The purpose of this note is to supplement the textbook's account of
basic probability.

Box Model
=========
A **box model** consists of a box that contains tickets, each of which is
equally likely to be selected on any single random draw from the box, as
in a fair lottery or raffle. In this course, every chance process can be
modeled in terms of a sequence of one or more random draws from a box
of tickets. The sequence of draws can be conducted in two different ways:
with replacement or **without replacement**. These terms refer to whether or
not each ticket drawn from the box is replaced in the box before the
next draw; either way, it is assumed that on any given draw, every
ticket in the box at that moment has an equal chance to be selected.

The **outcome** of a sequence of random draws is the list of specific
tickets that were drawn. That is, an outcome is a list of the form

  (1st ticket drawn), (2nd ticket drawn), ... , (last ticket drawn).

The assumption that each ticket is equally likely to be selected on any
given random draw implies that every draw sequence outcome is equally
likely. A set of outcomes is called an **event**. We say an event **happens** or
occurs if a sequence of random draws produces an outcome that belongs to
that event. The **probability** that an event occurs is the fraction

  (number of outcomes in the event) / (number of total possible outcomes).

We usually use capital letters to denote events. We write P(A) to
denote the probability of an event A. Here is the basic definition of
probability using these symbols.

(1) P(A) = (number of outcomes in A) / (total number of outcomes)

Example: Consider the box with tickets labeled x,y,z and consider the
chance process that selects two tickets, without replacement, from the
box. The set of all possible outcomes is

  {(x,y), (x,z), (y,x), (y,z), (z,x), (z,y)}

where the pair (y,x), for example, denotes the outcome where y is
selected on the first draw, and x is selected on the second draw. The
event A = "get an x" is the set

  {(x,y), (x,z), (z,x), (y,x)}

and the event B = "get an x on draw 1" is the set

  {(x,y), (x,z)}

so have the following probabilities.

  P(A) = P(get an x) = 4/6 = 2/3
  P(B) = P(get an x on draw 1) = 2/6 = 1/3

The point of the rest of this note is that we want to avoid having to
explicitly list and count outcomes in order to calculate probabilities;
we want to use shortcuts instead. First, we need more vocabulary.

Complement, Intersection, Union of Events
==

The **opposite** of an event U (also called the **complement** of U) is the
event that consists of all the outcomes not in U. For two events U,V,
the event "U AND V" (also called the **intersection** of U and V) is the set
of all the outcomes shared by both U and V. The event "U OR V" (also
called the **union** of U an V) is the set of all the outcomes that are in U
or in V or both. Events U,V are called **mutually exclusive** if the event "U
AND V" is empty.

Continuing the example above, consider these events.

  C = "get an x on draw 2" = {(y,x), (z,x)}
  D = "get a y" = {(x,y), (z,y), (y,x), (y,z)}

The we have the following.

```
    (OPPOSITE of A) = {(y,z), (z,y)}
    (A AND D) = {(x,y), (y,x)}
    (C OR D) = {(x,y), (z,y), (y,x), (y,z), (z,x)}
```

The event (B AND C) is the empty set, so B,C are mutually exclusive.

Conditional Probability
======================
Suppose we are about to make a sequence of draws from a box model and we are considering the probability that the outcome will be in some event V. Using the basic definition (1), we have this.

(2) P(V) = (number of outcomes in V) / (number of total possible outcomes)

Suppose now that we are somehow guaranteed that the outcome of our draws will be in a particular event U, and that all of the outcomes in U are equally likely. This effectively reduces the pool of possible outcomes: any outcome that is not in U should not be considered in the top or the bottom of the fraction (2). We define the **conditional probability of V given U**, denoted P(V|U), to be

(3) P(V|U) = (number of outcomes in U AND V) / (number of outcomes in U).

The conditional probability P(V|U) reflects our updated assessment of the likelihood that the outcome of our sequence of draws will be in V, in the presence of the knowledge that the outcome is in U. Dividing the top and bottom of the fraction (3) by the total number of possible outcomes yields the following expression for conditional probability.

(4) P(V|U) = P(U AND V) / P(U)

Continuing the example from above, we have the following.

```
  P(A|B) = 1
  P(B|C) = 0
  P(B|A) = 1/2
  P(A|D) = 1/2
```

Independence and Dependence
============================
If P(V) = P(V|U), the events U,V are called **independent**. Otherwise, the events are called **dependent**. Using the word "independent" to describe events U,V reflects the fact that knowing whether or not an outcome is guaranteed to be in event U does not affect the way we assess the probability of event V.

Probability Rules
=================
The usefulness of the following equations is when something you want to know can be put in the form of the expressions on the left sides, and the expressions on the right sides are easier to compute. Most of the time, these rules save you from having to make explicit lists and counts of outcomes.

Rearranging (4) we have the following equation, called the **multiplication rule**.

  P(U AND V) = P(U) * P(V|U)

If U,V are independent, the multiplication rule simplifies a little.

  P(U AND V) = P(U) * P(V)   (U,V independent)

If U,V are mutually exclusive, we have the following equation, called the **addition rule for mutually exclusive events**.

  P(U OR V) = P(U) + P(V)   (U,V mutually exclusive)

A special case of the addition rule is when V is the opposite of U. In this case, (U OR V) is the set of all possible outcomes, so P(U OR V) = 1. Rearranging the addition rule terms, we have the **opposite rule**.

  P(U) = 1 - P(opposite of U)

Binomial Formula

Consider a box model in which the tickets are divided into two categories: W (for winning tickets) and L for (losing tickets). Let p be the probability that a single random draw is a winning ticket. In symbols we have the following.

    P(W) = p
    P(L) = 1 - p

Now consider a sequence of random draws, with replacement. Let n be the number of draws in the sequence (n is some positive whole number), and let S be the event

    S = "get exactly k W's and (n-k) L's in n draws"

where k is some whole number in the range from 0 to n. For example, the outcome WLWWL belongs to S for n=5 and k=3. The **binomial formula** is

    P(S) = (n choose k) * p^k * (1-p)^(n-k)

where (n choose k) is the quantity

    (n choose k) = n! / [k! * (n-k)!].

The number (n choose k), called a **binomial coefficient**, is the number of different n-letter words you can write with k W's and (n-k) L's.

For example, the number of different 5-letter words you can write with k=3 W's and n-k=2 L's is

    (5 choose 3) = 5! / [3!*2!] = 10.

Indeed, here is the list of the 10 words.

    WWWLL WWLWL WLWWL LWWWL WWLLW WLWLW LWWLW WLLWW LWLWW LLWWW

For the box with 1 W and 3 L's (so p=P(W)=1/4), the chance of getting exactly k=3 W's in n=5 draws is

    (5 choose 3) * (1/4)^3 * (3/4)^2 = (approx) 8.79%.

Here is the reasoning for why the binomial formula works, in brief.

  There are (n choose k) outcomes with k W's and (n-k) L's. All of these outcomes are mutually exclusive to one another, so the probability of getting a outcome sequence with exactly k W's in in draws is the sum of the probabilities of each individual outcome (by the addition rule). Any particular sequence of k W's and (n-k) L's has a probability of p^k times (1-p)^(n-k) (by the multiplication rule for independent events). Put all this together and you get the binomial formula.