Mathematical Probability

STA 281 Fall 2004

1 Introduction

Probability is designed to describe the results of *random experiments*, which is a fancy name for anything unpredictable. A random experiment consists of drawing a card from a shuffled deck, rolling a die, or determining how long an applicance will last until it fails. We begin our discussion of probability with some definitions.

1. **Experiment** - any random event
2. **Outcome** - any possible result of an experiment. For example, if we roll a die the outcomes are 1,2,3,4,5, or 6. Each individual number is an outcome.
3. **Sample Space** - the collection of all possible outcomes of an experiment.
4. **Event** - Any subset of the sample space. Each outcome by itself is an event. The sample space S is also an event. The empty set, $\emptyset$, is also an event. Typical events for rolling a die might be "roll an even number", consisting of the outcomes 2, 4, and 6, or "roll a prime number", consisting of the outcomes 2, 3, and 5.

2 Set Theory

2.1 Set Operations

Often we will want to manipulate or combine events. For the purposes of discussion, suppose we are rolling a die and the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Let $A = \{2, 4, 6\}$ be the event "roll an even number" and $B = \{2, 3, 5\}$ be the event "roll a prime number". The three methods for combining events are **complement, union, and intersection**. In English these loosely correspond to *not*, *or*, and *and*.

1. The *complement* of a set A, A^c, is the set of outcomes not in A. In our die example $A^c = \{1, 3, 5\}$ and $B^c = \{1, 4, 6\}$. Note $S^c = \emptyset$ and that in general $(A^c)^c = A$.
2. The *union* of two events A and B, $A \bigcup B$, is the set of outcomes that are in either or both of the events. In our die example $A \bigcup B = \{2, 3, 4, 5, 6\}$. In general $A \bigcup A^c = S$, $A \bigcup S = S$, $A \bigcup \emptyset = A$
3. The *intersection* of two events A and B, $A \bigcap B$, is the set of outcomes common to both events. In our die example $A \bigcap B = \{2\}$. In general $A \bigcap A^c = \emptyset$, $A \bigcap S = A$, $A \bigcap \emptyset = \emptyset$

2.2 DeMorgan's Laws

There are two famous laws in set theory called DeMorgan's laws. These state

$$(A\bigcup B)^c = A^c\bigcap B^c \quad (A\bigcap B)^c = A^c\bigcup B^c$$

In English, the first rule states that the complement of the event "A occurs or B occurs" is "A doesn't occur and B doesn't occur". The second rule states the complement of the event "A occurs and B occurs" is "A doesn't occur or B doesn't occur". These rules can be generalized. Let $A_1, \ldots, A_n$ be n different events. Then

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c \quad \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

The first rule states the complement of "at least one of the A_i occurs" is "none of the A_i occur". The second rule states the complement of "all the A_i occur" is "at least one A_i does not occur".

3 Definition and Axioms of Probability

Our main goal will be to assign probabilities to events. We will write the probability of A as $P(A)$ or $\Pr(A)$. So what do we mean by probability? A random experiment cannot be predicted in the short run. For example, if we roll a die we cannot tell which number will be rolled. However, we can say a lot about the long run frequency of each of the outcomes. If we rolled the die a million times, we know about a sixth of the rolls will be a 1. This long term frequency is the *probability* of the outcome.

There are three *axioms* of probability, or rules that probability values must follow.

1. $P(S) = 1$. The sample space is the set of all possible outcomes. Therefore, some outcome within S occurs each time, so its relative frequency must be 100%, or 1.

2. $P(A) \geq 0$. The relative frequency must be greater than 0 since an event cannot occur a negative amount.

3. If A and B are disjoint events, meaning they share no outcomes ($A\bigcap B = \emptyset$), then $P(A) + P(B) = 0$. This is actually assumed more generally. For any finite collection of disjoint sets $A_1, A_2, \ldots, A_n$

   $$P(A_1\bigcup A_2\bigcup \ldots \bigcup A_n) = \sum_{i=1}^{n} P(A_i)$$

   and for any countably infinite collection of disjoint sets $A_1, A_2, \ldots$,

   $$P(A_1\bigcup A_2\bigcup \ldots) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Axiom 3 allows a very basic method for finding probabilities. Individual outcomes are by definition disjoint. If you can list the outcomes in an event A and have the probabilities of each outcome, you can find the probability of A by summing the probabilities of each outcome in A. For example, suppose the probabilities of observing a particular M&M color are

Color	Brown	Red	Yellow	Green	Orange	Tan
Probability	0.4	0.15	0.1	0.1	0.15	0.1

In this problem the sample space S is the set of possible colors, (Brown, Red, Yellow, Green, Orange, Tan). Let us define two events A = "A color with 6 or more letters is observed" and B = "A color whose second letter is r is observed"? A consists of the outcomes $\{Yellow, Orange\}$ while B consists of the outcomes $\{Brown, Green, Orange\}$. Since we can add the probabilities of outcomes to find event probabilities $P(A) = P(Yellow) + P(Orange) = 0.1 + 0.15 = 0.25$. By similar reasoning $P(B) = 0.4 + 0.1 + 0.15 = 0.65$. The event A^c is the set of all outcomes *not* in A, thus $A^c = \{Brown, Red, Green, Tan\}$ and $P(A^c) = 0.4 + 0.15 + 0.1 + 0.1 = 0.75$. The event $A \bigcup B$ is the set of all outcomes in either A or B, here $\{Brown, Yellow, Green, Orange\}$. The union has probability $P(A \bigcup B) = 0.4 + 0.1 + 0.1 + 0.15 = 0.75$. Finally, $A \bigcap B$ is the set of all outcomes in both A and B. The only outcome in both A and B is $\{Orange\}$, with probability 0.15.

3.1 Equally Likely Outcomes and Combinatorics

A common situation is *equally likely outcomes*. Often we try to make sure that all outcomes in the sample space have an equal chance of occuring. Forms of gambling such as cards, dice, or roulette rely on a shuffled deck, a fair die, and a balanced wheel. More seriously, most sampling methods are designed to make sure each individual has an equal chance of being in the sample. This helps to avoid biases in the results.

If all of N outcomes are equally likely, then each has probability $1/N$. If A is an event containing m outcomes, then $P(A) = m/N$, since we may find probabilities of events by adding the probabilities of the outcomes. If we roll a die, then each of 1 through 6 is equally likely, with probability 1/6. If $A = \{1, 2, 3, 5\}$, then A consists of four outcomes and has probability 4/6.

Therefore, when we have equally likely outcomes, the main task for determining the probabilities of an event is counting the number of outcomes in the event. Sometimes this is easy. If all 52 cards in a deck are equally likely to be drawn and A is the event "draw a spade", then $P(A) = 13/52$ since there are 13 spades in the deck. Often, however, there are large numbers of outcomes, and direct counting becomes difficult. For example, if we draw 5 cards from a deck to construct a poker hand, there are 2598960 possible outcomes. We count outcomes in these more complicated experiments using *combinatorics*.

We will focus on two basic results in combinatorics, the *product rule* and *combinatorial coefficients*. The product rule determines how many ways one item from each of several different sets may be combined in pairs, triples, and so on. For example, if you have 5 shirts and 3 pants, then there are 15 possible combinations of clothes you can wear. If you also have 10 pairs of shoes, then there are $5 \times 3 \times 10 = 150$ combinations. The product rule states that if you have n sets each containing m_i items and you want to pick one item from each set, there are $m_1 \times m_2 \times \ldots \times m_n$ ways to do this. One common application of the product rule is to determine how many ways N individuals may be ordered. For example, suppose $N = 5$. There are 5 people available to be first. After that person has been selected, there are 4 people remaining to be second, then 3 to be third, 2 to be fourth, and finally 1 to be fifth. Thus, there are $5 \times 4 \times 3 \times 2 \times 1 = 5! = 120$ possible orders. In general, N individuals may be ordered in $N!$ ways.

Combinatorial coefficients are used to determine how many ways k individuals may be selected from a group of n individuals. For example, suppose we have 10 people and we want to select 3 of them. (**Unless explicitly stated in this course, order does not matter. Typically order is irrelevant in statistical contexts, such as polling**). If order DID matter, we could solve the problem by realizing that we pick one of the ten people first (there are 10 possibilities). There are then 9 possibilities for the second person, and 8 for the third. So $10 \times 9 \times 8 = 720$. But, we ordered the people, since there was a first person, a second person, and a third person. When order doesn't matter, we have to remove the redundant orderings. There are $3 \times 2 \times 1 = 6$ ways to order the three people we selected, so there are $720/6 = 120$ actual groups of 3 people that may be selected.

Usually the solution to this problem is written as a *combinatorial coefficient*. Notice that $10 \times 9 \times 8 = 720$ may be written

$$\frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{10!}{7!} = 720$$

We then divided by $3 \times 2 \times 1 = 3! = 6$. The result was

$$\frac{10!}{3!7!} = 120$$

In general, the number of ways k people may be selected from n individuals is

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

This is usually stated "n choose k".

Combinatorial coefficients can be combined with the product rule to compute probabilities. Suppose we have a group of 50 animals, 10 males and 40 females. We select a group of 5 animals at random (meaning that every group is equally likely to be chosen). What is the probability that exactly 3 females are chosen?

First, let's determine how many groups there are overall. We have 50 animals and are selecting 5. There are 50 choose 5, or 2118760 possible groups of animals. Since each group is equally likely to be selected, each group has a 1/2118760 chance of being selected.

Any group with exactly 3 females contains 3 females and 2 males. Therefore we choose 2 males from the 10 available (there are 10 choose 2 possible ways to do this) and 3 females from the 40 available (there are 40 choose 3 ways to do this). Since each group of 3 females can be picked with each group of 2 males, we use the product rule to determine there are

$$\binom{10}{2}\binom{40}{3} = (45)(9880) = 444600$$

possible groups with exactly 3 females. Thus, 444600 of the 2118760 equally likely outcomes result in exactly 3 females being chosen, and therefore the probability of exactly 3 females being chosen is

$$\frac{\binom{10}{2}\binom{40}{3}}{\binom{50}{5}} = \frac{(45)(9880)}{2118760} = \frac{444600}{2118760} = 0.2098$$

In general, suppose we have a population of N individuals, M of which are labelled S while the other $N - M$ are labelled F. Suppose we draw a group of n individuals at random from the population. The probability we draw x individuals labelled S and $n - x$ labelled F is

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

4 Some useful, simple theorems

Although we only have three basic axioms of probability, we have already solved some apparently difficult problems. Here are some other basic formulas that follow from the axioms of probability. For each A, B, and C are arbitrary events.

1. $P(A^c) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. $if\ A \subseteq B, then\ P(A) \leq P(B)$
4. $P(A) \leq 1$
5. $if\ A, B\ disjoint,\ P(A \bigcap B) = 0$
6. $P(A) = P(A \bigcap B) + P(A \bigcap B^c)$
7. $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
8. $P(A \bigcup B \bigcup C) = P(A) + P(B) + P(C) - P(A \bigcap B) - P(A \bigcap C) - P(B \bigcap C) + P(A \bigcap B \bigcap C)$

Example

Some common drinks are juice, milk, and water. Suppose that 45% drink juice, 50% drink milk, and 50% drink water. Also suppose 25% drink milk and juice, 15% drink milk and water, 20% drink water and juice, and 5% drink all three.

Find (Hint: Use a Venn Diagram)

a) the proportion who drink only water. 0.20
b) the proportion who drink all of the drinks. 0.05
c) the proportion who drink at least two of the drinks. 0.50
d) the proportion who drink exactly one of the drinks. 0.40