The topic of probability often receives little emphasis in the math classes that are part of the program for biologists -- until they are required to take a serious course in statistical methods. At that point, the focus may be on the rules and formal tools for generating a statistical result rather than on making sense of what probability and statistics are telling us. This is extremely unfortunate, since the concept of probability is fundamental in a variety of situations that are of extreme importance to many biological professionals. In this essay we discuss briefly why we are interested in probability in the context of a physics class, why researchers (of any ilk) need to know probability, and why medical professionals need to understand probability.

# The basic idea of probability

The basic idea of probability is about situations that occur multiple times but have factors that we cannot control. When you flip a coin, the laws of Newtonian physics could tell you where it is going to go – and whether it will land heads or tails. That is, if you knew the initial position, upward speed, initial angular orientation, and rotational velocity to a very high accuracy. AND if you knew that the coin were a fair coin (perfectly symmetric and balanced), AND that there was no breeze, etc., etc.

Even in the context of systems that are well-described by Newtonian physics, there are many systems that we cannot predict well. Their motion is just too sensitive to factors that we cannot control.

In such a situation, what do we do? We might just give up and say: “that’s not predictable,” but another approach has developed, driven by mathematicians responding to questions from gamblers in the 17th century. (Really.) In this approach we carry out the following two steps:

1. Determine what results are equally probable;
2. Count the number of ways that a result we are looking for can be made up of the different equally probable results.

For example, consider throwing two cubical dice, each with 6 sides and the sides having 1, 2, 3, 4, 5, and 6 spots respectively. When the dice are thrown so that they bounce around in uncontrollable ways, one result comes up on each. The total will range from 2 (a one comes up on each) to 12 (a six comes up on each). But each total is not equally probable – each face of each die is assumed to be equally probable. As a result there is only one way to create the result of “2” – each die has to show one spot.  But there are six ways to create a total of “7” – 1+6, 2+5, 3+4, 4+3, 5+2, and 6+1, with the first number showing the result on the first die, the second the result on the second. This means, that *if we throw the dice many times we expect to get the result 7 six times as often as the result 2.*Understanding this ratio is crucial is you are going to not lose too much money playing dice!

Note a few key ideas:

* The result given by a probabilistic law does NOT tell you what will happen in any given experiment (trial); it will only what will tell you if you REPEAT the experiment many times. And then it will only tell you what fraction of the time you can expect different results.
* The states that are the result of our experiment do not specify every variable. There are “hidden” uncontrolled variables that we do not specify
* The model we have of the system is crucial – what are the hidden variable states (microstates) that are equally probable, and how many different ways can a result state (macrostate) be made up from different hidden variable states.

So the very nature of the “law” we are creating is different from many of the laws we are accustomed to learning in science classes – at least in the intro classes. They only tell the result of many equivalent experiments – an *ensemble* – not of an individual one.