Probability Theory

Definition: There are two different categories of probability "definitions" or "interpretations:" objective and subjective.

Subjective: In subjective definitions of probability, the probability of an event A, denoted by $P(A)$, measures our certitude that A did or will occur. If $P(A) = 1$ then we are completely certain A did or will occur. If $P(A) = 0$ then we are completely certain A did not or will not occur. Values of $P(A)$ between 0 and 1 denote varying degrees of certainty. Note that subjective probability is all about us, how certain we are about the event A. Thus $P(A)$ varies from person to person, hence it is subjective.

Objective: In objective definitions of probability, $P(A)$ is a constant of nature independent of the observer and hence objective. Of particular interest is the frequentist definition of probability. Let A be some event associated with a repeatable experiment. Then the probability that A occurs is

$$P(A) = \lim_{n\to\infty} \frac{\text{times } A \text{ occurs}}{n}$$

where n is the number of times we repeat the experiment. Note that $P(A)$ will be the same for different observers and hence is objective.

Bayesian & Frequentist Statistics: Luckily probability theory works the same for all the various definitions of probability. Unfortunately, how one does statistics depends on the definition used. There are two different ways of doing statistics depending on which definition of probability you use:

Bayesian Statisics: The Bayesian approach to statistics uses the subjective definition of probability.

Frequenist Statistics: Frequentist statistics uses the objective frequentist definition of probability.

Each approach has its advantages and weaknesses. Given sufficient data, both methods will give similar answers. In this course we will use the frequentist approach to statistics. It is more widely used and computationally easier than Bayesian statistics but also less intuitive.

Probability Models: Throughout this course we will construct and use probability models. A probability model of some random phenomenon provides a

I. description of all possible events, and an

II. assignment of probabilities to those events.

I. Describing Events

Events in probability theory are defined and manipulated using set theory. We begin with some definitions:

Simple Outcome: A simple outcome is a simplest event corresponding to a **complete** execution of the experiment.

Sample Space: The set of all simple outcomes, denoted by S.

Event: A subset of the sample space.

Exercise 1: Suppose our experiment consists of tossing a fair coin twice. Do the following using a Venn diagram:

1. Give the sample space S, i.e., list all the simple outcomes.
2. List the simple outcomes corresponding to the event $C =$ "exactly one tail."
3. List the simple outcomes for $D =$ "both tosses tails."

Mutually Exclusive: Note that events C and D can't both occur in one execution of the experiment because they don't share any simple outcomes. We say events C and D are **mutually exclusive**.

Exercise 2: Simple outcomes are mutually exclusive (True/False).

Combining Events: Since events are sets of simple outcomes, we can create new events using the set operations union ($\cup$), intersection ($\cap$), and complement (^{c}). These set operations comprise the basic "grammar" of probability theory.

Union: The **union** of events A and B - denoted by $A \cup B$ - consists of all simple outcomes belonging to A, to B, or to both. The event $A \cup B$ occurs when at least one of events A or B occur. More generally, $A_1 \cup A_2 \cup \ldots \cup A_k$ occurs if at least one of events A_1, A_2, $\ldots$, A_k occurs.

Exercise 3: List the simple outcomes in $C \cup D$. Describe the event $C \cup D$ in words.

Intersection: The **intersection** of events A and B - denoted by $A \cap B$ - consists of all simple outcomes belonging to both A and B. The event $A \cap B$ occurs when both events A and B occur. More generally, $A_1 \cap A_2 \cap \ldots \cap A_k$ occurs if all events A_1, A_2, $\ldots A_k$ occur.

Exercise 4: How many simple outcomes are in $C \cap D$? How does this relate to the fact that C and D are mutually exclusive?

Complement: The complement of event A - denoted by A^c - is the set of simple outcomes in S not in A. Event A^c occurs when A does not occur. It is convenient to use $\bar{A}$ to denote the complement of A. Also, note that $(A^c)^c = A$.

II. Assignment of Probabilities to Events

The Russian mathematician A.N. Kolmogorov rigorously defined probability theory in 1933. The assignment of probabilities to events must satisfy Kolmogorov's Axioms:

1. $P(S) = 1$.
2. For any event A, $0 \leq P(A) \leq 1$.
3. If A and B are mutually exclusive, $P(A \cup B) = P(A) + P(B)$. More generally, if A_1, A_2, $A_3, \ldots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup A_3 \ldots) = P(A_1) + P(A_2) + P(A_3) \ldots$

Axioms 1 and 2 are fairly obvious, especially for a frequentist (why?). Axiom 3 I call the **Additive Rule for Mutually Exclusive Events**.

Equally Likely Events: If the sample space consists of N equally likely simple outcomes, then for an event A containing k simple outcomes, $P(A) = k/N$.

Exercise 5. Reconsider our experiment in which we toss a fair coin twice. Since the coin is fair, each simple outcome is equally likely. Using this fact compute $P(C)$ and $P(D)$.

Randomly Selecting an Item: A population from which an item is sampled at random can be thought of as a sample space with equally likely events.

Exercise 6: There are currently 146 students taking 223 this spring. There are 20 students in section 3 and 15 in section 4, both of which are taught by Professor McSweeney. What's the probability a randomly selected 223 student is in section 3? What's the probability a randomly selected 223 student is in a section taught by McSweeney?

Exercise 7: Let A and B be two events associated with some experiment such that $P(A) = 0.6$ and $P(B) = 0.7$. Compute $P(A \cup B)$.

General Additive Rule: Suppose events A and B are not mutually exclusive. Then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

You couldn't compute $P(A \cup B)$ in exercise 7 because A and B aren't mutually exclusive. You can't determine $P(A \cup B)$ using the general additive rule since $P(A \cap B)$ is unknown. However, you can compute the smallest possible value for $P(A \cap B)$; do so.

Complement Rule: $P(A^c) = 1 - P(A)$. This rule is very useful when computing $P(A)$ directly is difficult but computing $P(A^c)$ is straightforward.