Probability Part 2: Random Variables

Random Variables

A number associated to the outcome of a random experiment.

- The sum of two dice rolls
- The number of heads in 5 flips of a coin
- The average height of a sample of 10 people

Discrete: Outcomes can be listed

Continuous: Can take on any value in an interval

Example: The Binomial Distribution

Gives the probability of k successes out of n independent Bernoulli trial.

Example: The Binomial Distribution

Gives the probability of *k* successes out of *n* independent Bernoulli trial. Requires:

- Binary outcome (yes/no)
- A fixed number (*n*) of repeated **independent** trials
- A fixed probability of “success” (*p*)

Example: The Binomial Distribution

Question: Is this a discrete distribution or a continuous distribution?

Example: The Binomial Distribution

Question: Is this a discrete distribution or a continuous distribution?

Answer - A discrete distribution since we can list the possible outcomes (0, 1, 2, 3, ..., n)

Example: The Binomial Distribution

Uses:

- Toy Example - Flipping a coin
- Polling
- A/B Testing

Discrete Probability Distributions

Probability Mass Function (pmf): For each possible outcome, gives the probability of that outcome occurring.

Example: The Binomial Distribution

In general, a binomial distribution with *n* trials and probability of success *p* has a pmf of:

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Example: The Binomial Distribution

In general, a binomial distribution with *n* trials and probability of success *p* has a pmf of:

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Number of ways to choose *x* success out of *n* trials

NASHVILLE SOFTWARE SCHOOL

Example: The Binomial Distribution

In general, a binomial distribution with *n* trials and probability of success *p* has a pmf of:

$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Number of ways to choose *x* success out of *n* trials

Probability of success

Example: The Binomial Distribution

In general, a binomial distribution with *n* trials and probability of success *p* has a pmf of:

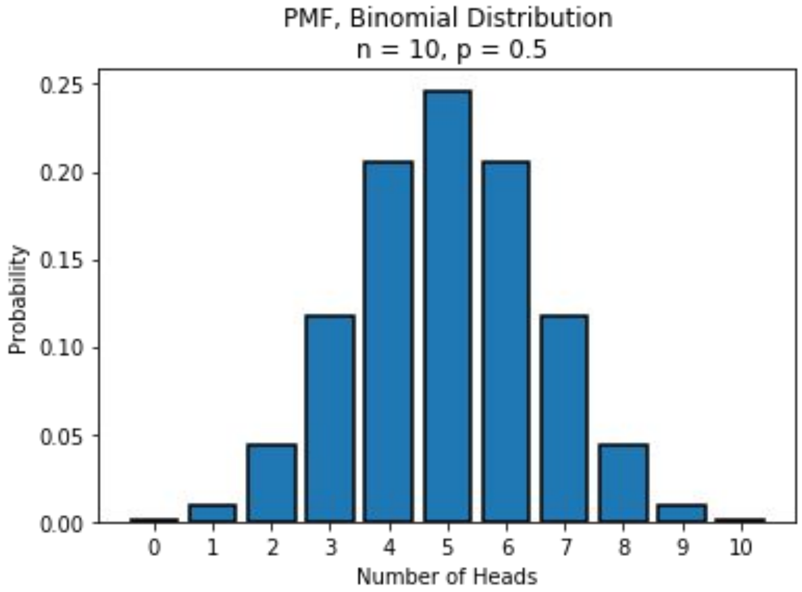
$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Number of ways to choose *x* success out of *n* trials

Probability of success

Probability of failure

The Binomial Distribution - Probability Mass Function



PMF for ten flips of a fair coin

The Binomial Distribution

To carry out these types of calculations, we will usually use the *scipy stats* library rather than calculating the probabilities explicitly.

See the corresponding notebook to see what this looks like in practice.

Continuous Random Variables

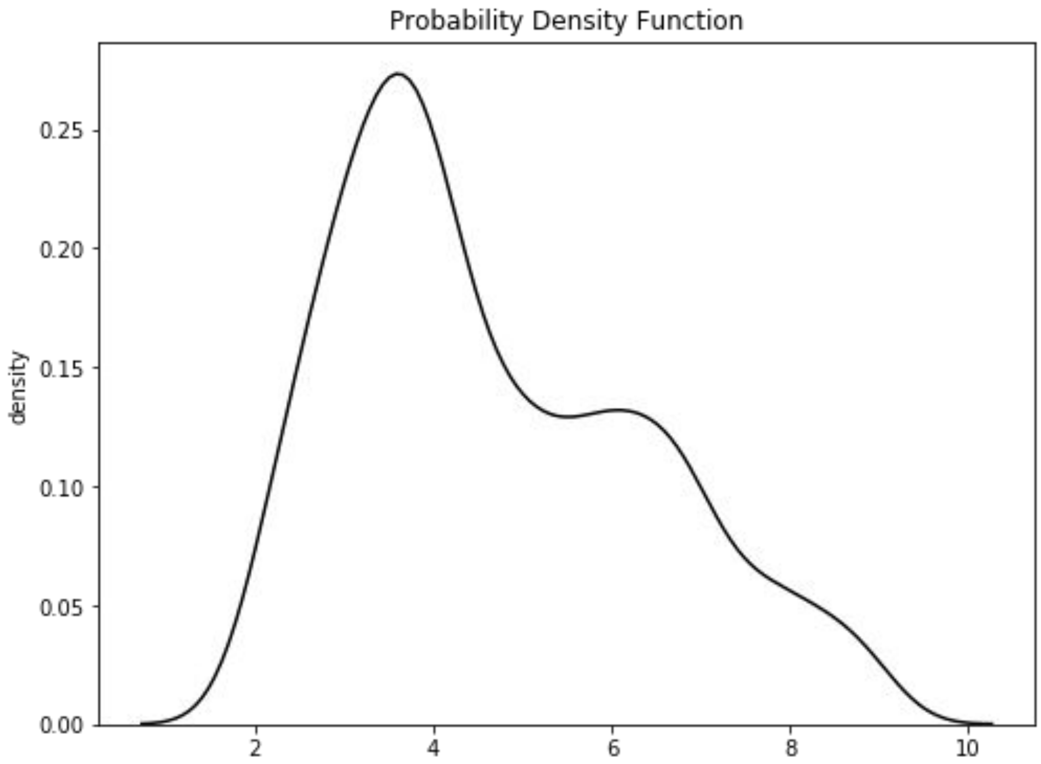
What about random variables which can take on any value in a range?

We can no longer talk about the exact probability for any particular value, but instead can talk about the probability *density* at a particular point.

To find probabilities, we can only find probabilities for the value landing in a particular *range* of values.

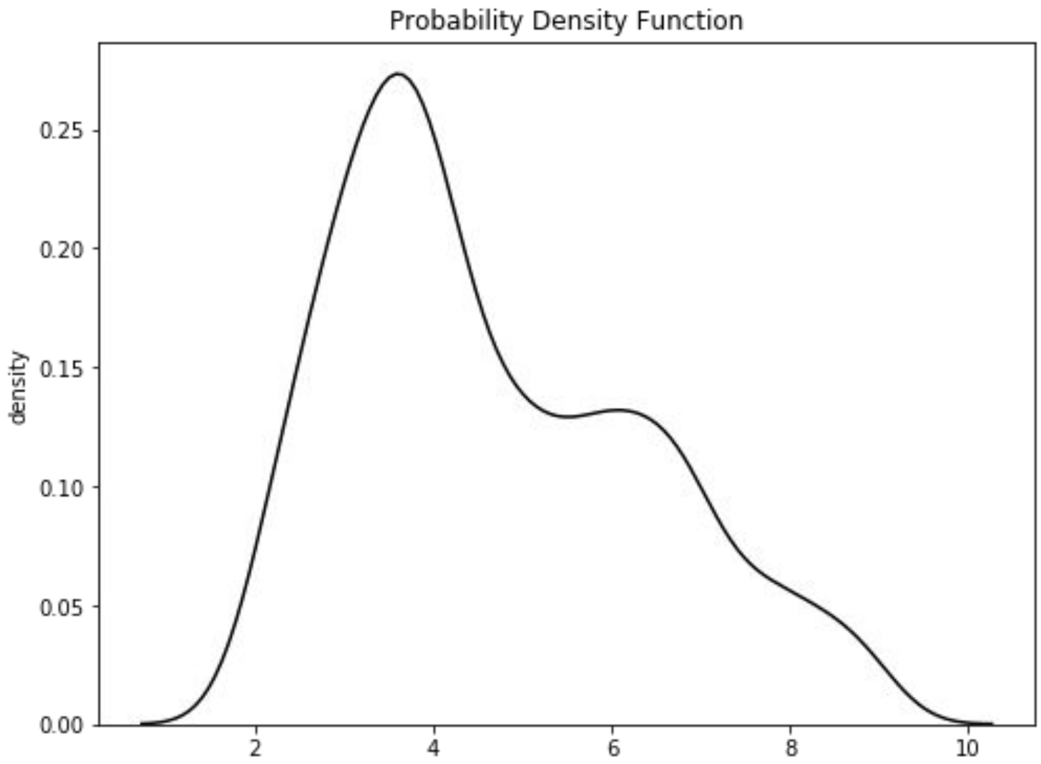
Probability Density Functions

The probability density at each possible value can be specified by a **probability density function**, or **PDF.**



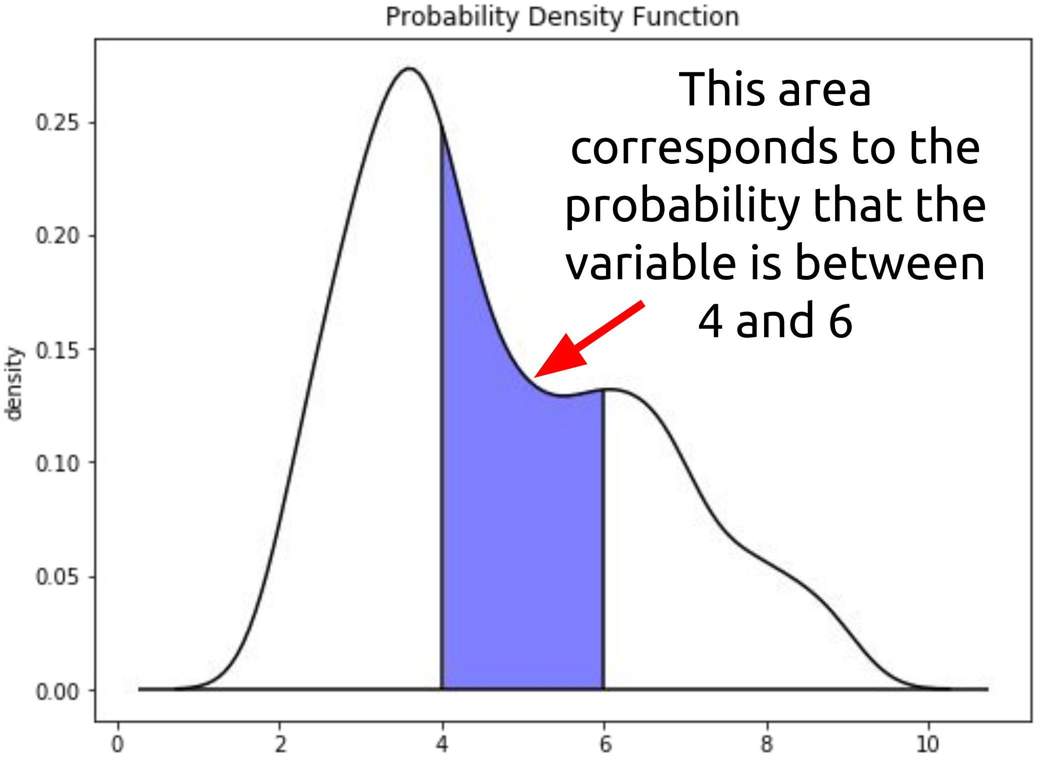
Probability Density Functions

Rather than talk about the probability at a specific value, we can calculate the probability of the variable being in a particular range.



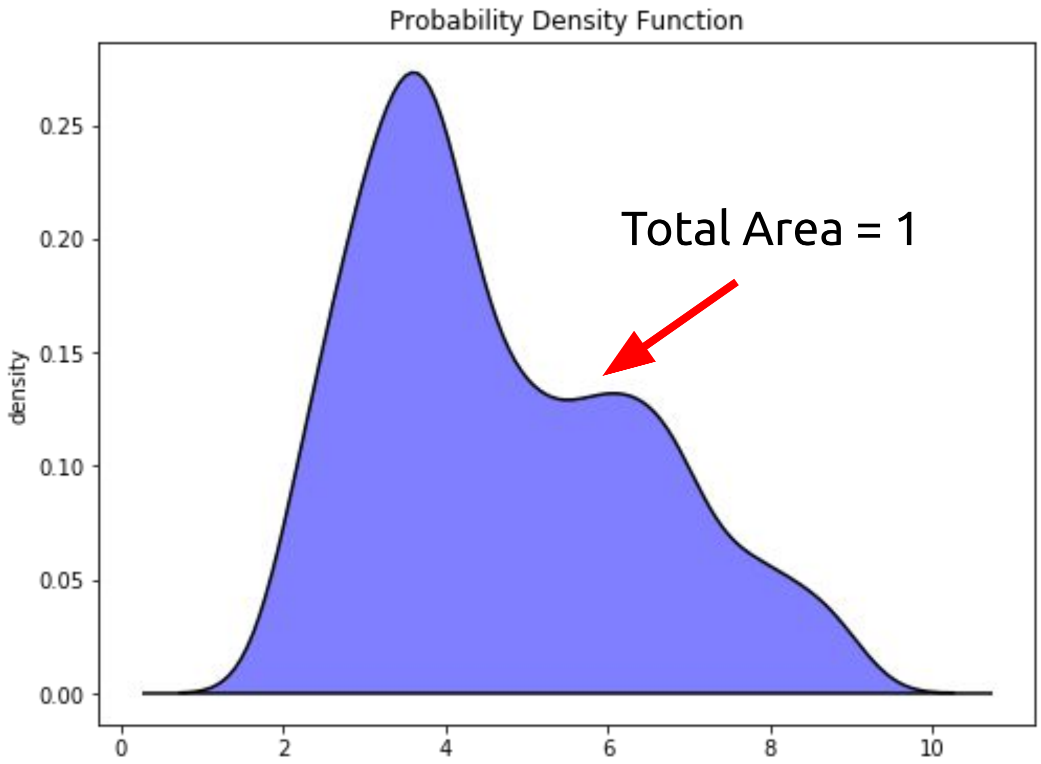
Probability Density Functions

The probability of the variable being in a particular range corresponds to the area under the PDF in that range.



Probability Density Functions

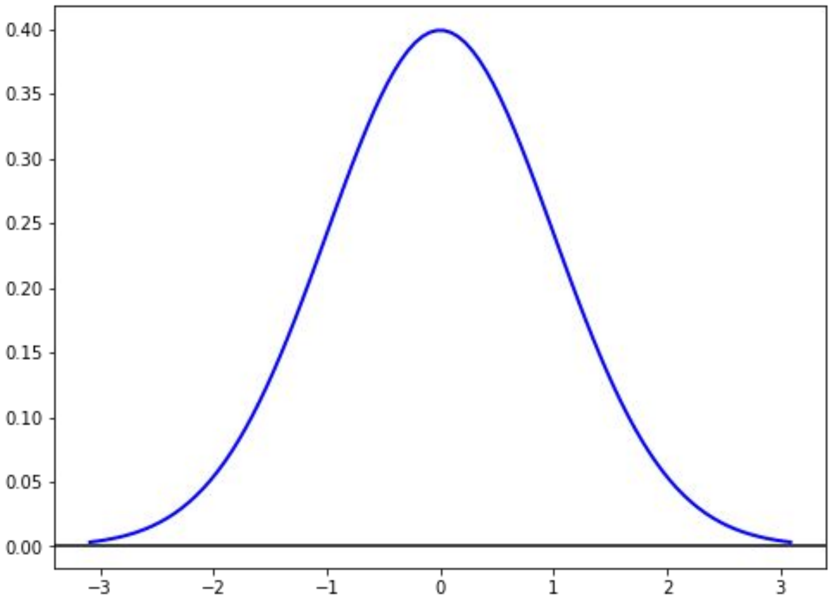
This means that the total area under the curve is 1.



The Normal Distribution

Perhaps the most well-known distribution is the Normal distribution, aka, the “Gaussian” distribution.

It is a symmetric, bell-shaped distribution.



The Standard Normal Distribution

The Normal Distribution

(It is thought that) many things can be described by a normal distribution.

Eg. IQs, test scores, heights, weights, random variations in industrial processes

However, these can only be approximately true (the normal distribution has nonzero density for all real numbers, including negatives)

The Normal Distribution

Bell-shaped distribution, described by two parameters: mean μ and standard deviation σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The **standard normal** distribution has mean 0 and standard deviation 1.

See the notebook for a demo of how μ and σ affect the distribution.