CS 240: Reasoning Under Uncertainty (Fall 21) September 13

Lecture 1-3

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

1 Introduction

Welcome to CS240 Reasoning Under Uncertainty! This class is about learning mathematical tools that help you reason under uncertain circumstances.

This worksheet reviews the materials discussed in the first three lectures.

The first lecture introduced set theory, set algebra, and probabilistic models.

The second lecture discussed the inclusion and exclusion principle, discrete probability models, odds, and conditional probability.

The third lecture discussed the Monty Hall problem, the total probability theorem, and the Bayes' Rule.

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

$$S \cup T = \{x | x \in S \; or \; x \in T\}$$

3.1.2 Intersection

$$S \cap T = \{x | x \in S \; and \; x \in T\}$$

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 (easily) tricked into believing that choosing either of the 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 gained no additional information by choosing 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 Practice problems

1. Think about how to prove DeMorgan's laws (hint: think from an element's perspective).