Intro to Probability

Why is Probability Important in Data Science?

- Probability lets us model uncertainties.
- Probabilities can be used to make classifications.
- Probability lets us make statements about a population.
- Many ML methods depend on probability distribution assumptions
  - Linear Regression assumes the residuals are normally distributed
  - Gaussian Naive Bayes assumes the predictors are normally distributed
- Probability distributions can be useful to create ‘toy’ datasets to test out methods.

What is Probability?

$$P(A) = \frac{\#\ of\ ways\ A\ can\ happen}{total\ number\ of\ outcomes}$$

A is an event. We read P(A) as “the probability of event A” or “the probability of A”. Note that the probability of a single event is called a **marginal probability**.

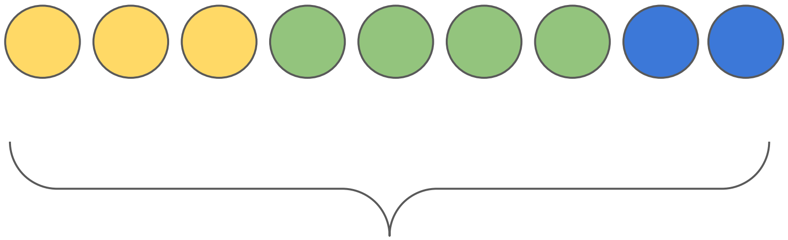
Basic Probability Examples

What is the probability of pulling out a blue marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

Basic Probability Examples

What is the probability of pulling out a blue marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

$$P(Blue) = \frac{2}{9}$$



Sample Space

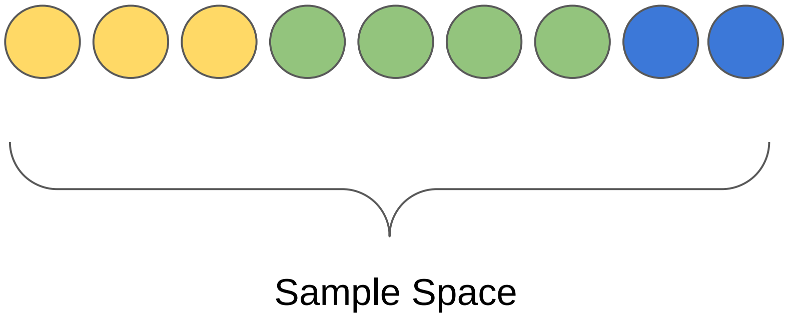
Basic Probability Examples

What is the probability of pulling out a **non**-blue marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

Basic Probability Examples

What is the probability of pulling out a **non**-blue marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

$$P(NonBlue) = \frac{7}{9}$$



Basic Probability Examples

What is the probability of rolling doubles on two six-sided die?

Basic Probability Examples

What is the probability of rolling doubles on two six-sided die?

Roll 1

Roll 2

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Total number of outcomes:
6 x 6 = 36

Basic Probability Examples

What is the probability of rolling doubles on two six sided die?

Roll 1

Roll 2

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Total number of outcomes:
6 x 6 = 36

Number ways:
6

Basic Probability Examples

What is the probability of rolling doubles on two six sided die?

Roll 1

Roll 2

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Total number of outcomes:
6 x 6 = 36

Number ways:
6

Number ways ÷
Total number of outcomes
= 6 / 36

= 1/6

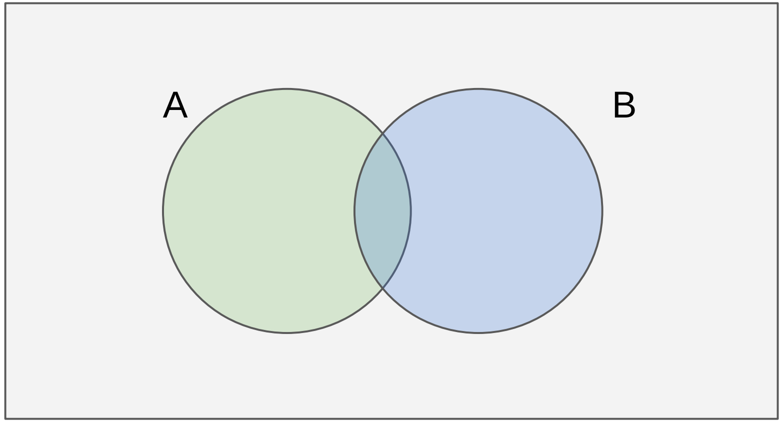
Probability Rules

- **Addition Rule (“OR” events)**
- **Multiplication Rule (“AND” events)**

Probability Rules

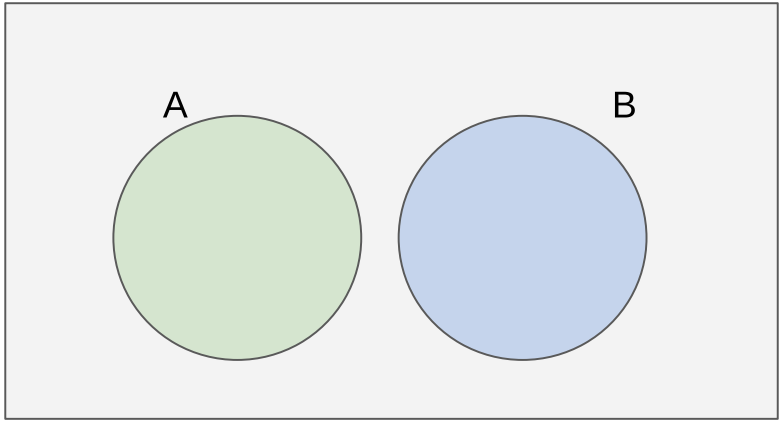
Addition Rule

P(A or B) = P(A) + P(B) - P(A and B)



Addition Rule (Mutually Exclusive)

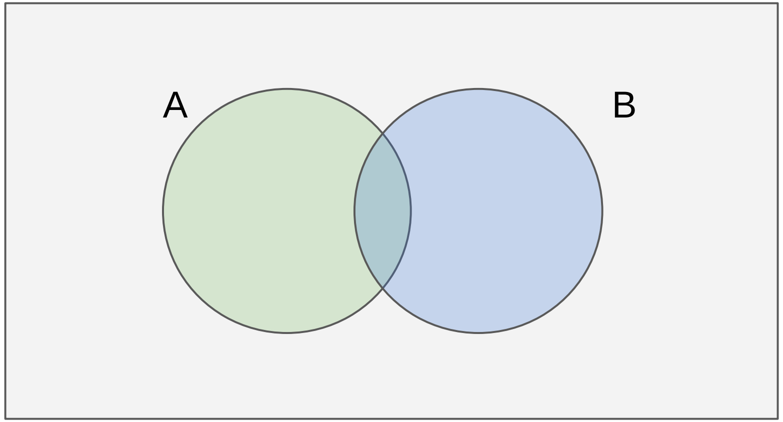
P(A or B) = P(A) + P(B)



Probability Rules

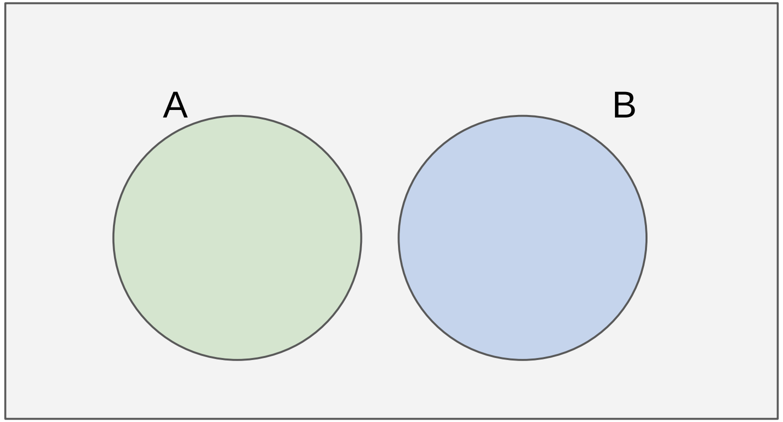
Addition Rule

P(A or B) = P(A) + P(B) - P(A and B)



Addition Rule (Mutually Exclusive)

P(A or B) = P(A) + P(B)



If not sure, use the non-mutually exclusive formula. P(A and B) == 0, for mutually exclusive events.

Addition Rule Examples

What is the probability of pulling out a blue marble **OR** a yellow marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

Addition Rule Examples

What is the probability of pulling out a blue marble **OR** a yellow marble from a bag that contains 3 yellow marbles, 4 green marbles, and 2 blue marbles?

P(Blue or Yellow)
= P(Blue) + P(Yellow)
= 2/9 + 3/9

= 5/9

Addition Rule Examples

If you draw a single card from a deck of cards, what is the probability the card is either a diamond **OR** an ace?

Addition Rule Examples

If you draw a single card from a deck of cards, what is the probability the card is either a diamond **OR** an ace?

P(Diamond) =

P(Ace) =

P(Diamond and Ace) =

P(Diamond or Ace)

Addition Rule Examples

If you draw a single card from a deck of cards, what is the probability the card is either a diamond **OR** an ace?

P(Diamond) = 13/52
P(Ace) = 4/52
P(Diamond and Ace) = (13/52) x (4/52) = 1/52

P(Diamond or Ace)
= P(Diamond) + P(Ace) - P(Diamond and Ace)
= 13/52 + 4/52 - 1/52

= 16/52

Probability Rules

Multiplication Rule - Note: This rule only applies for independent events!

If the first event happening impacts the probability of the second event, the events are **dependent**.

If the first event happening does not impact the probability of the second event, the events are **independent**.

P(A and B) = P(A)*P(B)

Note that the probability of two events occurring is called a **joint probability**.

Multiplication Rule Examples

If you roll two six-sided die, what is the probability that you roll two 5s?

Multiplication Rule Examples

If you roll two six-sided die, what is the probability that you roll two 5s?

P(First Roll = 5) =
P(Second Roll = 5) =

P(First & Second Rolls = 5) =

Multiplication Rule Examples

If you roll two six-sided die, what is the probability that you roll two 5s?

P(First Roll = 5) = 1/6
P(Second Roll = 5) = 1/6

P(First & Second Rolls = 5) = 1/6 x 1/6

= 1/36

Probability Questions

1. A bag contains 2 orange, 3 purple and 2 yellow balls. Two of the balls are randomly drawn **with replacement**. What is the probability that **none** of the balls drawn is yellow?

Probability Questions

1. A bag contains 2 orange, 3 purple and 2 yellow balls. Two of the balls are randomly drawn **with replacement**. What is the probability that **none** of the balls drawn is yellow?

Total number of balls = 2 + 3 + 2 = 7
P(draw_1 = Yellow) = 2/7 => P(draw_1 != Yellow) = 5/7
P(draw_2 = Yellow) = 2/7 => P(draw_2 != Yellow) = 5/7

P(none yellow) = P(draw_1 != Yellow) x P(draw_2 != Yellow)
= 5/7 x 5/7

= 25/49

Probability Questions

2. A normal deck of cards has 52 cards in it. Assuming that you **DON'T** replace the card you had drawn before the next draw, what is the probability that you will draw three kings in a row?

Probability Questions

2. A normal deck of cards has 52 cards in it. Assuming that you **DON'T** replace the card you had drawn before the next draw, what is the probability that you will draw three kings in a row?

P(draw_1 = king) = 4/52
P(draw_2 = king) = 3/51
P(draw_3 = king) = 2/50

P(3 kings) = P(draw_1 = king) x P(draw_2 = king) x P(draw_3 = king)
= 4/52 x 3/51 x 2/50

= 1/5,525

Probability Questions

3. Around 0.8% of men are blue-green colour-blind (the figure is slightly different for women) and roughly 1 in 5 men is left-handed. Assuming these characteristics are inherited independently, calculate the probability that a man chosen at random will:

(i) be both colour-blind and left-handed
(ii) be colour-blind and not left-handed
(iii) be colour-blind or left-handed
(iv) be neither colour-blind nor left-handed.

Probability Questions

3. Around 0.8% of men are blue-green colour-blind (the figure is slightly different for women) and roughly 1 in 5 men is left-handed. Assuming these characteristics are inherited independently, calculate the probability that a man chosen at random will:

Let's assume 1,000 men …

(i) be both colour-blind and left-handed

0.16% = 8/1000 x 200/1000 = 1.6/1000

(ii) be colour-blind and not left-handed

0.64% = 8/1000 x (1000 - 200)/1000 = 64/1000

(iii) be colour-blind or left-handed

20.64% = 8/1000 + 200/1000 - 1.6/1000 = 206.4/1000

(iv) be neither colour-blind nor left-handed.

79.36% = (1000 - 8)/1000 * (1000 - 200)/1000 = 793.6/1000

Probability Questions

3. Around 0.8% of men are blue-green colour-blind (the figure is slightly different for women) and roughly 1 in 5 men is left-handed. Assuming these characteristics are inherited independently, calculate the probability that a man chosen at random will:

(i) be both colour-blind and left-handed

$$0.16\% = 0.008 \times 0.2$$

(ii) be colour-blind and not left-handed

$$0.64\% = 0.008 \times (1 - 0.2)$$

(iii) be colour-blind or left-handed

$$20.64\% = 0.008 + 0.2 - 0.0016$$

(iv) be neither colour-blind nor left-handed.

$$79.36\% = (1 - 0.008) \times (1 - 0.2) = 1 - \text{(iii)}$$

Probability Questions

3. Around 0.8% of men are blue-green colour-blind (the figure is slightly different for women) and roughly 1 in 5 men is left-handed. Assuming these characteristics are inherited independently, calculate the probability that a man chosen at random will:

(i) be both colour-blind and left-handed

(ii) be colour-blind and not left-handed

(iii) be colour-blind or left-handed

(iv) be neither colour-blind nor left-handed.

	Color blind	Not color blind	
Left handed	0.16% (0.2*0.008)	19.84 (0.2*0.992)	20%
Not left handed	0.64% (0.8*0.008)	79.36% (0.8*0.992)	80%
	0.8%	99.2%	100%

Random Variables

A **random variable** is a variable that can take on more than one value.
When dealing with random variables, it is often useful to think about the **sample space** for that variable.

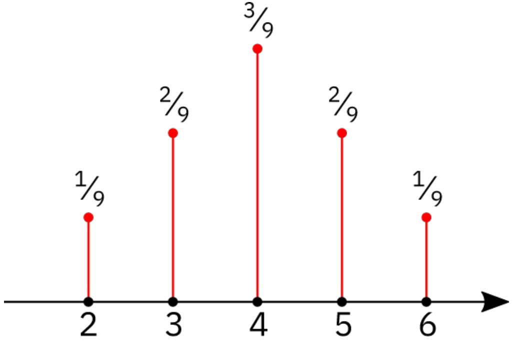
Examples

- Rolling a die
- Height of men

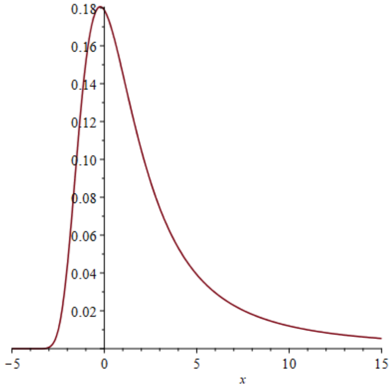
A **probability distribution** is a function that provides the likelihood of seeing each value in a random variable’s sample space.

Discrete vs. Continuous Probabilities

Discrete random variables are defined by a **probability mass function.**



Continuous random variables are defined by a **probability density function**.



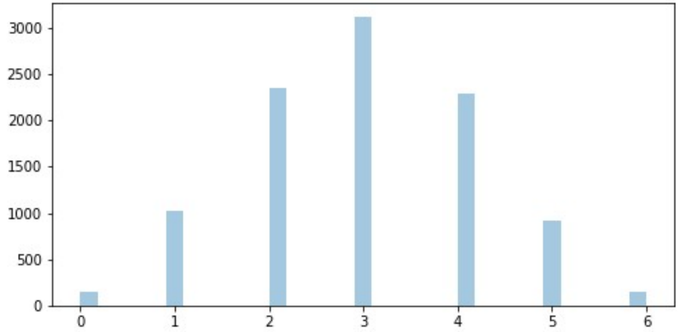
Probability Distributions

Probability distributions have a functional form that is defined by a set of parameters.

Binomial Distribution

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

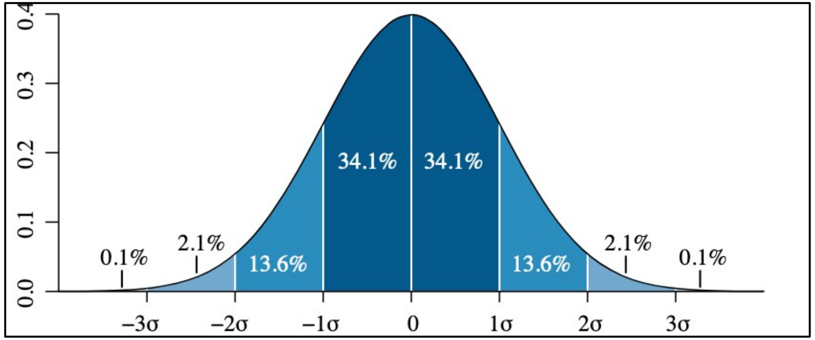
p is the probability of success (e.g., heads)
n is the number of trials



Normal Distribution

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

μ is the sample mean and
σ is the sample standard deviation



Probability Distributions

Binomial Distribution

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

Given n, choose x.
Also, n choose x, nCx,
nCk, nCr

Math Combination

Normal Distribution

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

e raised to the value

Euler's constant

Probability Distributions

There are many other probability distributions. We will discuss them and how to generate samples from probability distributions in Python in a later lecture.