DSCI 551 Lecture 1

Depicting Uncertainty

Hello and welcome!

Vincent Liu

G. Alexi Rodríguez-Arelis

Figure 1: Teaching team for **Section 1** (*Vincent Liu*) and **2** (*G. Alexi Rodríguez-Arelis*).

DSCI 551 Syllabus

High-Level Goals for DSCI 551

- Provide fundamental concepts in probability.
- Develop a statistical view of data coming from a probability distribution.

Course Essentials

- **Eight** lectures, **four** labs (12% each), and **two** quizzes (25% each).
- **iClicker** will be used for in-class polls (2% for participation) beginning `lecture2`.
- MDS general policies can be found **here**.
- Course content/logistics can be found in the **GitHub repo**.
- We will use R in lectures and labs.

Lecture Overview

- You are required to do **some reading in advance** (except for this lecture).
- You can find the lecture notes **here**.

Lab Overview

- You will be split in two lab sessions:
  - `L01` on Tuesday and `L02` on Monday
- Assignments to be submitted as **R markdowns** via **Gradescope**.

Communication

- We will use the course's **Slack channels**: one for **Section 1** and another for **Section 2**.
- Try to post all your course general inquiries on your corresponding channel.

Questions?

Outline of Lecture 1

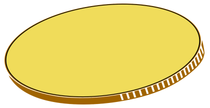
1. Thinking About of Probability
2. Probability Distributions
3. Measures of Central Tendency and Uncertainty

1. Thinking About of Probability

- You will find Probability throughout the MDS program and in many data science topics.
  - Regression, Bayesian, Supervised Learning, Causal Inference, …
- But what is Probability?

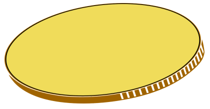
Experiments

- At the heart of probability is the concept of an experiment such as tossing a coin, rolling a die, …
- Each experiment has an outcome
  - Example: Heads or tails
  - Example: 1 or 2 or 3 or 4 or 5 or 6



Sample Space and Events

- Sample space (S) is the collection of **all the possible outcomes** of an experiment.
  - $S = \{\text{Heads}, \text{Tails}\}$
  - $S = \{1, 2, 3, 4, 5, 6\}$
- An event is a subset of the sample space.
  - The event that coin flip is heads: $E = \{\text{Heads}\}$
  - The event of rolling an even number: $E = \{2, 4, 6\}$



1.1 Defining Probability

- Let E be an event of interest
- Suppose we can perform n trials of the experiment which could result in "event E" occurring

- Its probability is defined as

$$P(E) = \frac{\text{Number of times event } E \text{ is observed}}{\text{Total number of experiments}}$$

as the *total number of experiments* goes to infinity.

The Coin Toss System

- Let us illustrate the idea with the typical coin toss example.
- The coin toss represents an experiment of two possible outcomes:

$$\begin{aligned} H &= \{\text{Getting heads}\} \\ T &= \{\text{Getting tails}\}. \end{aligned}$$

- The experiment has the following unknown parameters:

$$\begin{aligned} P(H) &= \text{Probability of getting heads} \\ P(T) &= \text{Probability of getting tails}. \end{aligned}$$

Think about the following question:

Suppose this coin is unfair, i.e., $P(H) \neq P(T) \neq \frac{1}{2}$,

how would you estimate these two unknown parameters?

- **Hint:** think about how we define probability.
- Tossing the coin a given number of times n and obtaining the proportions for both H and T.
- Ideally, they will more accurate as n tends to infinity.

1.2. Calculating Probabilities

- We will introduce axioms of probability and two fundamental laws that will allow us to exercise our probabilistic reasoning:
  - **Law of Total Probability**
  - **Inclusion-Exclusion Principle**

Axiom of Probability: Sample Space

- Recall that the sample space is the collection of **all the possible outcomes** of an experiment.
- Axiom 1:

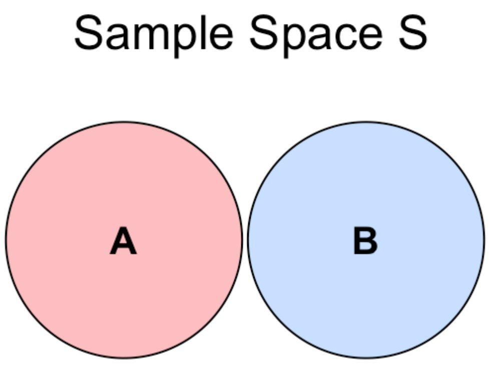
$$P(S) = 1.$$

  That is, all outcomes must be from the sample space.

Axiom of Probability: Mutually Exclusive (or Disjoint) Events

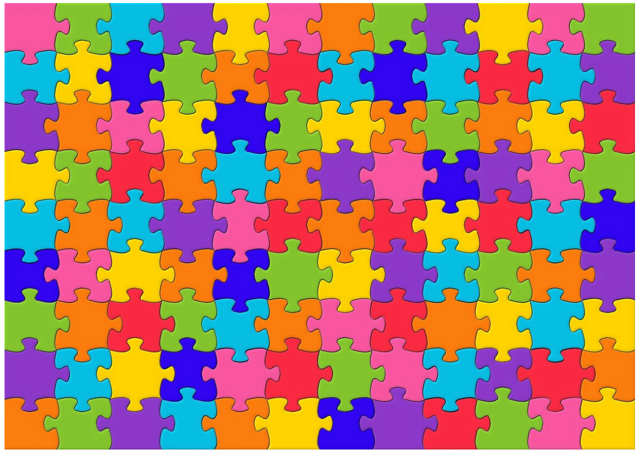
- Two events are mutually exclusive (or disjoint) if they cannot happen at the same time in the sample space S
- Axiom 2: For disjoint events,

$$P(A \cup B) = P(A) + P(B).$$



Law of Total Probability

- The Law of Total Probability allows us to break down an event E into disjoint parts.
  - $P(E)$ = the sum of its partitions' probabilities
- If we are interested in computing the probability of an event, then we can use its partitions to do so.



The Mario Kart Example

Item	Name	Probability
	Banana	0.12
	Bob-omb	0.05
	Coin	0.75
	Horn	0.03
	Shell	0.05

> **Attribution:** Images from pngkey.

The Mario Kart Example

Sample Space of Mario Kart

- The sample space contains only these 5 items, so

$$
\begin{aligned}
& P(S) \\
& = P(\text{Banana}) + P(\text{Bob-omb}) + P(\text{Coin}) + P(\text{Horn}) + P(\text{Shell}) \\
& = 1
\end{aligned}
$$

- It is NOT possible we could encounter any other item in this example

Complement of an event

- For a sample space S and an event A, the complement of A is the subset of other outcomes that do not belong to event A:

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

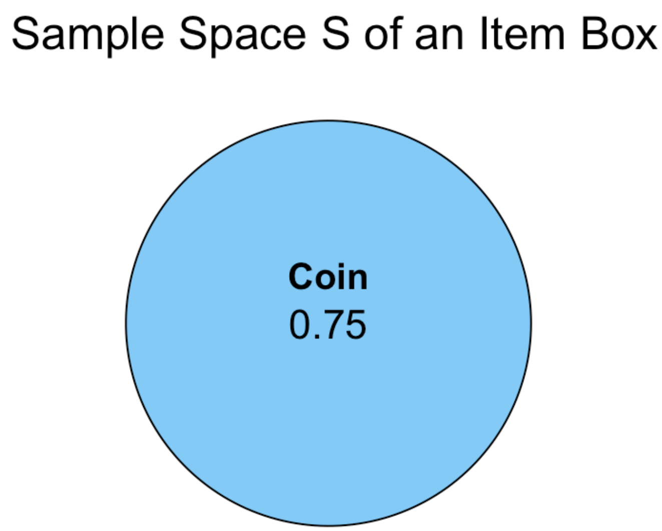
Exercise 1: Complement

- What is the probability of getting something other than a coin?

Item	Name	Probability
	Banana	0.12
	Bob-omb	0.05
	Coin	0.75
	Horn	0.03
	Shell	0.05

Answer

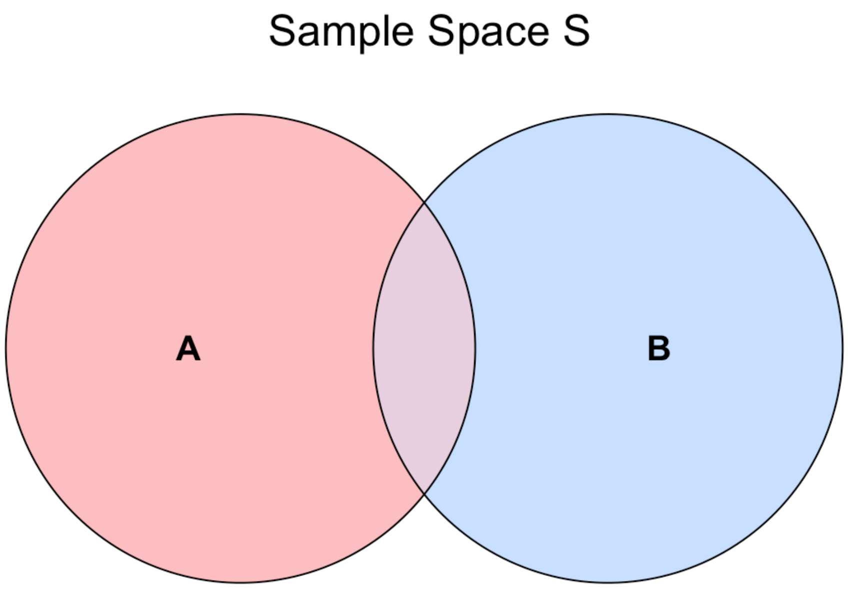
$$P(\text{Coin}^c) = 1 - P(\text{Coin}) = 1 - 0.75 = 0.25.$$



Inclusion and Exclusion Principle

- Let A and B be two events of interest:

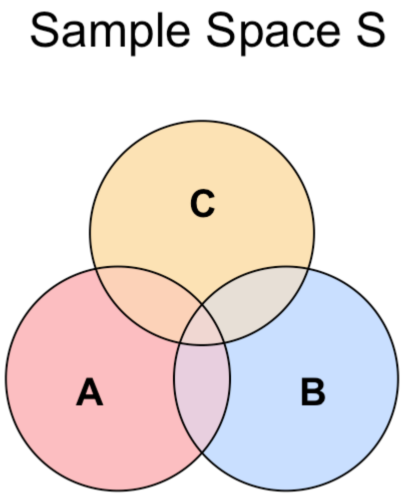
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Extension to Three Events

- Let A, B, and C be three events of interest in the sample space S:

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



The Mario Kart Example

Item	Name	Probability	Combat Type	Defeats Blue Shells
	Banana	0.12	contact	no
	Bob-omb	0.05	explosion	no
	Coin	0.75	ineffective	no
	Horn	0.03	explosion	yes
	Shell	0.05	contact	no

Exercise 2: Disjoint Events

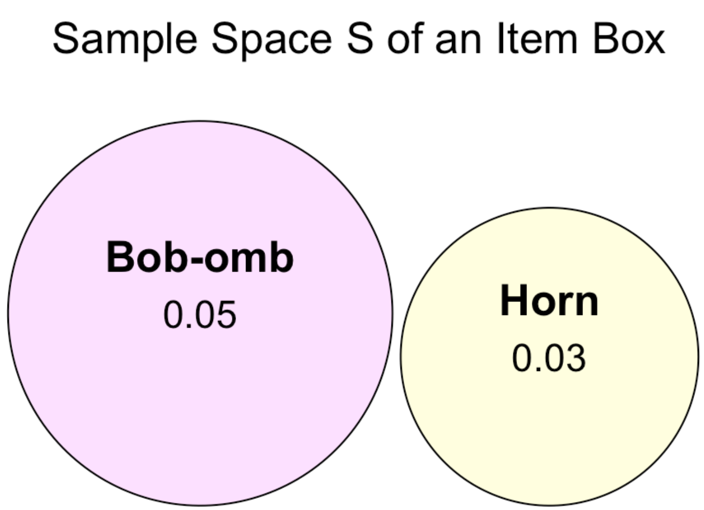
- What is the probability of getting an item with an explosion combat type (event E)?

Item	Name	Probability	Combat Type	Defeats Blue Shells
	Banana	0.12	contact	no
	Bob-omb	0.05	explosion	no
	Coin	0.75	ineffective	no
	Horn	0.03	explosion	yes
	Shell	0.05	contact	no

Answer

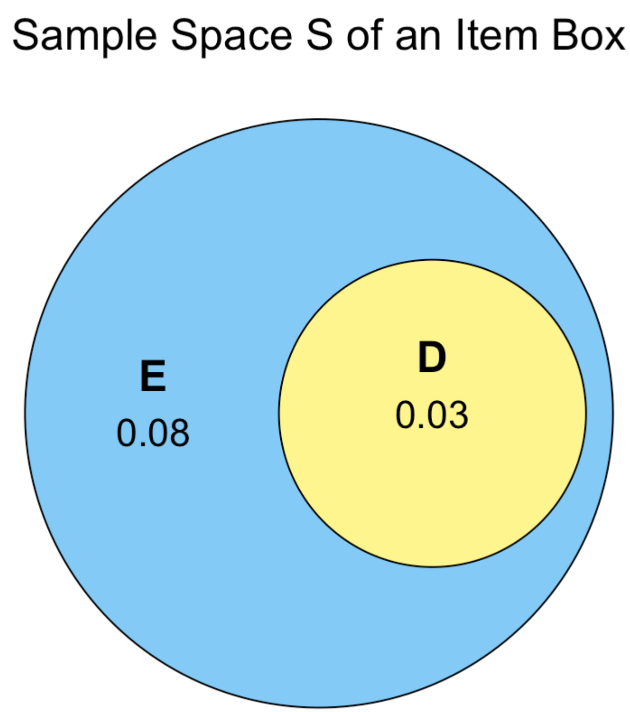
We can compute this probability by adding up the following probabilities:

$$P(E) = P(\text{Bob-omb}) + P(\text{Horn}) = 0.05 + 0.03 = 0.08.$$



Exercise 3: Inclusion-Exclusion Principle

- What is the probability of getting an item that is an explosion item (event E) or an item that defeats blue shells (event D)?



Answer

First note that

$$P(E \cap D) = P(\text{Horn}) = 0.03.$$

By the Inclusion-Exclusion Principle, we have that

$$\begin{aligned} P(E \cup D) &= P(E) + P(D) - P(E \cap D) \\ &= 0.08 + 0.03 - 0.03 \\ &= 0.08. \end{aligned}$$

Independent Events

- Two events are independent if the occurrence of one of them does not affect the probability of the other.
  - A = {Raining at UBC}
  - B = {Getting Coin from an item box}
- Their intersection satisfies:

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

- Independent events are different from mutually exclusive (or disjoint) events.

1.3. Comparing Probabilities

- We might be interested in comparing two probabilities.
- We will introduce a concept called odds.

The Odds

- Suppose an event has a probability p of happening.
- The odds o are defined as the ratio of this probability to the probability of not happening $1 - p$:

$$o = \frac{p}{1 - p}.$$

- With some algebraic rearrangements, we can obtain p with the odds:

$$p = \frac{o}{o + 1}.$$

Example

- Odds are commonly used in gambling.
- If you win 80% of the times at Poker, i.e., $p = 0.8$; then your odds are:

$$o = \frac{p}{1-p} = \frac{0.8}{0.2} = 4$$

- This is sometimes written as 4:1 odds – that is, *four wins for every loss*.

2. Probability Distributions

- A Random Variable (RV) is a variable that takes on a value (an outcome) based on some underlying probabilities.
  - Let X denote the number of heads in 10 coin tosses.
  - X can take a value from $0, 1, 2, \ldots, 10$.
- A probability distribution is a function that describes the probability for each possible value that the random variable can take.
  - $P(X = 0), P(X = 1), \ldots, P(X = 10)$

Types of Random Variables

In general, random variables are classified as:

- Discrete: it can take on a set of countable outcomes.
- Continuous: it can take on a set of uncountable outcomes.

- A discrete RV has a probability mass function (PMF).
  - $P(X = x)$ for all possible values x
- A continuous RV has a probability density function (PDF).

Example of a Discrete and Categorical Random Variable

$$Y = \text{Item obtained from the box.}$$

Item	Y	Probability
	Banana	0.12
	Bob-omb	0.05
	Coin	0.75
	Horn	0.03
	Shell	0.05

Example of a Discrete and Count Random Variable

$$P = \text{Slices of pizza you can eat in a MDS event.}$$

P	**Probability**
1	0.25
2	0.50
3	0.15
4	0.10

3. Measures of Central Tendency and Uncertainty

- These measures summarize the information of a probability distribution.
- There are two commom classes of metrics:
  - Central tendency: a “typical” value in a random variable.
  - Uncertainty: a measure of how “spread” the random variable is.

3.1. Mean and Variance

- The most common and useful are mean and variance.
- Note that both metrics apply to both discrete and continuous random variables (as long as they are numeric).

The Mean

- It is a measure of central tendency.
- If X is discrete, with $P(X = x)$ as a PMF, then

$$\mathbb{E}(X) = \sum_x x \cdot P(X = x).$$

- If X is continuous, with $f_X(x)$ as a PDF, then

$$\mathbb{E}(X) = \int_x x \cdot f_X(x) \mathrm{d}x.$$

The Variance

- It is a measure of uncertainty.

$$\mathrm{Var}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

- It is the expected squared deviation from the mean.
- The variance cannot be negative.
- The square root of variance is called the Standard Deviation

3.2. Mode and Entropy in Discrete Random Variables

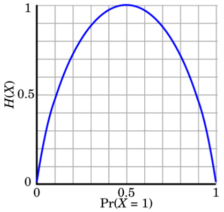
- These two metrics are commonly used with **discrete** random variables.
- The mode is a measure of central tendency. It is the outcome having the highest probability.

The Entropy

- It is a measure of uncertainty defined as

$$H(X) = -\sum_x P(X = x) \log[P(X = x)].$$

- It is the expected negative log probability (information).
- It is a nonnegative measure of uncertainty.
- If its value equals to zero, then there is no randomness.



Example

- What is the mode for $Y = \text{Item obtained from the box?}$

Item	Y	Probability
	Banana	0.12
	Bob-omb	0.05
	Coin	0.75
	Horn	0.03
	Shell	0.05

> **Attribution:** Images from pngkey.

How About the Entropy?

$$
\begin{aligned}
H(X) &= -\sum_x P(X = x) \log[P(X = x)] \\
&= -[0.12 \log(0.12) + 0.05 \log(0.05) + \\
&\qquad 0.75 \log(0.75) + 0.03 \log(0.03) + 0.05 \log(0.05) \\
&= 0.87
\end{aligned}
$$

Today’s Learning Objectives

Now you should be able to…

- Define probability as a proportion that converges to the truth as you perform more experiments.
- Calculate probabilities using the Inclusion-Exclusion Principle, and the Law of Total Probability.
- Convert between odds and probability.
- Interpret the concepts of random variables and probability distributions.
- Calculate and interpret mean, mode, entropy, and variance.