Hypothesis Testing: *p*-values

Hypothesis Testing

Recall: The idea behind a hypothesis test is that we want to see if it is likely that something observed in our sample will also be present in the larger population.

The setup is to start by a skeptical null hypothesis H_0.

So in our coin flipping example, we will assume that the coin is fair. That is, the probability of it landing on heads is 0.5.

Then, we look at the evidence and see how unusual our observation is. If what we observed is very unlikely under the assumption of the null hypothesis, then we will reject it in favor of the alternative.

p-values

A *p*-value is a way to quantify how unusual what we observe is, under the null hypothesis.

To calculate a *p*-value, we need to choose some statistic that we can calculate from our sample. In this case, it makes sense to look at the number of times the coin landed on heads.

The *p*-value in this case will measure how likely it is, under the assumptions of the null hypothesis, to see a sample number of heads at least as extreme as what we observed in our sample.

p-values

To determine whether or not to reject the null hypothesis, we must establish a threshold for how extreme our observation is. This threshold is called the **significance level**.

Traditionally, the significance level used has been 0.05, meaning that if we calculate a *p*-value less than 0.05, we will reject the null hypothesis.

The significance level determines the chance of a Type I error (incorrectly rejecting the null hypothesis) in the event that the null hypothesis is true.

p-values

In the case of coin flips, we know what the exact data generation process would be under the assumptions of the null hypothesis.

Namely, we would have a binomial distribution.

Under the assumptions of the null hypothesis, the binomial distribution would have $p = 0.5$ and $n = 100$.

If we are trying to show that the coin is not a fair coin, then our alternative hypothesis is that $p \neq 0.5$.

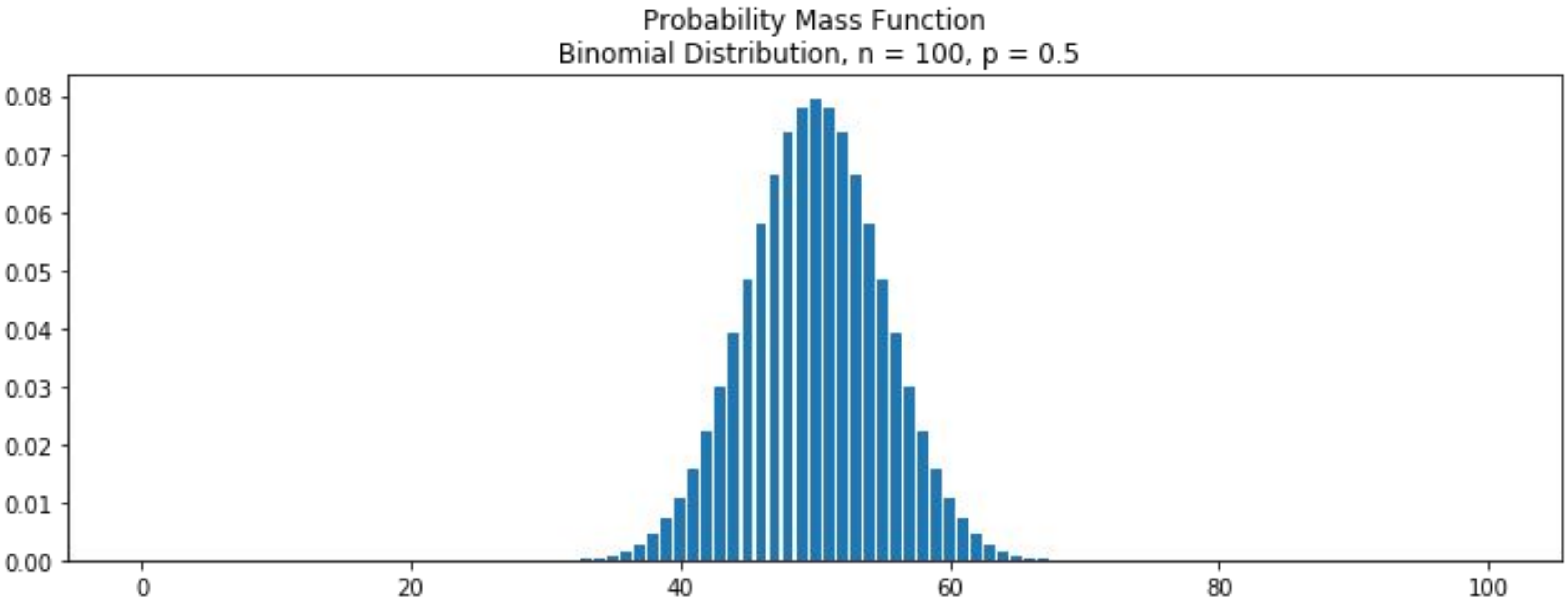
Example: Scenario 1

Recall that in scenario 1, we flipped the coin 100 times, and it lands on heads 47 out of those 100 times.

Let’s look at the probability mass function for the the corresponding binomial distribution.

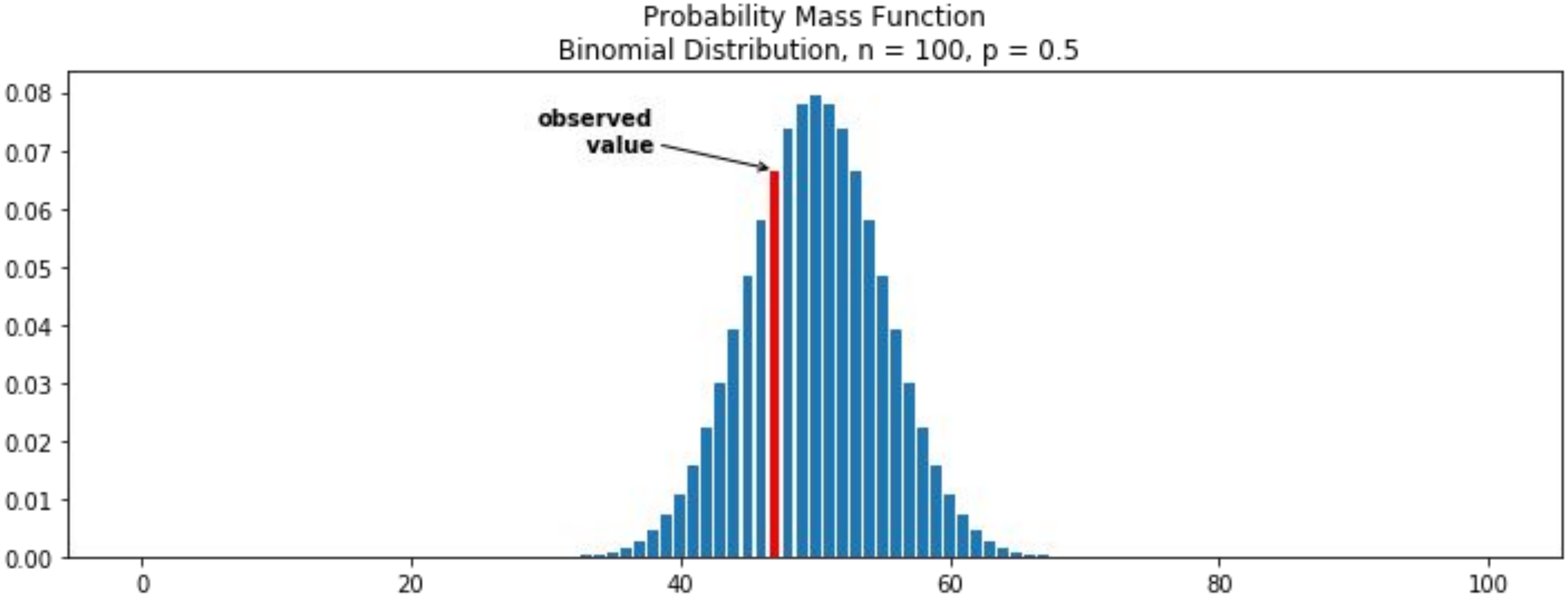
Example: Scenario 1

First, we assume the null hypothesis and see what the pmf looks like in this universe.



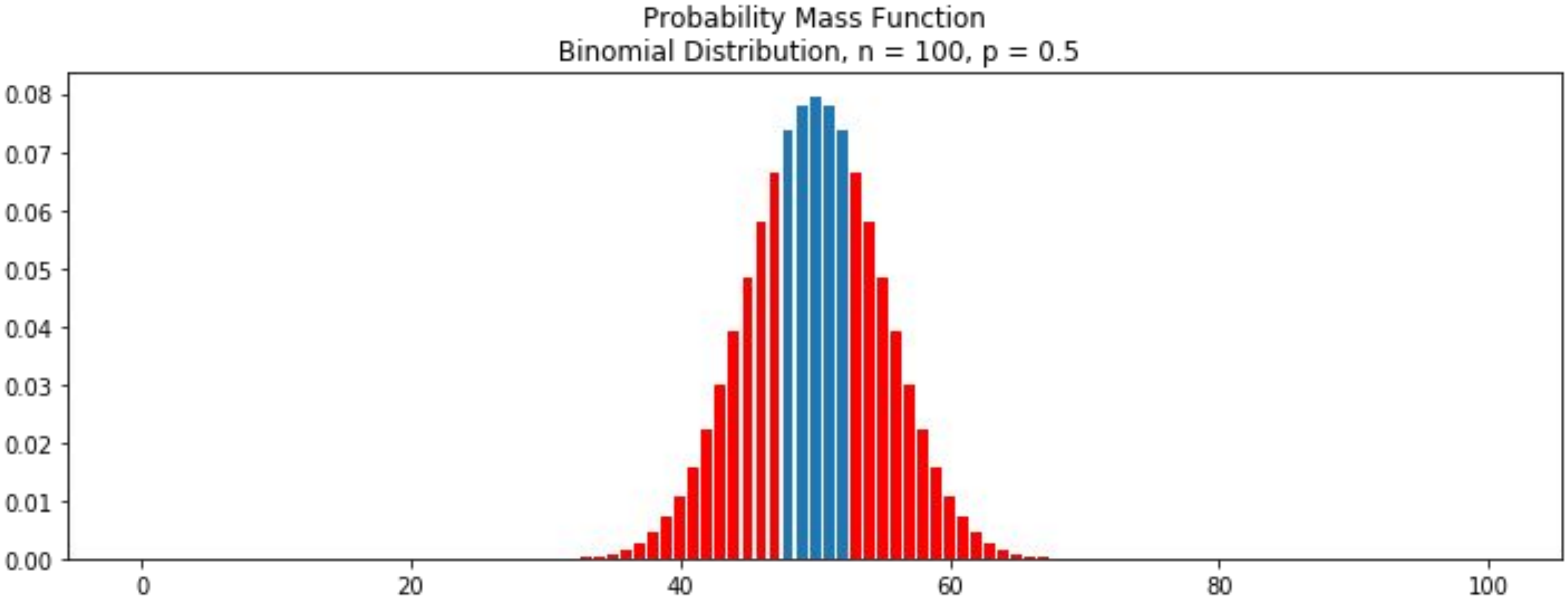
Example: Scenario 1

Let’s see where our observed value lands.



Example: Scenario 1

And then let's look at all the possible values that are at least as extreme as what we observed. That is, cases where we get no more than 47 heads, or 53 or more heads.



Example: Scenario 1

The observations which would have been at least as extreme as what we observed are flipping 47 or less heads or 53 or more.

How likely are those outcomes? To answer this, we can use the cumulative distribution function.

The resulting probability is approximately 0.617.

That means that getting at least as extreme an observation happens more than 60% of the time. This is not very compelling evidence, so in this scenario, we will not reject the null hypothesis.

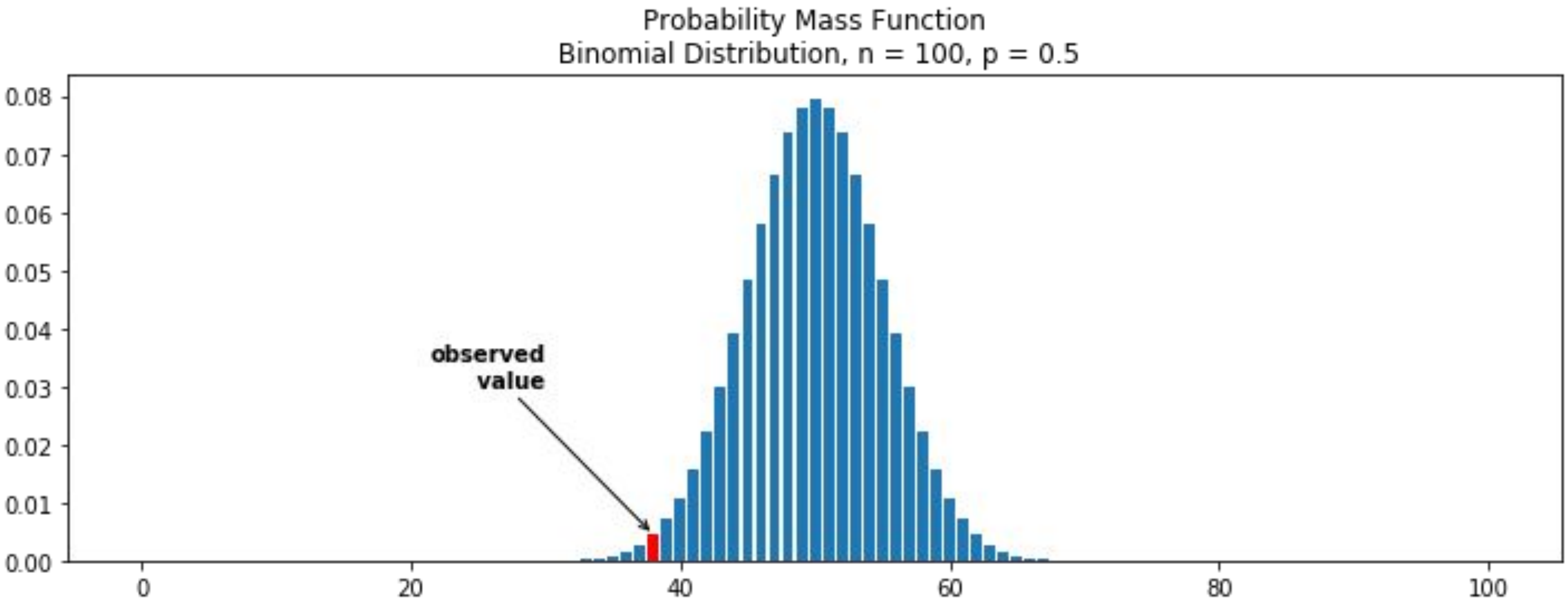
Example: Scenario 2

Recall that in scenario 1, we flipped the coin 100 times, and it lands on heads 38 out of those 100 times.

Let’s look at the probability mass function in this scenario.

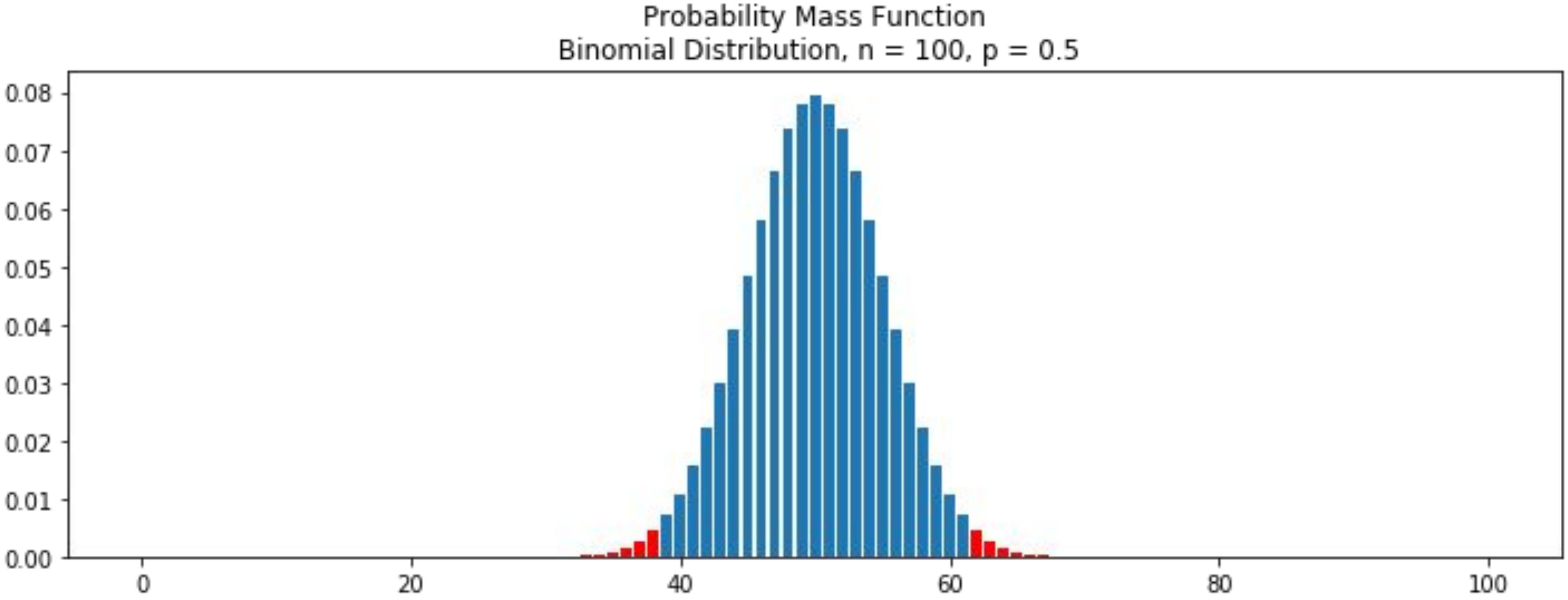
Example: Scenario 2

Let’s see where our observed value lands.



Example: Scenario 2

And then let's look at all the possible values that are at least as extreme as what we observed. That is, cases where we get no more than 38 heads, or 62 or more heads.



Example: Scenario 2

Using the cumulative distribution function, we can see that the probability of observations at least as extreme is only 0.021.

This probability is below the threshold value of 0.05, so in this case, we can reject the null hypothesis.

It seems unlikely that the extremeness of our observation was due only to random chance.

Hypothesis Testing

To recap:

- Create your null and alternative hypothesis:
  - Null hypothesis, H_0 is the skeptical view
  - Alternative hypothesis, H_A is that the effect you are testing is present in the population
- Assume that the null hypothesis is true, and choose a statistic to calculate
- Determine/estimate how your chosen statistic is distributed under the null hypothesis
- Calculate how often you would see a sample statistic as extreme or more extreme than the one you observed