Concepts of Probability

Trial question: we are given a die. How can we determine the probability that any given throw results in a six?

Try doing many tosses:

- Plot cumulative proportion of sixes

- Also look at other features that may or may not indicate fairness, e.g. are there three consecutive sixes in 100 throws?
Law of Large Numbers

Over very many trials, the proportion of times a given outcome occurs becomes very close to the true probability of that outcome.

It’s a useful way to check up on whether probabilities are correct, but it can also serve as a definition of probability.

Independent trials:

Successive trials of some random event (e.g. tosses of a coin, throws of a die) are said to be “independent” if the outcome of any one trial does not affect the outcomes of any others.
Defining Probabilities

- **Relative frequency definition** based on the proportion of times an event occurs in a large population. Implicit use of Law of Large Numbers.

- **Equally likely definition** applies when there are a finite number of possible outcomes and there is no reason to assume one occurs more frequently than others. Examples: throwing a die, drawing cards from a deck.

- **Subjective probability definition** applies when we use our own judgment to assess probabilities. Examples include betting on the outcome of a sporting event or the kinds of judgments business executives make all the time.
Calculating Probabilities

In this section we describe some of the methods for calculating probabilities and illustrate some of the mathematical issues associated with finding probabilities of complicated events.

Sample Space

The first step is to define the sample space. This is just a list of all the possible outcomes of an experiment.
Example 1. One die, possible outcomes \{1, 2, 3, 4, 5, 6\}.

Example 2. Two dice. We could list the possible outcomes by giving the total for each die, e.g. (3,4) would mean that the first die showed a 3 while the second showed a 4. Then the sample space would be all 36 possible combinations — (1,1), (1,2), (1,3), and so on all the way through to (6,6).

Alternatively, we might just be interested in the total. In this case the sample space is \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.

Note, however, that whereas in the first case we could still assume that all 36 outcomes are equally likely, in the second case this would not be true. That might be an argument for preferring the first way of writing the sample space.
Example 3. Sample space for a multi-choice quiz (see text for more discussion):
{CCC, CCI, CIC, CII, ICC, ICI, IIC, III}.
Example 4. Suppose we were interested in all possible outcomes of the 2009 UNC football season (see table). How would we define the sample space in this case?

<table>
<thead>
<tr>
<th>Opponent</th>
<th>H/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke</td>
<td>H</td>
</tr>
<tr>
<td>Miami</td>
<td>H</td>
</tr>
<tr>
<td>Virginia</td>
<td>H</td>
</tr>
<tr>
<td>Florida State</td>
<td>H</td>
</tr>
<tr>
<td>The Citadel</td>
<td>H</td>
</tr>
<tr>
<td>East Carolina</td>
<td>H</td>
</tr>
<tr>
<td>Georgia Tech</td>
<td>A</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td>A</td>
</tr>
<tr>
<td>NC State</td>
<td>A</td>
</tr>
<tr>
<td>Boston College</td>
<td>A</td>
</tr>
<tr>
<td>Connecticut</td>
<td>A</td>
</tr>
</tbody>
</table>
For example, we could define it just be number of wins, so the sample space is \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}.

Or we could define the sample space by the win-loss record, e.g. \textit{WLWLLWWLWLW}L would be one of 2048 possible outcomes. Although we would not in practice try to write out the complete sample space, there is no conceptual problem about defining the sample space in this way.

Or, we could go further and try to list even more information, e.g. the actual scores in each of the games. This would lead to a very large (hypothetically infinite) sample space, but again, so long as we made it clear what possible outcomes are permitted, there is nothing conceptually difficult about defining the sample space in this way.
Events

An event is technically defined to be any subset of the sample space. Usually events are denoted by capital letters, $A$, $B$, etc.

Two possible events are

- $A$: UNC wins at least five football games, and
- $B$: In two throws of a die, the total is at least 10.

Need to represent an event as a subset of the sample space. With UNC football games, if we define sample space

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$$

then the event $A$ is

$$A = \{5, 6, 7, 8, 9, 10, 11\}$$
With dice, if we represent sample space as 

\[ S = \{(1, 1), (1, 2), (1, 3), \ldots, (6, 6)\}. \]

then the subset that defines the event \( B \) is 

\[ B = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}. \]
Defining Probabilities

In the case of *equally likely outcomes*, the *probability* of the event $E$ is defined as

$$P(E) = \frac{\text{number of outcomes in event } E}{\text{number of outcomes in sample space } S}.$$ 

For example, in the case of the event $B$ defined above,

$$P(B) = \frac{6}{36} = \frac{1}{6}.$$ 

UNC football example: no reason to assume that all 12 possible outcomes are equally likely. However if we assigned them probabilities $p_0, p_1, \ldots, p_{11}$ (by whatever means) we can write case that

$$P(A) = p_5 + p_6 + p_7 + p_8 + p_9 + p_{10} + p_{11} + p_{12}.$$
The moral: We can build up probabilities for complicated events by starting with probabilities for elementary events and then manipulating them according to the rules of probability. However, we still need to think carefully about those probabilities for elementary events.