Finding probabilities

SMWadgymar

2022-08-12

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.

####Finding probabilities through simulation

6) Simulate 100000 draws each with 10 fair coins and save the outcome as 'flips'.

```
flips<-rbinom(100000, 10, 0.5)
```

Q: How many values are stored in the variable called flips? What is your interpretation of each value?

A: Flips has 10000 values, and each value reflects the number of flips out of 10 that landed on heads (i.e. that equal 1).

7) Let's find the fraction of draws where 5 heads occurred.

```
mean(flips==5)
```

```
## [1] 0.2463
```

Q: Which do you expect will be lower: the probability that 3 out of 10 flips will be heads or the probability that 5 out of 10 flips will be heads? Why?

A: With a fair coin, you would expect that half of the flips being heads would be the most common outcome. The further you deviate from half of the flips being heads, the lower the probability of that outcome.

8) Next, calculate the probability that 3 out of 10 flips will be heads. This reflects the probability that this outcome will occur, and is called the density of this distribution at the point where x=3.

```
mean(flips==3)
```

```
## [1] 0.1163
```

Q: How does the probability that 3/10 flips will be heads compare to the probability that 5/10 flips will be heads?

A: It is lower, as I expected.

9) Let's find the cumulative probability that heads will occur 4 or fewer times during each draw. This is referred to as the cumulative probability because it sums up the probability that the number of heads is 0, 1, 2, 3, or 4 out of 10 flips.

```
mean(flips<=4)
```

```
## [1] 0.37776
```

Q: How would you adjust this line of code to estimate the cumulative probability that heads will occur 6 or more times?

A:

```
mean(flips>=6)
```

```
## [1] 0.37594
```

####Finding the exact probabilities

10) The dbinom function estimates the exact probability density at a given point. The arguments here are the density being estimated, the number of coins, and the probability of producing a head. Note that the order and type of arguments used here do not match the order and type of arguments that you used for the 'rbinom' function.

```
dbinom(5, 10, 0.5)
```

```
## [1] 0.2460938
```

Q: How does this command compare to the steps you took for question 6 and 7 above? Are the estimates comparable? Why or why not?

A: The values are extremely similar and only start to differ at the thousandths decimal place. The reason they are very slighly different is that our initial simulation had a fixed number of trials (100000) to estimate the exact probability whereas the dbinom function returns the exact probability.

11) Now let's estimate the exact cumulative probability density for a given range. The arguments here are teh density being estimated, the number of coins, and the probability of producing a head. Note that the order and type of arguments used here do not match the order and type of arguments that you used for the 'rbinom' function.

```r
pbinom(4, 10, 0.5)
```

```
## [1] 0.3769531
```

Q: How does this command compare to the steps you took for question 9? Are the estimates comparable? Why or why not?

A: Again, this output is extremely similar as what we saw with mean(flips<=4). As with dbinom, pbinom returns an exact probability (in this case, a cumulative probability including all outcomes of 4 or less).