Chapter 4: Continued

Some Definitions (revisited)

Random phenomenon: Something whose outcome is uncertain: flipping a coin, tossing a die, a baseball game, selecting a person randomly from the class and recording the person’s height in inches.

Outcome: one possible result of a random phenomenon (e.g., “heads” and “tails” are the two possible outcomes of a coin flip; possible outcomes for measuring a random person’s height in inches are numbers, almost certainly between 0 and 100)

Event: a collection of outcomes (e.g., getting an even number on the roll of a die; this event is a collection of three outcomes: 2, 4, and 6).

Trial: a single realization of a random phenomenon (e.g., a single coin flip or die roll)

Independent trials: trials where the outcome of one trial doesn’t affect the outcome of any other trial

Probability of an event: long-run relative frequency in independent trials of a random phenomenon

Law of Large Numbers: The long-run relative frequency of an event in repeated independent trials gets closer and closer to the true relative frequency (the true probability) as the number of trials increases.

Simulation of a large number of coin tosses:

Note: a simulation model for a random phenomenon is just that: a model for the random phenomenon, not necessarily an exact representation of reality. In reality, the probability of a heads may not be exactly .5 and may depend on the coin and who flips it. In addition, the trials may not be independent (why might they not be?) But the model is generally a very good approximation to the reality of coin flipping.

There is no “Law of Averages”!! One misinterpretation of the Law of Large Numbers is the so-called “Law of Averages.” Essentially, this “law” states that, in independent trials, if an outcome has not occurred in recent trials then it has a higher chance of occurring in subsequent trials. This is based on the belief that this must happen in order to make the Law of Large Numbers work – that if the
cumulative relative frequency of an event has deviated from what is expected then the process must “make up” for this in order to make the long-run relative frequency equal to what is expected. This belief is false as illustrated below.

Suppose that, just by chance, we obtain 100 heads on the first 100 flips of a fair coin. This is an incredibly unlikely event, but it is possible. How can the Law of Large Numbers work in this case? How can the long-run relative frequency of heads ever approach .5 unless we obtain a greater than expected number of tails in subsequent trials?

Suppose we get exactly what we would expect to get in the next 100 trials: 50 heads and 50 tails. What is the cumulative proportion of heads after 200 trials?

Suppose that we get 500 heads and 500 tails in the next 1000 trials. What is the cumulative proportion of heads after the 1100 trials?

Suppose that we get 5000 heads and 5000 tails in the next 10000 trials. What is the cumulative proportion of heads after the 10100 trials?

Suppose that we get 50000 heads and 50000 tails in the next 100000 trials. What is the cumulative proportion of heads after the 100100 trials?

The coin-flipping model -- independent trials with a constant probability of heads (not necessarily .5) - is sometimes used as a model for binary outcomes in sports: shots in basketball (made or missed), at-bats in baseball (hit or out). The model is obviously not exactly correct (for example, the probability of a hit in baseball changes depending on the pitcher), but has been shown to be remarkably consistent with actual data from all sorts of sports phenomena like this. That is, the patterns of outcomes look remarkably like independent trials. This means that the hot and cold streaks players go through can be explained by pure chance, like coin flips (remember how long streaks of heads and tails occurred in our coin-flipping, both actual and simulated). Simply because a model is consistent with the results doesn’t mean it is correct, but, using the principle of parsimony (Occam’s razor), the burden is on those who don’t believe the model to demonstrate that it is wrong. The theory that players go through streaks where their true probability of success changes for a while (they get “hot” or go “cold”) is called the “hot hand” theory. See the link on the class website for more discussion and analysis of a variety of sports data.

Formal Probability
Notation and definitions:

- S represents the set of all possible outcomes of a random phenomenon
- The letter A (and other capital letters B, C, etc.) represents an event which is a subset of the outcomes in S. A could be the whole set S, the empty set ∅, or anything in between).
- $A^C$ represents the complement of the event A – all outcomes in S which are not in A.
- Events A and B are disjoint (or mutually exclusive) if they have no outcomes in common. That means that disjoint events cannot both occur.
A and B are independent if whether or not one occurs does not change the probability that the other occurs. We often assume that repeated trails of a random phenomenon (like coin tosses or dice rolls) are independent.

Rules for probabilities:
1. \(0 \leq P(A) \leq 1\)
2. \(P(S) = 1\)
3. \(P(A^C) = 1 - P(A)\)
4. If A and B are disjoint events, then \(P(A \cup B) = P(A) + P(B)\)
5. If A and B are independent events, then \(P(A \cap B) = P(A)P(B)\)  
   (this is the Multiplication Rule for independent events and can be extended to more than two events)
6. A partition of the sample space \(S\), is a collection of mutually exclusive events such that \(S = B_1 \cup B_2 \cup \ldots \cup B_k\) and \(B_i \cap B_j = \emptyset\)

Note: disjoint and independent are not the same. Disjoint events can’t be independent because if one of them occurs then the other one cannot occur.

Example: Suppose that in your city, 37% of the voters are registered as Democrats, 29% as Republicans, and 11% as members of other parties (Liberal, Right to Life, Green, etc.) Voters not aligned with any official party are termed “Independent.” You are conducting a poll by calling registered voters at random. In your first three phone calls, what is the probability you talk to

a) all Republicans?

b) no Democrats?

c) at least one Independent?

Equally likely outcomes: if there are \(k\) outcomes in a sample space for an experiment and the outcomes are equally likely, then the probability of each outcome is \(1/k\).
General Addition Rule: If A and B are any two events in a sample space, then
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

Conditional Probability:
\[ P(B \mid A) = \frac{P(A \cap B)}{P(A)} \]

General Multiplication Rule
\[ P(A \cap B) = P(B \mid A)P(A) \]
This follows from the definition of conditional probability and simple algebra.

Total Probability: Let A be an event and B₁,…,Bₖ be a partition of the sample space, then
\[ P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i} P(A \mid B_i)P(B_i) \]
**Example:** The table shows the political affiliation of American voters and their positions on the death penalty.

<table>
<thead>
<tr>
<th></th>
<th>Death Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Favor</td>
</tr>
<tr>
<td>Republican</td>
<td>0.26</td>
</tr>
<tr>
<td>Democrat</td>
<td>0.12</td>
</tr>
<tr>
<td>Other</td>
<td>0.24</td>
</tr>
</tbody>
</table>

a) What’s the probability that a randomly chosen voter favors the death penalty?

b) What’s the probability that a Republican favors the death penalty?

c) What’s the probability that a voter who favors the death penalty is a Democrat?

d) A candidate thinks she has a good chance of gaining the votes of anyone who is a Republican or in favor of the death penalty. What portion of the voters is that?

---

**Independence**

Events A and B are independent whenever \( P(B|A) = P(B) \), or, equivalently, whenever \( P(A \cap B) = P(A)P(B) \). The equivalence follows from the General Multiplication Rule above.

Example 1: If you randomly draw a card from a standard deck of 52 cards, are the events that the card is black card and the event that it is an ace independent?

By far, the most important use of the concept of independence is not in deciding whether two events are independent given a set of probabilities, but in deciding whether events in a model you are building can be reasonably assumed to be independent. This is an important consideration in modeling complex real stochastic processes. For example, in engineering it is common to put in redundant components so that the system fails only if all the components fail. If a component has a 1% chance of failing, then the chance that 3 identical components all fail is \((.01)^3 = .000001 \) or .0001% if we assume that the success or failure of the components are independent of each other. This was the reasoning in having redundant O-rings in the space shuttle booster, the O-rings that were breached in the Challenger disaster of 1986. The problem was that the cold temperatures affected all the rings so that their failures were not independent of each other. This is referred to as a “common mode failure.”
**Drawing without replacement**
Suppose two cards are drawn at random from a standard deck. What is the probability that both cards are aces? What’s the probability that you draw a red card and then a black card?

**Calculating Conditional Probabilities:**
Example: Random Testing for AIDS. In the mid 1980’s, it was suggested that the entire U.S. population be tested for the presence of HIV. Would this be an efficient way of identifying HIV positive individuals?

One of the most reliable tests for HIV is the ELISA test. It gives a positive result for 95% of people who are HIV positive (the sensitivity of the test) and a negative result for 99% of people who are HIV negative (the specificity of the test). The 5% false negative rate in detecting HIV carriers may be due to low concentrations of the virus or antibodies to the virus. The 1% false positive rate may be caused by errors in running the test or by unknown idiosyncrasies in the body chemistry of that 1% of the population. It is estimated that in the U.S. about 1 person in 1000 is HIV positive.

If a randomly selected person in the U.S. tests HIV positive, what is the probability that he or she is actually HIV positive?

**Tree Diagrams:**
Reversing the conditioning using Baye’s Rule:

**Baye’s Rule**: If \( A_1, A_2, \ldots, A_i, \ldots, A_k \) be a partition of the sample space and \( B \) any event in the sample space such that \( P(B) > 0 \), then

\[
\Pr(A_s \mid B) = \frac{\Pr(B \mid A_s) \Pr(A_s)}{\sum_{i=1}^{k} \Pr(B \mid A_i) \Pr(A_i)}
\]