Chapter 1

Set Theory and Probability Models

1.1 Set Theory

Set A set is a collection of elements. ex: A={red, blue, green}. Moreover, B is a **subset** of A if every element of B is also in A. ex: B={red, blue}. Notation: $B \subseteq A$ (B is a subset of A) or $A \supseteq B$ (A contains B).

In probability theory, set theory is used to describe possible outcomes. The **state space** S, is a set containing all **states** (possible outcomes). Exactly one state will occur. Each subset of S is called an **event**.

Set Theory	Probability Theory
set	state space
elements	states, possible outcomes
subsects	events

ex: Die Roll $S = \{1, 2, 3, 4, 5, 6\}$, $F = \{1, 2, 3\}$, $E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$.

Four Operations

Name	Notation	Example
Complement	A^c	$F^c = \{4, 5, 6\}$
Union	$A \cup B$	$F \cup E = \{1, 2, 3, 4, 6\}$
Intersection	$A \cap B$	$F \cap E = \{2\}$
Relative Complement	$A - B = A \cap B^c$	$F - E = \{1, 3\}$

Venn Diagram A Venn Diagram is a way to graphically represent sets and set operations. Each diagram begins with a rectangle representing the state space (universal set). Then each set is represented by a circle.

Empty Set The empty set is the set containing no elements. We denote it as $\emptyset$. (Do not misread it as the Greek Alphabet ϕ. It is called "empty set").

Disjoint (Mutually Exclusive) Two sets, A and B, are said to be disjoint (mutually exclusive) if their intersection is empty:

$$A \cap B = \emptyset$$

Partition $A_1, \dots, A_n \subseteq S$ is said to be a partition of S if

1. $A_i \cap A_j = \emptyset$ for $i \neq j$ (disjoint)
2. $A_1 \cup A_2 \cup \cdots \cup A_n = S$

Example Suppose $S = \{1, 2, 3, 4, 5, 6\}$. Consider the following subsets of S: A_1, A_2 and A_3.

1. $A_1 = \{1, 2\}$, $A_2 = \{3, 6\}$, $A_3 = \{4, 5\}$
2. $A_1 = \{1, 2\}$, $A_2 = \{5\}$, $A_3 = \{4, 6\}$
3. $A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$, $A_3 = \{2, 5, 6\}$

Check if A_1, A_2 and A_3 form a partition of S for each case.

A List of Connections between the Set Operations

1. Complementation:

$$(A^c)^c = A, \quad \emptyset^c = S, \quad S^c = \emptyset$$

2. Commutativity of set union and intersection:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

3. Associativity of union and intersection:

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

4. *De Morgan's laws*:

$$(A_1 \cup A_2 \cup \cdots \cup A_n)^c = A_1^c \cap A_2^c \cap \cdots \cap A_n^c$$

$$(A_1 \cap A_2 \cap \cdots \cap A_n)^c = A_1^c \cup A_2^c \cup \cdots \cup A_n^c$$

   (i) $(A \cap B)^c \subseteq (A^c \cup B^c)$

   (ii) $(A \cap B)^c \supseteq (A^c \cup B^c)$

5. Distributivity laws:

$$B \cap (A_1 \cup A_2 \cup \cdots \cup A_n) = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$$

$$B \cup (A_1 \cap A_2 \cap \cdots \cap A_n) = (B \cup A_1) \cap (B \cup A_2) \cap \cdots \cap (B \cup A_n)$$

6. As a consequence of the definitions:

$$\begin{aligned} A \cup A &= A, & A \cap A &= A, \\ A \cup \emptyset &= A, & A \cap \emptyset &= \emptyset, \\ A \cup S &= S, & A \cap S &= A, \\ A \cup A^c &= S, & A \cap A^c &= \emptyset. \end{aligned}$$

7. Some other important properties:

   (I1)

$$A = (A \cap B) \cup (A \cap B^c) \tag{1.1}$$

   proof.

$$\begin{aligned} A &= A \cap S \\ &= A \cap (B \cup B^c) \\ &= (A \cap B) \cup (A \cap B^c) \end{aligned}$$

(I2)

$$(A \cap B) \quad \text{and} \quad (A \cap B^c) \quad \text{are disjoint.} \tag{1.2}$$

proof.

$$\begin{aligned}(A \cap B) \cap (A \cap B^c) &= A \cap (B \cap B^c) \\ &= A \cap \emptyset \\ &= \emptyset\end{aligned}$$

(I3)

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

proof.

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

1.2 Probability Models

Probability Measure P is a probability measure on the state space S if all events $A \subseteq S$ are assigned numbers $P(A)$ satisfying

(a) $P(\emptyset) = 0$

(b) $P(S) = 1$

(c) $P(A) \geq 0$ for all $A \subseteq S$

(d) $P(A \cup B) = P(A) + P(B)$ for all disjoint $A, B \subseteq S$

The pair (S, P) is called a probability model.[1] [2]

[1] This axiom approach to constructing the probability model was done by the famous Russian mathematician A.N. Kolmogorov (1903-1987) in his fundamental monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung (Fundamental Ideas of the Probability Calculation)*, which appeared in 1933.

[2] A more thorough discussion of probability theory is based on measure theory by introducing the *probability space* $(S, \mathcal{F}, P)$. Where, $\mathcal{F}$ is the collection ($\sigma-$algebra) of measurable subsets of S. It is, however, beyond the scope of this course.

ex: Die Roll $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = 1/6$. We can then use (d) to assign probabilities to other events. For instance,

1. $P(\{1\} \cup \{2\}) = P(\{1\}) + P(\{2\}) = 1/6 + 1/6 = 1/3$.

2.

$$\begin{aligned} P(\{\leq 2\} \cup \{\geq 4\}) &= P(\{\leq 2\}) + P(\{\geq 4\}) \\ &= P(\{1\} \cup \{2\}) + P(\{4\} \cup \{5\} \cup \{6\}) \\ &= P(\{1\}) + P(\{2\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) \\ &= 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 5/6 \end{aligned}$$

Other properties of probability models are implications of (a)-(d):

(p1) $P(A) + P(A^c) = 1$

(p2) $A \subseteq B$ implies that $P(A) \leq P(B)$

(p3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (law of addition)

Exercise 1. Prove (p1), (p2) and (p3) according to (a)-(d).

Anything we would want to call a probability measure should satisfy (a)-(d). This includes both "objective probability" (ex: die roll) and "subjective probability" (ex: your beliefs about baseball game). This distinction (objective v.s. subjective) will become important when we talk about statistics.

1.3 Conditional Probability

How should probability assessments change in the face of new information? The conditional probability $P(A|B)$ means "the probability of event A given (conditional on) event B". We can think of $P(\cdot|B)$ as a new probability measure on S.

ex: Die Roll $P(\{1\}|\text{Odd}) = P(\{1\}|\{1,3,5\}) = 1/3$.

We would like the conditional probability to satisfy two properties:

(C1) $P(B|B) = 1$

(C2) If $C, D \subseteq B$ and $P(D) \neq 0$, then

$$\frac{P(C|B)}{P(D|B)} = \frac{P(C)}{P(D)}$$

i.e. fixed relative probability for subsets of B. If D was twice as likely as C, it remains twice as likely as C while given B.

Example

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

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

$$C = \{1,2\}$$

$$D = \{2,3,4\}$$

$$P(C) = 1/3 \qquad P(D) = 1/2$$

$$P(C|B) = 2/5 \qquad P(D|B) = 3/5$$

Definition (Conditional Probability)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ whenever} P(B) \neq 0.$$

Or

$$P(A \cap B) = P(A|B) \times P(B)$$

Therefore, according to the definition,

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

$$\frac{P(C|B)}{P(D|B)} = \frac{\frac{P(C \cap B)}{P(B)}}{\frac{P(D \cap B)}{P(B)}} = \frac{P(C)}{P(D)}$$

In fact, the definition above is the <u>only</u> one which satisfy (C1) and (C2).

ex: Die Roll

$$P(\{1\}|\text{Odd}) = P(\{1\}|\{1,3,5\}) = \frac{P(\{1\} \cap \{1,3,5\})}{P(\{1,3,5\})} = \frac{P(\{1\})}{P(\{1,3,5\})} = \frac{1/6}{1/2} = 1/3$$

$$P(\text{Odd}|\text{Lowest 3}) = P(\{1,3,5\}|\{1,2,3\}) = \frac{P(\{1,3,5\} \cap \{1,2,3\})}{P(\{1,2,3\})} = \frac{P(\{1,3\})}{P(\{1,2,3\})} = \frac{1/3}{1/2} = 2/3$$

1.4 Conditional and Marginal Probabilities

Senate vote on a bill to remove financial regulations

	Y	N	
Democrat	3	48	51
Republican	35	14	49
	38	62	100

Suppose we select a senator at random.

1. divide by 100 to get probabilities
2. sum to get marginal probabilities: $P(D)$, $P(R)$, $P(Y)$, and $P(N)$.
3. after obtaining marginal probabilities, we can compute conditional probabilities. For instance,
   - $P(N|D) = \frac{P(N \cap D)}{P(D)} = \frac{0.48}{0.51} \approx 0.9412$
   - $P(R|Y) = \frac{P(R \cap Y)}{P(Y)} = \frac{0.35}{0.38} \approx 0.9211$

A General Representation: Let $A_1, \ldots, A_n \subseteq S_A$ be a partition of S_A, and $B_1, \ldots, B_n \subseteq S_B$ be a partition of S_B.

	B_1	$\cdots$	B_n	
A_1	$P(A_1 \cap B_1)$	$\cdots$	$P(A_1 \cap B_n)$	$P(A_1)$
$\vdots$	$\vdots$	$\ddots$	$\cdots$	$\vdots$
A_n	$P(A_n \cap B_1)$	$\cdots$	$P(A_n \cap B_n)$	$P(A_n)$
	$P(B_1)$	$\cdots$	$P(B_n)$	1

1.5 The Doctor's Problem Revisited

Let

- T={randomly selected person tests positive}
- D={randomly selected person has lovesickness}
- D^c={randomly selected person does not have lovesickness}

According to the information we have

- 1 in 1000 have disease $\Rightarrow P(D) = 0.001$
- Prob(test positive|disease) = 0.99 $\Rightarrow P(T|D) = 0.99$
- Prob(test negative|no disease) = 0.95 $\Rightarrow P(T^c|D^c) = 0.95$

We can further get

- $P(D^c) = 0.999$
- $P(T|D^c) = 0.05$

using the fact that

$$P(D) + P(D^c) = 1,$$

and

$$P(T|D^c) + P(T^c|D^c) = 1.$$

Here is the proof of the second equation:

$$\begin{aligned} P(T|D^c) + P(T^c|D^c) &= \frac{P(T \cap D^c)}{P(D^c)} + \frac{P(T^c \cap D^c)}{P(D^c)} \\ &= \frac{1}{P(D^c)}[P(T \cap D^c) + P(T^c \cap D^c)] \\ &= \frac{1}{P(D^c)}[P(D^c)] \quad \text{by equations (1.1) and (1.2)} \\ &= 1 \end{aligned}$$

What we want to know is

$$\text{Prob(disease|test positive)} = P(D|T)$$

First note that:

$$P(T) = P(T \cap D) + P(T \cap D^c)$$

by equations (1.1) and (1.2).

$$\begin{aligned} P(D|T) &= \frac{P(D \cap T)}{P(T)} \\ &= \frac{P(D \cap T)}{P(D \cap T) + P(D^c \cap T)} \\ &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \end{aligned} \tag{1.3}$$

Thus,

$$P(D|T) = \frac{(0.99) \cdot (0.001)}{(0.99) \cdot (0.001) + (0.05) \cdot (0.999)} = \frac{0.00099}{0.00099 + 0.09995} = \frac{0.00099}{0.05094} = 0.0194$$

Remarks: Equation (1.3) is a special case of *Bayes' Rule (Bayes' Formula)*.

Bayes' Rule Let $T \subseteq S$ be a test event. Let $A_1, A_2, \dots A_n \subseteq S$ be a group of categories. Suppose

- $P(T) > 0$ (the positive tests can happen)
- $A_1, A_2, \dots, A_n$ be a partition of S (they are disjoint, $A_1 \cup A_2 \cup \cdots A_n = S$)

And suppose we know

- $P(A_1), P(A_2), \dots, P(A_n)$ (the probability of each category)
- $P(T|A_1), P(T|A_2), \dots, P(T|A_n)$ (how well each category performs on the test)

We want to know

- $P(A_i|T)$ how likely category i is if the test comes up positive)

Bayes' Rule:

$$P(A_i|T) = \frac{P(A_i \cap T)}{P(T)} = \frac{P(T|A_i)P(A_i)}{\sum_{j=1}^{n} P(T|A_j)P(A_j)}$$

Where

$$P(T) = \sum_{j=1}^{n} P(T|A_j)P(A_j)$$

is called **Law of Total Probability**.

Proof: The Bayes' Rule

Example A softball team has three pitchers, X and Y, with winning percentages of 0.4 and 0.8, respectively. They pitch with frequency 3, and 7 out of every 10 games, respectively. In other words, for a randomly selected game, $P(X) = 0.3$ and $P(Y) = 0.7$. Find:

1. P(team wins game)=$P(W)$.
2. P(X pitched game|team won)=$P(X|W)$.

$$P(W) = P(W|X)P(X) + P(W|Y)P(Y) = (0.4) \cdot (0.3) + (0.8) \cdot (0.7) = 0.68$$

$$P(X|W) = \frac{P(X \cap W)}{P(W)} = \frac{P(W|X)P(X)}{P(W)} = 0.12/0.68 = 3/17$$

1.6 Independence

Intuition Events are independent if the occurrence of some of them does not provide information about the occurrence of the others.

Definition Two events $A, B \subseteq S$ are independent if $P(A \cap B) = P(A)P(B)$.

Remarks

- If $P(B) \neq 0$, $P(A|B) = \frac{A \cap B}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$.
- As well, if $P(A) \neq 0$, $P(B|A) = P(B)$.

$\Longrightarrow$ Finding out that one event occurred does not change your assessment of the likelihood of the other event.

An Example of Independent Events Tossing a fair coin twice:

	Head	Tail	
Head	1/4	1/4	1/2
Tail	1/4	1/4	1/2
	1/2	1/2	1

However, in the example of Senate Vote, the event that the senator we selected at random is Democrat, D and the event that the senator voted for financial deregulation, Y are NOT independent. (Please Check!)

In many real world and experimental situations, we have good reasons to believe certain events are independent. Thus, knowing the probabilities of each individual event allows us to calculate all sorts of joint probabilities. For example, suppose $P(A) = 0.2$, $P(B) = 0.5$. What is $P(A \cup B)$?

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \\ &= 0.2 + 0.5 - 0.1 = 0.6 \end{aligned} \tag{1.4}$$

Definition The event $A_1, A_2, \ldots, A_n$ are independent if for all $I \subseteq \{1, 2, \ldots, n\}$

$$P\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} P(A_i) \tag{1.5}$$