Unit 4: Probability

Merrick Fanning

November 24, 2025

Unit 4 Outline: Probability

1. Basic Set Theory and Counting
2. Conditional probability
3. Estimating probability with simulation
4. Discrete Random Variables
5. Continuous Random Variables

What is a Set?

- A **set** is a collection of distinct objects, called **elements**.
- Elements can be numbers, categories, or other types of data.
- Sets are usually denoted with capital letters like A, B, or S.

Cardinality

The **cardinality** of a set is the number of elements it contains. For a set A, it is written as $|A|$.

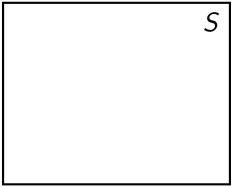
Example: How many elements are in $M = \{m_1, m_2, \ldots, m_n\}$?
There are n elements in M, so $|M| = n$.

The Universal Set and the Empty Set

Universal Set

The **universal set** contains all possible elements under consideration for a given context.

- Denoted by S in probability (stands for *sample space*).



Empty Set

The **empty set**, denoted by $\emptyset$ or $\{\}$, is the set with no elements.

- It is a subset of every set, including S.
- $|\emptyset| = 0$

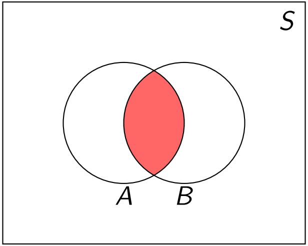
Intersection of Sets

Intersection

For two sets A and B, the **intersection** is the set of elements that are in both A **and** B.
It is denoted: $A \cap B$

Example: Let $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{1, 2, 14\}$
Then:

$$A \cap B = \{1, 2\}$$



Union of Sets

Union

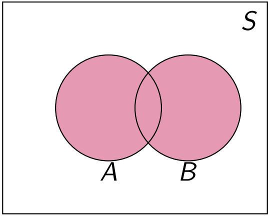
The **union** of two sets A and B includes all elements that are in A **or** in B (or in both).
It is denoted: $A \cup B$

Example: Let

$$A = \{\text{Apple}, \text{Banana}, \text{Orange}\}, \quad B = \{\text{Banana}, \text{Orange}, \text{Potato}\}$$

Then:

$$A \cup B = \{\text{Apple}, \text{Banana}, \text{Orange}, \text{Potato}\}$$



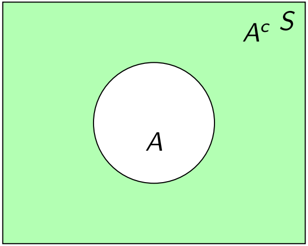
Complement of a Set

Complement

The **complement** of a set A includes all elements in the universal set S that are **not** in A.
It is denoted: A^c

Example: Let $S = \{1, 2, 3, 4, 5, 6\}$ and $A = \{2, 4, 6\}$
Then:

$$A^c = \{1, 3, 5\}$$



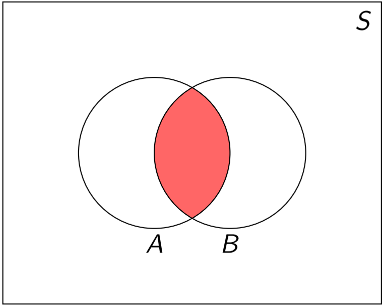
Set Laws: Commutative Laws

Commutative Laws

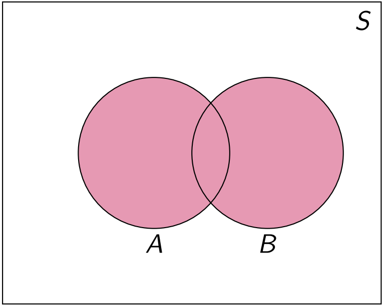
Order doesn't matter when taking the union or intersection of sets:

$$A \cup B = B \cup A \qquad \text{and} \qquad A \cap B = B \cap A$$

Intersection: $A \cap B$



Union: $A \cup B$



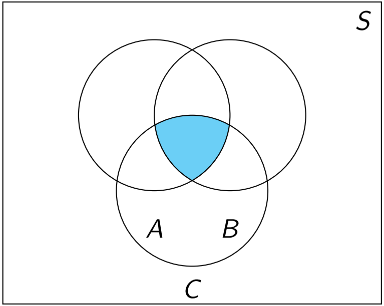
Set Laws: Associative Laws

Associative Laws

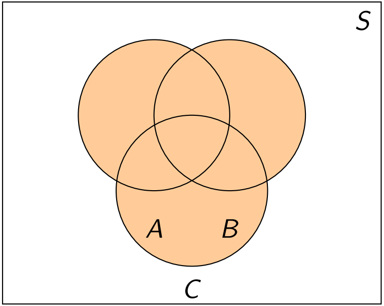
The grouping of sets does not affect the result of union or intersection:

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

Intersection: $(A \cap B) \cap C$



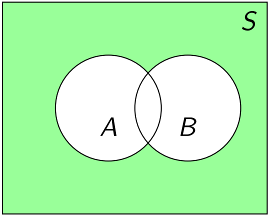
Union: $(A \cup B) \cup C$



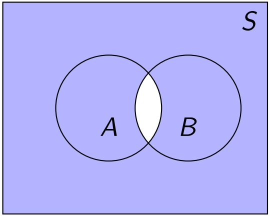
DeMorgan's Laws

For sets A and B, fill in the Euler diagrams to represent:

$(A \cup B)^c$



$(A \cap B)^c$



What is the result that follows?

DeMorgan's Laws

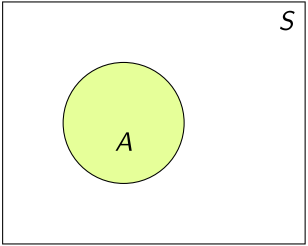
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Set Laws: Identity Laws

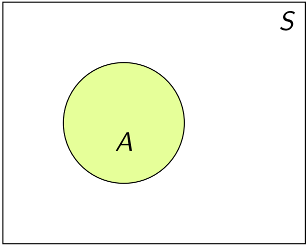
Identity Laws

$$A \cup \emptyset = A$$
$$A \cap S = A$$

$A \cup \emptyset$



$A \cap S$

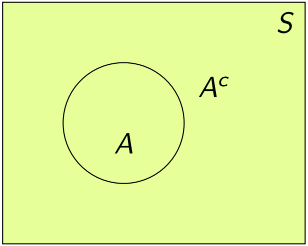


Set Laws: Complement Laws

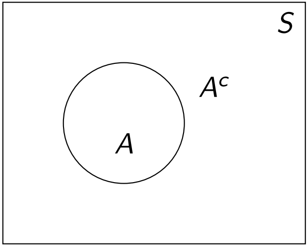
Complement Laws

$$A \cup A^c = S$$
$$A \cap A^c = \emptyset$$
$$(A^c)^c = A$$

$A \cup A^c$



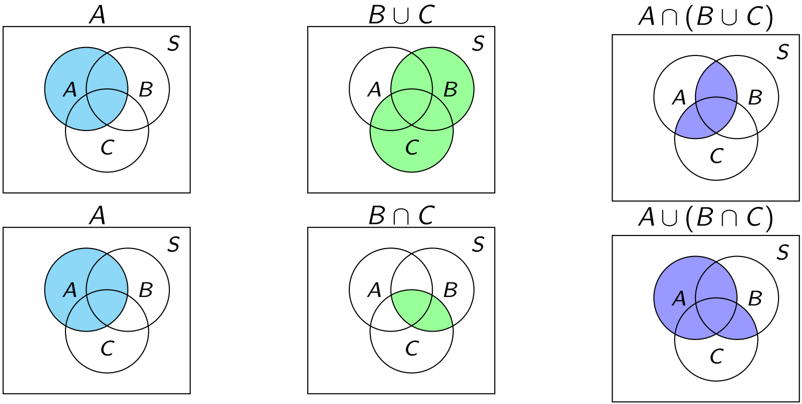
$A \cap A^c$



Set Laws: Distributive Laws

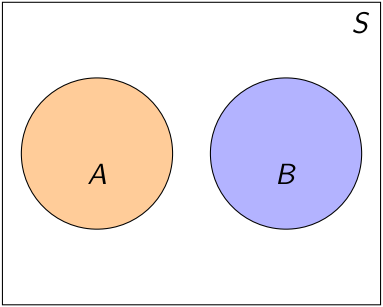
Distributive Laws

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



Disjoint Sets

Example: Draw two sets such that $A \cap B = \emptyset$



Definition: Disjoint Sets

Two sets A and B are called **disjoint** if they have no elements in common, i.e.,

$$A \cap B = \emptyset$$

Partition of a Set

Definition: Partition of a Set

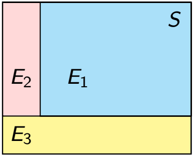
A **partition** of a set S is a collection of subsets $E_1, E_2, \ldots, E_n$ such that:

- Each $E_i \subseteq S$
- The subsets are **disjoint**: $E_i \cap E_j = \emptyset$ for $i \neq j$
- Their union covers all of S: $E_1 \cup E_2 \cup \cdots \cup E_n = S$

Example: Let $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
A valid partition into 3 subsets:

$$E_1 = \{2\}, \quad E_2 = \{3\}, \quad E_3 = \{4, \ldots, 13\}$$

Partition of S



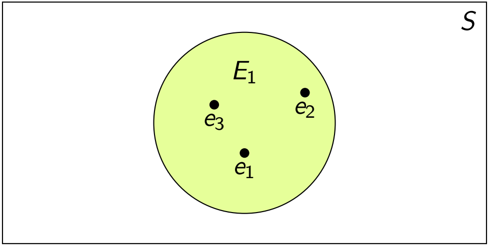
Probability and Sets

- **Random Experiment:** A process that leads to a single but unpredictable outcome.
- **Sample Space** (S): The set of all possible outcomes of a random experiment. It plays the role of the *universal set* in probability.
  - *Discrete:* countable outcomes (e.g., rolling a die)
  - *Continuous:* uncountable outcomes (e.g., measuring height)
- **Sample Points:** The individual outcomes in the sample space.
- **Simple Events:** Events consisting of exactly one sample point.
- **Compound Events:** Events made up of multiple sample points - i.e., subsets of S with more than one element.
- **Probability** ($P(X)$): A number between 0 and 1 that measures the likelihood that event X occurs when the experiment is run.

Probability and Sets

Set notation gives us a convenient way to represent probability in the context of random experiments:

- The **sample space** S is the set of all possible outcomes (the universal set).
- **Simple events** are individual outcomes, denoted by e, and are elements of S.
- **Compound events** are subsets of the sample space, denoted by E.
- Events in probability are treated as sets, so the laws of sets apply to probability.



The Axioms of Probability

Let A be an event in a sample space S:

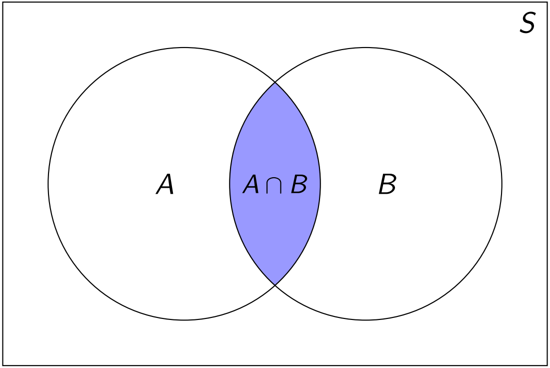
- **Axiom 1:**
$$0 \leq P(A) \leq 1$$
The probability of any event is a number between 0 and 1.
- **Axiom 2:**
$$P(S) = 1$$
The probability of the entire sample space is 1.
- **Axiom 3 (Additivity):** If $A_1, A_2, \ldots, A_n$ are *pairwise mutually exclusive events*, then:
$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

Probability for the Intersection of Two Events

Intersection of Events

$$P(A \cap B)$$

This represents the probability that **both** events A **and** B occur.



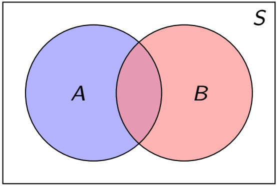
Probability for the Union of Two Events

Rule for the Union of Two Events

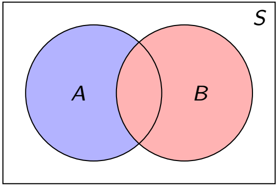
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This represents the probability of event A **or** event B occurring.

$P(A) + P(B)$



$P(A \cup B)$



If we add $P(A) + P(B)$, we count $P(A \cap B)$ twice. Subtracting it once gives the correct union.

$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$
$$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1. When would $P(A \cup B) = P(A) + P(B)$?
2. Give a formula for $P(A \cup B \cup C)$

Law of Total Probability

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

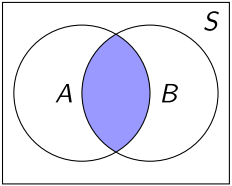
To get the total probability for event A, we consider:

Either A occurs with B, or A occurs with B^c

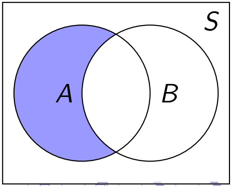
Since these events are disjoint, we can add their probabilities:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$P(A \cap B)$



$P(A \cap B^c)$



Other Laws of Probability

DeMorgan's Laws

$$P((A \cup B)^c) = P(A^c \cap B^c) \qquad P((A \cap B)^c) = P(A^c \cup B^c)$$

Distributive Laws

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

Mutually Exclusive Events

Two events A and B are **mutually exclusive** if:

$$P(A \cap B) = 0$$

This means it's impossible for both events to occur at the same time.

Example: What is the probability of rolling a 2?

Question: Two three-sided dice (numbered 1 to 3) are rolled. What is the probability that a 2 appears on at least one die?

Step 1: What is the sample space?

$$S = \{11, 12, 13, 21, 22, 23, 31, 32, 33\} \quad \text{(9 outcomes)}$$

Step 2: Which outcomes include at least one 2?

$$E = \{12, 21, 22, 23, 32\}$$

Step 3: Calculate the probability.

$$P(\text{At least one 2}) = \frac{5}{9}$$

Counting: Multiplicative Rule & Factorials

Multiplicative Rule

If a task can be done in m ways, and a second task in n ways, then the total number of outcomes is:

$$m \times n$$

Example: If you choose 1 of 3 shirts and 1 of 2 pants, total outfits $= 3 \times 2 = 6$.

Factorials

$$n! = n \times (n-1) \times \cdots \times 1 \qquad \text{(by definition: } 0! = 1\text{)}$$

Example: $4! = 4 \times 3 \times 2 \times 1 = 24$

Counting: Permutations & Combinations

Permutations (Order Matters)

$$P(n, r) = \frac{n!}{(n-r)!}$$

Example: Number of ways to assign 1st, 2nd, 3rd place from 5 people:

$$P(5, 3) = \frac{5!}{2!} = 60$$

Combinations (Order Doesn't Matter)

$$C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example: Number of ways to choose 3 people from 5:

$$\binom{5}{3} = \frac{5!}{3!2!} = 10$$

Counting: Repeated Elements

Permutations with Repeated Elements

When elements repeat, divide to correct for overcounting:

$$\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$$

Example: "BALLOON" has 7 letters with repeats: 2 L's, 2 O's

$$\frac{7!}{1! \cdot 1! \cdot 2! \cdot 2! \cdot 1!} = \frac{5040}{4} = 1260 \text{ arrangements}$$

Simulating Probabilities In An Experiment

Up to this point, we've focused on *theoretical probabilities* derived mathematically.

Theoretical probability predicts what should happen, but it doesn't tell us what *will* happen in one experiment.

We can also use **simulation** to estimate probabilities by performing experiments and observing outcomes.

Key Idea

Estimate the probability of an event by simulating the experiment many times and observing the relative frequency of the event.

The Process of Simulation

We can summarize the steps for simulating an experiment as follows:

1. **State the Problem:** Clearly define the random process or phenomenon you are trying to simulate.
2. **Identify Assumptions:** Describe any assumptions you are making (e.g., independence, fixed probabilities).
3. **Assign Digits:** Use random digits (e.g., 00-99) to represent outcomes based on their probabilities.
4. **Simulate Repetitions:** Run many trials of the simulated experiment using random number tables or technology (e.g., `randInt` on a calculator).
5. **Draw Conclusions:** Estimate probabilities or answer the question based on the proportion of trials that meet the event criteria.

The Process of Simulation

We can summarize the steps for simulating an experiment as follows:

1. **State the Problem:** Clearly define the random process or phenomenon you are trying to simulate.
2. **Identify Assumptions:** Describe any assumptions you are making (e.g., independence, fixed probabilities).
3. **Assign Digits:** Use random digits (e.g., 00-99) to represent outcomes based on their probabilities.
4. **Simulate Repetitions:** Run many trials of the simulated experiment using random number tables or technology (e.g., `randInt` on a calculator).
5. **Draw Conclusions:** Estimate probabilities or answer the question based on the proportion of trials that meet the event criteria.

This method is especially useful when theoretical calculations are too complex or unknown.

Simulation: How Many Boxes Until 2 Dumbledore Cards?

In *Harry Potter and the Philosopher's Stone*, Ron Weasley says: "I got about 6 of him," referring to Dumbledore cards in chocolate frogs.
Suppose 12% of boxes contain a Dumbledore card. Simulate how many boxes it takes to get 2 Dumbledore cards.

1. **Define:** Count how many boxes until 2 Dumbledore cards appear.
2. **Assumptions:** Trials are independent; $P(\text{Dumbledore}) = 0.12$
3. **Assign Digits:**
   - 00-11 $\rightarrow$ Dumbledore (12 numbers)
   - 12-99 $\rightarrow$ Other (88 numbers)
4. **Simulate using TI-84:** `randInt(0,99,n)`

Simulation: How Many Boxes Until 2 Dumbledore Cards?

In *Harry Potter and the Philosopher's Stone*, Ron Weasley says: "I got about 6 of him," referring to Dumbledore cards in chocolate frogs.
Suppose 12% of boxes contain a Dumbledore card. Simulate how many boxes it takes to get 2 Dumbledore cards.

1. **Define:** Count how many boxes until 2 Dumbledore cards appear.
2. **Assumptions:** Trials are independent; $P(\text{Dumbledore}) = 0.12$
3. **Assign Digits:**
   - 00-11 $\rightarrow$ Dumbledore (12 numbers)
   - 12-99 $\rightarrow$ Other (88 numbers)
4. **Simulate using TI-84:** `randInt(0,99,n)`

Simulation 1: 30 35 36 85 59 76 80 82 25 99 26 73 72 71 61 **9** 96 **3** $\rightarrow$ **18 boxes**
Simulation 2: 53 43 **5** 34 33 63 **7** $\rightarrow$ **7 boxes**
Simulation 3: 89 31 61 **9** 81 83 23 41 81 97 14 **11** $\rightarrow$ **12 boxes**
Estimated average: $\frac{18+7+12}{3} = \mathbf{12.33}$ boxes
Computer simulation (10,000 trials): $\approx$ **16.63** boxes

Conditional Probability

Conditional probability describes the probability of an event occurring *given that* another event has already occurred.

Examples:

- What's the probability someone is taller than $6'4''$ given they play in the NBA?
- Given someone is a heavy smoker, what is the probability they develop lung cancer?
- What is the probability of passing a test, given you studied for more than 5 hours?

Key idea: Conditional probability *restricts* the sample space. You're no longer considering all possible outcomes, but only those for which the given condition is true.

Conditional Probability

Conditional Probability:

The conditional probability of event A given that event B has occurred is denoted:

$$P(A \mid B)$$

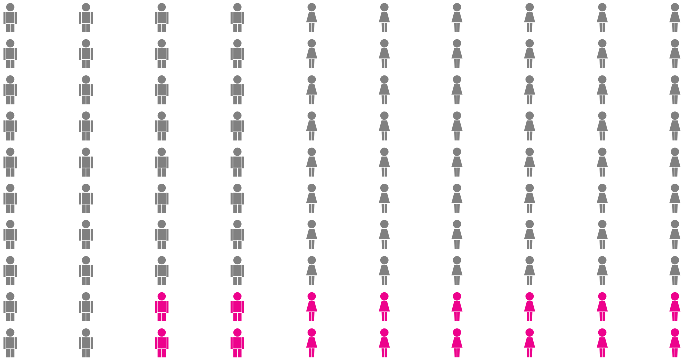
It is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Important Distinctions:

- $P(A \cap B)$: Probability that both A and B occur - an intersection.
- $P(A \mid B)$: Probability that A occurs, *given* that B has already occurred (condition).
- $P(B \mid A)$: Probability that B occurs, given A has occurred - note the reversal in condition.

Example: Conditional Probability



A town has 100 people: 40 are male and 60 are female. Out of the 100 people, 16 have competed in international math competitions.

1. What is the probability someone is a female given they have competed in a competition?
2. What is the probability of haven competed in a math competition, given the person selected is a male?

Conditional Probability: Sports Preferences

A statistician surveys 324 people:

- 37 prefer hockey, the rest prefer soccer.
- Of the 37 who prefer hockey, 32 are from Canada.
- 48 people in total are from Canada, the rest (276) from Mexico.

(a) What is the probability someone prefers Hockey given they are from Mexico?

	Canada	Mexico	Total
Hockey	32	5	37
Soccer	16	271	287
Total	48	276	324

We restrict to the 276 people from Mexico. Out of these, 5 prefer hockey.

$$P(\text{Hockey} \mid \text{Mexico}) = \frac{5}{276}$$

(b) What is the probability someone is both from Canada and prefers Hockey?

$$P(\text{Canada} \cap \text{Hockey}) = \frac{32}{324}$$

Independent Events

Two events A and B are **independent** if the occurrence of one does **not** affect the probability of the other.

Test for Independence:

- $P(A \mid B) = P(A)$
- $P(B \mid A) = P(B)$
- Equivalent condition:

$$P(A \cap B) = P(A) \cdot P(B)$$

If this condition is not met, the events are said to be **dependent**.

Bayes' Theorem

We know from the definition of conditional probability:

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

Rearranging gives us Bayes' Theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Interpretation: Bayes' Theorem allows us to reverse conditional probabilities - finding $P(A|B)$ when $P(B|A)$ is known.

It is especially useful when:

- The probability of a cause given an outcome is needed.
- Diagnostic reasoning is involved (e.g. medical testing, spam filters).

Example: Disease and Diagnostic Test (Tree Diagram)

Suppose the probability of carrying a certain disease is 0.023. A diagnostic test has been developed to test for the presence of the disease. If an individual has the disease, the test shows "positive" for the disease with a probability of 0.96. However, if the individual does not have the disease, the test shows "positive" for the disease with a probability of 0.02.

1. What is the probability that a person will test positive for the disease?

$P(D) = 0.023$

- $P(+|D) = 0.96$
- $P(-|D) = 0.04$

$P(D^c) = 0.977$

- $P(+|D^c) = 0.02$
- $P(-|D^c) = 0.98$

$$\begin{aligned} P(+) &= P(+ \cap D) + P(+ \cap D^c) \\ &= (.023)(.96) + (.977)(.02) \\ &= 0.04162 \end{aligned}$$

Example: Disease and Diagnostic Test (Tree Diagram)

Suppose the probability of carrying a certain disease is 0.023. A diagnostic test has been developed to test for the presence of the disease. If an individual has the disease, the test shows "positive" for the disease with a probability of 0.96. However, if the individual does not have the disease, the test shows "positive" for the disease with a probability of 0.02.

2. What is the probability that a person has the disease given that they test positive?

- $P(D) = 0.023$
  - $P(+|D) = 0.96$
  - $P(-|D) = 0.04$
- $P(D^c) = 0.977$
  - $P(+|D^c) = 0.02$
  - $P(-|D^c) = 0.98$

$$P(D|+) = \frac{P(D) \cdot P(+|D)}{P(+)}$$

Example: Alpine Skiing (Tree Diagram 2)

11.6% of Canadians live in Alberta, 13.5% in B.C., and 22.6% in Quebec. 74% of Albertans enjoy alpine skiing, 68% of B.C. residents enjoy it, and 33% of Quebec residents enjoy it. If a person is not from Alberta, B.C., or Quebec (i.e., Other = 52.3%), there is a 54% chance they do not enjoy alpine skiing.

(a) What is the probability someone is from Alberta given they do not enjoy Alpine skiing?

$$\begin{aligned} P(E^c) &= P(\text{Alberta} \cap E^c) + P(\text{B.C.} \cap E^c) + P(\text{Quebec} \cap E^c) + P(\text{Other} \cap E^c) \\ &= (0.116)(0.26) + (0.135)(0.32) + (0.226)(0.67) + (0.523)(0.54) \\ &= 0.03016 + 0.04320 + 0.15142 + 0.28242 = 0.5072 \end{aligned}$$

$$P(\text{Alberta} \mid E^c) = \frac{P(\text{Alberta} \cap E^c)}{P(E^c)} = \frac{(0.116)(0.26)}{0.5072} = \frac{0.03016}{0.5072} \approx 0.0595$$

Final Answer: $\boxed{P(\text{Alberta} \mid E^c) \approx 6\%}$

Random Variables

Random Variables offer a way to assign numeric values to outcomes of a random experiment.
Example: Roll two 3-sided dice. The sample space is:

$$S = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}$$

Let X be the number of ones rolled. X is a **random variable** that varies with each outcome.
We define the probabilities using the values of X:

$$P(X = 0) = P(22) + P(23) + P(32) + P(33) = \frac{4}{9}$$
$$P(X = 1) = P(12) + P(13) + P(21) + P(31) = \frac{4}{9}$$
$$P(X = 2) = P(11) = \frac{1}{9}$$

The values of X (0, 1, 2) are **mutually exclusive**, so:

$$P(X = 0 \text{ or } 1 \text{ or } 2) = \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1$$

Types of Random Variables

Random variables can be either **qualitative** (categorical) or **quantitative** (numerical). In this unit, we focus on **quantitative random variables**, which fall into two categories:

1. Discrete Random Variables
    - Binomial
    - Geometric
    - Hypergeometric
    - Poisson
2. Continuous Random Varaiables
    - Uniform
    - Normal
    - χ^2
    - t_{df}

We will explore both types throughout the chapter. For now, we'll begin with **discrete random variables**.

Probability Distributions for Discrete Random Variables

Example: Flip two coins. Let X be the number of heads. Then $X \in \{0, 1, 2\}$ and:

$$S = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\} \quad \Rightarrow \quad \begin{cases} P(X = 0) = 0.25 \\ P(X = 1) = 0.50 \\ P(X = 2) = 0.25 \end{cases}$$

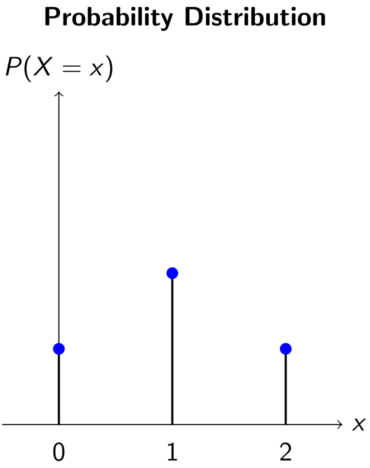
Probability Distribution Table:

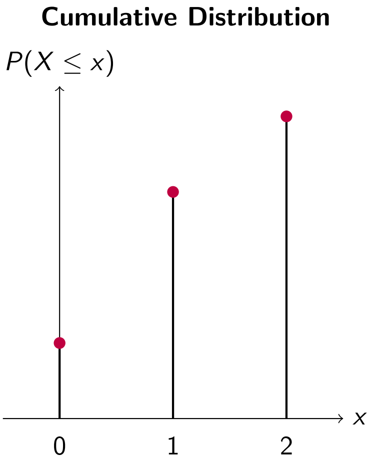
x	0	1	2
$P(X = x)$	0.25	0.50	0.25

Cumulative Distribution Table:

x	0	1	2
$P(X \leq x)$	0.25	0.75	1.00

Visualizing Distributions for X (Heads in Two Coin Flips)





Expected Value: Basketball Shot Simulation

A basketball player shoots 3 times. Historically, she makes 1 out of every 5 shots (20% success rate). Let X be the number of shots made.

(a) Estimate $E[X]$ using Simulation:

- Assign digits: 00-19 for made shots, 20-99 for missed shots.
- Use a random number generator (e.g., TI-84: `randInt(0,99,3)`).

Trial	Random Digits	X
1	76, 21, 91	0
2	57, 36, 38	0
3	54, 18, 00	2
4	05, 56, 54	1

Estimated $E[X] = \frac{0+0+2+1}{4} = 0.75$ (rough estimate from 4 trials)

Running 100,000 simulations in R gives: $E[X] \approx 0.6$

Probability Distribution: Basketball Shooter

A basketball player makes a shot with probability 0.2. Suppose they take 3 shots. Let the random variable X represent the number of shots made.

We compute the theoretical probabilities:

$$\begin{aligned} P(X=0) &= (0.8)^3 = 0.512 \\ P(X=1) &= 3(0.2)(0.64) = 0.384 \\ P(X=2) &= 3(0.04)(0.8) = 0.096 \\ P(X=3) &= (0.2)^3 = 0.008 \end{aligned}$$

Summary Table:

x	0	1	2	3
$P(X=x)$	0.512	0.384	0.096	0.008

Expected Value via Simulation Logic

Let X be the number of shots made out of 3 attempts. The distribution is:

x	0	1	2	3
$P(X = x)$	0.512	0.384	0.096	0.008
Expected count (n trials)	$0.512n$	$0.384n$	$0.096n$	$0.008n$

Expected Value via Simulation Logic

Let X be the number of shots made out of 3 attempts. The distribution is:

x	0	1	2	3
$P(X = x)$	0.512	0.384	0.096	0.008
Expected count (n trials)	$0.512n$	$0.384n$	$0.096n$	$0.008n$

Average value across n trials:

$$E[X] = \frac{0.512n(0) + 0.384n(1) + 0.096n(2) + 0.008n(3)}{n}$$

Expected Value via Simulation Logic

Let X be the number of shots made out of 3 attempts. The distribution is:

x	0	1	2	3
$P(X = x)$	0.512	0.384	0.096	0.008
Expected count (n trials)	$0.512n$	$0.384n$	$0.096n$	$0.008n$

Average value across n trials:

$$E[X] = \frac{0.512n(0) + 0.384n(1) + 0.096n(2) + 0.008n(3)}{n}$$

Cancel n and simplify:

$$E[X] = 0(0.512) + 1(0.384) + 2(0.096) + 3(0.008) = \boxed{0.6}$$

Expected Value of a Discrete Random Variable

Definition: The expected value of a discrete random variable X (denoted $E(X)$ or μ_X) represents the long-run average value of X over many repetitions of an experiment.
Formula:

$$E[X] = x_1 \cdot p(x_1) + x_2 \cdot p(x_2) + \cdots + x_n \cdot p(x_n) = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

Alternatively written as:

$$E[X] = \sum_{\text{all } x} x \cdot p(x)$$

Prove these key properties:

- $E(c) = c$ (constant)
- $E(aX) = a \cdot E(X)$ (scaling)
- $E(aX + c) = a \cdot E(X) + c$ (scaling + shifting)

Example: Expected Value Transformation

In a trivia contest you win \$5 for each correct answer, but must pay a flat \$2 fee. Let X be the number of correct answers out of 3. the distribution of X is shown below:

x	0	1	2	3
$P(X = x)$	0.3	0.4	0.2	0.1

1. Determine the expected number of correct answers out of three.

$$E(X) = 0(0.3) + 1(0.4) + 2(0.2) + 3(0.1) = 1.1$$

2. Determine the expected net winnings.

$$E(Y) = E(5X - 2) = 5E(X) - 2 = 5(1.1) - 2 = \boxed{3.5}$$

Conclusion: Expected net earnings are \$3.50 per game.

Distribution and Expected Value: Red Cards in Poker

In a game of poker, you are dealt 5 cards from a standard deck. Let Y be the number of red cards in your hand.

(a) Find the Distribution for Y:

y	0	1	2	3	4	5
$P(Y=y)$	0.0253	0.1496	0.3251	0.3251	0.1496	0.0253

These values come from:

$$P(Y=y) = \frac{\binom{26}{y}\binom{26}{5-y}}{\binom{52}{5}}$$

(b) Determine the Expected Value for Y:

$$E[Y] = \sum y \cdot P(Y=y) = 0(0.0253) + 1(0.1496) + \cdots + 5(0.0253) = \boxed{2.5}$$

Variance via Simulation Logic

(c) Estimate Var(Y) using simulation.

- 26 of 52 cards are red $\Rightarrow P(\text{Red}) = 0.5$
- Assign 0–25 to red cards, 26–99 to black cards
- Sample without replacement

Run simulations and compute deviations:

Sim #	Random Draw	Y	$Y - \mu_Y$	$(Y - \mu_Y)^2$
1	41 12 25 35 44	2	-0.5	0.25
2	36 47 06 04 28	2	-0.5	0.25
3	12 17 14 38 06	4	1.5	2.25
4	34 26 10 15 03	3	0.5	0.25

Estimate the variance:

$$\text{Var}(Y) \approx \frac{0.25 + 0.25 + 2.25 + 0.25}{4} = \boxed{0.75}$$

Why not use average distance from the mean?

Variance via Simulation Logic

Let Y be the number of red cards in a 5-card hand. From earlier:

y	0	1	2	3	4	5
$P(Y=y)$	0.0253	0.1496	0.3251	0.3251	0.1496	0.0253

We previously calculated $\mu_Y = 2.5$

If we simulate n trials, we expect about $np(y)$ values of each y.

Average of squared deviations:

$$\begin{aligned}\text{Var}(Y) &= \frac{1}{n}\sum np(y)(y-\mu_Y)^2 \\ &= \sum p(y)(y-2.5)^2 \\ &= \boxed{1.151961}\end{aligned}$$

So: $\text{Var}(Y) = E\left[(Y-\mu_Y)^2\right]$

Variance and Standard Deviation

Variance: Measures the average squared deviation from the mean.

$$\text{Var}(X) = E\left[(X - \mu_X)^2\right]$$

- Variance gives a sense of how *spread out* the values of a random variable are.
- Units of variance are the *square* of the units of the original variable.

Standard Deviation: The square root of the variance. It brings the measure of spread back to the original units.

$$SD(X) = \sqrt{\text{Var}(X)} \quad \text{or} \quad \sigma_X = \sqrt{\sigma_X^2}$$

- Easier to interpret because it matches the units of X.
- Commonly used to describe typical deviation from the mean.

An Alternative Formula for Variance

Another useful formula for variance.

$$\boxed{\text{Var}(Y) = E[Y^2] - (E[Y])^2}$$

Proof:

$$\begin{aligned}
\text{Var}(Y) &= E[Y^2 - 2\mu_Y Y + \mu_Y^2] \\
&= E[Y^2] - 2\mu_Y E[Y] + \mu_Y^2 \\
&= E[Y^2] - 2\mu_Y^2 + \mu_Y^2 \\
&= E[Y^2] - \mu_Y^2 \\
&= E[Y^2] - (E[Y])^2
\end{aligned}$$

Example: Variance Using Shortcut Formula

Let X be the number of heads when flipping two coins. Then:

x	x^2	$P(X = x)$
0	0	0.25
1	1	0.50
2	4	0.25

1. Compute $E[X]$ and $E[X^2]$

$$E[X] = 0(0.25) + 1(0.50) + 2(0.25) = 1$$

$$E[X^2] = 0(0.25) + 1(0.50) + 4(0.25) = 1.5$$

2. Calculate $Var(x)$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 1.5 - (1)^2 = \boxed{0.5}$$

Discrete Random Variables in This Course

In this course, we will study four important discrete random variables. Only the first two are part of the **AP Statistics Curriculum**.

- **Binomial Distribution** (*AP Required*)
  Describes the number of successes in a fixed number of independent trials, each with the same probability of success.
- **Geometric Distribution** (*AP Required*)
  Describes the number of trials needed to get the first success in a sequence of independent trials with the same probability of success.
- **Hypergeometric Distribution** (*Pseudo - Enrichment*)
  Like the binomial, but without replacement. Describes the number of successes in a sample drawn without replacement from a finite population.
- **Poisson Distribution** (*Enrichment*)
  Describes the number of occurrences of an event in a fixed interval of time or space when events occur independently at a constant average rate.

We will focus deeply on binomial and geometric, and briefly explore the others for

Binomial Random Variables

Example: A basketball player makes 75% of free throws. During practice, she takes 8 shots. Let X be the number of shots she makes.
This situation can be modeled by a **binomial random variable**. Why?

- Each shot is either a **make (success)** or a **miss (failure)**.
- The number of shots ($n = 8$) is **fixed**.
- The probability of success on each shot is **constant**, $p = 0.75$.
- Each shot is **independent** of the others.

Definition: A random variable X is **binomial** if it counts the number of successes in n independent trials, each with the same probability of success p.

Notation: $X \sim \text{Binomial}(n, p)$

Probability Mass Function:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{for } k = 0, 1, 2, \ldots, n$$

Mean and Variance of a Binomial Variable

If $X \sim \text{Binomial}(n, p)$, then:

- **Mean (Expected Value):**

$$\mu = E[X] = np$$

- **Variance:**

$$\sigma^2 = \text{Var}(X) = np(1-p)$$

- **Standard Deviation:**

$$\sigma = \sqrt{np(1-p)}$$

Interpretation:

- $E[X]$ is the average number of successes in n trials.
- $\text{Var}(X)$ and $\text{SD}(X)$ measure the variability in the number of successes.

Proving Expected Value and Variance for Binomial Random Variables

We will prove the expected value and variance for binomial random variables using three different methods (see class notes):

1. Direct algebraic proof
2. Sum of Bernoulli Random Variables
3. Moment Generating Functions

Moment Generating Functions (MGFs)

Definition: The **moment generating function** (MGF) of a random variable X is defined as:

$$M_X(t) = E[e^{tX}]$$

if the expectation exists for values of t in an open interval around 0.

Why it's useful:

- MGFs encode all moments (like mean and variance) of a random variable.
- The kth moment of X is given by:

$$E[X^k] = M_X^{(k)}(0)$$

  That is, the kth derivative of the MGF evaluated at $t = 0$.

To find the expected value (mean):

$$E[X] = M'_X(0)$$

Example: If $M_X(t) = e^{3t+2t^2}$, then:

$$M'_X(t) = 3e^{3t+2t^2} + 4te^{3t+2t^2}$$

$$\Rightarrow E[X] = M'_X(0) = 3$$

Geometric Random Variables

Example: A basketball player makes 75% of free throws. She shoots until she makes her first basket. Let X be the number of shots it takes to make her first successful free throw.

This situation is modeled by a **geometric random variable**. Why?

- Each shot is either a **success (make)** or **failure (miss)**.
- The probability of success on each shot is **constant**, $p = 0.75$.
- The trials are **independent**.
- We are counting the number of trials **until the first success**.

Definition: A random variable X is **geometric** if it counts the number of independent trials until the first success.

Notation: $X \sim \text{Geometric}(p)$

Probability Mass Function (PMF):

$$P(X = k) = (1 - p)^{k-1} \cdot p \quad \text{for } k = 1, 2, 3, \ldots$$

Proving Expected Value and Variance for Geometric Random Variable

We will prove the expected value and variance for binomial random variables using three different methods (see class notes):

1. Geometric Series
2. Moment Generating Functions
3. Method from Calculus

Mean, Variance, and SD of Geometric Random Variables

Let $X \sim \text{Geometric}(p)$, where p is the probability of success on each trial.

Mean (Expected Value):

$$E[X] = \frac{1}{p}$$

Variance:

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

Standard Deviation:

$$\text{SD}(X) = \sqrt{\text{Var}(X)} = \frac{\sqrt{1-p}}{p}$$

Example: If $p = 0.75$ (free throw success rate):

$$E[X] = \frac{1}{0.75} = \boxed{1.\overline{3}}, \quad \text{SD}(X) = \frac{\sqrt{0.25}}{0.75} \approx \boxed{0.666}$$

Introducing the Hypergeometric Distribution

Example: A box contains 10 marbles: 4 are red and 6 are blue. You randomly draw 3 marbles **without replacement**. Let X be the number of red marbles drawn.

This scenario is modeled by a **hypergeometric random variable**. Why?

- The population size is fixed: $N = 10$
- The number of successes in the population is fixed: $r = 4$ red marbles
- The sample size is fixed: $n = 3$ draws
- Sampling is done **without replacement**

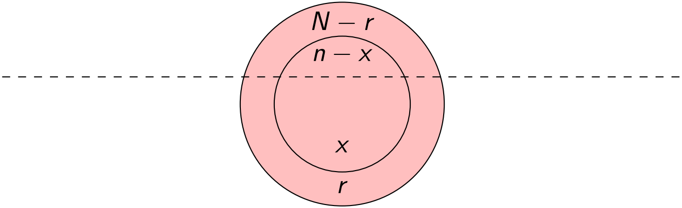
Therefore: $X \sim \text{Hypergeometric}(N = 10, r = 4, n = 3)$

$$P(X = x) = \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}, \quad \text{for } x = 0, 1, 2, 3$$

Let $X \sim \text{Hypergeometric}(N, r, n)$, where:

- N: Population size
- r: Number of successes in the population
- n: Sample size drawn without replacement

X counts the number of successes in a random sample of size n drawn *without replacement* from a population of size N that contains r successes.



1. Probability Mass Function:

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, n$$

Hypergeometric: Mean, Variance, and Standard Deviation

Let $X \sim \text{Hypergeometric}(N, r, n)$, where:

- N: Population size
- r: Number of successes in the population
- n: Sample size drawn without replacement

Mean (Expected Value):

$$E(X) = n \cdot \frac{r}{N}$$

Variance:

$$\text{Var}(X) = n \cdot \frac{r}{N} \cdot \left(1 - \frac{r}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$$

Standard Deviation:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Note: This is similar to the binomial variance, but includes a correction factor $\frac{N-n}{N-1}$ because the draws are without replacement.

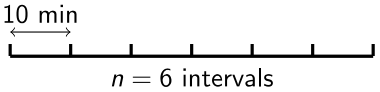
Poisson Random Variables from Binomial (Enrichment)

Example: A police officer observes cars for 1 hour. On average, 5 cars speed per hour. Let X be the number of cars speeding per hour.

Poisson Random Variables from Binomial (Enrichment)

Example: A police officer observes cars for 1 hour. On average, 5 cars speed per hour. Let X be the number of cars speeding per hour.

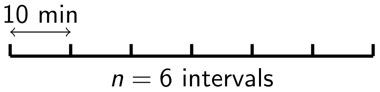
Break the hour into 6 equal intervals (10 minutes each). Assume at most one speeder per interval. An approximation is $X \sim \text{Binomial}(n = 6, p = \frac{5}{6})$.



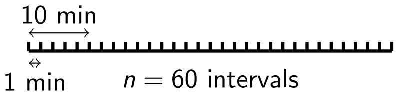
Poisson Random Variables from Binomial (Enrichment)

Example: A police officer observes cars for 1 hour. On average, 5 cars speed per hour. Let X be the number of cars speeding per hour.

Break the hour into 6 equal intervals (10 minutes each). Assume at most one speeder per interval. An approximation is $X \sim \text{Binomial}(n = 6, p = \frac{5}{6})$.



Now refine to 60 one-minute intervals. Then $X \sim \text{Binomial}(n = 60, p = \frac{5}{60})$.



Take the limit:

$$P(X = x) = \lim_{n \to \infty} \binom{n}{x} \left(\frac{5}{n}\right)^x \left(1 - \frac{5}{n}\right)^{n-x} = \frac{5^x e^{-5}}{x!}, \quad x = 0, 1, 2, \ldots$$

Poisson Distribution Summary

The **Poisson distribution** models the number of events that occur in a fixed interval of time or space, assuming:

- Events occur **independently**
- Events occur **at a constant average rate** λ
- Two events cannot occur at exactly the same instant (i.e., events are **rare and discrete**)

Probability Mass Function (PMF):

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

Parameter: $\lambda > 0$, the average number of occurrences in the interval

Expected Value: $E(X) = \lambda$
Variance: $\text{Var}(X) = \lambda$
Standard Deviation: $\sigma_X = \sqrt{\lambda}$

Which Random Variable?

For each situation, identify the appropriate random variable type.

1. A basketball player shoots 10 free throws. Let X be the number of shots made. **Binomial**
2. A call center receives an average of 3 calls per minute. Let X be the number of calls in one minute. **Poisson**
3. Cards are drawn from a deck without replacement. Let X be the number of red cards in a hand of 5 cards. **Hypergeometric**
4. A factory tests items until the first defective one is found. Let X be the number of items tested. **Geometric**
5. A quiz has 5 multiple choice questions. Let X be the number answered correctly by random guessing. **Binomial**
6. A traffic officer counts the number of speeding violations in a 2-hour period. **Poisson**

Sums and Differences of Independent Random Variables

Let X and Y be two **independent** random variables. Then for any constants a and b:

Mean of a Linear Combination

$$E(aX + bY) = aE(X) + bE(Y)$$

Variance of a Linear Combination

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Important: These formulas *only apply if X and Y are independent.*

Note: We can compute the mean and variance of $Z = aX + bY$, but we may not know the full probability distribution of Z.

Example: Combining Independent Random Variables

Let $X \sim \text{Binomial}(n = 10, p = 0.4)$, and $Y \sim \text{Binomial}(n = 15, p = 0.6)$, with X and Y independent.

Let $Z = X + Y$. Then:

	X	Y
Mean	$10(0.4) = 4$	$15(0.6) = 9$
Variance	$10(0.4)(0.6) = 2.4$	$15(0.6)(0.4) = 3.6$

Then:

- $E(Z) = E(X) + E(Y) = 4 + 9 = \boxed{13}$
- $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) = 2.4 + 3.6 = \boxed{6.0}$
- $\text{SD}(Z) = \sqrt{6.0} \approx \boxed{2.45}$

Sums and Differences Without Independence (Enrichment)

Let X and Y be two random variables (not necessarily independent). For constants a and b:

Mean (still the same)

$$E(aX + bY) = aE(X) + bE(Y)$$

Variance (requires covariance)

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\,\text{Cov}(X, Y)$$

- $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
- If X and Y are independent, then $\text{Cov}(X, Y) = 0$

Important: The independence assumption simplifies calculations-without it, we must account for how the two variables move together.

Discrete vs. Continuous Random Variables

Discrete Random Variables take on a finite or countable number of values.

- Example: $X = \{\text{number of heads in 5 coin flips}\}$
- Visualized using a probability **mass** function (pmf)

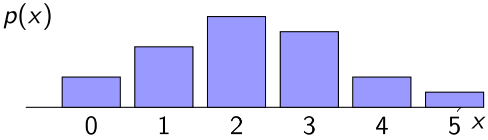
Continuous Random Variables take on an infinite (uncountable) number of values.

- Example: $Y =$ time it takes for a light to change
- Visualized using a probability **density** function (pdf)

We must now rethink how we assign and visualize probabilities when the number of possible outcomes is infinite.

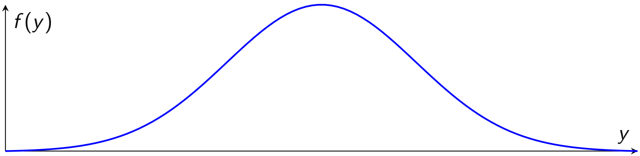
Comparing PMF and PDF

Discrete Random Variable (PMF)



PMF: Assigns probability to each countable outcome. $\sum p(x) = 1$

Continuous Random Variable (PDF)



PDF: Probability is area under curve. $\int f(y)\,dy = 1$

Probability Density Functions (PDFs)

A **probability density function (PDF)** is a function that describes the likelihood of a continuous random variable taking on values within a certain range.

Properties of PDFs:

1. $f(y) \geq 0$ for all values of y
   - A density function can never be negative.
2. $\int_{-\infty}^{\infty} f(y)\, dy = 1$
   - The total area under the curve is always 1.
   - This represents the total probability across all outcomes.

Important Reminder:

- The **probability at a single point** is always zero: $P(Y = a) = 0$
- Only ranges of values (areas under the curve) have nonzero probability.

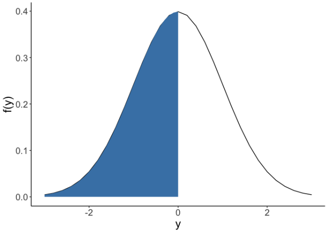
Cumulative Distribution Functions (CDFs)

Definition: The **cumulative distribution function (CDF)** of a random variable Y gives the probability that Y is less than or equal to a value a:

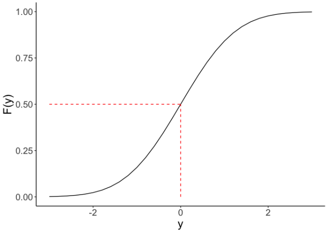
$$F(a) = P(Y \leq a)$$

$F(y)$ is the total area under $f(y)$ from $-\infty$ to y:

$$F(y) = \int_{-\infty}^{y} f(t)\,dt \qquad f(y) = \frac{d}{dy}F(y)$$



PDF $f(y)$



CDF $F(y)$

Mean and Variance for Continuous Random Variables

Expected Value: Just like in the discrete case, the expected value is the theoretical long-run average of the variable. For discrete random variable X:

$$E(X) = \sum_{\text{all } x} x \cdot p(x)$$

For continuous random variable Y with pdf $f(y)$:

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y)\, dy$$

Variance: Measures how spread out values of Y are from its mean μ_Y.

$$\text{Var}(Y) = E[(Y - \mu_Y)^2] = E(Y^2) - [E(Y)]^2$$

Note: These formulas are direct analogues to the discrete case, just replacing sums with integrals.

Uniformly Distributed Random Variables

A continuous random variable Y is **uniformly distributed** on the interval $[a, b]$ if it has constant density throughout the interval.

1. Derive the pdf for X that is uniformly distributed on the interval $[a, b]$.
2. Derive the mean and variance for X directly
3. Determine the moment generating function for X

Uniform Random Variables

Uniform Random Variables model outcomes that are equally likely over an interval $[a, b]$. We write:

$$Y \sim \mathrm{Uniform}(a, b)$$

Probability Density Function (PDF):

$$f(y) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function (CDF):

$$F(y) = \begin{cases} 0 & \text{for } y < a \\ \dfrac{y-a}{b-a} & \text{for } a \leq y \leq b \\ 1 & \text{for } y > b \end{cases}$$

Expected Value and Variance:

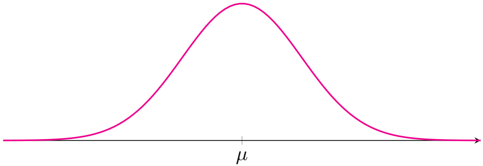
$$E(Y) = \frac{a+b}{2}, \qquad \mathrm{Var}(Y) = \frac{(b-a)^2}{12}$$

Normal Random Variables

Normally Distributed Random Variables are the most widely used type of random variable in statistics. The normal (or Gaussian) distribution forms the familiar bell curve.

$$Y \sim \text{Normal}(\mu, \sigma) \qquad f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- μ is the **mean** (center) of the distribution.
- σ is the **standard deviation** (spread).
- $E(Y) = \mu, \quad \text{Var}(Y) = \sigma^2$



The Standard Normal Distribution

The **standard normal distribution** is a normal distribution with:

$$Z \sim \text{Normal}(0, 1)$$

That is, mean $\mu = 0$ and standard deviation $\sigma = 1$.

Any normal random variable $X \sim \text{Normal}(\mu, \sigma)$ can be transformed into a standard normal variable Z using:

$$Z = \frac{X - \mu}{\sigma}$$

What is a z-score?

- A z-score tells us how many standard deviations a value is from the mean.
- A value x can be rewritten as: $x = \mu + z\sigma$
- Rearranging gives: $z = \dfrac{x - \mu}{\sigma}$

Sums and Differences of Normal Random Variables

Suppose we have normal random variables $X_1, X_2, \ldots, X_n$ where $X_i \sim \text{Normal}(\mu_i, \sigma_i)$.

Let:

$$U = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

Then U is also normally distributed:

$$U \sim \text{Normal}\left(\sum_{i=1}^{n} a_i\mu_i\,,\ \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Example:

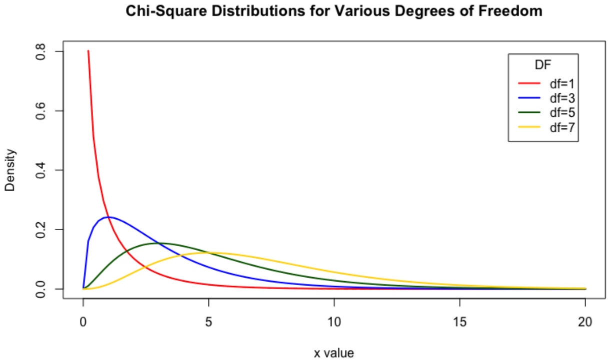
- Let $X \sim \text{Normal}(5, 3)$ and $Y \sim \text{Normal}(2, 2)$
- Find the distribution of $2X + 3Y$

$$2X + 3Y \sim \text{Normal}\left(2(5) + 3(2),\ 2^2(3^2) + 3^2(2^2)\right) = \text{Normal}(16, 72)$$

The χ^2 Distribution

Definition: χ^2_k denotes a χ^2 distribution with k **degrees of freedom**.

- Mean: $E(X) = k$
- Variance: $Var(X) = 2k$
- The shape is **right-skewed**, especially for small k
- Becomes more symmetric as k increases



The t-Distribution

A t_k distribution can be defined as the ratio of a standard normal variable Z and the square root of a chi-square variable divided by its degrees of freedom:

$$t = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}}$$

Key Properties:

- Symmetric, bell-shaped, and centered at 0.
- Heavier tails than the normal distribution (more probability in the extremes).
- As degrees of freedom increase, it approaches the normal distribution.
- Used when estimating population means with small samples or unknown standard deviations.

William Sealy Gosset and the t-Distribution

William Sealy Gosset (1876-1937) was a chemist and statistician who worked at the **Guinness Brewery** in Dublin.

- At Guinness, Gosset was tasked with improving the quality of stout using better statistical methods for small samples.
- He developed the t-**distribution** to solve problems of inference when sample sizes were small and population standard deviation was unknown.
- Guinness had strict rules against employees publishing research, so Gosset wrote under the pseudonym **"Student"**.
- His 1908 paper *"The Probable Error of a Mean"* introduced the now-famous **Student's t-distribution**.
- The t-distribution became one of the cornerstones of modern statistics, particularly in small-sample hypothesis testing.

Fun fact: Gosset and R.A. Fisher were close collaborators, and their work laid the foundations for modern statistical inference.

Beyond AP: More Continuous Distributions

In university-level probability and statistics, students explore a variety of continuous distributions beyond those required in AP Statistics. These are often used in modeling, simulation, and theoretical work.

- **Exponential Distribution:**
  - Models waiting times between independent events (e.g., time until the next customer arrives).
  - Memoryless property.
- **Gamma Distribution:**
  - Generalization of the exponential distribution.
  - Used to model wait times for multiple events.
- **Beta Distribution:**
  - Used for modeling random variables that are constrained to an interval $[0, 1]$.
  - Appears in Bayesian statistics.
- **Weibull Distribution:**
  - Often used in reliability engineering and survival analysis.