

10.6: Probability Distributions

A *probability distribution* describes the probability of all of the possible outcomes in an experiment. For example, on Jan 20 2018, the basketball player Steph Curry hit only 2 out of 4 free throws in a game against the Houston Rockets. We know that Curry's overall probability of hitting free throws across the entire season was 0.91, so it seems pretty unlikely that he would hit only 50% of his free throws in a game, but exactly how unlikely is it? We can determine this using a theoretical probability distribution; during this course we will encounter a number of these probability distributions, each of which is appropriate to describe different types of data. In this case, we use the *binomial* distribution, which provides a way to compute the probability of some number of successes out of a number of trials on which there is either success or failure and nothing in between (known as "Bernoulli trials") given some known probability of success on each trial. This distribution is defined as:

$$P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (10.6.1)$$

This refers to the probability of k successes on n trials when the probability of success is p . You may not be familiar with $\binom{n}{k}$, which is referred to as the *binomial coefficient*. The binomial coefficient is also referred to as "n-choose-k" because it describes the number of different ways that one can choose k items out of n total items. The binomial coefficient is computed as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (10.6.2)$$

where the explanation point (!) refers to the *factorial* of the number:

$$n! = \prod_{i=1}^n i = n * (n-1) * \dots * 2 * 1 \quad (10.6.3)$$

In the example of Steph Curry's free throws:

$$P(2; 4, 0.91) = \binom{4}{2} 0.91^2 (1 - 0.91)^{4-2} = 0.040 \quad (10.6.4)$$

This shows that given Curry's overall free throw percentage, it is very unlikely that he would hit only 2 out of 4 free throws. Which just goes to show that unlikely things do actually happen in the real world.

10.3.1 Cumulative probability distributions

Often we want to know not just how likely a specific value is, but how likely it is to find a value that is as extreme or more than a particular value; this will become very important when we discuss hypothesis testing in a later chapter. To answer this question, we can use a **cumulative probability distribution**; whereas a standard probability distribution tells us the probability of some specific value, the cumulative distribution tells us the probability of a value as large or larger (or as small or smaller) than some specific value.

In the free throw example, we might want to know: What is the probability that Steph Curry hits 2 or fewer free throws out of four, given his overall free throw probability of 0.91. To determine this, we could simply use the binomial probability equation and plug in all of the possible values of k and add them together:

$$\begin{aligned} P(k \leq 2) &= P(k=2) + P(k=1) + P(k=0) \\ &= 6e^{-5} + 0.002 + 0.040 \\ &= 0.043 \end{aligned}$$

In many cases the number of possible outcomes would be too large for us to compute the cumulative probability by enumerating all possible values; fortunately, it can be computed directly. For the binomial, we can do this in R using the `pbinom()` function:

Table 10.1: Cumulative probability distribution for number of successful free throws by Steph Curry in 4 attempts.

numSuccesses	probability
--------------	-------------

numSuccesses	probability
0	0.00
1	0.00
2	0.04
3	0.31
4	1.00

From the table we can see that the probability of Curry landing 2 or fewer free throws out of 4 attempts is 0.043.

This page titled [10.6: Probability Distributions](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.