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: Bernoulli Distribution

Setup: An experiment with exactly two outcomes, labeled "success" (denoted by 1) and "failure" (denoted by 0).

Probability of success = p
Probability of failure = $1 - p$

Example: Bernoulli Distribution

Setup: An experiment with exactly two outcomes, labeled "success" (denoted by 1) and "failure" (denoted by 0).

Probability of success = p
Probability of failure = $1 - p$

Example: A marketing company knows that historically, search ads have a click-through rate of 1.5%.
We can view each interaction as a Bernoulli trial with $p = 0.015$

Example: The Binomial Distribution

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

Example: The Binomial Distribution

Gives the probability of *k* successes out of *n* independent Bernoulli trials.

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

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Let S correspond to a hit and F correspond to not getting a hit.

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Let S correspond to a hit and F correspond to not getting a hit.

Possible sequences involving two hits:

SSFFF	SFSFF	SFFSF	SFFFS	FSSFF
FSFSF	FSFFS	FFSSF	FFSFS	FFFSS

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Let S correspond to a hit and F correspond to not getting a hit.

Possible sequences involving two hits:

SSFFF	SFSFF	SFFSF	SFFFS	FSSFF
FSFSF	FSFFS	FFSSF	FFSFS	FFFSS

Each sequence has two successes and three failures, so each has probability

$$(0.325)^2 \cdot (1 - 0.325)^3$$

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Let S correspond to a hit and F correspond to not getting a hit.

Possible sequences involving two hits:

SSFFF	SFSFF	SFFSF	SFFFS	FSSFF
FSFSF	FSFFS	FFSSF	FFSFS	FFFSS

Each sequence has two successes and three failures, so each has probability

$$(0.325)^2 \cdot (1 - 0.325)^3$$

Total Probability: $10 \cdot (0.325)^2 \cdot (1 - 0.325)^3 \approx 0.3248$

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Rather than listing out all sequences, we can use a shortcut.

In general, the number of sequences of 5 at-bats with *x* hits is equal to

$$\binom{5}{x} = \frac{5!}{x! \cdot (5-x)!}$$

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly two hits?

Total Probability: $10 \cdot (0.325)^2 \cdot (1 - 0.325)^3 \approx 0.3248$

$$= \binom{5}{2} \cdot (0.325)^2 \cdot (1 - 0.325)^3$$

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

Exactly 2: $$\binom{5}{2} \cdot (0.325)^2 \cdot (1 - 0.325)^3$$

Exactly *x*:

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

Exactly 2: $$\binom{5}{2} \cdot (0.325)^2 \cdot (1 - 0.325)^3$$

Exactly *x*: $$\binom{5}{x} \cdot (0.325)^x \cdot (1 - 0.325)^{(5-x)}$$

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

x	P(x hits)
0	
1	
2	
3	
4	
5	

Binomial Distribution

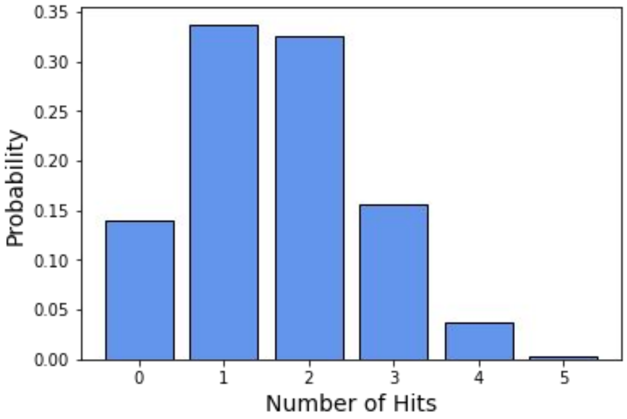
Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

x	P(*x* hits)
0	0.1401
1	0.3373
2	0.3248
3	0.1564
4	0.0377
5	0.0036

Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

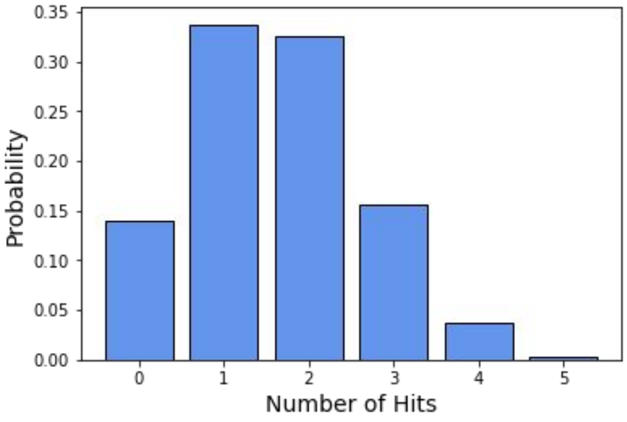
x	P(*x* hits)
0	0.1401
1	0.3373
2	0.3248
3	0.1564
4	0.0377
5	0.0036



Binomial Distribution

Example: A baseball player has a batting average of 0.325. In 5 at-bats, how likely is it that he gets exactly ***x*** hits?

x	P(*x* hits)
0	0.1401
1	0.3373
2	0.3248
3	0.1564
4	0.0377
5	0.0036



We are looking at the **probability mass function (pmf)**, which gives, for each possible outcome, 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

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:

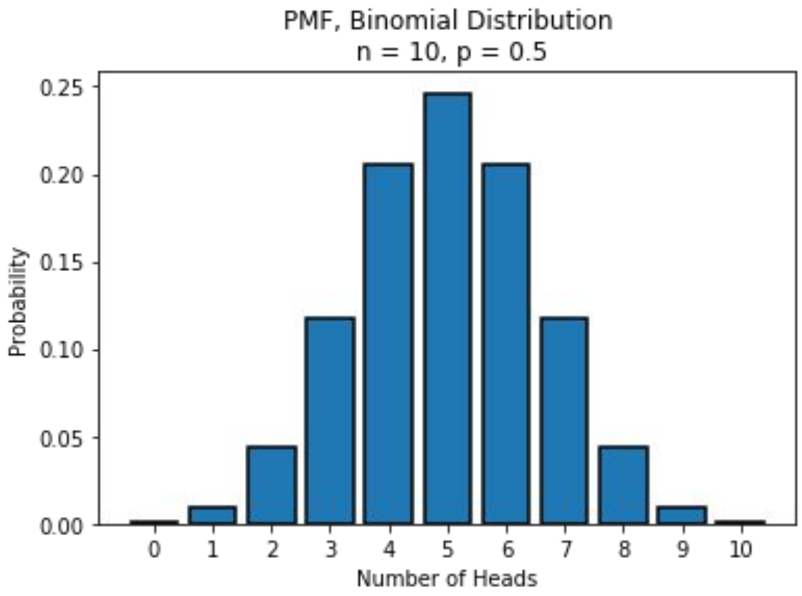
$$f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

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

NASHVILLE SOFTWARE SCHOOL

The Binomial Distribution

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

Let’s see this in practice in the notebook.

Continuous Random Variables

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

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.

Continuous Random Variables

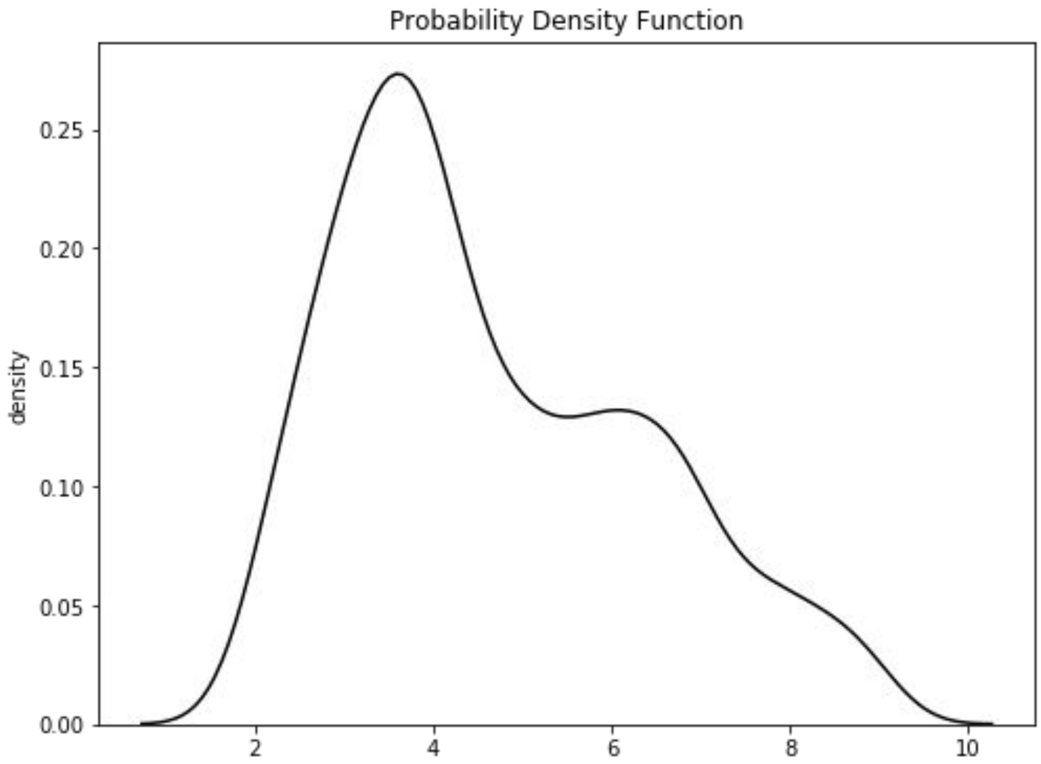
What about random variables which can take on any value in a range?

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.

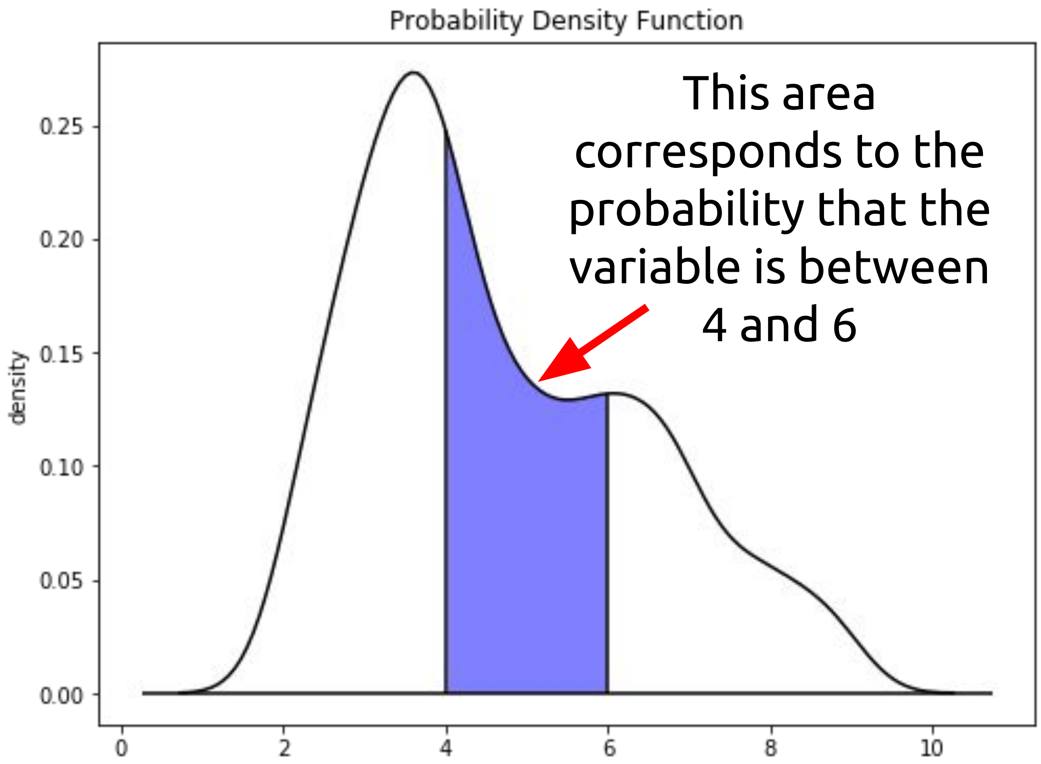
Probability Density Functions

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



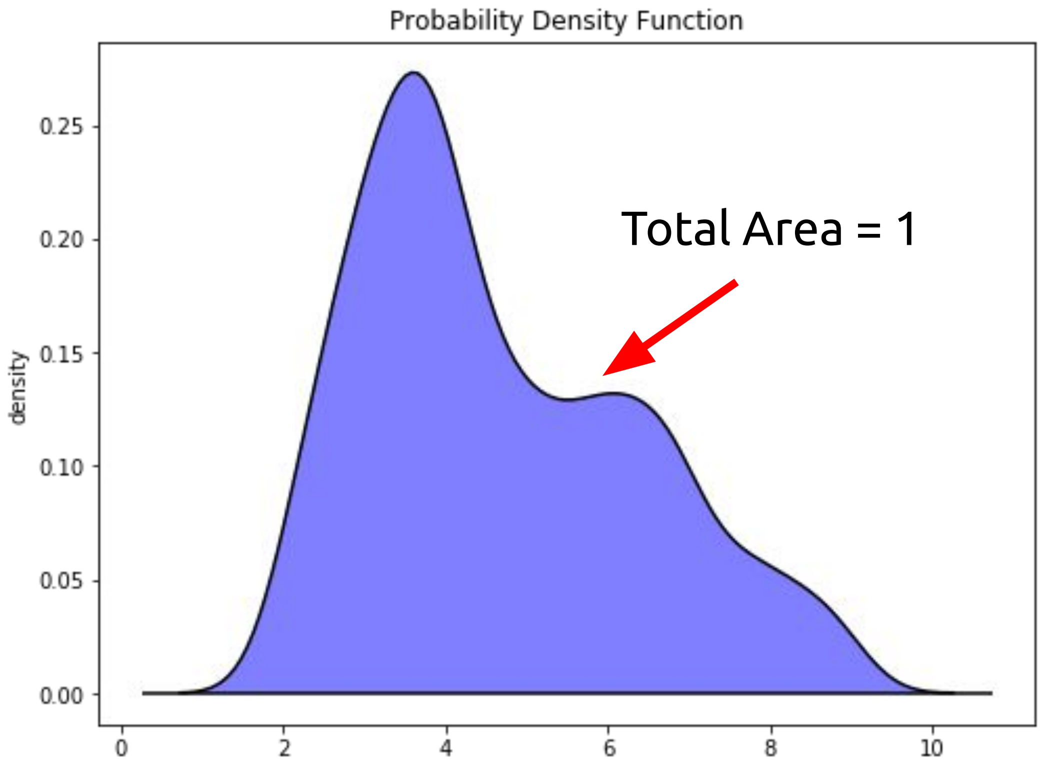
Probability Density Functions

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



Probability Density Functions

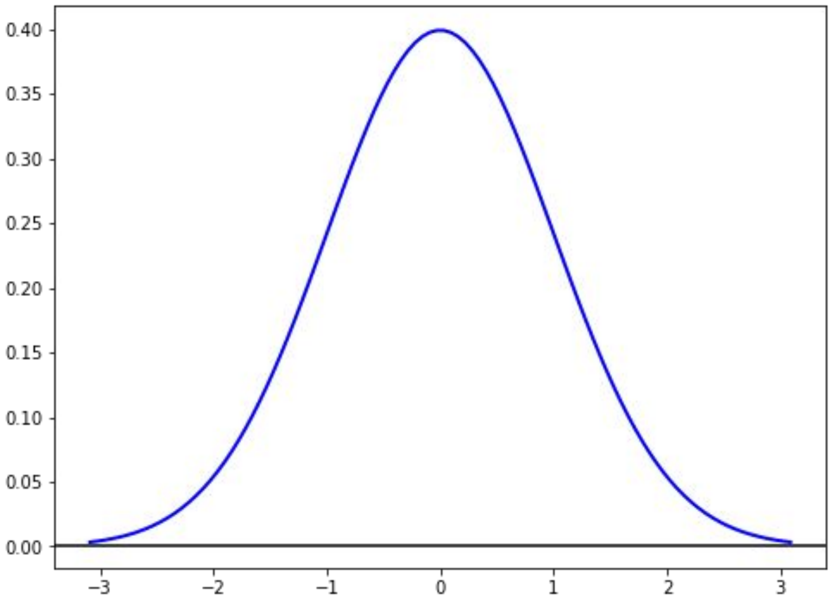
This means that the total area under the curve is 1.



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)

NASHVILLE SOFTWARE SCHOOL

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.