Recall:
- A **random variable** assigns a number to each outcome in the sample space of a random process.
- A random variable is **discrete** if it can take on only isolated points on a number line, for example if it can only take on a finite number of values.
- The **probability distribution** of a discrete random variable lists all possible values of the random variable and the probability of each.

Today we study a particular kind of random variable/probability distribution that has many applications: the *binomial* distribution.

**Example 12-1: Pop Quiz!**

a) Circle your answers to the following five multiple-choice questions:

1) A  B  C
2) A  B  C
3) A  B  C
4) A  B  C
5) A  B  C

Let the random variable $X$ be the number of questions that a student (who is completely guessing at random on each question) answers correctly.

b) What are the possible values of $X$?

We can approximate the probability distribution of $X$ based on our class simulation.

c) Record the class results in the following table:

<table>
<thead>
<tr>
<th># correct</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td># students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d) Suppose that you need to get more than half of the questions right to pass. Approximate the probability that a guessing student would pass this quiz.

Now we will determine the theoretical probability distribution of $X$, one value at a time.

e) Determine the probability that a (guessing) student gets all 5 questions correct. Also identify the probability rule that you use in this calculation.
f) Determine the probability that a (guessing) student gets 0 questions correct.

g) In order to consider the probability of getting exactly one question correct, consider one particular sequence that contains 1 correct (S) and 4 incorrect (F) answers: F1F2F3S4F5. First find the probability of this particular sequence. Then determine how many such sequences (with 1 correct and 4 incorrect answers) there are. Finally, use these answers to determine the probability of answering exactly 1 question correctly.

h) Use a similar process as in g) to determine the probability that a (guessing) student answers exactly 2 questions correct.

i) Repeat this process to find the remaining probabilities, and record them in the table below:

<table>
<thead>
<tr>
<th>Number correct</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

j) Use these probabilities to determine the probability of passing the quiz (getting more than half correct).

This probability distribution is known as the **binomial distribution**. It applies to situations that consists of trials for which:

- Each trial has *two* possible outcomes (typically referred to as “success” and “failure”).
- The outcomes of the trials are *independent*.
- The probability of “success” remains *constant* on each trial (call it \( p \)).
- The random variable of interest (call it \( X \)) is the number of *successes* in a fixed number (call it \( n \)) of trials.
The probability distribution of a binomial random variable with parameters \( n \) and \( \pi \) is given by
\[
P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n.
\]

- Notation: \( X \sim \text{Binomial}(n, p) \)
- It turns out that the expected value of a binomial random variable is \( E(X) = np \) and the variance is \( \text{Var}(X) = np(1-p) \).
- You can calculate these probabilities by hand or with Minitab (Graph > Probability Distribution Plot > Binomial) or with Excel (Insert > Function > BINOMDIST) or with an applet.

k) Use Minitab, Excel, or the applet to verify the calculations above for a guessing student on the five-question, three-choice quiz. Also graph the probability distribution.

l) Now suppose that you were to take a multiple choice quiz with three options on each of 15 questions, rather than 5 questions. Let the random variable \( Y \) be the number of questions that you answer correctly. Would you expect the probability of passing (getting at least half of the questions correct) to be smaller, larger, or the same as before? Explain.

m) Does \( Y \) follow a binomial distribution? Explain, and also indicate the values of \( n \) and \( p \).

n) Use Minitab to graph the probability distribution of \( Y \) and to determine the probability that you pass this quiz. Do you need to re-think your answer to l)? Explain.

One of the primary uses of probability is to assess whether a result is statistically significant.
- A statistically significant result is one that is unlikely to occur by chance alone.

Example 12-2: Naughty vs. Nice
We all recognize the difference between naughty and nice, right? What about children less than a year old: Do they recognize the difference and show a preference for nice over naughty? In a study reported in the November 2007 issue of *Nature*, researchers investigated whether infants take into account an individual’s actions towards others in evaluating that individual as appealing or aversive (Hamlin, Wynn, and Bloom, 2007). In one component of the study, 10-month-old infants were shown a “climber” character (a piece of wood with “google” eyes glued onto it) that could not make it up a hill in two tries. Then they were alternately shown two scenarios for the climber’s next try, one where the climber was pushed to the top of the hill by another character (“helper”) and one where the climber was pushed back down the hill by another character (“hinderer”). The infant was alternately shown these two scenarios several times. Then the child was presented with both pieces of wood (the helper and the hinderer) and asked to pick one to play with. The researchers found that the 14 of the 16 infants chose the helper over the hinderer.
Suppose for the moment that the researchers’ conjecture is wrong, and infants actually have no preference for either type of toy.

a) Is it possible, even if infants actually have no genuine preference, that 14 out of 16 infants in the study would have chosen the helper toy just by chance?

b) Do you think it would be very surprising, if infants actually have no genuine preference, that 14 out of 16 infants in the study would have chosen the helper toy just by chance?

The key question here is to determine whether the observed result is very surprising under the assumption that infants actually have no preference. (We will call this assumption of no genuine preference the null model.) We could answer this question in two ways:

- first with simulation, and then
- with an exact probability calculation.

c) Describe how you could simulate the infants’ selections, assuming that infants have no genuine preference for either toy.

d) Now let the random variable \( X \) represent the number of infants who would choose the helper toy, if in fact infants have no genuine preference for either toy. Describe the probability distribution of \( X \).

e) Determine the probability, assuming that infants have no genuine preference, that 14 or more infants in the group of 16 would have chosen the helper toy.

- This probability is called a p-value. A p-value is the probability of obtaining a result at least as extreme as the one observed, assuming that there is no genuine preference/difference (i.e., assuming the null model is true).
- A small p-value casts doubt on the null model/hypothesis used to perform the calculation (in this case, that infants have no genuine preference).
  - A p-value of .10 or less is generally considered to provide some evidence against the null model/hypothesis.
  - A p-value of .05 or less is generally considered to provide fairly strong evidence against the null model/hypothesis.
  - A p-value of .01 or less is generally considered to provide very strong evidence against the null model/hypothesis.
f) Is this probability small enough to consider the actual result obtained by the researchers surprising, assuming the null model that infants have no preference and so choose blindly?

g) Would you conclude that the experimental data obtained by the researchers provide strong evidence that infants in general have a genuine preference for the helper toy over the hinderer toy? Explain the reasoning process behind your answer.

Example 12-3: Flat Tire?
A legendary story on college campuses concerns two students who miss a chemistry exam because of excessive partying but blame their absence on a flat tire. The professor allowed them to take a make-up exam, and he sent them to separate rooms to take it. The first question, worth five points, was quite easy. The second question, worth ninety-five points, asked: Which tire was it? I will ask each of you to indicate which tire you would pick. Do not confer with anyone else before answering.

a) Which tire would you pick?

b) Name the tire that I predict to be the most popular choice.

c) Record the counts for the class data below.

<table>
<thead>
<tr>
<th>Left front</th>
<th>Left rear</th>
<th>Right front</th>
<th>Right rear</th>
</tr>
</thead>
</table>


d) The null model asserts that my conjecture is wrong and there is nothing special about the right front tire, so it is equally likely to be picked as any other tire. Let the random variable X be the number of students in our class who would choose the right front tire. Under the null model of “nothing special” about this tire, describe the probability distribution of X.

f) Indicate how to calculate the p-value, and then use Minitab, Excel, or an applet to calculate it.

g) Is this p-value small enough that the class data provide fairly strong or very strong evidence against the null model?