Probability, Statistics, and Bayes' Theorem
Session 1

1 Introduction

These sessions are intended to give a brief introduction to some of the main ideas in probability and statistics as they occur in daily life. The emphasis will be on conditional probability and the role of Bayes' Theorem in both probability and statistics. These notes are intended to give the formal mathematical backbone that the main content of the sessions will be built on.

We live in a world that is filled with data. Almost every time that we make a decision, we consult the appropriate data and base our decision on it in some way. Often, we don't have (or don't want) direct access to the data itself, but instead are dealing with certain numbers that represent some important feature of the data. These are statistics. Being able to interpret these statistics is a vital part of our daily life. Unfortunately, it is very, very easy to either misunderstand the statistics or to be mislead by others who claim to understand them. It is my hope that after these three sessions, you will have a firmer grasp on what these statistics mean, and just as importantly, what they don't mean.

The general breakdown will be as follows:

- Session 1
  - Mathematical preliminaries
  - Finite probability
- Session 2
  - Conditional probability
  - Bayes' Theorem
- Session 3
  - Statistical inference
  - Statistical fallacies/paradoxes

2 Mathematical Preliminaries

The two main preliminaries that will be needed are some basic set theory and combinatorics. Set theory gives us a unified way to talk about the properties that various objects have and the relationship between various classes of objects. Combinatorics gives us a way to count the number of ways that certain things can happen.

2.1 Set Theory

Intuitively, we can think of the set as being specified by some collection of properties that objects either can or can't have, and then the object is included in the set if it has all of the properties. For example, if the set is specified by the property that it's members are male, then I would be in that set, while my wife would not be in it. More abstractly, we can define the set by the objects it contains. For example, we could define $A = \{x, y, z\}$ to be the set that contains the objects x, y, z. It is important to note at the outset that objects are only listed once in a set, or to put it another way, they are **not** repeated within the set. So $\{1, 4, 9\}$ is a valid set, but $\{1, 4, 9, 9\}$ is not because the object "9" is repeated (this would be called a *multiset*).

The basic notion of set theory is *inclusion*. This states whether a certain object belongs to a certain set or not. In symbols, we write $x \in A$ if the object x is in the set A, and we write $y \notin A$ if y is not in A. We denote by $|A|$ the number of objects in A. There is one important set that has no objects at all in it. It is called the *empty set* and is denoted by $\emptyset$ (sometimes you will see it written as $\{\emptyset\}$, but I will usually omit the brackets).

Example Let $A = \{1, 2, 3, 4, 5\}$, $B = \{1, 2, 3\}$, and $C = \{3, 4, 5\}$. Then we have $1 \in A, 1 \in B, 1 \notin C$ and $3 \in A, 3 \in B, 3 \in C$, for two examples, and $|A| = 5, |B| = |C| = 3$.

Given any non-empty set A, we can form other sets that contain some, but not necessarily all, of the elements in A. These are called *subsets* of A. If B is a subset of A, this is denoted by $B \subseteq A$. As with inclusion, if B is **not** a subset of A, we write $B \nsubseteq A$. Often times I will not write the horizontal line, instead just writing $B \subset A$. Every set has at least two subsets, the empty set and the set itself. These are called *trivial* subsets.

Example Returning to the sets A, B, C from before, we see that $B \subset A$ and $C \subset A$, but $B \nsubseteq C$.

Suppose that we are given a set A and a subset B. We call the set of all elements that are in A but are not in B the *complement* of B in A, and we denote it by B^c. Formally, $B^c = \{x | x \in A \text{ but } x \notin B\}$.

Example Working with the same sets as before, $B^c = \{4, 5\}$ and $C^c = \{1, 2\}$.

There are two very important ways to form new sets from any two given sets. Let A and B be arbitrary sets. Define the *union* of A and B, denoted $A \cup B$, to be the set that contains every element of A and B, where we only included each element once. Formally, $A \cup B = \{x | x \in A \text{ or } x \in B\}$. Define the *intersection* of A and B , denoted $A \cap B$, to be the set that contains all the elements that are in both A and B. Formally, $A \cap B = \{x | x \in A \text{ and } x \in B\}$. Since the union and intersection of two sets produces another set, you can take unions or intersection of as many sets as you'd like.

Example Still working with the same sets, $B \cup C = A$, $B \cap C = \{3\}$, and $A \cup B \cup C = A$.

Two sets A and B are said to be *disjoint* if $A \cap B = \emptyset$. In words, two sets are disjoint if they do not have any elements in common. Let A be a given set and let A_i be subsets for $i = 1, \ldots, n$. The collection of subsets is called a *partition* of A if $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $A = A_1 \cup A_2 \cup \cdots \cup A_n$ (this is often abbreviated to $\cup_{i=1}^{n} A_i$).

2.1.1 Exercises

1. Show that $A \cap \emptyset = \emptyset$ for any set A.
2. Show that if $B \subset A$, then $A \cup B = A$ and $A \cap B = B$.
3. Let A, B, C be any three sets. Show that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

2.2 Combinatorics

One of the basic things that we do everyday is count. Combinatorics is the branch of mathematics concerned with counting things.

In what follows, the idea is that we have a urn (bowl) with n marbles in it, and we are going to draw r marbles out of the urn. We will also assume that the marbles are labeled from 1 to n. If the order in which we draw the r marbles out matters, then we say that it's ordered. If the order doesn't matter, then we say that it's unordered. If we put the marble back into the urn after we draw it out, we say that it is with replacement. If we keep the marble out once it's been drawn, we say that it's without replacement. We want to count the total number of possible outcomes of drawing r marbles given the various conditions outlined above. This can be summarized as follows:

1. Ordered, without replacement.
2. Ordered, with replacement.
3. Unordered, without replacement.
4. Unordered, with replacement.

It turns out that all 4 of these situations have very simple answers, but we have to introduce a little notation first.

We define the *factorial* of a non-negative integer n, written $n!$, to be the numbers from 1 to n multiplied together: $n! = n \times n-1 \times n-2 \times \cdots \times 3 \times 2 \times 1$. For convience, we define $0! = 1$. We define the *binomial coefficient* as follows:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}, \text{ if } 0 \leq r \leq n, \text{ and } \binom{n}{r} = 0 \text{ if } r > n.$$

While it's not at all obvious, it's a very important fact that $\binom{n}{r}$ is always a non-negative integer (more specifically, even though you are dividing, the answer is never a fraction).

Given these, the formulas for the quantities we're interested in are given by

1. Ordered, without replacement $=\frac{n!}{(n-r)!}$.
2. Ordered, with replacement $= n^r$.
3. Unordered, without replacement $= \binom{n}{r}$.
4. Unordered, with replacement $= \binom{n+r-1}{r}$.

Example Suppose we are interested in knowing how many different hands of 4 cards can be dealt from a standard deck of 52. The cards are now the marbles, and the urn is the shuffled deck. In this example, $n = 52$ and $r = 4$. Now all we have to do is figure out which of the 4 possibilities we are dealing with. Since we are free to rearrange the cards we've been dealt in our hand, we see that order doesn't matter, and since once a card has been dealt it is removed from the deck, we see that this is without replacement. That means the number of ways is given by $\binom{52}{4} = 270725$.

2.2.1 Exercises

1. Show that $\binom{n}{r} = \binom{n}{n-r}$.
2. Give a concrete example of a situation that would be counted by each of the 4 ways listed above, and explicitly calculate the corresponding value.

3 Probability

We all have some intuitive idea of what probability is. This is all well and good, but the problem is that math can't be built upon intuitions alone for what I hope are obvious reasons. So we need some formal definitions.

3.1 Formal Definition of Probability

The framework that we will be using is built in the following way. Imagine that we are conducting some kind of experiment, and before the experiment we already know all of the possible outcomes that can occur. We call the set of all possible outcomes the *sample space* and it is customary to denote it by the capital Greek letter omega, Ω. Each element of the sample space is called an *event*, and is denoted usually by a lowercase Greek omega, ω. In the framework we are using, each ω is a possible outcome of our experiment. We will always assume that there are only a finite number of ω in Ω.

Given this, a *probability measure*, or just probability, is a function defined on all subsets of Ω that satisfies the following three axioms:

1. $0 \leq P(A) \leq 1$ for any $A \subset \Omega$.
2. $P(\Omega) = 1$.
3. If $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

This might seem like enough information to do anything useful, but let's look at some examples of what we can do with just these three axioms.

Example Let $A \subset \Omega$ be an arbitrary subset of our sample space. It follows from the definitions that $A \cap A^c = \emptyset$ and that $A \cup A^c = \Omega$, so that A and A^c form a partition of Ω. Axiom 2 tells us that $P(A \cup A^c) = P(\Omega) = 1$, while axiom 3 tells us that $P(A \cup A^c) = P(A) + P(A^c)$. Putting these two together gives $P(A) + P(A^c) = 1$, which means $P(A^c) = 1 - P(A)$.

Example Let A and B be any two subsets of Ω. Consider $A \cap B$ and $B \cap A^c$. It's not hard to see that $(A \cap B) \cap (B \cap A^c) = \emptyset$ and that $(A \cap B) \cup (B \cap A^c) = B$. Then by the same steps we used in the previous example, we get that $P(B) = P(A \cap B) + P(B \cap A^c)$, which gives $P(B \cap A^c) = P(B) - P(A \cap B)$.

Example With A and B as before, notice that $A \cup B = A \cup (B \cap A^c)$. But we also have that A and $B \cap A^c$ are disjoint, so axiom 3 gives $P(A \cup B) = P(A) + P(B \cap A^c)$. Using the result from the previous example, this becomes $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

3.1.1 Exercises

1. Show that $P(\emptyset) = 0$.
2. Show that if $A \subset B$, then $P(A) \leq P(B)$.

3.2 Calculating Finite Probabilities

Since we are only considering sample spaces that have a finite number of events, most of our probabilities will come from counting the number of outcomes of a certain type and dividing this by the total number of possible outcomes. This means that for a given $A \subset \Omega$,

$$P(A) = \frac{\text{number of } \omega \in A}{\text{total size of } \Omega} = \frac{|A|}{|\Omega|}.$$

This is where the tools we have from combinatorics come in very handy.

Example Let's investigate various probabilities associated with 4 card hands dealt from a shuffled deck of 52 cards. From the example in the combinatorics section, we know that there are $\binom{52}{4}$ possible 4 card hands, so this means that our sample space, Ω, has $\binom{52}{4}$ events in it. Each event, ω, represents a specific 4 card hand. So, for example, the probability of being dealt 4 aces is given by

$$\frac{1}{\binom{52}{4}} = \frac{1}{270725}.$$

Example What about the probability of being dealt exactly 2 aces? Here we have to count all the possible ways to have exactly 2 aces in our 4 cards. This is easiest to think about if you break things up into aces, of which there are 4 in the deck, and non-aces, of which there are 48 in the deck. Of those 4 aces, we need to get exactly 2 of them, so this is unordered drawing of 2 objects from 4 without replacement, and is given by $\binom{4}{2}$. Since we want exactly 2 aces, the other 2 cards have to be non-aces, of which there are 48. So we are drawing 2 from 48, and again it's unordered and without replacement. This is then given by $\binom{48}{2}$. Now we have to multiply these two together to get all possible cases of exactly 2 aces:

$$\frac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}} = \frac{6768}{270725}.$$

This says that it is almost 7000 times more likely to be dealt a pair of aces than all 4 aces in a 4 card hand.

We are cheating a little here by using our intuitive notion of "independence" to multiply the probability of getting 2 aces and the probability of getting two non-aces. That's okay because in this example our intuition is correct, though great care must be taken in general when assuming that two or more events are independent. We will discuss independence next time in the context of conditional probability.

It's worth pointing out that this formula, and the reasoning behind it can be generalized to give a formula for the probability of being dealt exactly i aces in a hand of 4 cards, for $i = 0, 1, 2, 3, 4$:

$$P(\text{exactly } i \text{ aces}) = \frac{\binom{4}{i}\binom{48}{4-i}}{\binom{52}{4}}, \quad \text{for } i = 0, 1, 2, 3, 4.$$

3.2.1 Exercises

1. What is the probability of being dealt exactly 3 aces in a 4 card hand?
2. What is the probability of being dealt at least 2 aces in a 4 card hand?
3. What is the probability of **not** being dealt 4 aces in a 4 card hand?