

6.4: Types of probability

Introduction

By **probability**, we mean a number that quantifies the uncertainty associated with a particular event or outcome. If an outcome is certain to occur, the probability is equal to 1; if an outcome cannot happen, the probability is equal to zero. Most events we talk about in biology have probability somewhere in between.

A basic understanding, or at least an appreciation for probability is important to your education in biostatistics. Simply put, there is no certainty in biology beyond simple expectations like Benjamin Franklin's famous quip about "death and taxes..."

Discrete probability

You are probably already familiar with **discrete probability**. For example, what is the probability that a single toss of a fair coin will result in a "heads?" The outcomes of the coin toss are discrete, categorical, either "heads" or "tails."

Note:

Obviously, this statement assumes a few things — coin tossed not in a vacuum; although possible, we ignore the possibility of the coin landing on edge.

And for ten tosses of the coin? And more cogently, what is the probability that you will toss a coin ten times and get all ten "heads?" While different in tone, these are discrete outcomes. An important concept is independence. Are the multiple events independent? In other words, does the toss of coin on the first attempt affect the toss of the coin on the second attempt and so on up to the tenth toss? At least in principle, the repeated tosses are independent, so to find the probability you just multiply each event's probability to get the total. In contrast, if one or more events are not independent but somehow influence the behavior of the next event, then you add the probabilities for each dependent event. We can do better than simply multiply or add events one at a time; depending on the number of discrete outcomes, it is very likely that someone has already calculated all possible outcomes and come up with an equation. In the case of the tossing of two coins, this is a binomial equation problem and repeat tosses can be modeled by use of the Bernoulli distribution.

Now, try this on for size. What is the probability that the next child born at Kapiolani Medical Center in Honolulu will be assigned female?

We just described a discrete random variable, which can only take on discrete or "countable" numbers. This distribution of values is the **probability mass function**. The probability of one fatal airline accident in a year exactly on 20.1 is practically zero (the area under a point along the curve is zero), so we can get the probability of a range of values around the point as our answer.

Continuous probability

Many events in biology are of degree, not kind. It is kind of awkward to think about it, but for a sample of adult house mice drawn from a population, what is the probability of obtaining a mouse that is 20.0000 grams (g) in weight? Each possible value of body mass for a mouse is considered an event, just like in our example of tossing a coin. But clearly, we don't expect to get a lot of mice that are exactly 20.0000 g in weight. For variables like body mass, the type of data we collect is continuous, and the probability values need to be rethought along a continuum of possible values and, in turn, how likely each value is for a mouse. Although it is theoretically possible that a mouse could weigh ten pounds, we know by experience that this is impossible. Adult mice weigh between 15 and 50 g or thereabouts.

We just described a continuous random variable, which can take on any value between a specific interval of values. This distribution of values is the **probability density function**. The probability of a mouse's weight falling exactly on 20.0 is practically zero (the area under a point along the curve is zero), so we can get the probability of a range of values around the point as our answer.

In statistical inference, following our measurements of the variables from our sample drawn from a population, we make conclusions with the following kind of caveat: "the mean body mass for this strain of mouse is 20 g." That is our best estimate of the mean (middle) for the population of mice, more specifically, for the body mass of the mice. Here, the variable body mass is more formally termed a **random variable**. This implies that there is in fact a true population mean body mass for the mice and that any deviations from that mean are due to chance. In statistics we don't settle for a single point estimate of the population mean. You will find that most reporting of estimates of random variables is accompanied by a statement like "the mean was 20 g with a

95% probability that the true population mean is between 18.9 and 21 g.” This is called the 95% **confidence interval** for the mean and it takes into account how good of an estimate our sample is likely to be relative to the true population value. Not only are we saying that we think the population mean is 20 g, but we’re willing to say that we are 95% certain that the true value must be between a lower limit (18.9 g) and an upper limit (21 g). In order to make this kind of statement, we have to assume a distribution that will describe the probability of mouse weights. For many reasons we usually assume a normal distribution. Once we make this assumption we can calculate how probable a particular weight is for a mouse.

We introduced how to calculate Confidence Intervals in [Chapter 3.4](#) and will extend this in [Chapter 7.6](#) and [Chapter 8.6](#).

Types of probability

To begin refining our concept of probability, it is sometimes useful to distinguish among kinds of probabilities:

- between **theoretical** and **empirical**;
- between **subjective** and **objective**.

In most cases, including your statistics book, we would begin our discussion of probability by talking of some probabilities for events we’re familiar with.

1. The theoretical probability of heads appearing the next time you flip a fair coin is $1/2$ or 50%. As long as we’re talking about a fair coin, the probability of a heads appearing each time you flip the coin remains 50%. We can check this by conducting an experiment: out of 10 tosses, how many heads appear? The answer would be an empirical probability, and we understand the chance in an objective manner (no interpretation needed).
2. The theoretical probability that a “5” will appear on the face of a fair die after a toss is $1/6$ or 16.667%. Again, as long as we’re talking about a fair, standard 6-sided die, the probability of a “5” appearing each time you roll the dice remains 16.667%.
3. The probability that at birth, a human baby’s sex will be male is about $1/2$ or 50%. This is an empirical probability based on millions of observations. Changes in technology and ethical standards notwithstanding, the probability will remain the same.
4. The probability of the birth of a Downs syndrome baby is $1/800$, but increases with age of the mother until by age 45, the chance is $1/12$. Again, these are empirical and objective.
5. The probability of winning the Publisher’s Clearing House Sweepstakes is about 1 in 100 million. This probability is theoretical, it is also objective; however, by adding lots of twists to the game, by having multiple opportunities and by giving the appearance that a person must purchase a magazine, some players perceive their chances as increasing or decreasing by their efforts (=subjective).

R and distributions

R Commander: Distributions menus give four options

- Quantiles
- Probabilities
- Plot
- Sampling

Questions

1. Define and distinguish, with examples
 1. discrete and continuous probability
 2. theoretical and empirical probability
 3. subjective and objective probability

This page titled [6.4: Types of probability](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.