Module 6: Introduction to Probability

California State University - Northridge

Ovande Furtado, Jr., Ph.D.

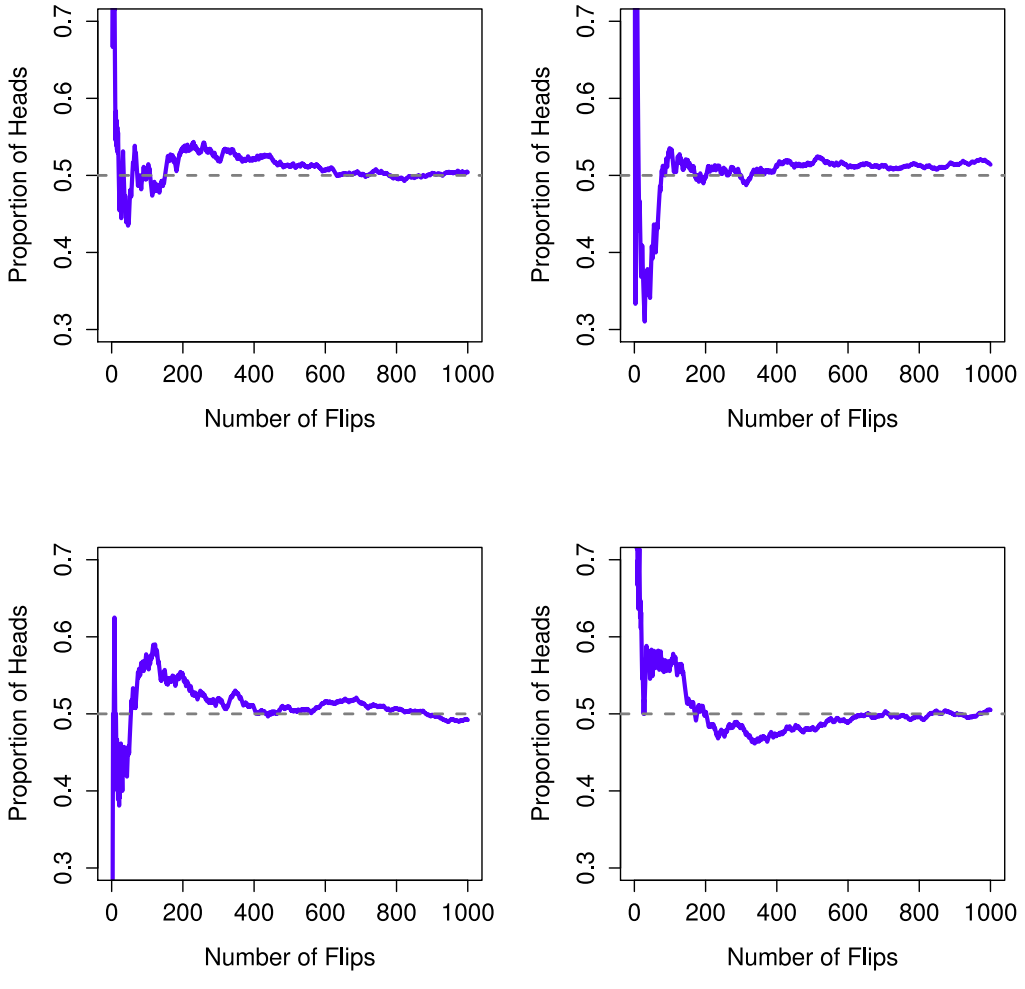
22 February, 2021

CSUN | CALIFORNIA STATE UNIVERSITY NORTHRIDGE

What is probability?

The frequentist view

- more dominant approach in statistics and **defines probability as a long-run frequency**
- flip a fair coin over and over again, and as *N* grows large (approaches infinity, denoted $N \rightarrow \infty$) the proportion of heads will converge to 50%.
- the proportion of heads fluctuates wildly, starting at 0.00 and rising as high as 0.80.
- with more and more of the values actually being pretty close to the “right” answer of 0.50



What is probability?

The Bayesian view

- is often called the subjectivist view
- an event is seen as the **degree of belief** that an intelligent and rational agent assigns to that truth of that event
- probabilities don’t exist in the world but rather in the thoughts and assumptions of people and other intelligent beings

Problem: how to operationalize “degree of belief”?

Example: If someone offers me a bet that if it rains tomorrow then I win $5, but if it doesn’t rain I lose $5. Clearly, from my perspective, this is a pretty good bet. On the other hand, if I think that the probability of rain is only 40% then it’s a bad bet to take. So we can operationalise the notion of a “subjective probability” in terms of what bets I’m willing to accept.

Probability and inferential stats

Do 7-9 year-old boys perform better than girls on the skill of kicking a stationary ball[1]

To answer the question above, we can randomly draw a sample from the population (N = 100), and using the tools of **probability**, try to infer what we learn from the sample to the population. This is called **inferential statistics**.

[1] At this age level, there is no reason for boys and girls perform differently in kicking

Statistics and probability

Probability theory is “the doctrine of chances”. It’s a branch of mathematics that tells you how often different kinds of events will happen. Questions one may ask:

- What are the chances of a fair coin coming up heads 10 times in a row?
  - if we know the coin is far - 50% chance a flip will result in “heads”
- If I roll a six sided dice twice, how likely is it that I’ll roll two sixes?
  - chance of rolling a 6 on a single die is 1 in 6
- probabilistic questions start with a `known model` of the world

Statistics and probability, cont.

- Model: *P*(heads) = 0.5
  - the probability of “heads” is 0.5[1]
- In `probability theory`, the model is known but the data are not

1. The probability of "heads" is known if ones flips a fair coin. But the data does not exist, at least one collects the data.

- In `statistics` we do not know the truth about the world. We have the data and we want to learn the truth about the world

1. If my friend flips a coin 10 times and gets 10 heads are they playing a trick on me?
2. Data suggest 7-9-year-old boys perform better than girls on the skill of kicking. Is it true in the population?

[1] probability ranges from 0 - 1

Probability distributions

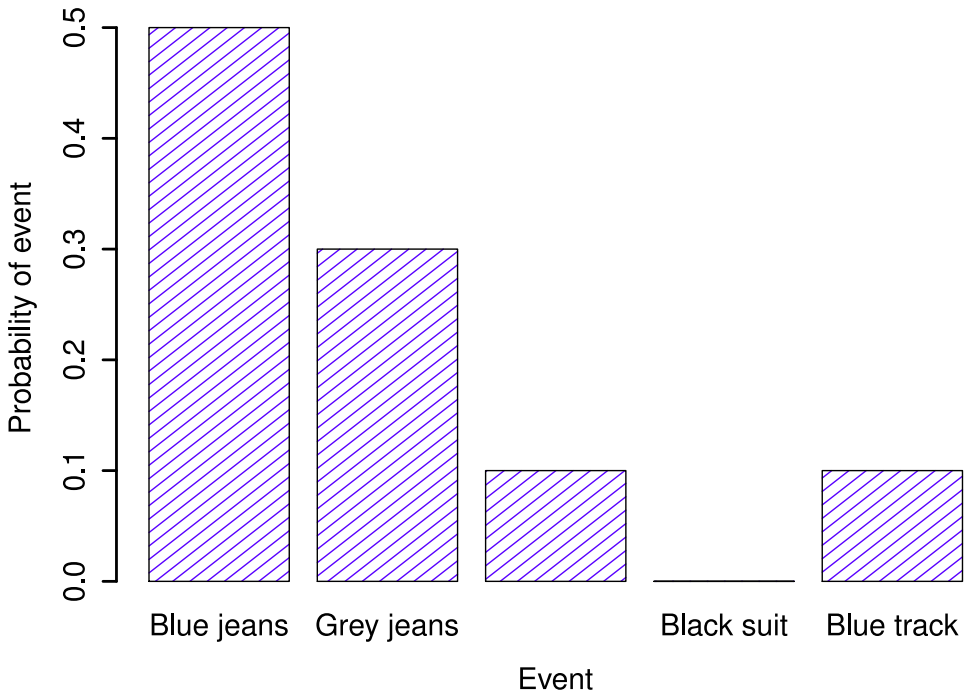
- elementary event (X) --> Pair of trousers
  - the outcome of any observation (i.e., put on a pair) will be one and only one of these events
- The set of possible events is called **sample space**
  - assigning probability to events:
    - the bigger the value of P(X), the more likely the event is to occur
    - P(X) = 0 it means the event X is impossible
    - P(X) = 1 it means that event X is certain to occur
    - P(X) = 0.5 it means that event X occurs half of the time.

Which trousers?	Label	Probability
Blue jeans	X1	P(X1) = 0.5
Grey jeans	X2	P(X2) = 0.3
Black jeans	X3	P(X3) = 0.1
Black suit	X4	P(X4) = 0
Blue tracksuit	X5	P(X5) = 0.1

Each of the events has a probability that lies between 0 and 1, and if we add up the probability of all events they sum to 1

The probabilities of the elementary events need to add up to 1, and this is known as the **law of total probability**.

How about distributions?



Visual depiction of the “trousers” probability distribution. There are five “elementary events”, corresponding to the five pairs of trousers that I own. Each event has some probability of occurring: this probability is a number between 0 to 1. The sum of these probabilities is 1.

===>

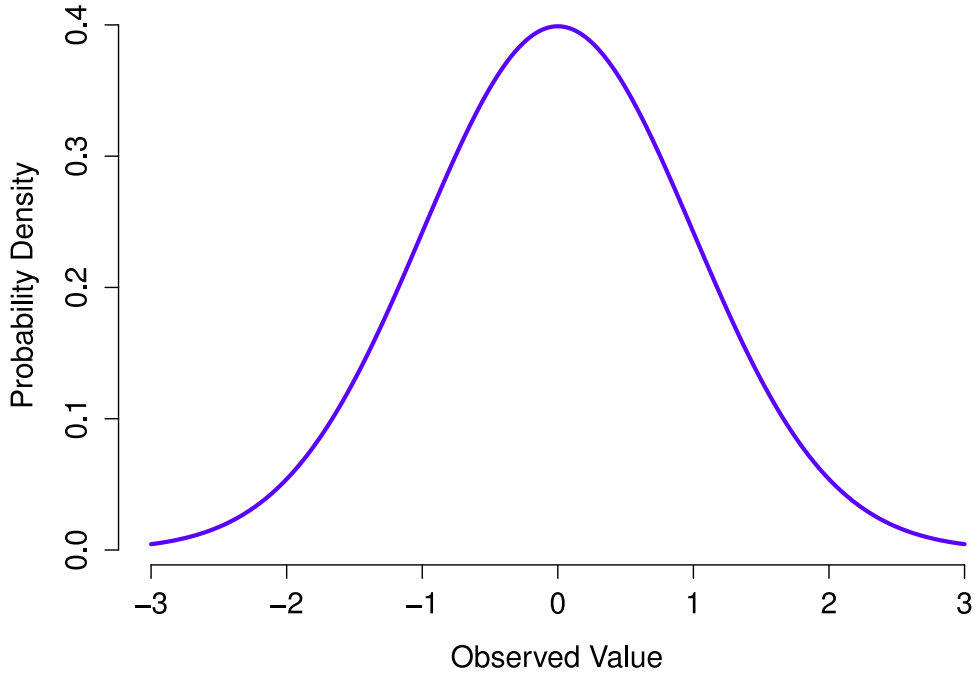
more about distribution on the next section...

The Normal Distribution

- is described using two parameters: the mean of the distribution μ and the standard deviation of the distribution σ.
- The notation that we sometimes use to say that a variable X is normally distributed is as follows:

```
X ~ Normal(μ, σ)
```

- smooth curve instead of “histogram-like” bars - continuous data
- different from the binomial distribution - Discrete variables allow only 3, 4, 5, not 3.8



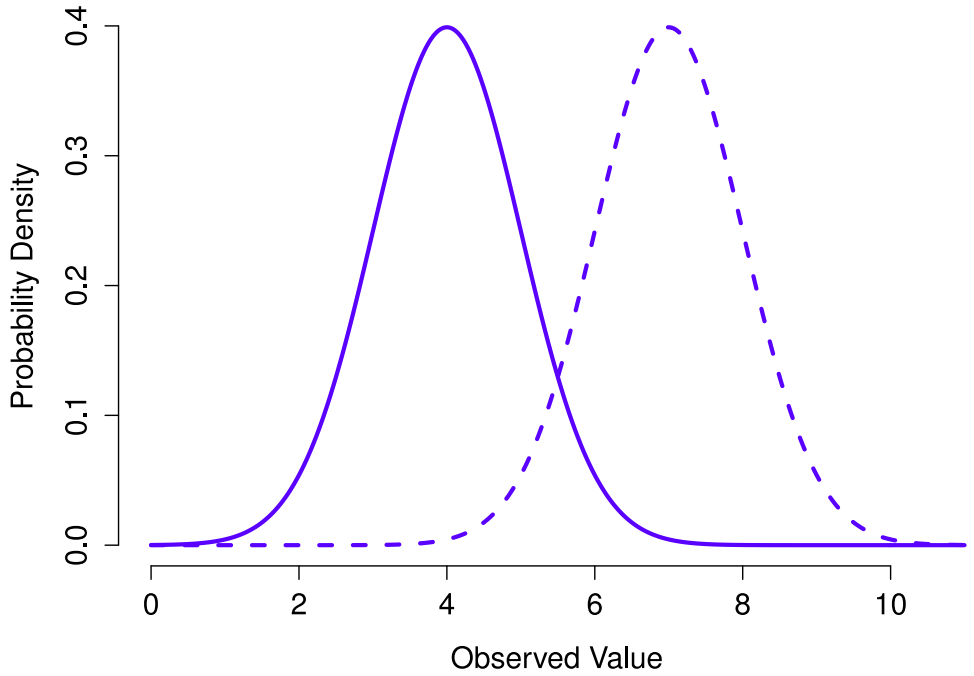
The normal distribution with mean μ = 0 and standard deviation σ = 1. The x-axis corresponds to the value of some variable, and the y-axis tells us something about how likely we are to observe that value.

The Normal Distribution

Illustration of what happens when you change the mean of a normal distribution. The solid line depicts a normal distribution with a mean of μ = 4.

The dashed line shows a normal distribution with a mean of μ = 7. In both cases, the standard deviation is σ = 1.

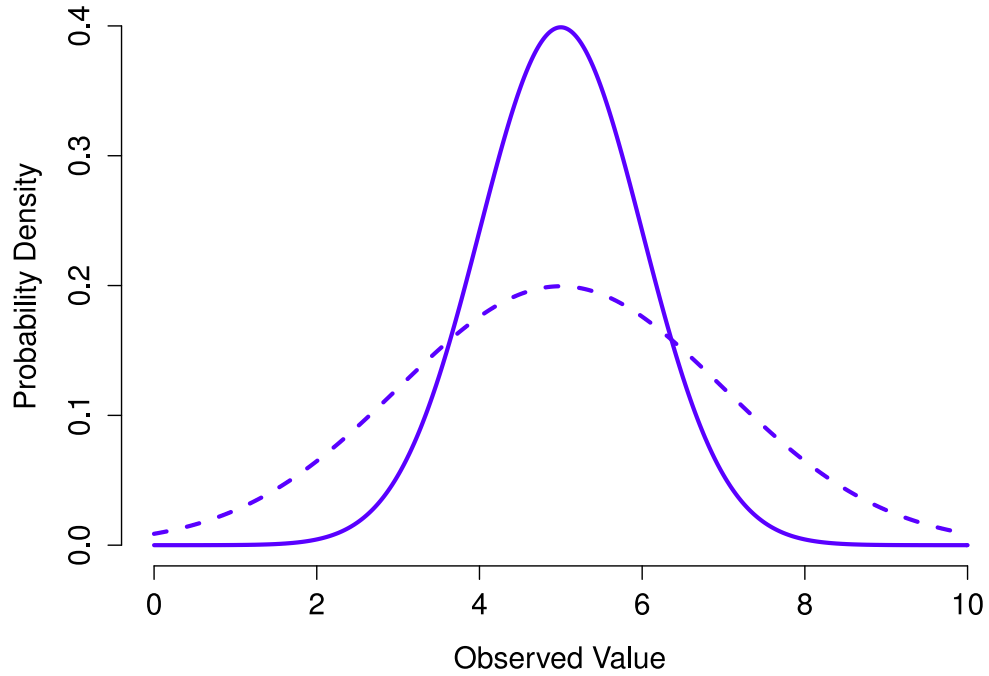
Not surprisingly, the two distributions have the same shape, but the dashed line is shifted to the right.



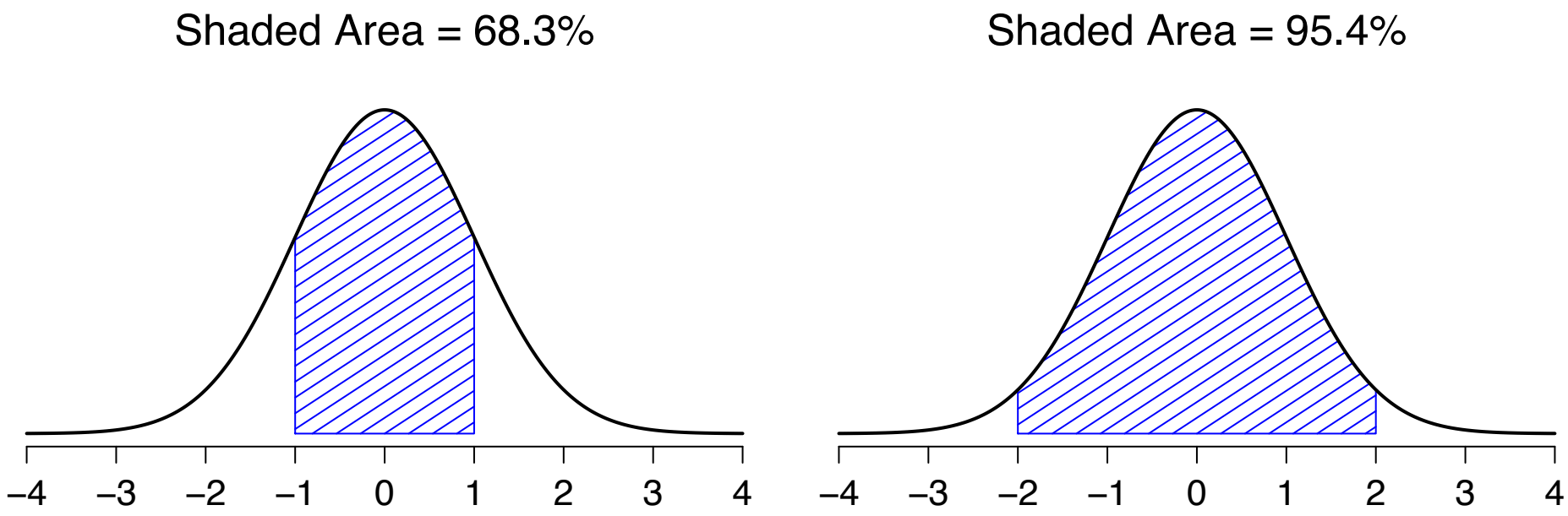
The Normal Distribution

Illustration of what happens when you change the standard deviation of a normal distribution. Both distributions plotted in this figure have a mean of μ = 5, but they have different standard deviations. The solid line plots a distribution with standard deviation σ = 1, and the dashed line shows a distribution with standard deviation σ = 2.

As a consequence, both distributions are “centred” on the same spot, but the dashed line is wider than the solid one.

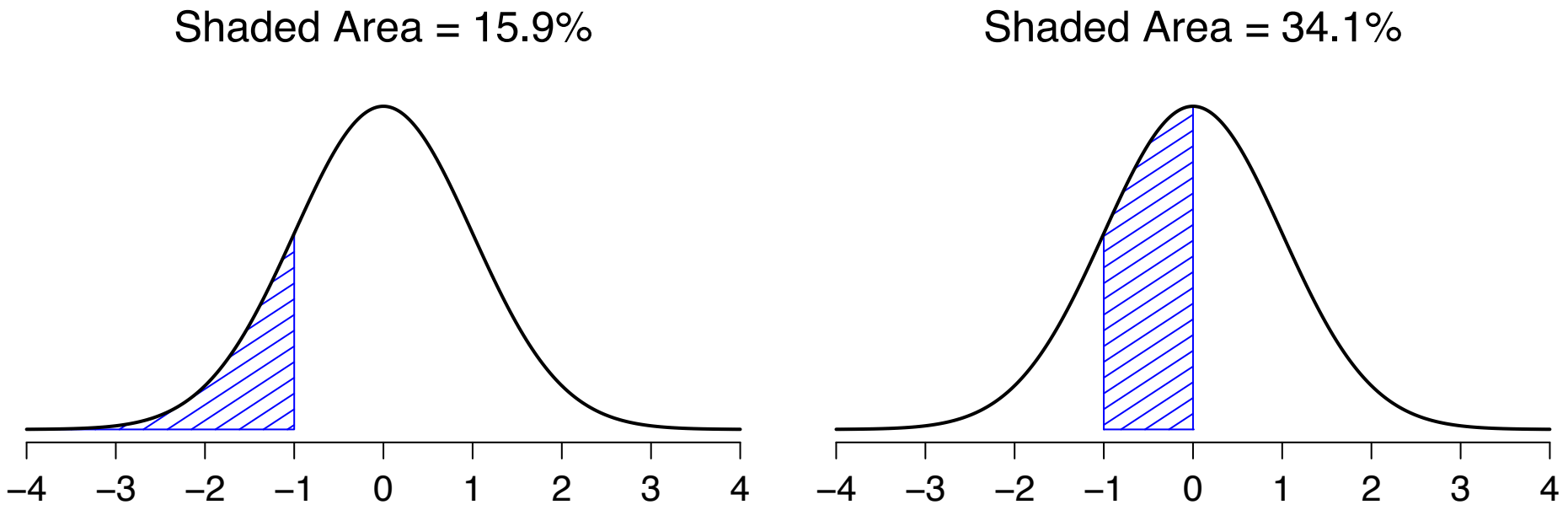


The Normal Distribution



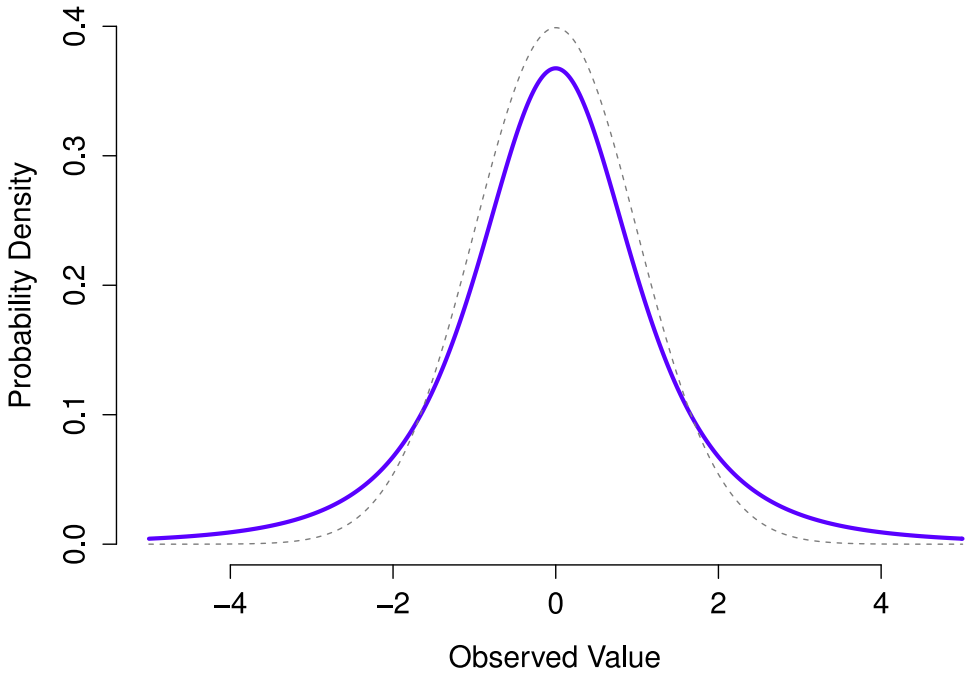
The area under the curve tells you the probability that an observation falls within a particular range. The solid lines plot normal distributions with mean μ = 0 and standard deviation σ = 1. The shaded areas illustrate “areas under the curve” for two important cases. In panel a, we can see that there is a 68.3% chance that an observation will fall within one standard deviation of the mean. In panel b, we see that there is a 95.4% chance that an observation will fall within two standard deviations of the mean.

The Normal Distribution



Two more examples of the “area under the curve” idea. There is a 15.9% chance that an observation is one standard deviation below the mean or smaller (left panel), and a 34.1% chance that the observation is somewhere between one standard deviation below the mean and the mean (right panel). Notice that if you add these two numbers together you get 15.9% + 34.1% = 50%. For normally distributed data, there is a 50% chance that an observation falls below the mean. And of course that also implies that there is a 50% chance that it falls above the mean.

The *t*-Distribution



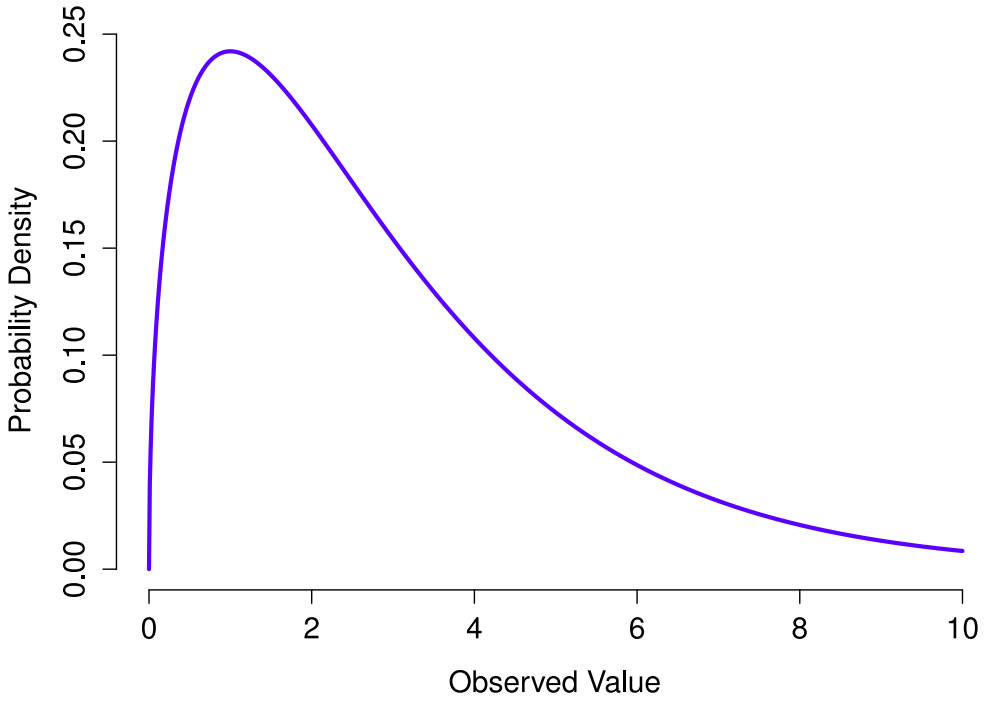
t-distribution with 3 degrees of freedom (solid line). It looks similar to a normal distribution, but it's not quite the same. For comparison purposes I've plotted a standard normal distribution as the dashed line.

The t-distribution is a continuous distribution that looks very similar to a normal distribution. Note that the "tails" of the t-distribution are "heavier" (i.e., extend further outwards) than the tails of the normal distribution).

That's the important difference between the two. This distribution tends to arise in situations where you think that the data actually follow a normal distribution, but you don't know the `mean` or `standard deviation`.

We'll run into this distribution again in chapter Comparing two means.

The $\tilde{\chi}^2$-Distribution

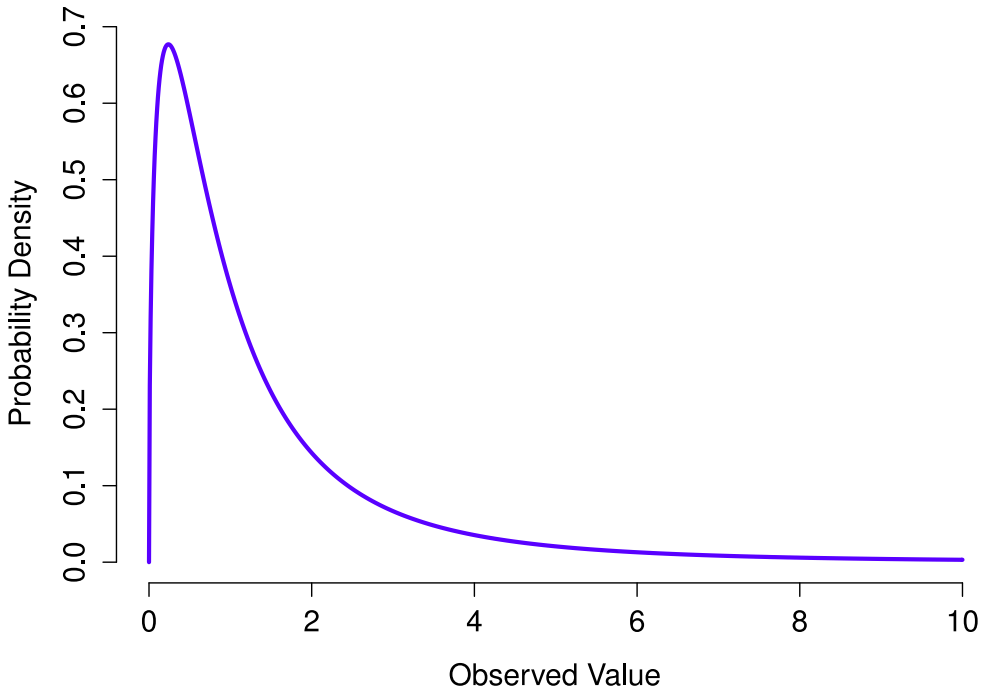


χ²-distribution with 3 degrees of freedom. Notice that the observed values must always be greater than zero, and that the distribution is pretty skewed. These are the key features of a χ²-distribution.

The χ²-distribution is is typically used with Categorical data analysis.

The χ²-distribution turns up all over the place is that if you have a bunch of variables that are normally distributed, square their values and then add them up (a procedure referred to as taking a “sum of squares”), this sum has a χ²-distribution

The F-distribution



F-distribution with 3 and 5 degrees of freedom. Qualitatively speaking, it looks pretty similar to a χ^2-distribution, but they're not quite the same in genera.

The F-distribution looks a bit like a χ^2-distribution, and it arises whenever you need to compare two χ^2-distributions to one another. Remember that χ^2 turns out to be the key distribution when we're taking a "sum of squares"? Well, what that means is if you want to compare two different "sums of squares", you're probably talking about something that has an F-distribution.

Later in chapter Comparing several means (one-way ANOVA), we'll run into the F-distribution.

Thanks!

CSUN CALIFORNIA STATE UNIVERSITY NORTHRIDGE

ovandef@csun.edu

@ofurtado

@drfurtado