CS 240: Reasoning Under Uncertainty (Fall 21)   Oct 11

Lecture 1-9

Instructor: *Profs Peter J. Hass and Jie Xiong*   SI Worksheet: *Juelin Liu*

1 Introduction

The first mid-term will cover materials from lecture 1 to 9. This worksheet intends to summarize these materials.

2 Set theory

2.1 Core Concepts

- An event is a **set** of outcomes.

- A set is a collection of objects.

- The objects are called elements of the set.

2.2 Notations

2.2.1 Empty Set

- $\emptyset$: empty set, the set with zero object.

2.2.2 Universal Set

- Ω: universal set, the set that contains all possible objects.

2.2.3 Element

We say $x \in S$ if the set S contains element x. (otherwise, $x \notin S$)

2.2.4 Subset

We say $S \subset T$ if all elements in S are also in T ($\emptyset \subset S \subset \Omega$.)
Note that in the textbook $\subset$ and $\subseteq$ have the same meaning. If $S \subset T$ then S can be equivalent to T.

2.2.5 Complement

The complement of a set S is defined as:

$$S^c = \{x \in \Omega \mid x \notin S\}$$

Some useful laws:

$$\Omega^c = \emptyset$$
$$\emptyset^c = \Omega$$
$$(S^c)^c = S$$

2.2.6 Power Set

The power set P^S is the set that contains all subsets of S.
The cardinality (size) of the power set $|P^S| = 2^{|S|}$.
For example, $S = \{1, 2\}$, its power set $P^S = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

2.2.7 Disjoint Sets

Two sets are disjoint if their intersection is the empty set.

$$A \cap B = \emptyset \iff \text{A and B are disjoint}$$

2.2.8 Partition

We say $S_1, S_2, \ldots, S_n$ form a partition of S if $i \neq j \implies S_i \cap S_j = \emptyset$ and $S_1 \cup S_2 \cup \ldots \cup S_n = S$.

3 Set algebra

3.1 Set Operation

3.1.1 Union

$$S \cup T = \{x \mid x \in S \text{ or } x \in T\}$$

3.1.2 Intersection

$$S \cap T = \{x \mid x \in S \text{ and } x \in T\}$$

3.1.3 Important Laws

DeMorgan's Laws:

$$(S \cap T)^c = S^c \cup T^c$$
$$(S \cup T)^c = S^c \cap T^c$$

Union Distributivity:

$$S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$$

(think about how to show these laws using the Venn diagrams)

4 Probabilistic models

We use probabilistic models to describe uncertain situations mathematically. The three key factors are:

- **Experiment**: an experiment with a set Ω of all possible outcomes.

- **Event**: a set of outcomes (which is a subset of Ω.)

- **Probability Law**: $P(A) \geq 0$ for $(A \subset \Omega)$

5 Inclusion-Exclusion Principle

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

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

6 Discrete Probability Models

We are dealing with discrete probability models if Ω consists of a finite number of possible outcomes.

6.1 Uniform Discrete Model

All possible outcomes in Ω are equally likely.

6.2 Odds

The odds are x to y in favor of A means $P(A) = \frac{x}{x+y}$

6.3 Odds Ratio

$$Odds(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

7 Conditional Probability

General Definition:

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

$P(A|B)$ denotes the conditional probability of A given B.
When B is given, the sample space changes to B and the event space changes to $A \cap B$.

7.1   Sequential Model

Before discussing the details, it is worth to mention the multiplication rule:

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots P(A_n|\bigcap_{i=1}^{n-1} A_i)$$

The intuition is instead of computing $P(\bigcap_{i=1}^{n} A)$ as a whole, we can break it into a series of events that depend on the previous outcomes. Then we can compute each event's probability and multiply them together to get the result.

We can use a sequential model when the future outcomes depend on the previous outcomes. The tree-based sequential description can help you visualize this process.

8   Monty Hall

The Monty Hall problem is a mind teaser. People can be tricked into believing that choosing either of the two remaining doors has the same chance of winning. But keep in mind that the revealing of one of the goat doors is not at random (which is difficult to understand.) Since the host never revealed the door picked by the contestant, we gain no more information for the door picked by the contestant. So the chances of winning stays at $\frac{1}{3}$ if not switching.

9   Total Probability Theorem

Definition:

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space Ω and assume $P(A_i) > 0$ for all i. Then, for any event B, we have:

$$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B)$$

Proof (from the textbook, p29):

The events $A_1, A_2, \ldots, A_n$ form a partition of the sample space so that the event B can be decomposed into the disjoint union of its intersections $A_i \cap B$ with the sets A_i:

$$B = (A_1 \cap B) \cup \ldots \cup (A_n \cap B)$$

Using the additivity axiom (see lecture 1, p42), it follows that:

$$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B)$$

10   Bayes' Rules

Intuition:

$A_1, A_2, \ldots, A_n$ are related to the "causes" and B represent the effect. We use Bayes' rule to infer the causes based on our observed effect. (textbook, p32)

Definition:

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space Ω and assume $P(A_i) > 0$ for all i. Then, for any event B such that $P(B) > 0$, we have:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)}$$

Proof:

The first equality comes from the definition of conditional probability and the second uses the total probability theorem to rewrite $P(B)$.

11 Independence

Definition:

We say events A and B are independent events if and only if:

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

In addition, if $P(B) > 0$:
$$P(A|B) = P(A)$$

Note: The definition says that the occurrence of event A does not change the probability that event B has occurred (and vice versa.)

It is easy to think that two disjoint events are independent but this is **not true**. The fact is two disjoint events A, B are not independent if $P(A) > 0$ and $P(B) > 0$. The best practice is to always verify the equation in the above definition.

12 Independence of Many Events

Definition:

We say that the events $A_1, A_2, \ldots, A_n$ are independent if and only if:

$$P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$$

for **every** subset S of $\{1, 2, \ldots, n\}$.

13 Conditional Independence

Definition:

We say A and B are conditionally independent given C ($P(C) > 0$) if and only if:

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

In addition, if $P(B \cap C) > 0$:

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

14 Counting

14.1 Counting Principle

Intuitively, if there are a ways to do the first thing and regardless of which way to choose there are b ways to do the second thing, then there are $a \times b$ ways to do both.

14.2 Permutation

Permutation is a method to count the number of ways to order objects.

$$P_n^n = n \times (n-1) \times \ldots \times 1 = n!$$

14.3 k-Permutation

k-Permutation is a method to count the number of ways to form a sequence of size k using k different objects from a set of n objects.

$$P_k^n = \frac{n!}{(n-k)!}$$

14.4 Combination

Combination counts the number of size k subsets of a size n set.

$$C_k^n = \left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k! \times (n-k)!} = C_{n-k}^n$$

14.5 Partition

Partition counts several ways to partition n objects into l groups of size $n_1, \ldots, n_l$.

$$\left(\begin{array}{c} n \\ n_1, \ldots, n_l \end{array} \right) = \frac{n!}{n_1! \times n_2! \times \ldots \times n_l!}$$

15 More on Counting

15.1 Letter Arrangement

15.1.1 Unique Letters

Think about how many different ways you can arrange the letters in the word "python". This is the simplest case because all the characters in the world are unique. The number of different arrangements is the permutation of the 6 letters {p,y,t,h,o,n}, which is $6! = 720$.

15.1.2 Duplicate Letters

Then think about how many different ways you can arrange the letters in the word "java". You have two a in the word which makes it tricky. I would recommend you list all the arrangements first.

The first method uses permutation to compute all possible arrangements of the 4 letters and then remove duplicates. Using permutation, we assume the two a are two different letters a_1 and a_2. We will find that $j - a_1 - v - a_2$ and $j - a_2 - v - a_1$ are counted as two different arrangements which are duplicated. The takeaway is the two a cause the same arrangement to be counted twice. To avoid double counting, we need to divide 4! (permutation of the four letters) by 2! (double counted due to two a) which gives $\frac{4!}{2!} = 12$.

Now, take a few seconds to think about what if there are **three** a in the word. In this case, the same arrangement will be counted for **3!** times. (**not 3 times**)

The second way is to think of this in terms of partition. Essentially, we are choosing the slots for each unique character in the word. If we fix the two slots for a, no matter how we arrange them inside the two slots, we will have the same arrangement if the others are the same. In other words, if we determine the slots for a letter then there is only one way to arrange this letter regardless of its number of occurrences in the word. Now the question becomes how many different ways we have to assign slots to each letter (two 'a' are considered as the same letter but take two slots.) This is the partition of the slots based on each unique letter's number of occurrences. So the result is $\frac{4!}{1! \times 2! \times 1!} = 12$

15.1.3 More Than One Duplicate Letters

Suppose there are k unique characters in the word with frequency $\{n_1, n_2, \ldots, n_k\}$, the number of different arrangements is the same as the number of partitions of the n elements into k sets with cardinality $\{n_1, n_2, \ldots, n_k\}$:

$$\binom{n}{n_1, n_2, \ldots, n_k}$$

15.2 Seat Arrangement

For this type of question, you have n slots and n people. You need to arrange the number of people in a row with certain constraints. The main difficulty is to interpret these constraints and translate them into correct formulas. While there is no such one-fits-all strategy, always think about the following:

1. How many **distinct objects** you have.

2. How many **slots** you can freely assign to the objects? (You might have n slots but due to the constraint you cannot freely assign each of them.)

3. How many different ways you can arrange each of the **distinct object** if its slot is fixed? (Each distinct object might have more than one element.)

Sometimes it is a good idea to fix the slots for certain people/groups before determining the rest. This is called the divide and conquer strategy.

15.3 Round Table

In this case, the n people are sitting in a round table. Solve it as it is a row problem first and then divide the results by the number of people. This is because, for the same arrangement of the n people, there are n ways to count it in the circle through rotation.

16 Expectation & Variance & Standard Deviation

For a discrete random variable, its expected value can be computed using the following formula:

$$E(X) = \sum_{x \in R} xP(X = x)$$

Its variance, which measures how far we expect this random variable to be from its expected value:

$$Var(X) = E[(X - E(X)^2)] = \sum_{k}(k - E(X))^2 \times P(X = k)$$

Or alternatively:

$$Var(X) = E(X^2) - E(X)^2$$

Its standard deviation is the nonnegative roots of its variance:

$$std(X) = Var(X)$$

16.1 Linearity of Expectation

If $Y = aX + b$ where a and b are real numbers:

$$
\begin{aligned}
E(Y) &= E(aX + b) = aE(X) + b \\
Var(Y) &= Var(aX + b) \\
&= E((aX + b)^2) - E(aX + b)^2 \\
&= E((a^2X^2 + 2abX + b^2) - (aE(X) + b)^2 \\
&= a^2E(X^2) + 2abE(X) + b^2 - a^2E(X)^2 - 2abE(X) - b^2 \\
&= a^2(E(X^2) - E(X)^2) \\
&= a^2Var(X)
\end{aligned}
$$

Note: it is **not** generally true that $E[g(X)] = g(E(X))$ unless $g(X)$ is a linear function.

16.2 Discrete Uniform

A discrete uniform random variable with range [a, b] takes on any integer value between a and b inclusive. (For simplicity we assume a and b are integers.) Each value has the same probability. Its PMF is:

$$P(X = k) = \frac{1}{b - a + 1} \; for \; k = a, \ldots, b$$

Its expected value is:

$$E(X) = \frac{a+b}{2}$$

Its variance is:

$$Var(X) = \frac{(b-a+1)^2 - 1}{12}$$

16.3 Bernoulli

A Bernoulli random variable has two possible outcomes. We use 1 (success) or 0 (fail) to represent the two outcomes. Its PMF is often described as:

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

Its expected value is:

$$E(X) = (1 - p) \times 0 + 1 \times p = p$$

Its variance is:

$$Var(X) = p - p^2$$

16.4 Binomial

A binomial random variable is associated with the number of successful trials in several independent Bernoulli trials. In this formula, n is the total number of independent Bernoulli trials, x is the number of successful trials and p is the probability that a Bernoulli trial is successful. Its PMF is:

$$P(X = k) = \binom{n}{x} \times p^k \times (1 - p)^{n-k}$$

Its expected value is:

$$E(X) = \sum_{k=0}^{n} k \times \binom{n}{k} p^k (1 - p)^{n-k} = np$$

Its variance is:

$$Var(X) = np(1 - p)$$

16.5 Gemometric

A geometric random variable is associated with the number of independent Bernoulli trials required to come up with the first successful trial.

$$P(X = k) = (1 - p)^{k-1} \times p$$

Its expected value is:

$$E(X) = \sum_{k=0}^{\infty} k(1 - p)^{k-1} p = \frac{1}{p}$$

Its variance is:

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

16.6   Poisson

A Poisson random variable is associated with the number of times a random and independent event occurs in a fixed interval. In this formula, λ is the expected number of occurrences, k is the actual number of occurrences and e is the Euler's number.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Its expected value is:

$$E(X) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \lambda$$

Its variance is:

$$Var(X) = \lambda$$