Getting to know rbinom

SMWadgymar

2022-08-16

A probability distribution is a mathematical description of the possible outcomes of a random variable. There are many types of distributions that have been characterized mathematically and are useful to us when analyzing data. In our simulations above, each flip of a coin was considered a draw, and the outcome of the flip was random. Our simulations used fair coins, which have a 50% chance of coming up heads or tails. These outcomes follow a binomial distribution.

The binomial distribution is the distribution that describes the frequency of outcomes for each set of trials where there are exactly two possible outcomes for each trial (e.g. heads or tails, survives or dies, etc.). The notation used to describe the binomial distribution, B(n,p), includes two parameters: n is the size (# flips) and p is the probability that each is heads (for now, 0.5). A variable is binomially distributed if (1) the trials are independent, (2) the number of trials is fixed, (3) each trial outcome can be classified as success or failure, and (4) the probability of success is the same for each trial. This is one of many distributions and we will touch on a few distributions (not all by any means!) in class.

R has many built-in functions that allow us to draw samples from well-characterized distributions. We'll make use of three of those functions for the binomial distribution here: 'rbinom', 'dbinom', and 'pbinom'. Importantly, we will make use of these functions to examine probabilities in two ways. First, we'll run simulations, as we did before. Simulations are digital experiments. We use them to conduct large, replicated experiments that would not be possible in real life, and as in real life, we want our outcomes to be the product of random chance. The goal is to run enough simulations that the probabilities that we estimate are likely very close approximations of the true probabilities (which are often unknown). For our second approach, we will calculate exact probabilities. We can do this because we are drawing our samples from a well-characterized distribution where the exact probabilities are known.

####Getting to know the 'rbinom' function

1) Let's simulate a random draw from a binomial distribution. The rbinom function has three arguments: the number of random draws, the number of "coins" being flipped on each draw, and the probability of a heads as the outcome. Heads will be assigned a value of 1 and tails a value of 0.

```
rbinom(1, 1, 0.5)
```

```
## [1] 1
```

Q: What does the output here mean?

A: The output was a single number, reflecting that you flipped one coin one time and it had a 50% chance of landing on heads (1) or tails (0).

2) How would you adjust the code above to simulate ten flips of a single coin?

```r
rbinom(10, 1, 0.5)
```

```
##  [1] 1 1 1 0 0 1 1 1 0 1
```

Q: What does the output here mean?

A: The output was a single set of 10 numbers, reflecting that you drew 10 coins and flipped each one once, with each having a 50% chance of landing on heads (1) or tails (0).

3) How about one draw of ten coins?

```r
rbinom(1, 10, 0.5)
```

```
## [1] 9
```

Q: What does the output here mean?

A: The output was a single number, reflecting the number of times your single draw of 10 coins landed on heads (1). To calculate the number that landed on tails, you subtract the value you get here from the total number of coins you drew in this trial.

4) How about ten draws with each having 10 coins?

```r
rbinom(10, 10, 0.5)
```

```
##  [1] 7 5 2 2 4 5 6 2 6 1
```

Q: What does the output here mean?

A: The output included ten values, possibly ranging from 0 to 10. Each of these values reflects the number of times a single draw of 10 coins landed on heads (1). There were ten draws of ten coins each, resulting in ten total counts in the outpout. The results of each draw vary because the outcome of each coin flip is random.

5) What would the outcome of 10 trials of 10 coins each be if you were using a biased coin that gave you an 80% chance of producing heads instead of a 50% chance?

```r
rbinom(10, 10, 0.8)
```

```
##  [1]  5  8  9  8  9 10 10  9 10  9
```

Q: How does the output of this line of code differ from the output you saw in the previous example that used a fair coin? Did you expect these differences?

A: The interpretation of the output is the same as in the last question, except the coins are biased towards producing a head 80% of the time instead of 50% of the time. As a result, the total number of heads is much higher across these trails than it was with a fair coin.