Topic 1: Basic probability

- Review of sets
- Sample space and probability measure
- Probability axioms
- Basic probability laws
- Conditional probability
- Bayes' rules
- Independence
- Counting

Definition of Sets

- A **set** S is a collection of objects, which are the **elements** of the set.
  - The number of elements in a set S can be *finite*

$$S = \{x_1, x_2, \ldots, x_n\}$$

  or *infinite* but countable

$$S = \{x_1, x_2, \ldots\}$$

  or *uncountably infinite*.
  - S can also contain elements with a certain property

$$S = \{x \mid x \text{ satisfies } P\}$$

- S is a **subset** of T if every element of S also belongs to T

$$S \subset T \text{ or } T \supset S$$

  If $S \subset T$ and $T \subset S$ then $S = T$.
- The **universal set** Ω is the set of all objects within a context. We then consider all sets $S \subset \Omega$.

Set Operations and Properties

- Set operations
  - Complement A^c: set of all elements not in A
  - Union $A \cap B$: set of all elements in A or B or both
  - Intersection $A \cup B$: set of all elements common in both A and B
  - Difference $A - B$: set containing all elements in A but not in B.
- Properties of set operations
  - Commutative: $A \cap B = B \cap A$ and $A \cup B = B \cup A$.
    (But $A - B \neq B - A$).
  - Associative: $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$. (also for $\cup$)
  - Distributive:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

  - DeMorgan's laws:

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

Elements of probability theory

A probabilistic model includes

- The sample space Ω of an *experiment*
  - set of all possible *outcomes*
  - *finite* or *infinite*
  - *discrete* or *continuous*
  - possibly multi-dimensional
- An event A is a set of outcomes
  - a subset of the sample space, $A \subset \Omega$.
  - special events: certain event: $A = \Omega$, null event: $A = \emptyset$

  The *set of events* $\mathcal{F}$ is the set of all possible subsets (events A) of Ω.
- A probability law $P(A)$ that defines the likelihood of an event A.

Formally, a probability space is the triplet $\{\Omega, \mathcal{F}, P(A)\}$.

The probability axioms

- A probability measure $P(A)$ must satisfy the following axioms:
  1. $P(A) \geq 0$ for every event A
  2. $P(\Omega) = 1$
  3. If $A_1, A_2, \ldots$ are disjoint events, $A_i \cap A_j = \emptyset$, then

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

- Notes:
  - These axioms are called non-negativity, normalization, and additivity, respectively.
  - The probability *measure* in a sense is like other measures such as *mass, length, volume* – all satisfy axioms 1 and 3
  - The probability measure, however, is bounded by 1 (axiom 3). It also has other aspects such as conditioning, independence that are unique to probability.
  - $P(\emptyset) = 0$, but $P(A) = 0$ does not necessarily imply $A = \emptyset$.

Discrete Probability Space

- The sample space Ω is discrete if it is *countable*.
  - It can be finite or infinite (countably infinite).
- Examples:
  - Rolling a dice: $\Omega = \{1, 2, \ldots, 6\}$
  - Flipping a coin until the first head appears: $\Omega = \{H, TH, TTH, \ldots\}$
  - Number of users connecting to the cellular network in 1 minute intervals: $\Omega = \{0, 1, 2, 3, \ldots\}$
- The probability measure $P(A)$ can be defined by assigning a probability to each single outcome event $\{s_i\}$ (or *elementary event*) such that

$$P(s_i) \geq 0 \quad \text{for every} \quad s_i \in \Omega$$
$$\sum_{s_i \in \Omega} P(s_i) = 1$$

  - Probability of any event $A = \{s_1, s_2, \ldots, s_k\}$ is

$$P(A) = P(s_1) + P(s_2) + \ldots + P(s_k)$$

  - If Ω consists of n equally likely outcomes, then $P(A) = k/n$.

Continuous Probability Space

- The sample space Ω is continuous if it is *uncountable infinite*.
- Examples:
  - Call arrival time: $\Omega = (0, \infty)$
  - Random dot in a unit-square image: $\Omega = (0, 1)^2$
- For continuous Ω, the probability measure $P(A)$ cannot be defined by assigning a probability to each outcome.
  - For any outcome $s \in \Omega$, $P(s) = 0$

    Note: A zero-probability event does not imply that the event cannot occur, rather it occurs *very infrequently*, given that the set of possible outcomes is infinite.
  - But we can assign the probability to an *interval*.

    For example, to define the uniform probability measure over $(0, 1)$, assign $P((a, b)) = b - a$ to all intervals with $0 < a, b < 1$.

Basic probability laws

- If $A \subset B$ then $P(A) \leq P(B)$
- Complement
$$P(A^c) = 1 - P(A)$$
- Joint probability
$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$
- Union
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
- Union of event bound
$$P\left(\bigcup_{i=1}^{N} A_i\right) \leq \sum_{i=1}^{N} P(A_i)$$
- Total probability law: Let $S_1, S_2, \ldots$ be events that partition Ω, that is, $S_i \cap S_j = \emptyset$ and $\bigcup_i S_i = \Omega$. Then for any event A
$$P(A) = \sum_i P(A \cap S_i)$$

Conditional Probability

- Conditional probability is the probability of an event A, given *partial information* in the form of an event B. It is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\ , \quad \text{with} \quad P(B) > 0$$

  - Conditional probability $P(.|B)$ can be viewed as a probability law on the new universe B.
  - $P(.|B)$ satisfies all the axioms of probability.

$$\begin{aligned} P(\Omega|B) &= 1 \\ P(A_1 \cup A_2|B) &= P(A_1|B) + P(A_2|B) \quad \text{for } A_1 \cap A_2 = \emptyset \end{aligned}$$

- The conditional probability of A given B – the *a posteriori* probability of A – is related to the unconditional probability of A – the *a priori* probability – as

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

- Chain rules:

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

$$P\left(\cap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P\left(A_3|A_1 \cap A_2\right) \dots P\left(A_n| \cap_{i=1}^{n-1} A_i\right)$$

- Examples: Radar detection, the false positive puzzle.

Bayes' rule

- Let $S_1, S_2, \ldots, S_n$ be a partition of the sample space Ω. We know $P(S_i)$.
- Suppose an event A occurs and we know $P(A|S_i)$. What is the a posteriori probability $P(S_i|A)$?
- Bayes' rule:

$$P(S_i|A) = \frac{P(S_i \cap A)}{P(A)} = \frac{P(A|S_i)}{\sum_{i=1}^{n} P(S_i)P(A|S_i)} P(S_i)$$

  - Prove by using the total probability law.
  - Bayes' rule also applies to a countably infinite partition ($n \to \infty$).
- Example: Binary communication channel.

Independence

- Two events A and B are independent if

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

  - In terms of conditional probability, if $P(B) \neq 0$, then

$$P(A|B) = P(A)$$

    That is, B does not provide any information about A.
  - Independence *does not* mean mutually exclusive.
    Mutually exclusive events with non-zero probability ($P(A) \neq 0$ and $P(B) \neq 0$) are not independent since

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

- Independence of multiple events: $\{A_k\}$, $k = 1, \ldots, n$ are independent iff for any set of m events ($2 \leq m \leq n$)

$$P(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_m}) = P(A_{k_1})P(A_{k_2}) \ldots P(A_{k_m})$$

  - For example, 3 events $\{A_1, A_2, A_3\}$ are independent if the following expressions *all* hold:

$$\begin{aligned} P(A \cap B \cap C) &= P(A)P(B)P(C) \\ P(A \cap B) &= P(A)P(B) \\ P(B \cap C) &= P(B)P(C) \\ P(A \cap C) &= P(A)P(C) \end{aligned}$$

- Note: It is possible to construct sets of 3 events where the last three equations hold but the first one does not.
  Example: Let $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$ where

  $$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{8}\,, \quad P(7) = \frac{1}{4}$$

  Now let $A = \{1, 2, 7\}$, $B = \{3, 4, 7\}$, and $C = \{5, 6, 7\}$. What are the probabilities of these events and their intersections?
- It is also possible for the first equation to hold while the last three do not.

• Pair-wise independence: If every pair (A_i, A_j) $(i \neq j)$ are independent, we say A_k are *pair-wise* independent.

  - Independence implies pair-wise independence, but not the reverse.

• Independent experiments: The most common application of the independence concept is to assume separate experiments are independent.

  - Example: A message of 3 bits is transmitted over a noisy line. Each bit is received with a probability of error $0 \leq p \leq \frac{1}{2}$, independent of all other bits.
    What is the probability of the receiving at least two bits correctly?

• Conditional independence: A and B are independent given C if

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

  - Independence *does not* imply conditional independence.
    Example: Consider 2 independent coin tosses, each with equally likely outcome of H and T. Define

$$\begin{aligned} A &= \{\text{ 1st toss is H }\} \\ B &= \{\text{ 2nd toss is H }\} \\ C &= \{\text{ Two tosses have different results }\} \end{aligned}$$

  - Vice-versa, conditional independence *does not* imply independence.

Counting

- In many experiments with finite sample spaces, the outcomes are equally likely.
- Then computing the probability of an event reduces to counting the number of outcomes in the event.
- Assume that there are n distinct objects. We want to count the number of sets A with k elements, denoted as N_k.
  - Counting is similar to sampling from a population.
  - The count N_k depends on
    * If the order of objects matters within the set A.
    * If repetition of objects is allowed within the set A (replacement within the population).
- The sampling problem
  - Ordered sampling with replacement: $N_k = n^k$
  - Ordered sampling without replacement:

$$N_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

Permutations: $n! = n(n-1)\dots 1$ (when $k = n$)

  - Unordered sampling without replacement:

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

  - Unordered sampling with replacement:

$$N_k = \binom{n+k-1}{k}$$