Applied Probability and Statistics I - STAT400

Tom Mitchell

Mestiyage - Fall 2024

Syllabus

Grading

- Homework — 28% (7% each)
- R Projects — 12% (4% each)
- Two exams — 30% (15% each)
- Final exam — 30%

Office Hours

- Tuesday: 1:00 PM - 1:50 PM (in person, MTH 4106)
- Wednesday: 11:00 AM - 11:50 AM (online)

Exams

- 2 midterms and a final exam

Lecture 1: Tuesday 8/27/2024

Course Overview: STAT400

We study:

1. Probability
2. Descriptive Statistics
3. Inferential Statistics

1. Probability

Probability is the mathematical study of uncertainty.

2. Descriptive Statistics

Descriptive Statistics involves methods for summarizing and describing the important characteristics of a dataset.

3. Inferential Statistics

Inferential Statistics involves methods for using data from a subset (sample) of a larger group (population) to make meaningful conclusions. Note that each time you pick a subset, you lose out on certain information, leading to uncertainty.

Setting: Conducting an Experiment

In this course, we will often assume we are about to conduct an experiment. The possible outcomes of the experiment are known, but the exact outcome is not.

Examples:

- Tossing a coin

- Rolling a die

To study these types of situations, we introduce a *model for probability*.

Ordered Triples $(\Omega, \mathcal{F}, P)$

- Ω is the **sample space**: The set of all possible outcomes of the experiment.

- $\mathcal{F}$ is the **event space**:

  - An **event** is a subset of Ω.
  - An event captures the idea that in many situations, we care about collections of outcomes rather than a single outcome.
  - The event space contains collections of outcomes for which we can assign probabilities.

- P is the **probability measure**:
  $$P : \mathcal{F} \to \mathbb{R}$$

  where P assigns probabilities to events.

Example 1: Tossing a Fair Coin

Consider the simple experiment of tossing a **fair** coin.

$$\Omega = \{H, T\}$$

$$\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$$

- $P(\emptyset) = 0$
  ("A probability of 0 represents an outcome/collection of outcomes that never takes place.")

- $P(\{H, T\}) = 1$
  ("A probability of 1 represents an outcome/collection of outcomes that always takes place.")

- $P(\{H\}) = \frac{1}{2}$
  (Requires the coin to be fair.)

- $P(\{T\}) = \frac{1}{2}$
  (Requires the coin to be fair.)

Example 2: Rolling a Fair Die

Consider the experiment of rolling a fair die. Note: ($\mathcal{P}(\Omega)$ is the Powerset of Ω)

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\mathcal{F} = \mathcal{P}(\Omega) =$$
$$\{\emptyset, \{1\}, \{2\}, \ldots, \{6\}, \{1, 2\}, \ldots, \{1, 6\}, \{2, 3\}, \ldots,$$
$$\{1, 2, 3\}, \{2, 3, 4\}, \ldots, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}, \ldots, \Omega = \{1, 2, 3, 4, 5, 6\}\}$$

- $P(\{i\}) = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$

- $P(\emptyset) = 0$

- $P(\Omega) = 1$

- $P(\{1, 2\}) =?$
  (Listing out probabilities for each event is **not** feasible.)

We would want a way to calculate the probability of a particular event by using the "base" probabilities.

We will look at these rules for computing probabilities of complicated events in terms of basic ones in a bit.

Example 3: Tossing a Fair Coin Until We See Heads for the First Time

Consider the experiment of tossing a fair coin repeatedly until we observe heads for the first time, then we stop.

$$\Omega = \{(H), (T, H), (T, T, H), (T, T, T, H), \ldots\}$$

$$\mathcal{F} = \mathcal{P}(\Omega)$$

Let A_i be the event of seeing heads on the ith toss.

$$P(A_i) = \frac{1}{2^i}$$

Question: What is the probability of seeing heads on an even-numbered toss?

$$Q = P(\text{seeing heads on an even-numbered toss}) = ?$$

Example 4: Picking a Number from the Interval [0, 1]

Consider picking a number from the interval $[0, 1]$ such that the probabilities of selecting numbers from two subintervals I_1 and I_2 of $[0, 1]$ are the same whenever I_1 and I_2 have the same length.

$$\Omega = [0, 1]$$

$$\mathcal{F} = \mathcal{P}(\Omega)$$

$$P(\text{selecting a number from a subinterval } I \text{ of } [0, 1]) = \text{length}(I)$$

$$P([0, \frac{1}{4}]) = \frac{1}{4} - 0 = \frac{1}{4}$$

$$P([0, \frac{1}{2}]) = \frac{1}{2} - 0 = \frac{1}{2}$$

$$P\left(\left\{\frac{1}{2}\right\}\right) = P\left(\left[\frac{1}{2}, \frac{1}{2}\right]\right) = \frac{1}{2} - \frac{1}{2} = 0$$

"The probability of 0 indicates an event that does not take place or an event that is so unlikely that the only reasonable value to assign to it is 0, i.e., an event that almost surely does not take place."

On Thursday will learn how to compute problems similar to $P(\mathbb{Q} \cap [0, 1]) = ?$

Lecture 2: Thursday 8/29/2024

Probability

- Mathematical study of uncertainty.

Probability Space

$$(\Omega, F, P)$$

- Ω - All possible outcomes of the experiment of interest.

- F - Contains events, i.e., collections of outcomes we would like to assign probabilities to.

- P - Probability measure.

$$P : F \to \mathbb{R}$$

On Tuesday, we discussed examples and introduced some formulas to compute probabilities of "basic" events.

Goal for Today

Explore how to compute probabilities of "complicated" events in terms of basic events.

Two Questions

1. How do we express a complicated event in terms of basic ones?

2. What rules can be used to compute probabilities once Question 1 has been answered?

- Question 1 is best answered by using tools from set theory.

$$\to \text{We treat } \Omega \text{ as the list of all possible outcomes.}$$

$$\to \text{Events are sublists of } \Omega.$$

Since we are only interested in outcomes (or combinations thereof) of the experiment, we treat Ω as the universal set (i.e., we pretend that nothing outside of Ω exists).
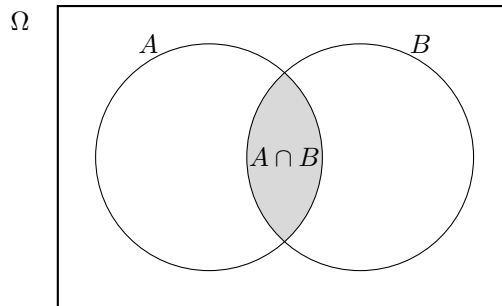
Pictorial Representation

Ω
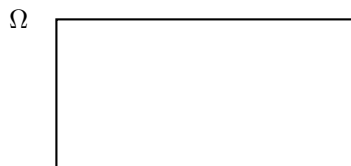
We can manipulate collections of events to obtain new events in various ways:

- Intersections

- Unions

- Relative complements

Given events A and B, the intersection is the event that contains all outcomes common to both A and B. We denote this by $A \cap B$.

Intersection of Two Events

Example

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Let $A = \{1, 3, 5\}$ and $B = \{4, 5\}$. Then, the intersection of A and B is:

$$A \cap B = \{5\}$$

Extending the Intersection to Families of Events

The idea of the intersection of events can be extended to families of events.

$$\bigcap_{i=1}^{3} A_i = (A_1 \cap A_2) \cap A_3$$

For an infinite family of events:

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 \cap \ldots$$

Disjoint Events

Two events are said to be disjoint if $A \cap B = \emptyset$.

- Events $A_1, \ldots, A_n$ are said to be disjoint if $\bigcap_{i=1}^{n} A_i = \emptyset$.

- Events $A_1, \ldots, A_n, \ldots$ are said to be disjoint if $\bigcap_{i=1}^{\infty} A_i = \emptyset$.

- Events $A_1, A_2, \ldots$ are said to be pairwise disjoint if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Example

Consider the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the following events:

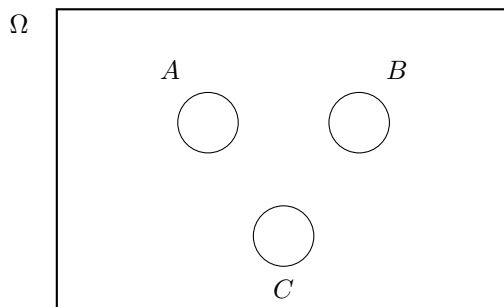$$A = \{1, 3, 5\}, \quad B = \{4, 5\}, \quad C = \{6\}$$

Here, A, B, and C are disjoint events, meaning:

$$A \cap B \cap C = \emptyset$$

However, they are not pairwise disjoint, because:

$$A \cap B = \{5\} \neq \emptyset$$
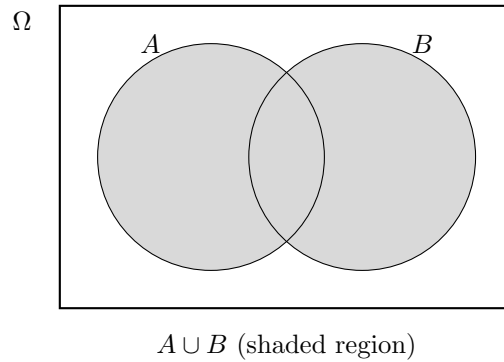
Pairwise Disjoint Sets

The word "and" usually translates to intersection when discussing sets or events. Conversely, the word "or" translates to a union.

Unions

The union of two events A and B, denoted by $A \cup B$, is the event that contains all outcomes that are in A, B, or both.

Venn Diagram: Union of Two Sets

Ω

$A \cup B$ (shaded region)

Example

Consider the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the following events:

$$A = \{1, 3, 5\}, \quad B = \{2, 3, 5\}$$

The union of A and B is:

$$A \cup B = \{1, 2, 3, 5\}$$

Infinite and Finite Unions

Infinite versions of unions and finite versions of unions with more than two sets can be similarly defined.

$$\bigcup_{i=1}^{3} A_i = A_1 \cup A_2 \cup A_3$$

Similarly, the union of more than two finite sets can be defined as:

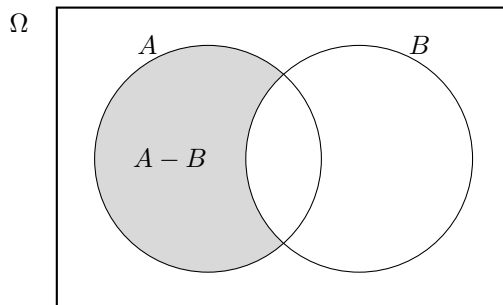$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$$

Unions can also be extended to infinite collections of sets. For example, the union of an infinite sequence of sets $A_1, A_2, \ldots$ is denoted as:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \ldots$$

Relative Complements

Given events A and B, the relative complement $A - B$ is the event that contains only the outcomes that are **unique** to A.

Venn Diagram: Unique to A $(A - B)$

Example

Consider the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the following events:

$$A = \{1, 2, 3\}, \quad B = \{2, 4, 5\}$$

The relative complements are:

$$A - B = \{1, 3\}, \quad B - A = \{4, 5\}$$

Complement of A

We use A^c to denote the complement of A within the universal set Ω, which is the set of outcomes that are not in A:

$$A^c = \Omega - A = \{4, 5, 6\}$$

The word **not** is associated with complements and the phrases "and not", "but not" are associated with relative compliments

Aside (for the HW): Verifying Set Identities with Venn Diagrams

We can use Venn diagrams to verify set identities. For example, let's verify the identity:

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

Venn diagrams representing set operations of (left hand side) LHS and (right hand side) RHS respectively are shown below:

LHS

RHS

$A \cap B$

A^c

$(A \cap B)^c$

B^c

$A^c \cup B^c$

Conclusion

Both sides of the identity are visually represented by the same shaded region, thus confirming that:

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

The shaded regions for $(A \cap B)^c$ and $A^c \cup B^c$ match. Therefore, $(A \cap B)^c = A^c \cup B^c$.

Note: (we are not worrying about edges cases, for example, where A and B are disjoint but the identity still holds true)

Now that the terminology is in place, we start to answer Q2.

Axioms

1. $P(\Omega) = 1$, $P(\emptyset) = 0$

2. $P(A) \geq 0$ for all $A \in \mathcal{F}$

3. If $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Example: Consider the problem of tossing a fair coin until we observe heads. We want to find P(heads on an even numbered toss)

Let A be the event that heads appears for the first time on an even-numbered toss.

Thus, A can be expressed as:

$$A = \{(T, H), (T, T, T, H), (T, T, T, T, T, H), \ldots\}$$

Recall, A_i is heads on the ith toss.

$$P(A_i) = \frac{1}{2^i}$$

Notice that we can describe A as the union of disjoint events where heads occurs for the first time on the $2i$-th toss:

$$A = A_2 \cup A_4 \cup A_6 \cup \cdots = \bigcup_{i=1}^{\infty} A_{2i}$$

where A_{2i} is the event that heads first appears on the $2i$-th toss. The probability of A_{2i} can be computed as:

$$P(A_{2i}) = \left(\frac{1}{2}\right)^{2i}$$

11

Since the events A_{2i} are pairwise disjoint, we can use axiom 3 to see the probability of A is given by:

$$P(A) = P\left(\bigcup_{i=1}^{\infty} A_{2i}\right) = \sum_{i=1}^{\infty} P(A_{2i}) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{i}$$

This series is an infinite geometric series with the first term $a = \frac{1}{4}$ and common ratio $r = \frac{1}{4}$:

Recall, that a the sum of a converging infinite geometric series where $|r| < 1$ is:

$$S = a + ar + ar^2 + ... = \sum_{i=0}^{\infty} (ar^n) = \frac{a}{1 - r}$$

Therefore:

$$\sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{i} = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Thus, the probability that the first occurrence of heads is on an even-numbered toss is $\boxed{\dfrac{1}{3}}$.

Derived rule: If $A_1, \ldots, A_N \in \mathcal{F}$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Example: Consider a fair die roll. What is the probability of rolling a 1, 3, or 5?

$$P(\{i\}) = \frac{1}{6} \quad \text{for } i = 1, \ldots, 6$$

$$P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Lecture 3: Thursday 9/3/2024

1   Probability

Definition: Probability is the mathematical framework for quantifying and analyzing uncertainty. A probability space is defined by the triple $(\Omega, \mathcal{F}, P)$, where:

- Ω is the sample space (set of all possible outcomes)

- $\mathcal{F}$ is the event space (sigma-algebra of subsets of Ω)

- P is the probability measure (function assigning probabilities to events)

12

Axioms of Probability

The probability measure P satisfies the following axioms:

1. $P(\Omega) = 1$ and $P(\emptyset) = 0$

2. For all $A \in \mathcal{F}$, $P(A) \geq 0$

3. (Countable additivity) If $A_1, A_2, \ldots \in \mathcal{F}$ are pairwise disjoint events (i.e $A_i \cap A_j = \emptyset$ for all $i \neq j$):
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Note: Understanding and being able to derive properties from these axioms is crucial for the first midterm and for building a solid foundation in probability theory.

Important Results

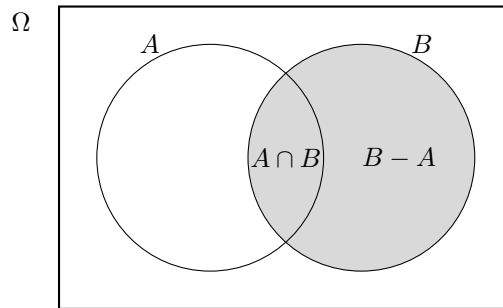1. **Decomposition of Probability:** For any two events $A, B \in \mathcal{F}$:
$$P(B) = P(B - A) + P(A \cap B)$$

Proof:

- Observe that B can be partitioned into two disjoint sets: $B = (B - A) \cup (A \cap B)$
- Note that $(B - A) \cap (A \cap B) = \emptyset$ (these sets are mutually exclusive)
- By the axiom of countable additivity (which implies finite additivity for disjoint events):
$$P(B) = P((B - A) \cup (A \cap B)) = P(B - A) + P(A \cap B)$$

This result is fundamental in probability theory and is often used in solving complex probability problems. It allows us to break down the probability of an event into mutually exclusive parts, which can be easier to calculate individually.
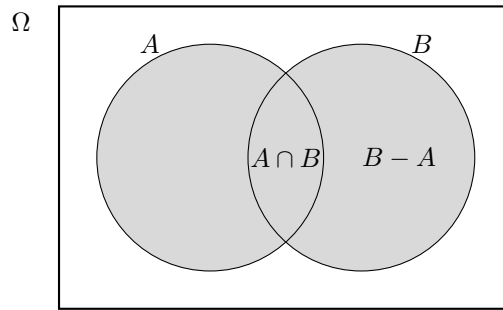
2. **Finite Additivity:** For any two events $A, B \in \mathcal{F}$:

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

Proof:

- Observe that $A \cup B = A \cup (B - A)$ and $A \cap (B - A) = \emptyset$
- Since A and $B - A$ are disjoint, $P(A \cup B) = P(A) + P(B - A)$ (finite additivity)
- From the decomposition result: $P(B) = P(B - A) + P(A \cap B)$
- Rearranging: $P(B - A) = P(B) - P(A \cap B)$
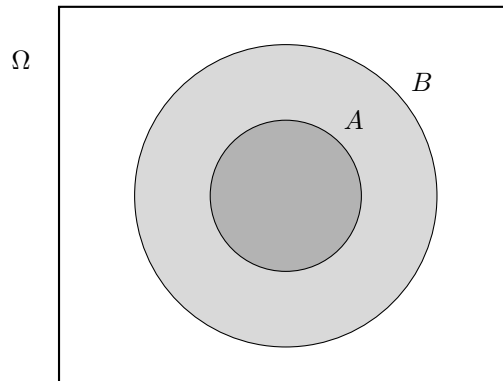- Substituting into our earlier equation:

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

3. **Monotonicity:** If $A \subseteq B$, then $P(A) \leq P(B)$

Proof:

- If $A \subseteq B$, then:
  - $B = A \cup (B - A)$
  - $A \cap (B - A) = \emptyset$
- By finite additivity: $P(B) = P(A \cup (B - A)) = P(A) + P(B - A)$
- Since probabilities are non-negative, $P(B - A) \geq 0$
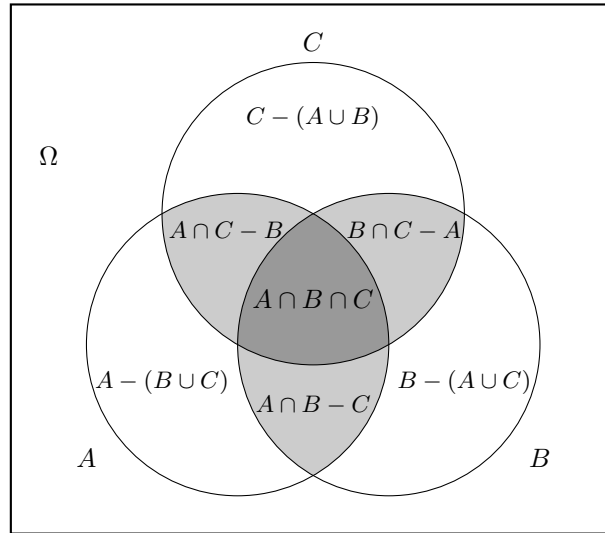- Therefore, $P(B) \geq P(A)$

4. **Union of Three Sets:** For $A, B, C \in \mathcal{F}$:

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

How to compute:

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\
&= P(A) + [P(B) + P(C) - P(B \cap C)] - P(A \cap (B \cup C)) \quad \text{(decompose the union)} \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P((A \cap B) \cup (A \cap C)) \quad \text{(by distributive law)} \\
&= P(A) + P(B) + P(C) - P(B \cap C) - (P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C))) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - (P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)
\end{aligned}
$$

Example: Suppose that A and B are events with $P(A) = 0.5$ and $P(B) = 0.6$. Find numbers a and b such that $a \leq P(A \cup B) \leq b$ and $a \leq P(A \cap B) \leq b$.

1. Using the addition rule of probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.6 - P(A \cap B) = 1.1 - P(A \cap B)$$

2. Since $A \cup B \subseteq \Omega$ and $P(\Omega) = 1$, we have:

$$P(A \cup B) \leq P(\Omega) \leq 1$$

3. From steps 1 and 2:

$$1.1 - P(A \cap B) \leq 1 \implies P(A \cap B) \geq 0.1$$

15

4. The maximum value for $P(A \cap B)$ is $\min(P(A), P(B)) = 0.5$, because:

   - $A \cap B \subseteq A$ and $A \cap B \subseteq B$
   - Therefore, $P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$
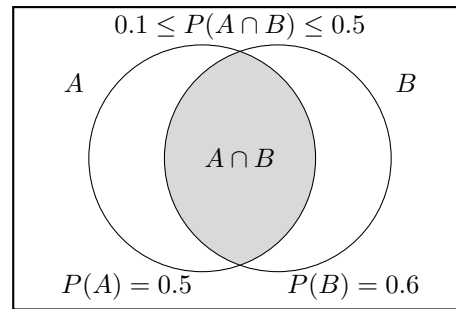
5. We can conclude:
$$0.1 \leq P(A \cap B) \leq 0.5$$
$$0.6 \leq P(A \cup B) \leq 1.0$$

6. To verify the lower bound for $P(A \cup B)$:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\geq P(A) + P(B) - \min(P(A), P(B)) \\ &= 0.5 + 0.6 - 0.5 = 0.6 \end{aligned}$$

Conclusion: We have found that $a = 0.1$ and $b = 0.5$ satisfy $a \leq P(A \cap B) \leq b$, while $a = 0.6$ and $b = 1.0$ satisfy $a \leq P(A \cup B) \leq b$.

Visualization: To better understand the relationship between A and B, consider the following Venn diagram:

This visualization illustrates why $0.1 \leq P(A \cap B) \leq 0.5$ and $0.6 \leq P(A \cup B) \leq 1.0$.

Example: When Alice visits a certain grocery store, she:

- buys apples with probability 0.4

- buys bananas with probability 0.7

- buys both with probability 0.25

Compute the probability that on her next visit to the store, Alice:

1. buys either apples or bananas

2. buys bananas but not apples

3. buys neither apples nor bananas

Step 1: Define notation and given probabilities

$$A : \text{Alice buys apples}$$
$$B : \text{Alice buys bananas}$$

$$P(A) = 0.4$$
$$P(B) = 0.7$$
$$P(A \cap B) = 0.25$$

Note: Symbol correspondences

- $\cup$ corresponds to "or" (union)

- $\cap$ corresponds to "and" (intersection)

- $\complement$ corresponds to "not" (complement)

- $-$ corresponds to "and not", or "but not" (relative complement)

Step 2: Calculate probabilities

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.7 - 0.25 = 0.85$

2. $P(B - A) = P(B) - P(A \cap B) = 0.7 - 0.25 = 0.45$

3. $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0.85 = 0.15$

Note: Important probability relationships:

$$P(A) = 1 - P(A^c)$$
$$\Omega = A \cup A^c$$
$$P(\Omega) = P(A) + P(A^c) = 1$$
$$P(A^c) = 1 - P(A)$$

Computing probabilities for finite, equally likely outcomes: For any event $A \subseteq \Omega$, where Ω is a finite sample space with equally likely outcomes, the probability of A is given by:

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega} = \frac{|A|}{|\Omega|}$$

where $|A|$ denotes the cardinality (number of elements) of set A.

Lecture 4: Thursday 9/5/2024