Week 3: Probability Theory

POP88162 Introduction to Quantitative Research Methods

Tom Paskhalis

Department of Political Science, Trinity College Dublin

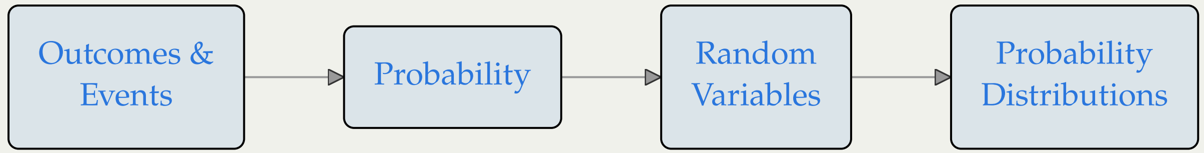
So Far

- Quantitative research involves collecting *data* - a *sample* of observations selected from a larger *population*, in which one or more *variables* are measured for each *observation*.
- The goal of collecting data is usually to calculate *statistics* which can be used to infer *parameters* of a population.
- Variables can be measure on different *scales*, which determine which statistics are applicable.
- *Measures of central tendency* describe a typical observation.
- *Measures of variability* describe the spread of the variable.

Topics for Today

- Probability
- Random variables
- Probability distributions
- Normal distribution

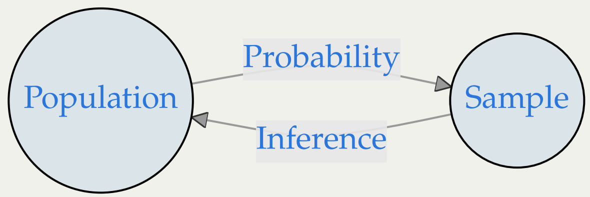
Today's Plan


Outcomes & Events
Probability
Random Variables
Probability Distributions

Probability

Why Probability?

- **Probability** describes the uncertainty about our sample.
- **Inference** (i.e. statistical inference) allows us to make conclusions about the population.



Origins of Probability Theory

Georges de La Tour, Louvre

Example: Sortition

Imagine a world in which political candidates are selected by lot (sortition).

Three Parties:

- Left Party 🛠️
- Right Party 🏛️
- Green Party 🌳

Candidates can be of two genders:

- Female ♀️
- Male ♂️

Six possible candidates:

- 👩‍🔧 (♀️🛠️)
- 👨🏿‍⚕️ (♂️🛠️)
- 👩🏾‍⚖️(♀️🏛️)
- 👨🏻‍🏫 (♂️🏛️)
- 👩🏽‍🌾 (♀️🌳)
- 👨🏾‍🎨 (♂️🌳)

Example Continued: Sortition

- Hypothetical trial: roll a 🎲 to pick a candidate.
- Uncertainty: we don't know which candidate will be selected.
- One possible outcome:

  picking 👩🏽‍🌾 (♀️🌳)
- Sample space S:

  {👩🏻‍🔧, 👨🏿‍⚕️, 👩🏾‍⚖️, 👨🏻‍🏫, 👩🏽‍🌾, 👨🏿‍🎨}
- An event A:

  selecting a ♀️
- Any of these outcomes would make us say that an event A has occurred:

  {👩🏻‍🔧, 👩🏾‍⚖️, 👩🏽‍🌾}

What is Probability?

- Probability $P(A)$ represents how likely is an event A to occur.
- If all outcomes are equally likely:

$$P(A) = \frac{\text{Number of elements in A}}{\text{Number of elements in } S}$$

- Sortition example:
  - Probability of selecting ♀️: $\frac{3}{6} = \frac{1}{2}$
  - Probability of selecting 🏛️: $\frac{2}{6} = \frac{1}{3}$

The Basics of Probability

- Probability is a property of **events**.
- The same event can occur when different outcomes are observed.
- One outcome is a draw from all possible outcomes (**sample space**).
- We are not interested in events per se, but the properties of the events.
- The more individual events we observe, the closer our estimates are to population parameters.

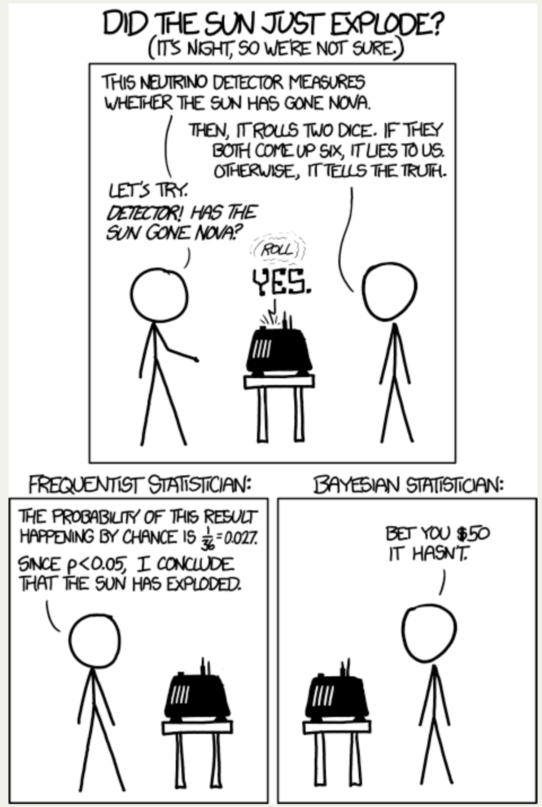
Approaches to Probability

- **Frequentist**: long-run frequency over a large number of repeated events.
- **Bayesian**: degree of belief about the event in question.
- In frequentist view population parameters are fixed (but unknown).
- In Bayesian view population parameters are themselves **random variables**.
- In the rest of this class we will focus only on frequentist approach.

Extra

Statistical Rethinking by Richard McElreath

Frequentists vs Bayesians



xkcd

Extra

Critique of this cartoon by Andrew Gelman

Probability Axioms

- Probabilities are always non-negative:
  - $P(A) \geq 0$ for any event A
- Probabilities of all possible outcomes add up to 1:
  - $P(S) = 1$
- If two events A and B are mutually exclusive:
  - $P(A \text{ or } B) = P(A) + P(B)$

Some Properties of Probability

- Probability of the complement:
  - $P(A^c) = P(\text{not } A) = 1 - P(A)$
  - E.g. $P(\text{not }$ 🏛️$) = 1 - P($🏛️$) = 1 - \frac{1}{3} = \frac{2}{3}$
  - "Probability of not selecting a candidate from the Right Party is $\frac{2}{3}$"
- General addition rule:
  - $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - E.g.
    $P(♀ \text{ or }$ 🌳$) = P(♀) + P($🌳$) - P(♀ \text{ and }$ 🌳$) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{3+2-1}{6} = \frac{4}{6} =$
  - "Probability of selecting a woman or a Green Party candidate is $\frac{2}{3}$"

Non-Naive Definition of Probability

- The definition of probability from above:

$$P(A) = \frac{\text{Number of elements in A}}{\text{Number of elements in } S}$$

is actually rather naive.

- There two big problems with it:
  - All outcomes are assumed to be equally likely.
  - All outcomes have to be listed.
- More generally, we can call probability any function that maps events to a real number between 0 and 1.

Random Variables

Random Variables

- How do we map the possible outcomes of sortition to numbers in our data?
- Using **random variables**.
- Consider the sample space:

  {👩🏻‍🔧, 👨🏿‍⚕️, 👩🏾‍⚖️, 👨🏻‍🏫, 👩🏽‍🌾, 👨🏿‍🎨}
- Let Y be the selection of a ♀️ candidate:

  Y(👩🏻‍🔧) = Y(👩🏾‍⚖️) = Y(👩🏽‍🌾) = 1

  Y(👨🏿‍⚕️) = Y(👨🏻‍🏫) = Y(👨🏿‍🎨) = 0
- These 0's and 1's are what we actually see in our data.
- In other words, random variable Y provides the numerical summary of the candidate draw with our question (selection of a ♀️ candidate) in mind.
- The source of randomness is that we don't know which candidate will be selected.

Random Variables Continued

- Imagine that instead of being interested of the selection of a ♀️ candidate we are interested in selection of a 🌳 candidate.
- We have the same sample space: {👩🏻‍🔧, 👨🏿‍⚕️, 👨🏾‍⚖️, 👩🏻‍🏫, 👩🏽‍🌾, 👨🏿‍🎨}
- But another random variable X maps the same outcomes differently than Y:

  X(👩🏻‍🔧) = X(👨🏾‍⚖️) = X(👨🏿‍⚕️) = X(👩🏻‍🏫) = 0

  X(👩🏽‍🌾) = X(👨🏿‍🎨) = 1
- Alternatively, rather than focussing on 🌳, we may choose the variable X to map the selection of a candidate from any party such that:

  X(👩🏻‍🔧) = X(👨🏿‍⚕️) = 1 for 🛠️

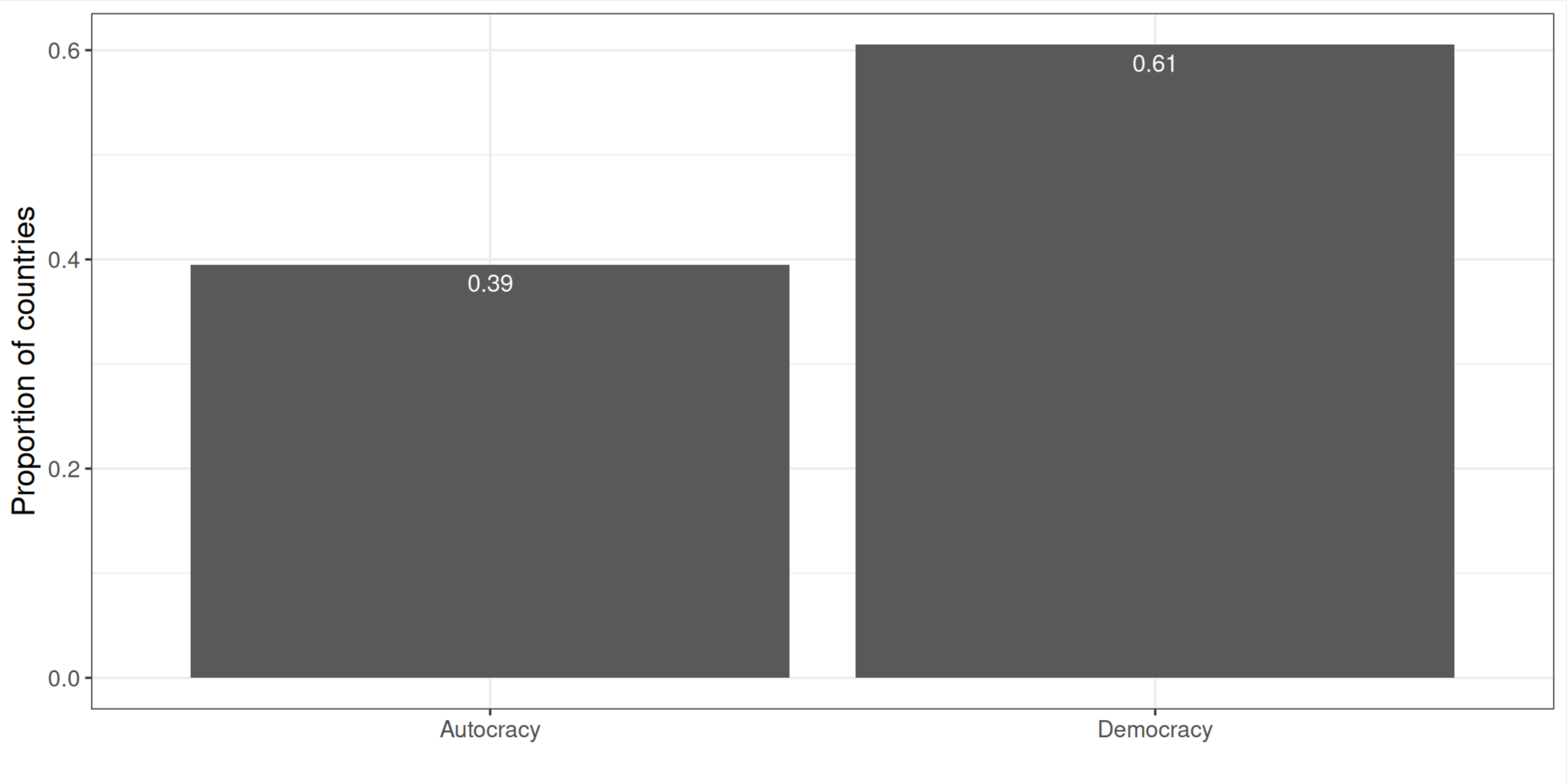
  X(👨🏾‍⚖️) = X(👩🏻‍🏫) = 2 for 🏛️

  X(👩🏽‍🌾) = X(👨🏿‍🎨) = 3 for 🌳
- Note that since these all are categorical variables the actual numbers assigned by random variables are somewhat arbitrary.

Discrete and Continuous Random Variables

- Imagine we observe one event.
- We want to know what is the probability that the random variable associated with this event takes on a certain value.
- But that depends on what are the potential values that this random variable can take.
- This hinges on whether the measurement scale is *discrete* or *continuous*.
- Probability works slightly differently for them.

Example: Discrete Random Variable

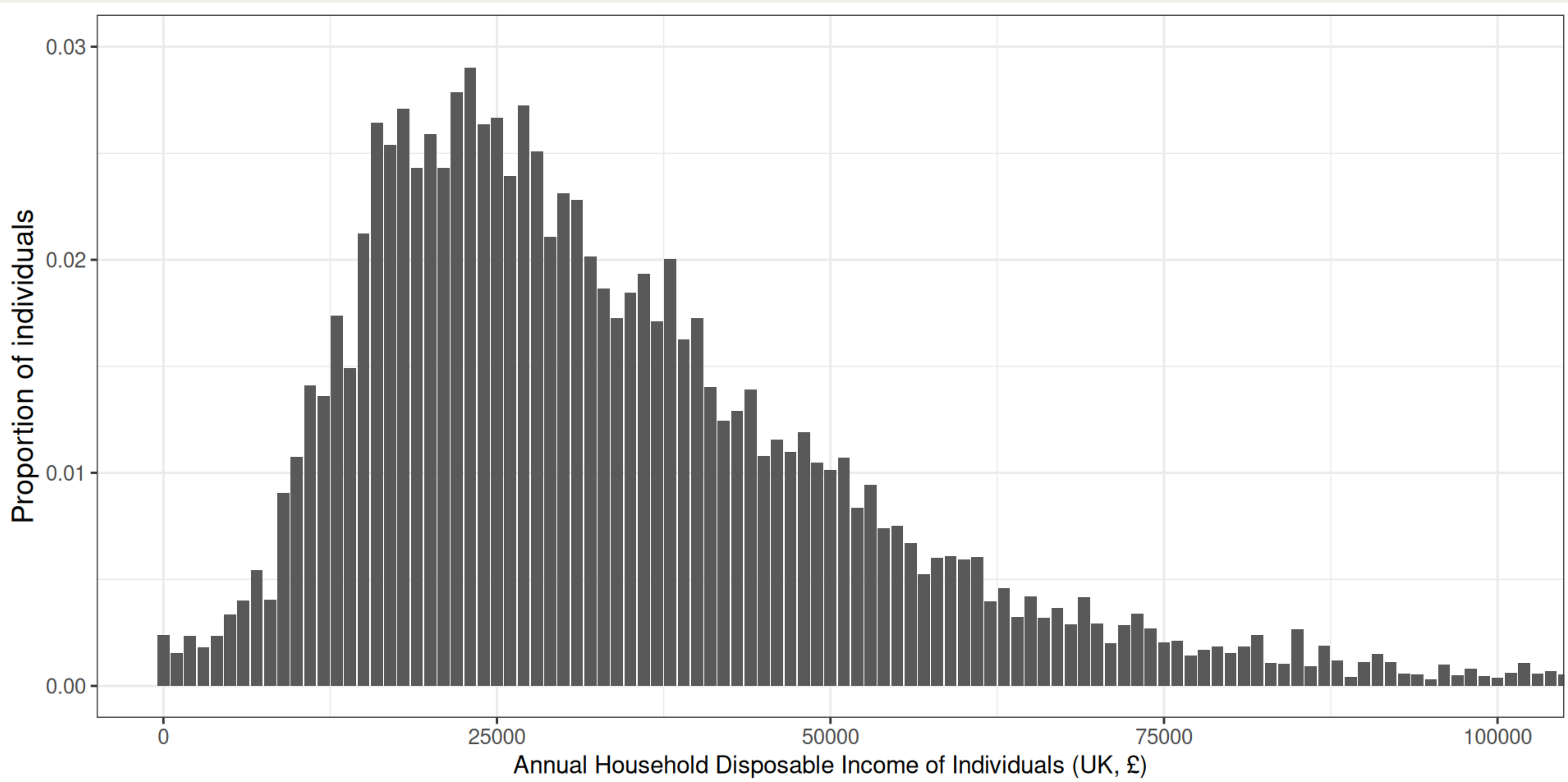


Discrete Random Variables

- A random variable that takes on a countable number of values.
- Describes the data measured on nominal and ordinal scales.
- **Probability distribution** of a discrete random variable assigns probability to each possible value of the variable.
  - E.g. $P(🌳) = \frac{1}{3}$
- We can write out all the individual probabilities for such variables:

Party	$P(Y)$
🛠️	0.33
🏛️	0.33
🌳	0.33

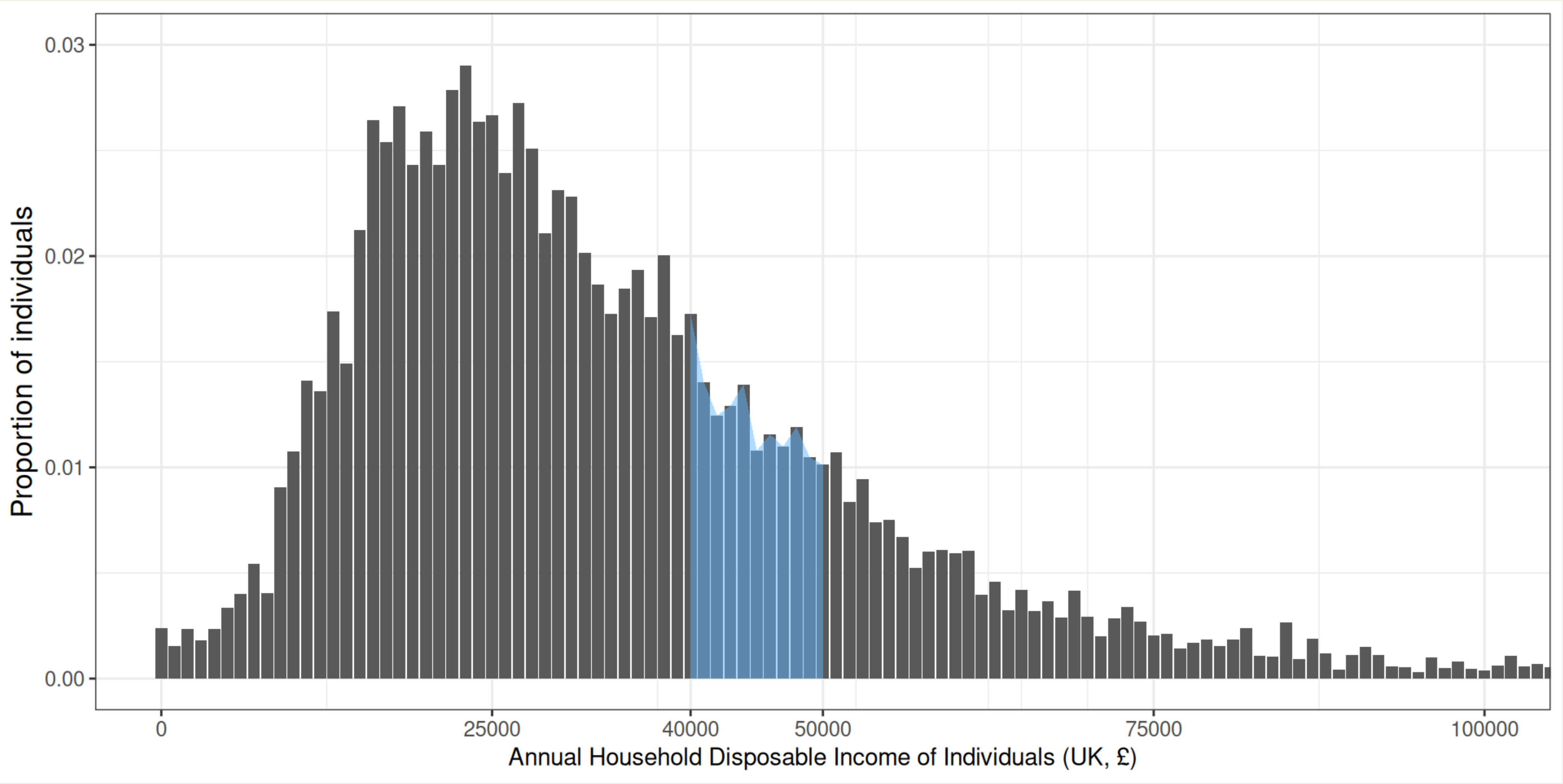
Example: Continuous Random Variable



Continuous Random Variables

- What is the probability that someone's income is exactly £$39,674.39$?
- For specific values it's always 0.
- Continuous random variables take an infinite number of possible values.
- Probability distribution for continuous variables assigns probabilities for *intervals*.
- So, we can calculate the probability that someone's income, for example, is between £$40,000$ and £$50,000$ or $> 30,000$.
- Those are defined with formulas and involve calculus, but we will use R instead (more in the workshop).

Example: Continuous Random Variable

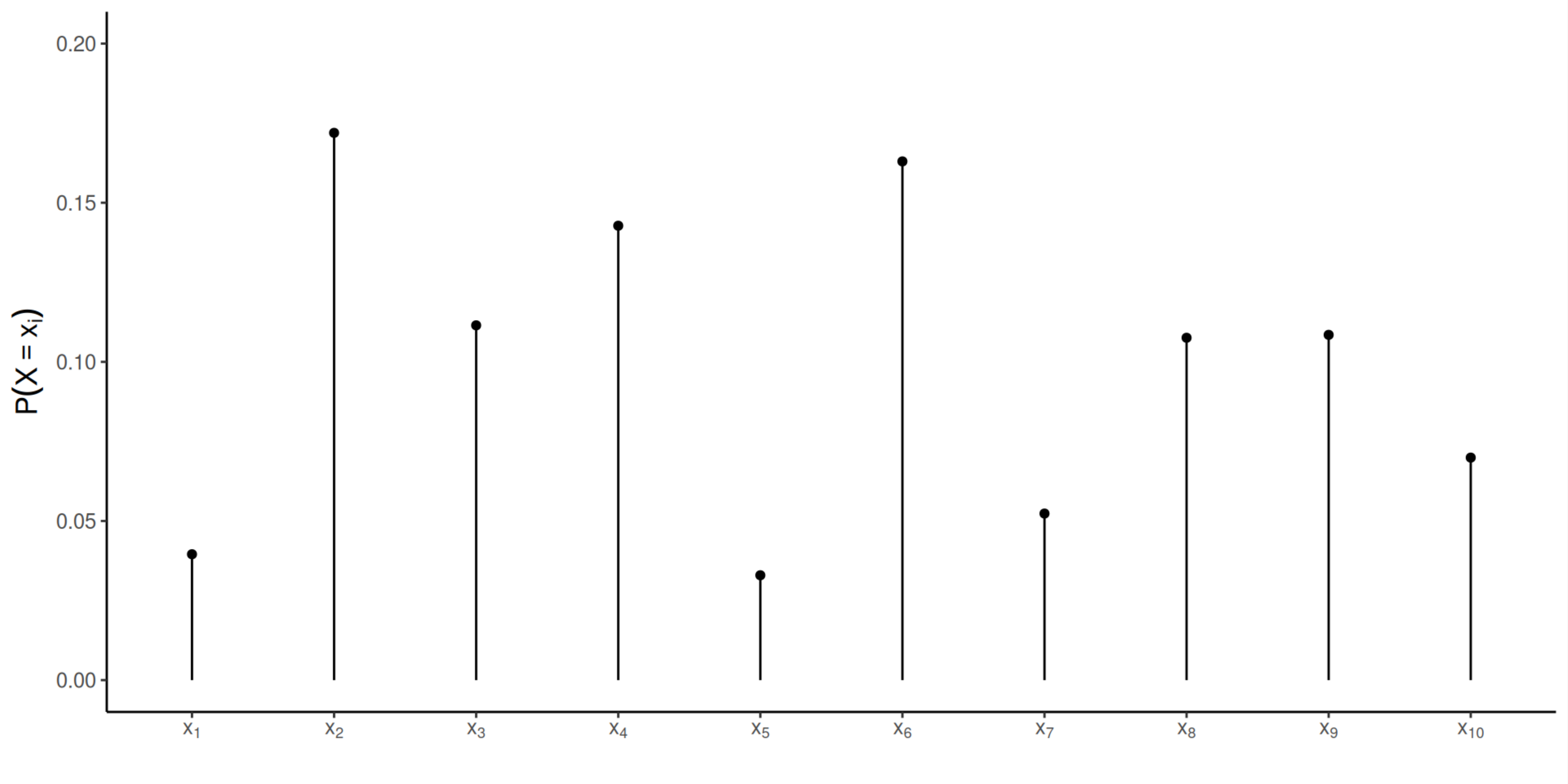


Probability Distributions

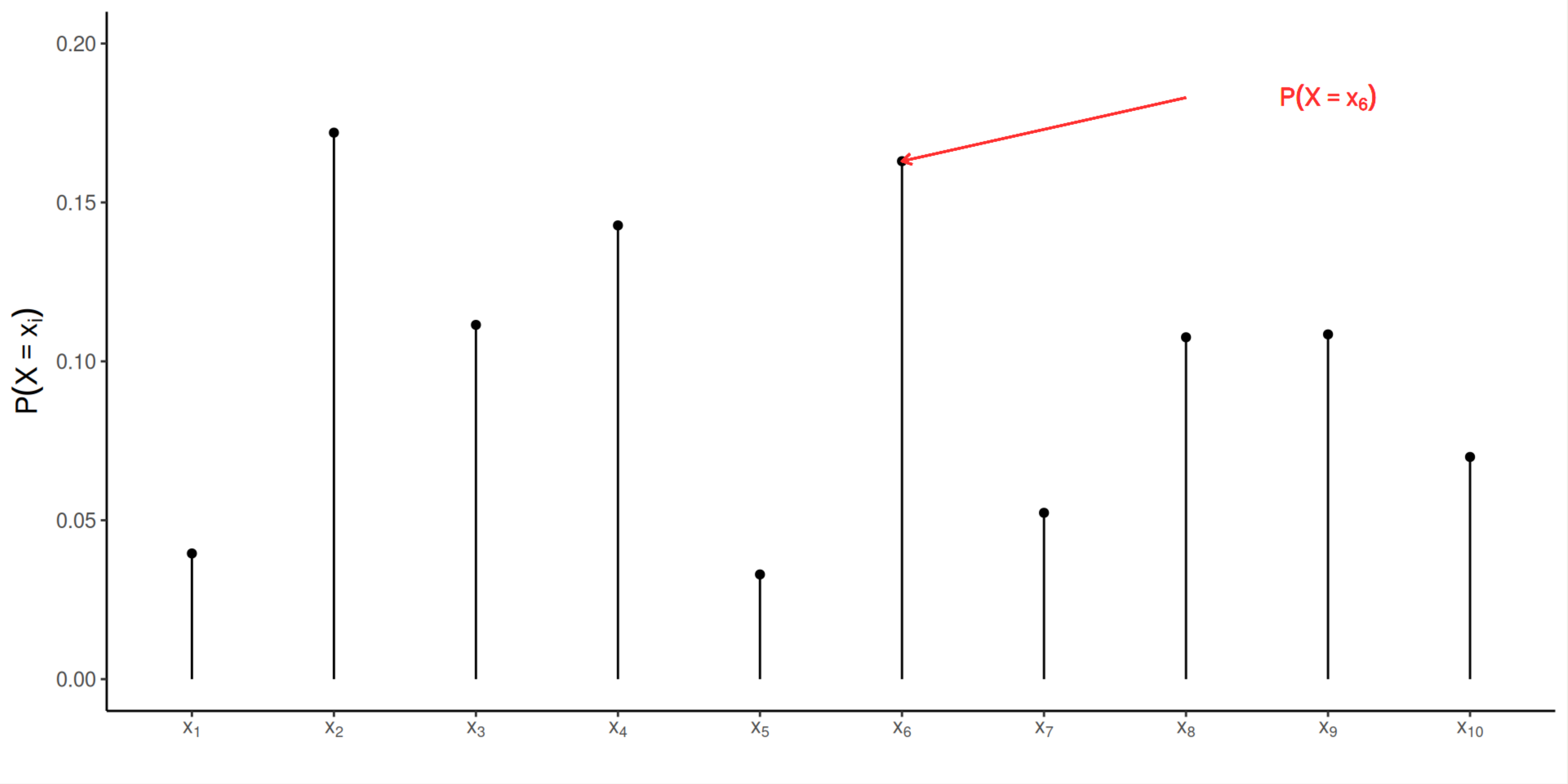
Probability Distribution

- **Probability distribution** assigns probabilities to values taken by random variables.
- *Discrete distributions* assign probabilities to individual values.
- *Continuous distributions* assign probability to intervals.
- Probability distributions are defined using mathematical formulas.
- But graphical representations provide a good intuition.

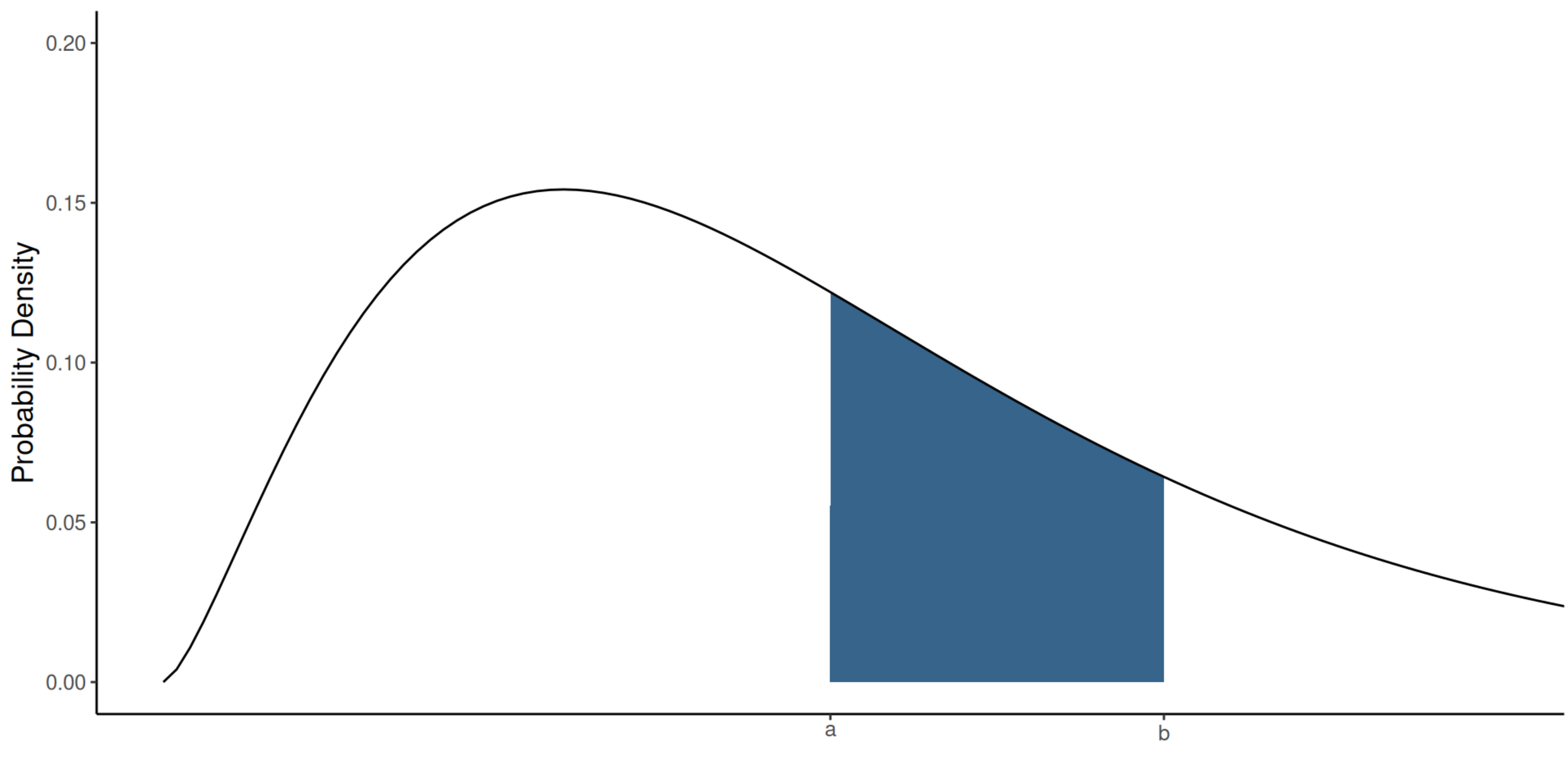
Discrete Distribution



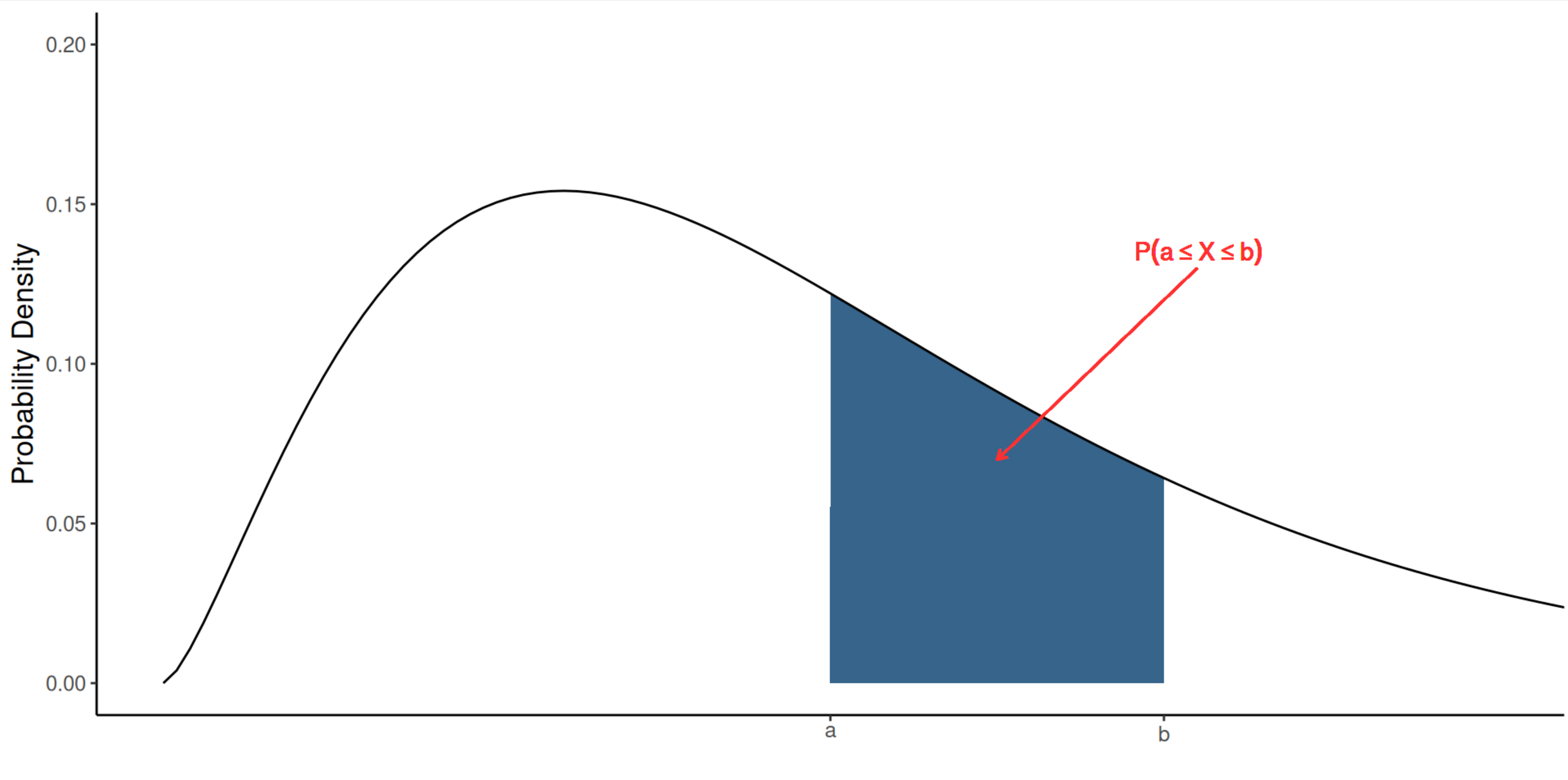
Discrete Distribution



Continuous Distribution



Continuous Distribution



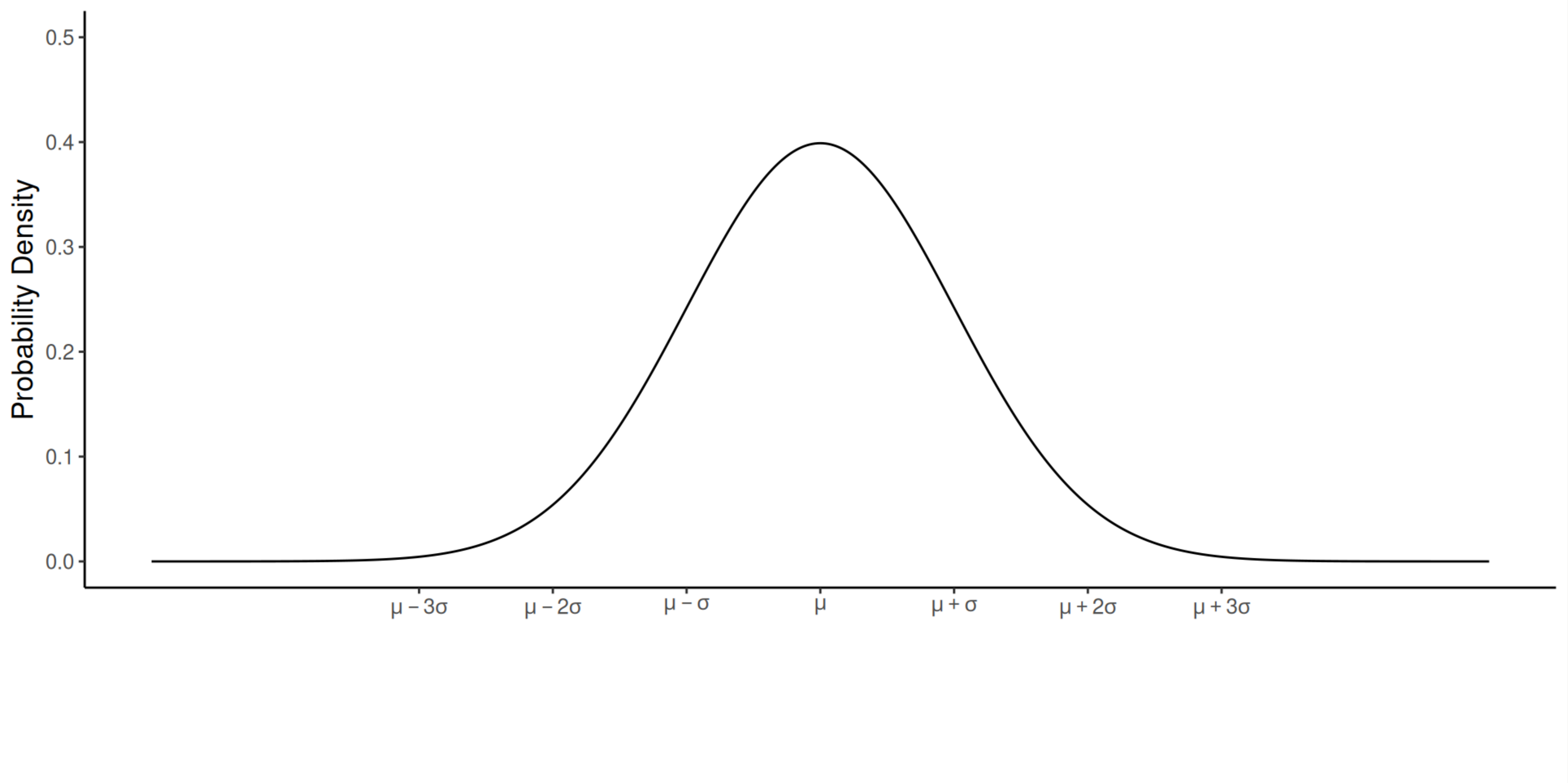
Normal Probability Distribution

- The most important probability distribution.
- As it:
  - Approximates the distribution of many variables in the real world.
  - Is used a lot in inferential statistics.
- It is *symmetric*, *bell-shaped* and fully described by its **mean** μ and **variance** σ^2.
- Can also be called **normal distribution** for short.
- We can denote it as:

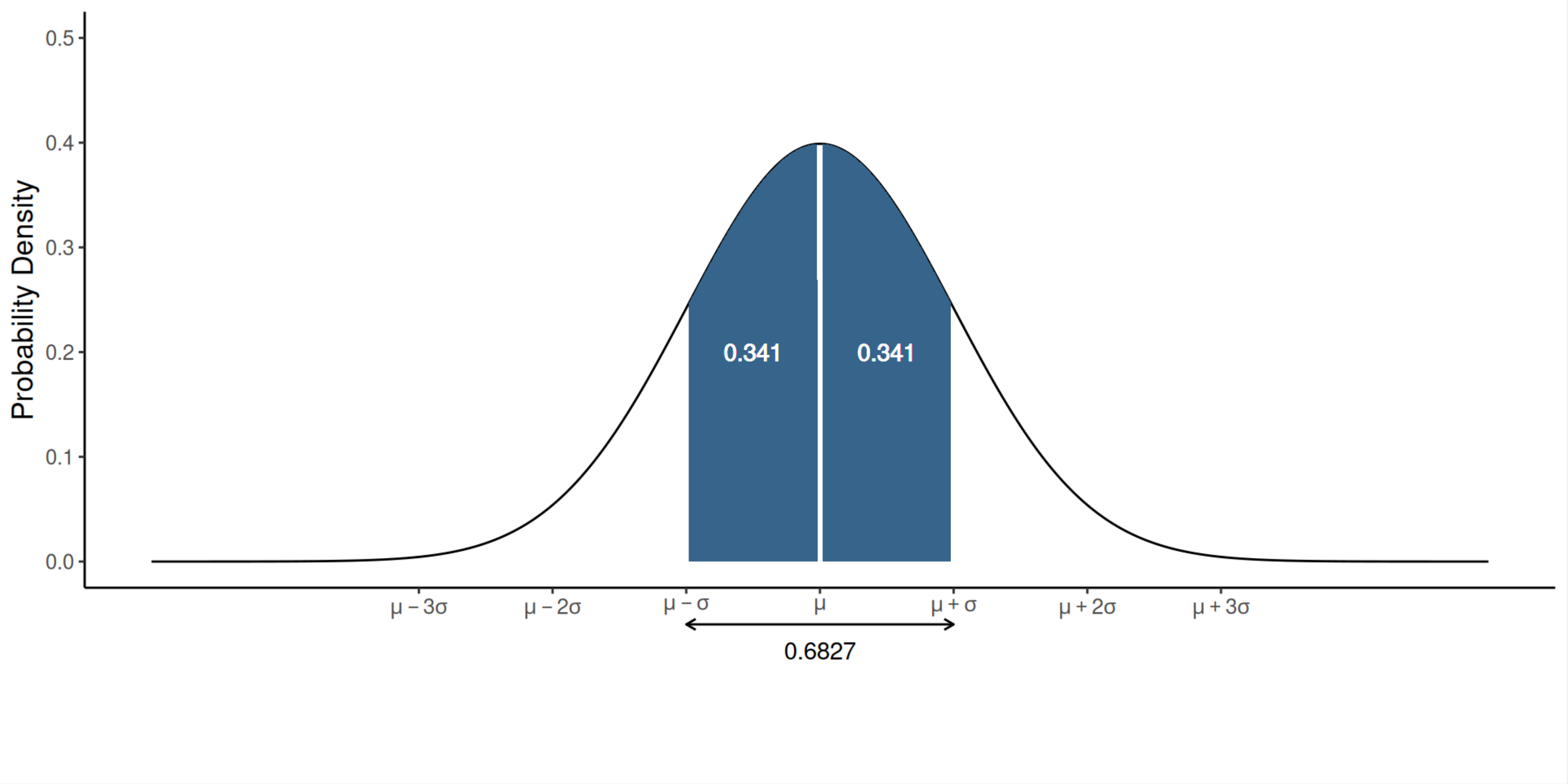
$$Y \sim N(\mu, \sigma^2)$$

"Y is distributed according to a normal distribution with mean μ and variance σ^2."

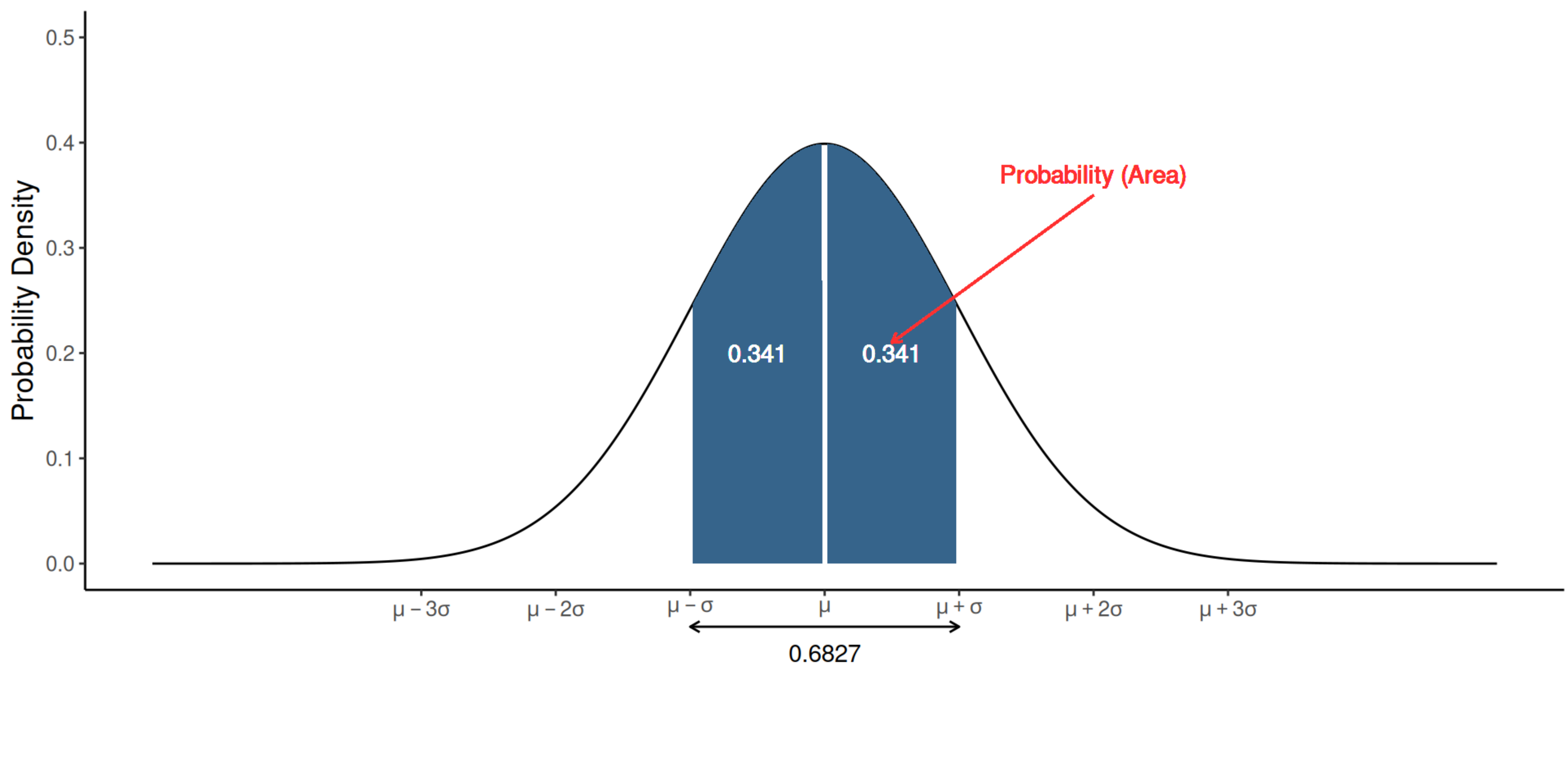
Normal Distribution



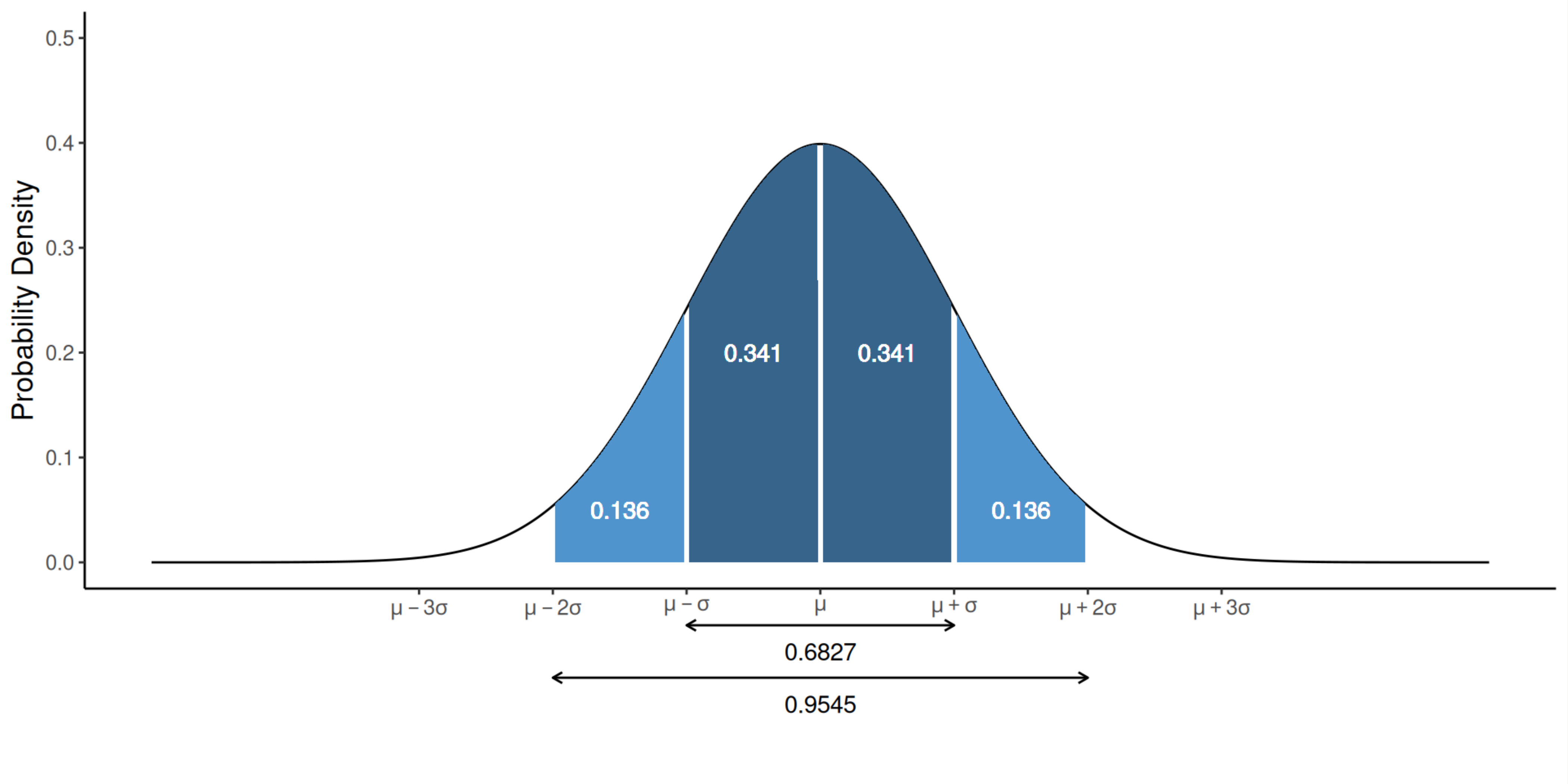
Normal Distribution



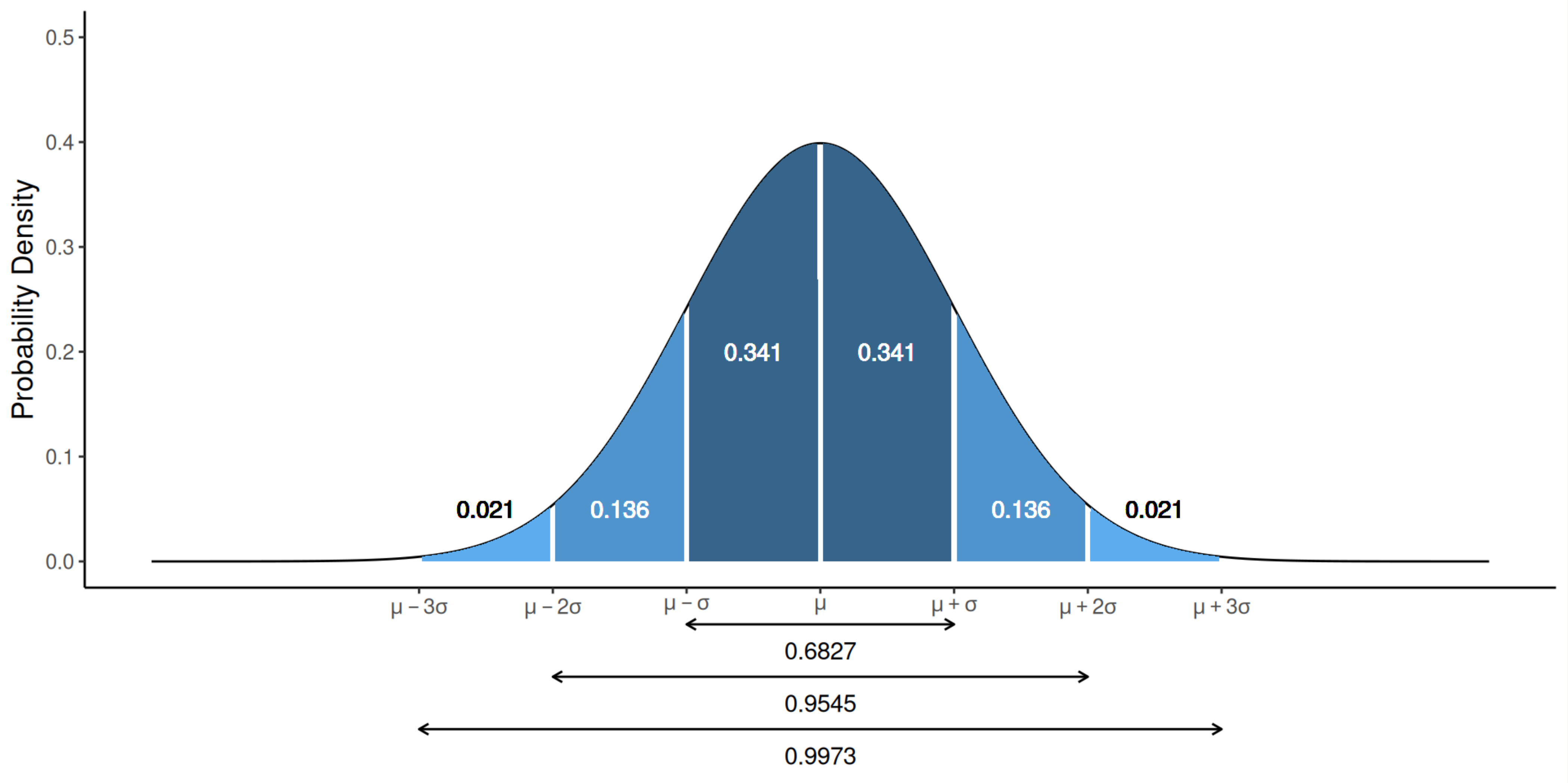
Normal Distribution



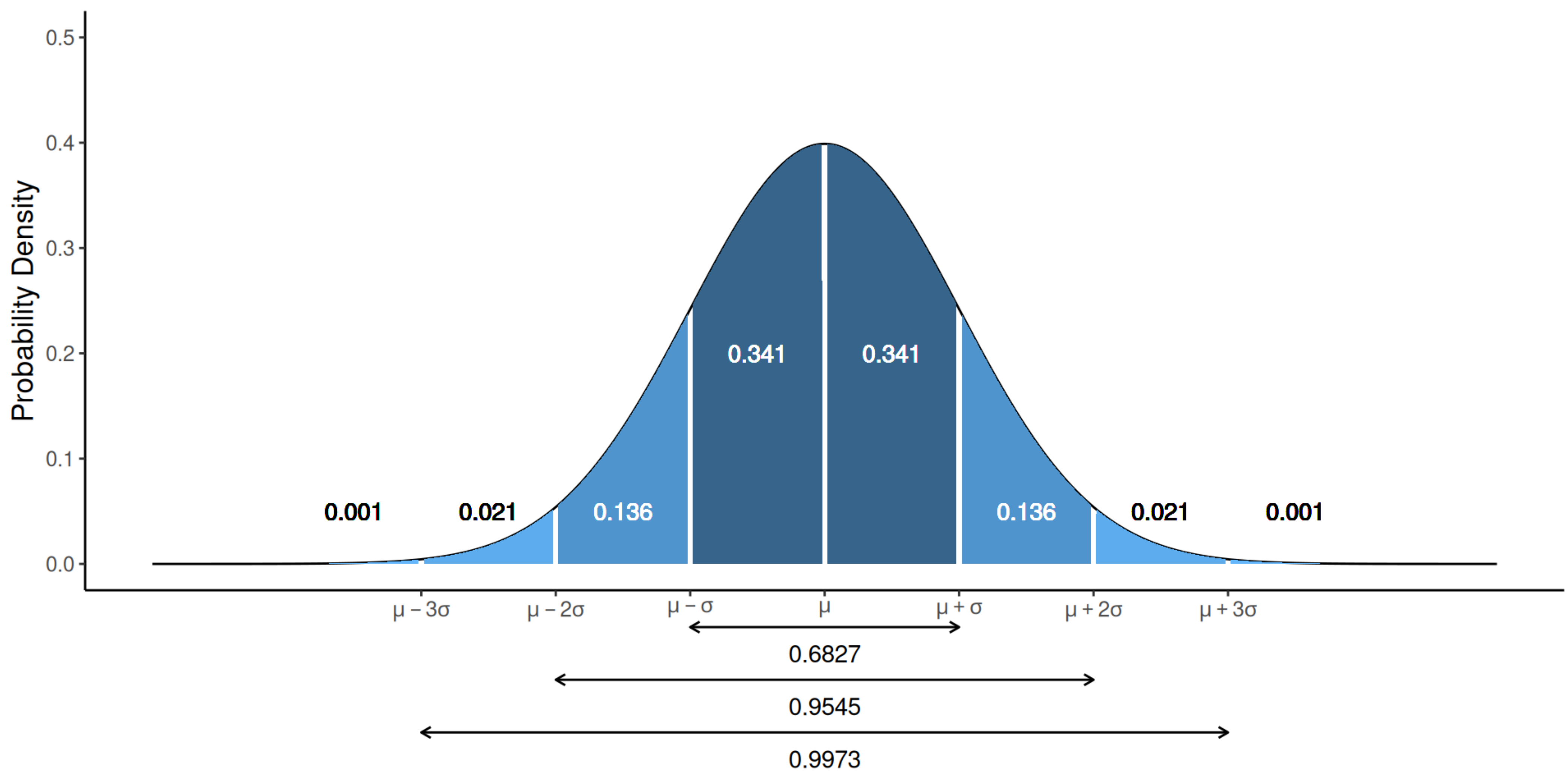
Normal Distribution



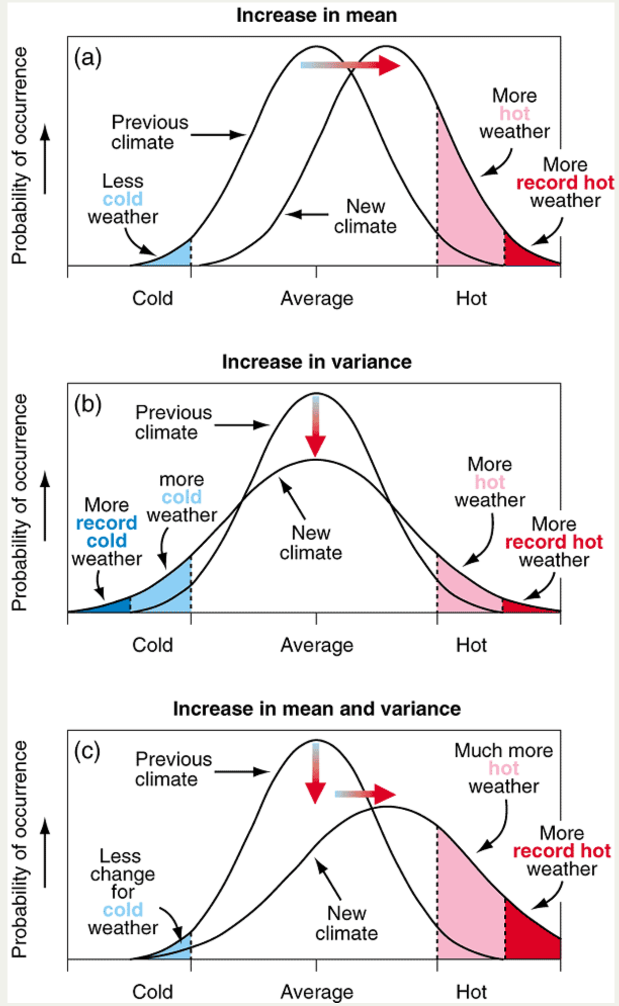
Normal Distribution



Normal Distribution



Example: Climate Change



IPCC - Intergovernmental Panel on Climate Change

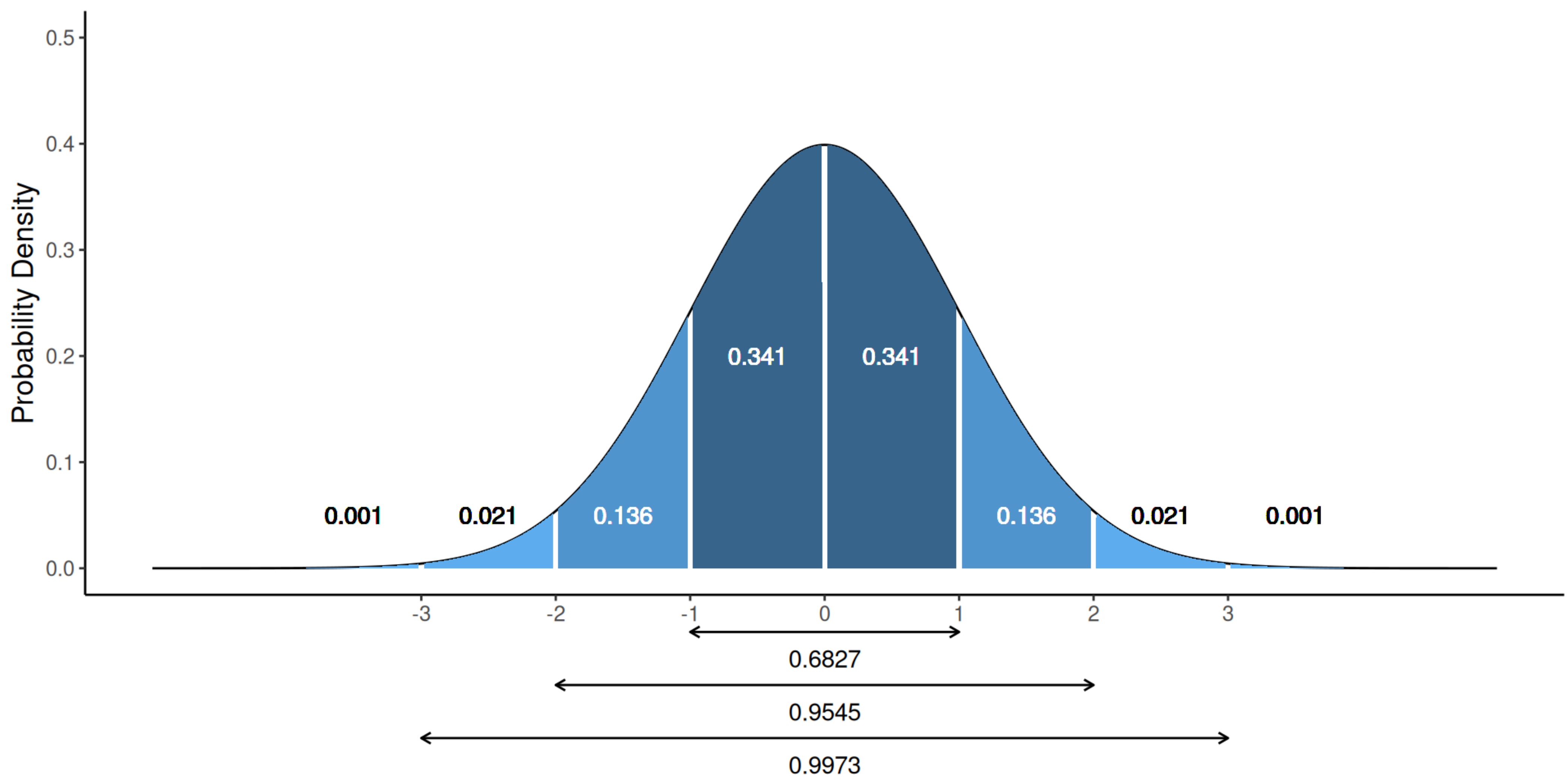
Standard Normal Distribution

- To calculate probabilities for a normal variable with a general mean and variance, we must standardise the variable by first subtracting the mean, then by dividing the result by the standard deviation:

$$z = \frac{x - \mu_x}{\sigma_x}$$

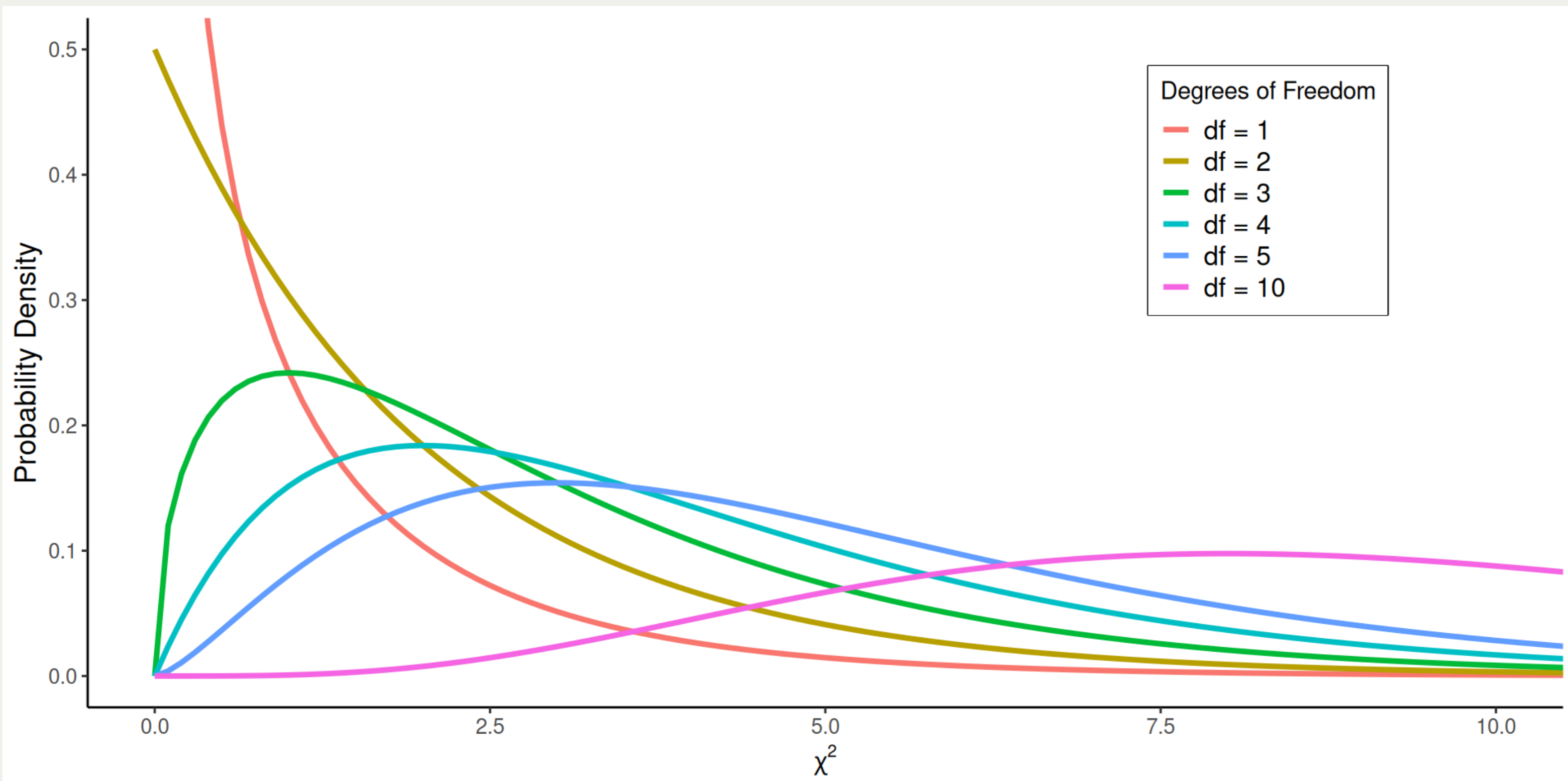
- Z-scores indicate the number of standard deviation units a value is from the mean of a distribution.
- **Standard normal distribution** is the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ and can be denoted as $N(0, 1)$.

Standard Normal Distribution



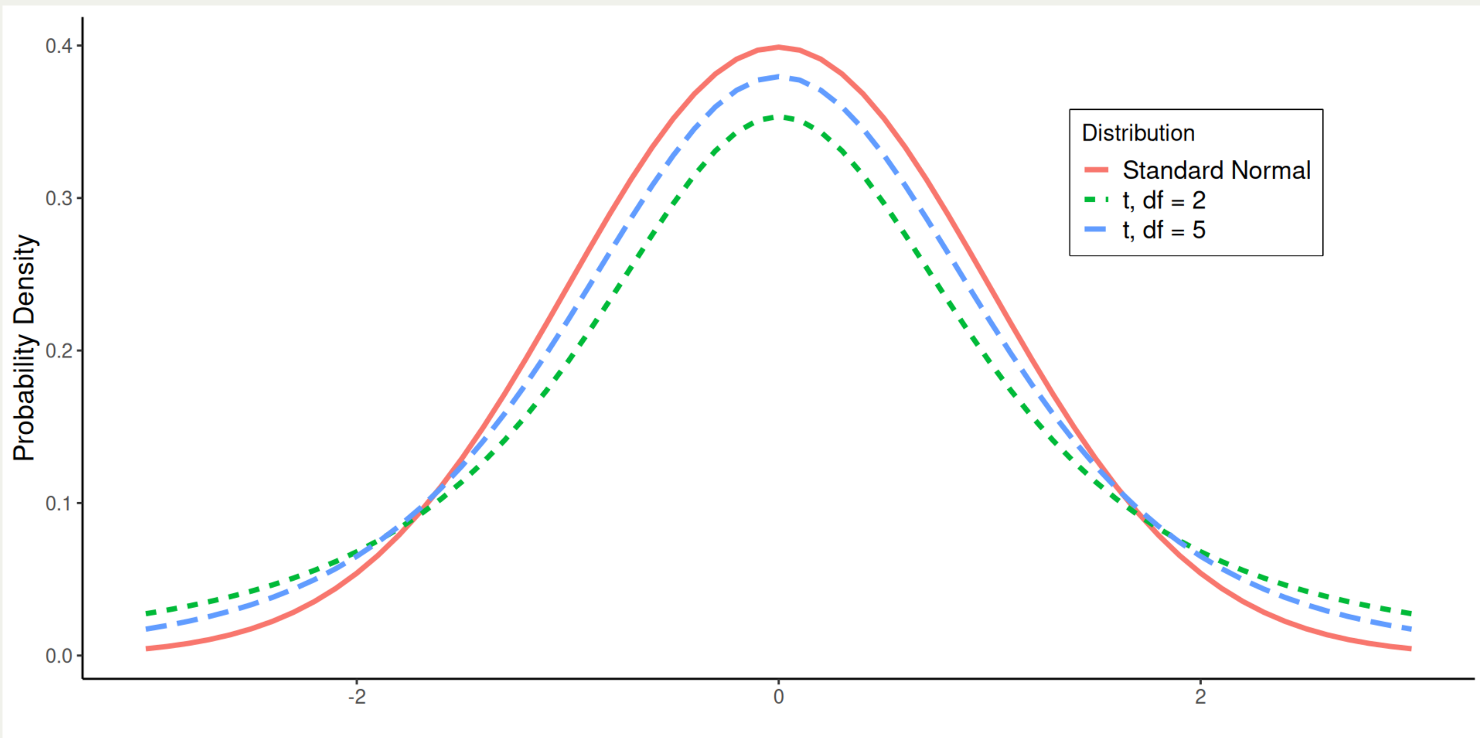
χ^2 Distribution

- χ^2 (pronounced chi-squared) distribution
- Shape depends on the degrees of freedom (more later)
- Used to analyse contingency tables.



t Distribution

- t distribution is bell shaped and and symmetric around the mean of 0.
- In comparison to the standard normal distribution its standard error is a bit larger than 1 and depends on the degrees of freedom.
- Used to compare means of variables between different groups (more later).



Next

- Workshop:
  - Probability Distributions
- Next week:
  - Hypothesis Testing