ \left(1 \;-\; \dfrac{(2-w)^2}{2}\right)' \;=\; 2 - w, & \text{if } 1 < w < 2, \\ 0, & \text{otherwise.} \end{cases}$$

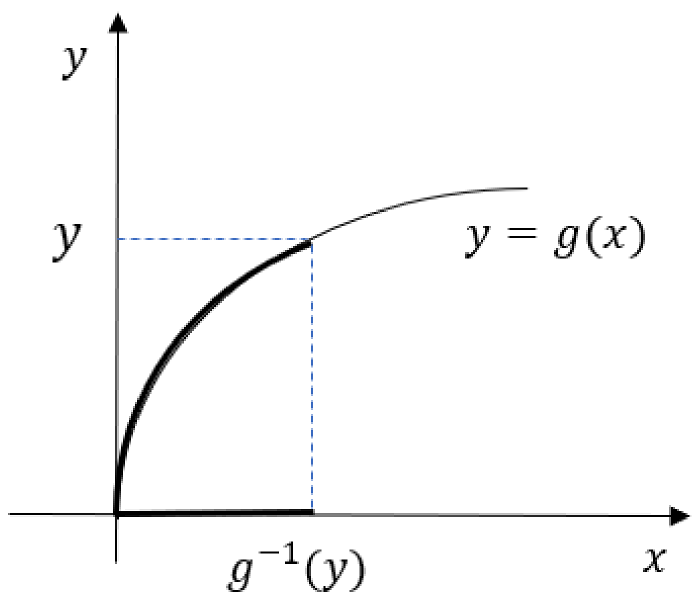
Here is the plot of the triangular density. Note that it indeed has the shape of a triangle and also note that the total area under the curve is equal to one, as it should be.



24. METHOD OF TRANSFORMATIONS

Proposition (Method of Transformations or Change of Variable Method). Let $X \sim f_X(x)$ and suppose $Y = g(X)$ where g is a **strictly increasing** or **strictly decreasing** function. We can write $X = g^{-1}(Y)$. Then $f_Y(y) = f_X(g^{-1}(y)) \left|\frac{d\,g^{-1}(y)}{d\,y}\right|$. Here $\left|\frac{d\,g^{-1}(y)}{d\,y}\right|$ is termed the **Jacobian of the transformation**.

Proof. We start with the cdf of Y. We first consider the case of a strictly increasing function g. Note that in this case, as depicted in the picture, $g(x) \le y$ if and only if $x \le g^{-1}(y)$.



We get

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}\big(X \le g^{-1}(y)\big) = F_X\big(g^{-1}(y)\big),$$

and thus, the pdf of Y is obtained as

$$f_Y(y) = F_Y'(y) = f_X\big(g^{-1}(y)\big)\,\frac{d\,g^{-1}(y)}{dy}.$$

Now we consider the case of a strictly decreasing function g. As shown in the picture, $g(x) \le y$ if and only if $x \ge g^{-1}(y)$. So, we write

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}\big(X \ge g^{-1}(y)\big) = 1 - F_X\big(g^{-1}(y)\big),$$

and thus, the pdf of Y is obtained as

$$f_Y(y) = F_Y'(y) = -\, f_X\big(g^{-1}(y)\big)\, \frac{d\, g^{-1}(y)}{dy}.$$

Since for a decreasing function, $\frac{d\, g^{-1}(y)}{dy}$ is negative, putting a minus in front of it, makes it positive. Combining both cases, of increasing and decreasing function g, we can write

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{d\, y} \right|.$$

Example. Suppose $X \sim Exp(mean = \beta)$, and $Y = \sqrt{X}$. Since a square root is a strictly increasing function, we can use the method of transformations to find the pdf of Y. We invert the function $y = g(x) = \sqrt{x}$, to get $x = g^{-1}(y) = y^2, y > 0$. Now we write the pdf of Y:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{d\, y} \right| = f_X(y^2) \cdot \frac{d\, y^2}{dy} = \frac{2y}{\beta}\, e^{-y^2/\beta}, \quad y > 0.$$

Example. Suppose $X \sim Exp(mean = \beta)$, and $Y = e^{-X}$. The function $y = g(x) = e^{-x}$ is strictly decreasing, and solving for x, we get $x = g^{-1}(y) = -\ln(y)$. By the method of transformations, we find the cdf of Y as follows:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{d\, y} \right| = f_X(-\ln(y)) \cdot \left| \frac{d\,(-\ln(y))}{dy} \right| = \frac{1}{y\beta}\, e^{\ln(y)/\beta}$$

where $-\ln(y) > 0$ or $0 < y < 1$.

Exercise 24.1. Let $X \sim Unif(0, 1)$. Use the method of transformations to show that $Y = -\beta \ln(X)$ has an exponential distribution with mean β. Answer: A negative natural logarithm is a strictly decreasing function. Its inverse is $x = g^{-1}(y) = e^{-y/\beta}$. The pdf is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{d\, y} \right| = f_X(e^{-y/\beta}) \cdot \left| \frac{d\,(e^{-y/\beta})}{dy} \right| = \frac{1}{\beta}\, e^{-y/\beta},$$

where $0 < e^{-y/\beta} < 1$ or $y > 0$.

Exercise 24.2. Suppose X has a **Pareto distribution** with parameter a, for which the pdf is given as $f_X(x) = \frac{a}{x^{a+1}}, \quad 1 \le x < \infty, \quad a > 0$. Use the method of transformations to find the pdf of Y if

(a) $Y = \ln(X)$. Answer: $\ln(x)$ is an increasing function with inverse $g^{-1}(y) = e^y$. Therefore, the pdf of Y is

$$f_Y(y) = f_X(e^y) \cdot \frac{d\,e^y}{dy} = \frac{a}{e^{(a+1)y}}\, e^y = a\, e^{-ay}$$

where $e^y > 1$ or $y > 0$. This is an exponential distribution with a mean $1/a$.

(b) $Y = 1/X$. Answer: $1/X$ is a decreasing function with inverse $g^{-1}(y) = 1/y$. Therefore, the pdf of Y is

$$f_Y(y) = f_X(1/y) \cdot \Big|\frac{d\,(1/y)}{dy}\Big| = ay^{a+1}\,\big| - \frac{1}{y^2}\big| = a\,y^{a-1}$$

where $1/y > 1$ or $0 < y < 1$.

Exercise 24.3. Consider $X \sim f_X(x) = 3x^2,\ \ 0 \le x \le 1$, and let$Y = 1/X^2$. Use the method of transformations to determine the pdf of Y. Answer: On the interval $[0, 1]$, $1/X^2$ is a decreasing function with the inverse $x = g^{-1}(y) = 1/\sqrt{y}$. The pdf of Y is

$$f_Y(y) = f_X(1/\sqrt{y}) \cdot \Big|\frac{d\,(1/\sqrt{y})}{dy}\Big| = 3y\,\big| - \frac{1}{2y\sqrt{y}}\big| = \frac{3}{2}y^{-5/2}$$

where $0 < 1/\sqrt{y} < 1$ or $y > 1$.

25. METHOD OF MOMENT GENERATING FUNCTIONS

Proposition. Suppose $X_1, \dots, X_n$ are independent identically distributed (iid) random variables with the mgfs $M(t),\ i = 1, \dots, n$. Then the mgf of the sum $S = X_1 + \cdots + X_n$ is $M_S(t) = M^n(t)$.

Proof. We write $M_S(t) = \mathbb{E}e^{tS} = \mathbb{E}\big[\exp\{t(X_1 + \cdots + X_n)\}\big] = \{\text{by independence}\}$ $= \mathbb{E}e^{tX_1}\mathbb{E}e^{tX_1} \cdots \mathbb{E}e^{tX_n} = M^n(t)$.

Example. Suppose $X_1, \dots, X_n \overset{iid}{\sim} Ber(p)$. The mgf is $M(t) = pe^t_+ 1 - p$. Then the sum $S = X_1 + \cdots + X_n$ has the mgf $M_S(t) = M^n(t) = (pe^t + 1 - p)^n$ which is the mgf for $Bi(n, p)$ distribution.

Exercise 25.1. Let $X_1, \dots, X_n \overset{iid}{\sim} Geom(p)$. Prove that $X_1 + \cdots + X_n \sim NB(n, p)$. Answer: The mgf of $Geom(p)$ is $M(t) = \dfrac{pe^t}{1 - (1-p)e^t}$. The mgf of the sum S is $M_S(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^n$ which means that the sum has a negative binomial distribution with parameters n and p.

Exercise 25.2. Consider $X_1, \dots, X_n \overset{iid}{\sim} Poi(\lambda)$. Show that $X_1 + \cdots + X_n \sim Poi(\lambda n)$. Answer: The mgf for a $Poi(\lambda)$ distribution is $M(t) = \exp(\lambda(e^t - 1))$, and so, the mgf of the sum is $M_S(t) = \Big(\exp(\lambda(e^t - 1))\Big)^n = \exp((\lambda\, n)(e^t - 1))$ which corresponds to a Poisson distribution with rate $\lambda\, n$.

Exercise 25.3. Let $X_1, \ldots, X_n$ are iid random variables having exponential distribution with mean β. Show that $X_1 + \cdots + X_n$ has a gamma distribution with parameters n and β. A gamma distribution with parameters α and β has density $f(x) = \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\,\beta^{\alpha}}\, e^{-x/\beta},\ x \geq 0,\ \alpha, \beta > 0$ where the gamma function $\Gamma(\alpha) = \displaystyle\int_0^{\infty} x^{\alpha-1}\, e^{-x}\, dx$ is the normalizing constant. The mean of this distribution is $\alpha\beta$, variance is $\alpha\beta^2$, and the mgf is $M(t) = \Big(\dfrac{1}{1 - \beta\, t}\Big)^{\alpha}$. Answer: By the method of moment generating functions, we have that the mgf of $X_1 + \cdots + X_n$ is $\Big(\frac{1}{1-\beta\, t}\Big)^n$ which corresponds to a gamma distribution with parameters n and β.

Exercise 25.4. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Verify that $X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)$. Answer: The mgf of a normal distribution with parameters μ and σ^2 is $M(t) = \exp\left\{\mu\, t + \frac{\sigma^2\, t^2}{2}\right\}$. The mgf of the sum is $M_S(t) = \Big(\exp\left\{\mu\, t + \frac{\sigma^2\, t^2}{2}\right\}\Big)^n = \exp\left\{n\mu\, t + \frac{n\sigma^2\, t^2}{2}\right\}$ which corresponds to a normal distribution with parameters $n\mu$ and $n\sigma^2$.

26. THE CENTRAL LIMIT THEOREM

Theorem (The Central Limit Theorem (CLM)). Let $X_1, X_2, \ldots, X_n$ be iid with a common mean μ and standard deviation σ. Denote by $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Then,

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to N(0, 1), \quad \text{as } n \to \infty.$$

In other words, for large n, $\bar{X}_n$ has an approximate normal distribution with mean μ and variance σ^2/n.

Proof. First we write

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\,\mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

The mgf of Z_n is $M_{Z_n}(t) = M^n\Big(\dfrac{t}{\sqrt{n}}\Big)$ where $M(t)$ is the mgf of $\dfrac{X_i - \mu}{\sigma}$ for $i = 1, \ldots, n$. The Taylor's expansion of $M(t)$ is $M(t) = \displaystyle\sum_{k=0}^{\infty} M^{(k)}(0)\frac{t^k}{k!}$. Here $M(0) = 1$, $M'(0) = \mathbb{E}\Big(\dfrac{X_i - \mu}{\sigma}\Big) = 0$, and $M''(0) = \mathbb{E}\Big(\dfrac{X_i - \mu}{\sigma}\Big)^2 = \mathbb{V}ar\Big(\dfrac{X_i - \mu}{\sigma}\Big) = 1.$

Thus, $M\Big(\frac{t}{\sqrt{n}}\Big) = 1 + \frac{t^2}{2n} +$ terms of higher order of $\frac{t}{\sqrt{n}}$. Putting it together, we obtain that

$$M_{Z_n}(t) \approx \Big(1+\frac{t^2}{2n}\Big)^n \to \exp\Big\{\frac{t^2}{2}\Big\}, \text{ as } n\to\infty,$$

which is the mgf of a $N(0,1)$ random variable.

Remark. In practice, the CLT is used if n is at least 30.

Example. The lifespan of a light bulb has a mean of 750 hours and a standard deviation of 90 hours. The probability that, for example, in a random sample of 100 bulbs, the average lifespan is less than 730 hours is found as follows. We use the Central Limit Theorem (CLT) to assert that $\bar{X}$ is approximately normally distributed with a mean of $\mu = 750$ hours and a standard deviation of $\frac{\sigma}{\sqrt{n}} = \frac{90}{\sqrt{100}} = 9$ hours. Thus, $\mathbb{P}(\bar{X} < 730) = \mathbb{P}\Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{730-750}{90/\sqrt{100}}\Big) = \mathbb{P}(Z \geq -2.22) = 1 - \mathbb{P}(Z < -2.22) = 1 - \mathbb{P}(Z > 2.22) = 0.0132$.

Exercise 26.1. The amount of time that a drive-through bank teller spends on a customer is a random variable with a mean of 3.2 minutes and a standard deviation of 1.6 minutes. If a random sample of 64 customers is observed, evaluate the approximate probability that their mean time at the teller's counter is at least 3 minutes. Answer: By the CLT, $\bar{X} \overset{approx}{\sim} N\big(3.2, (1.6/\sqrt{64})^2\big)$. So,

$$\mathbb{P}(\bar{X} \geq 3) = \mathbb{P}\Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{3-3.2}{1.6/\sqrt{64}}\Big)$$

$$= \mathbb{P}(Z \geq -1) = 1 - \mathbb{P}(Z < -1) = 1 - \mathbb{P}(Z > 1) = 1 - 0.1587 = 0.8413.$$

Exercise 26.2. Japan's birth rate is believed to be 1.57 births per woman with a population standard deviation of 0.8. If a random sample of 200 women is selected, approximate the probability that the sample mean falls above 1.6. Answer: By the CLT, $\bar{X} \overset{approx}{\sim} N\big(1.57, (0.8/\sqrt{200})^2\big)$. Hence,

$$\mathbb{P}(\bar{X} > 1.6) = \mathbb{P}\Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{1.6-1.57}{0.8/\sqrt{200}}\Big) = \mathbb{P}(Z \geq 0.53) = 0.2981.$$

Exercise 26.3. In a very large group of gifted children, the average IQ is 120 and the standard deviation is 15. If 100 children are randomly selected from this group, what is the approximate probability that the average IQ of the children in the sample will be less than 117? Answer: By the CLT, $\bar{X} \overset{approx}{\sim} N\big(120, (15/\sqrt{100})^2\big)$. Therefore, $\mathbb{P}(\bar{X} < 117) = \mathbb{P}\Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{117-120}{15/\sqrt{100}}\Big) = \mathbb{P}(Z < -2) = \mathbb{P}(Z > 2) = 0.0228.$

Exercise 26.4. Insurance claims have a mean of \$5,000 with a standard deviation of \$3,000. What is the approximate probability that in a random sample of 50 claims, the average is above \$6,000? Answer: By the CLT, $\bar{X} \overset{approx}{\sim} N\big(5000, (3000/\sqrt{50})^2\big)$. Thus, $\mathbb{P}(\bar{X} > 6000) = \mathbb{P}\Big(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \dfrac{6000 - 5000}{3000/\sqrt{50}}\Big) = \mathbb{P}(Z > 2.36) = 0.0091.$

Exercise 26.5. Coke bottles have a mean of 19.8 oz with a standard deviation of 1 oz. Find the approximate probability that in a random sample of 60 Coke bottles, the sample mean exceeds 20 oz. Answer: By the CLT, $\bar{X} \overset{approx}{\sim} N\big(19.8, (1/\sqrt{60})^2\big)$. So, $\mathbb{P}(\bar{X} > 20) = \mathbb{P}\Big(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \dfrac{20 - 19.8}{1/\sqrt{60}}\Big) = \mathbb{P}(Z > 1.55) = 0.0606.$