Chapter 5

Probability

Randomness

- A situation where we know what outcomes could happen, but we don’t know which particular outcome did or will happen.
- However, we can calculate the probability with which each outcome will happen.
- People are not good at identifying truly random samples or random experiments, so we need to rely on outside mechanisms such as coin flips or random number tables.

Simulating Randomness

- Flip a coin to generate random 0’s and 1’s
- Pick a card to generate random numbers
- Pick a number out of a hat
- Use a Random Number table

Simulating Randomness

- Real randomness is hard to achieve without a computer or some other device that produces a truly random result
- Example: Simulate rolling a die 10 times using the following random number table.

Line						
28	31498	85304	22393	21634	34560	77404
29	93074	27086	62559	86590	18420	33290
30	90549	53094	76282	53105	45531	90061
31	11373	96871	38157	98368	39536	08079
32	52022	59093	30647	33241	16027	70336

- Pick any line to start, let's say we start with line 30
- Go through the numbers, picking numbers 1-6 and ignoring 0, 7, 8, 9.
- The "random" numbers are: 5,4,5,3,4,6,2,2,5,3

16% of cars fail pollution tests in California

We can represent this in many ways:

- In a bag of 100 chips, 16 are red and represent cars that fail pollution tests, the remaining 84 are black and represent cars that pass the tests.

16 84

- In a bag with 50 marbles, 8 are orange and represent cars that fail pollution tests, the remaining 42 are blue and represent cars that pass the tests.

8 42

- On a random number table, numbers 00-15 represent cars that fail pollution tests and 16-99 represent cars that pass the tests.

Estimating Probabilities via Simulation

We would like to estimate the probability that an entire fleet of seven cars would pass using a random number table. We assume each car is independent.

- Start at row 1 of the random number table and read two digits at a time.
- If the random number is between 00-15, the car will fail the pollution test.
- If the number is between 16-99, the car will pass the test.
- A fleet of cars is comprised of seven cars, i.e. seven 2 -digit random numbers. If all seven cars pass the test, record a 1, if not record a 0.
- Repeat many times and calculate the proportion of 1s among the total number of run, i.e. number of fleets where all cars passed the pollution test.

Run 1

7 0 3 6 6   *2 3 3 8 9*   *6 0 6 1*7   8 2 8 6 8   0 2 2 5 3
8 4 5 4 2   9 6 1 1 2   7 8 1 3 8   9 5 3 9 3   8 8 2 6 7
0 7 9 0 1   4 3 2 6 2   1 8 9 3 3   6 2 4 8 9   8 8 9 1 6

Run #	Simulation	# failed	Outcome
Run 1:	$\underbrace{70}_{PASS}\ \underbrace{36}_{PASS}\ \underbrace{62}_{PASS}\ \underbrace{33}_{PASS}\ \underbrace{89}_{PASS}\ \underbrace{60}_{PASS}\ \underbrace{61}_{PASS}$	0	1

Run 2

7 0 3 6 6   2 3 3 8 9   6 0 6 1 *7*   *8 2 8 6 8*   *0 2 2 5 3*
8 4 5 4 2   9 6 1 1 2   7 8 1 3 8   9 5 3 9 3   8 8 2 6 7
0 7 9 0 1   4 3 2 6 2   1 8 9 3 3   6 2 4 8 9   8 8 9 1 6

Run #	*Simulation*	*# failed*	*Outcome*
Run 1:	$\underbrace{70}_{PASS}\ \underbrace{36}_{PASS}\ \underbrace{62}_{PASS}\ \underbrace{33}_{PASS}\ \underbrace{89}_{PASS}\ \underbrace{60}_{PASS}\ \underbrace{61}_{PASS}$	0	1
Run 2:	$\underbrace{78}_{PASS}\ \underbrace{28}_{PASS}\ \underbrace{68}_{PASS}\ \underbrace{02}_{FAIL}\ \underbrace{25}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{45}_{PASS}$	1	0

Run 3

Clicker!

7 0 3 6 6	2 3 3 8 9	6 0 6 1 7	8 2 8 6 8	0 2 2 5 3
8 4 5 *4 2*	*9 6 1 1 2*	*7 8 1 3 8*	*9 5* 3 9 3	8 8 2 6 7
0 7 9 0 1	4 3 2 6 2	1 8 9 3 3	6 2 4 8 9	8 8 9 1 6

Did all the cars pass in run 3?

Answer:

A. Yes

B. No

Run 3

7 0 3 6 6	2 3 3 8 9	6 0 6 1 7	8 2 8 6 8	0 2 2 5 3
8 4 5 *4 2*	*9 6 1 1 2*	*7 8 1 3 8*	*9 5* 3 9 3	8 8 2 6 7
0 7 9 0 1	4 3 2 6 2	1 8 9 3 3	6 2 4 8 9	8 8 9 1 6

Did all the cars pass in run 3?

Answer:
A. Yes
B. No

Run 4

Clicker!

7 0 3 6 6	2 3 3 8 9	6 0 6 1 7	8 2 8 6 8	0 2 2 5 3
8 4 5 4 2	9 6 1 1 2	7 8 1 3 8	9 5 *3 9 3*	*8 8 2 6 7*
0 7 9 0 1	*4* 3 2 6 2	1 8 9 3 3	6 2 4 8 9	8 8 9 1 6

Did all the cars pass in run 4?

Answer:

A. Yes

B. No

Run 4

7 0 3 6 6	2 3 3 8 9	6 0 6 1 7	8 2 8 6 8	0 2 2 5 3
8 4 5 4 2	9 6 1 1 2	7 8 1 3 8	9 5 *3 9 3*	*8 8 2 6 7*
0 7 9 0 1	*4* 3 2 6 2	1 8 9 3 3	6 2 4 8 9	8 8 9 1 6

Did all the cars pass in run 4?

Answer:

A. Yes

B. No

Estimation of Probability

Based on the simulation results, estimate the probability that an entire fleet of seven cars would pass the pollution test. What proportion of runs got outcome "1"?

Run #	Simulation	# failed	Outcome
Run 1:	$\underbrace{70}_{PASS}$ $\underbrace{36}_{PASS}$ $\underbrace{62}_{PASS}$ $\underbrace{33}_{PASS}$ $\underbrace{89}_{PASS}$ $\underbrace{60}_{PASS}$ $\underbrace{61}_{PASS}$	0	1
Run 2:	$\underbrace{78}_{PASS}$ $\underbrace{28}_{PASS}$ $\underbrace{68}_{PASS}$ $\underbrace{02}_{FAIL}$ $\underbrace{25}_{PASS}$ $\underbrace{38}_{PASS}$ $\underbrace{45}_{PASS}$	1	0
Run 3:	$\underbrace{42}_{PASS}$ $\underbrace{96}_{PASS}$ $\underbrace{11}_{FAIL}$ $\underbrace{27}_{PASS}$ $\underbrace{81}_{PASS}$ $\underbrace{38}_{PASS}$ $\underbrace{95}_{PASS}$	1	0
Run 4:	$\underbrace{39}_{PASS}$ $\underbrace{38}_{PASS}$ $\underbrace{82}_{PASS}$ $\underbrace{67}_{PASS}$ $\underbrace{07}_{FAIL}$ $\underbrace{90}_{PASS}$ $\underbrace{14}_{FAIL}$	2	0
Total		4	1

Estimation of Probability

What proportion of runs got outcome “1”?

$$= \frac{\text{Outcome "1"}}{\text{Total \# Runs}} = \frac{1}{4}$$

Run #	Simulation	# failed	Outcome
Run 1:	$\underbrace{70}_{PASS}\ \underbrace{36}_{PASS}\ \underbrace{62}_{PASS}\ \underbrace{33}_{PASS}\ \underbrace{89}_{PASS}\ \underbrace{60}_{PASS}\ \underbrace{61}_{PASS}$	0	1
Run 2:	$\underbrace{78}_{PASS}\ \underbrace{28}_{PASS}\ \underbrace{68}_{PASS}\ \underbrace{02}_{FAIL}\ \underbrace{25}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{45}_{PASS}$	1	0
Run 3:	$\underbrace{42}_{PASS}\ \underbrace{96}_{PASS}\ \underbrace{11}_{FAIL}\ \underbrace{27}_{PASS}\ \underbrace{81}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{95}_{PASS}$	1	0
Run 4:	$\underbrace{39}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{82}_{PASS}\ \underbrace{67}_{PASS}\ \underbrace{07}_{FAIL}\ \underbrace{90}_{PASS}\ \underbrace{14}_{FAIL}$	2	0
Total		4	1

Estimation of Probability

Based on the simulation results, estimate the average number of cars that fail the pollution test per run.

Run #	Simulation	# failed	Outcome
Run 1:	$\underbrace{70}_{PASS}\ \underbrace{36}_{PASS}\ \underbrace{62}_{PASS}\ \underbrace{33}_{PASS}\ \underbrace{89}_{PASS}\ \underbrace{60}_{PASS}\ \underbrace{61}_{PASS}$	0	1
Run 2:	$\underbrace{78}_{PASS}\ \underbrace{28}_{PASS}\ \underbrace{68}_{PASS}\ \underbrace{02}_{FAIL}\ \underbrace{25}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{45}_{PASS}$	1	0
Run 3:	$\underbrace{42}_{PASS}\ \underbrace{96}_{PASS}\ \underbrace{11}_{FAIL}\ \underbrace{27}_{PASS}\ \underbrace{81}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{95}_{PASS}$	1	0
Run 4:	$\underbrace{39}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{82}_{PASS}\ \underbrace{67}_{PASS}\ \underbrace{07}_{FAIL}\ \underbrace{90}_{PASS}\ \underbrace{14}_{FAIL}$	2	0
Total		4	1

Estimation of Probability

The average number of cars that fail the pollution test per run is

$$= \frac{0+1+1+2}{\text{Total \# Runs}} = \frac{4}{4} = 1$$

Run #	*Simulation*	*# failed*	*Outcome*
Run 1:	$\underbrace{70}_{PASS}\ \underbrace{36}_{PASS}\ \underbrace{62}_{PASS}\ \underbrace{33}_{PASS}\ \underbrace{89}_{PASS}\ \underbrace{60}_{PASS}\ \underbrace{61}_{PASS}$	0	1
Run 2:	$\underbrace{78}_{PASS}\ \underbrace{28}_{PASS}\ \underbrace{68}_{PASS}\ \underbrace{02}_{FAIL}\ \underbrace{25}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{45}_{PASS}$	1	0
Run 3:	$\underbrace{42}_{PASS}\ \underbrace{96}_{PASS}\ \underbrace{11}_{FAIL}\ \underbrace{27}_{PASS}\ \underbrace{81}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{95}_{PASS}$	1	0
Run 4:	$\underbrace{39}_{PASS}\ \underbrace{38}_{PASS}\ \underbrace{82}_{PASS}\ \underbrace{67}_{PASS}\ \underbrace{07}_{FAIL}\ \underbrace{90}_{PASS}\ \underbrace{14}_{FAIL}$	2	0
Total		4	1

Notes

- 4 runs is usually not considered sufficient. For this class use at least 5 runs.
- You can start at any row you like on the random number table, but you should make sure to note it when you're writing up your simulation scheme.
- You should not arbitrarily pick numbers from the random number table. Just pick a row and follow across. Otherwise the number you're using won't be random (they will instead be your choices).

Randomness to Probability

- Let's say there are 1000 songs in your iTunes library. 20 of these songs are by Beyonce.

- What is the probability that the next time you hit shuffle you get a Beyonce song?

Answer:

Clicker!

A. 0.04
B. 0.02
C. 0.40
D. 0.20

Randomness to Probability

- Let's say there are 1000 songs in your iTunes library. 20 of these songs are by Beyonce.
- What is the probability that the next time you hit shuffle you get a Beyonce song?

Prob(Beyonce song) = $\frac{20}{1000} = 0.02$

Definitions of Probability

- Probability is used to measure how often random events occur.
- There are two kinds of probabilities: theoretical and empirical.
- Theoretical probability is the relative frequency at which an event happens after infinitely many repetitions.
    - We may not be able to predict which song we play each time we hit shuffle, however we know that in the long run 20 out of 1000 songs will be Beyonce songs.
- Empirical probability is the relative frequency based on an experiment or on observations of a real-life process.
    - We listened to 100 songs on shuffle and 4 of them are Beyonce songs. The empirical probability is 4/100 = 0.04.

Definitions of Probability

- Theoretical probabilities are very abstract and can take forever to compute.
- Hence, we can use empirical probabilities to estimate and test theoretical probabilities.
- Simulations are experiments used to produce empirical probabilities.

Definitions of Probability

- For any random phenomenon, each attempt is called a trial and each trial generates an outcome.
    - Each time you hit shuffle is a trial.
    - The song that plays as a result of hitting shuffle is an outcome.
- Sample space is the collection of all possible outcomes of a trial.
    - Sample space is the entire iTunes library of 1000 songs.
- A collection of outcomes is called an event.
    - For example, playing two Beyonce songs in a row in shuffle.

Sample Space

Clicker!

- A couple has two kids (no twins), what is the sample space for the sex of these kids?
  (a) $S = \{B, G\}$
  (b) $S = \{BB, GG\}$
  (c) $S = \{BB, GG, GB\}$
  (d) $S = \{BB, GG, GB, BG\}$

Sample Space

- A couple has two kids (no twins), what is the sample space for the sex of these kids?

  (a) S = {B, G}

  (b) S = {BB, GG}

  (c) S = {BB, GG, GB}

  (d) S = {BB, GG, GB, BG}

Calculating Probabilities

Probability of A is:

$$P(A) = \frac{\text{Number of outcomes in A}}{\text{Number of all possible outcomes}}$$

This is true only for equally likely outcomes.

Calculating Probabilities

If a couple has two kids, what is the probability that both are boys?

If a couple has two kids, the list of possible scenarios (sample space) is:

BB, GG, BG, GB

Since each scenario is equally likely:

Prob(BB) = 1/4 = 0.25

Clicker Test

Clicker!

- Freebie! Click in choice C!

Independence

- When thinking about what happens with combinations of outcomes, things are simplified if the individual trials are independent.
- Roughly speaking, this means that the outcome of one trial doesn't influence or change the outcome of another trial.
- If the iTunes shuffle is truly random then the songs played are independent of each other.
- In other words, the iTunes shuffle is memoryless, it doesn't say to itself "Wait, I just played a Beyonce song, I shouldn't play another one."
- Similarly, if the genders of kids a couple has are independent, then the probability of having a boy for a second child doesn't change based on whether or not the first child of the couple was a girl.

Coin Toss Example

- Trial: Each coin toss
- Outcome: Heads or Tails
- Probability: P(Heads) = 0.5 and P(Tails) = 0.5

  Note: Each time we toss a coin we can't tell which side will come up, however in the long run tails will come up 50% of the time and heads will come up 50% of the time.

- Sample Space:

  Tossed once: S = {H, T}

  Tossed twice: S = {HT, TH, HH, TT}

- Independence: The outcome of one coin toss does not affect the outcome of the next coin toss.

Dice Example

- Trial:
- Outcome:
- Probability:
- Sample Space:
- Independence:

Clicker!

Click A when you're done!

Dice Example

- Trial: Each die roll
- Outcome: A die has six sides
- Probability: Prob(rolling a side)=1/6

  Note: With a fair die the probability of getting each side is the same and is 1/6.

- Sample Space:

  Rolled once: S = {1, 2, 3, 4, 5, 6}

- Independence: The outcome of one die roll does not affect the outcome of the next die roll.

Deck of Cards Example

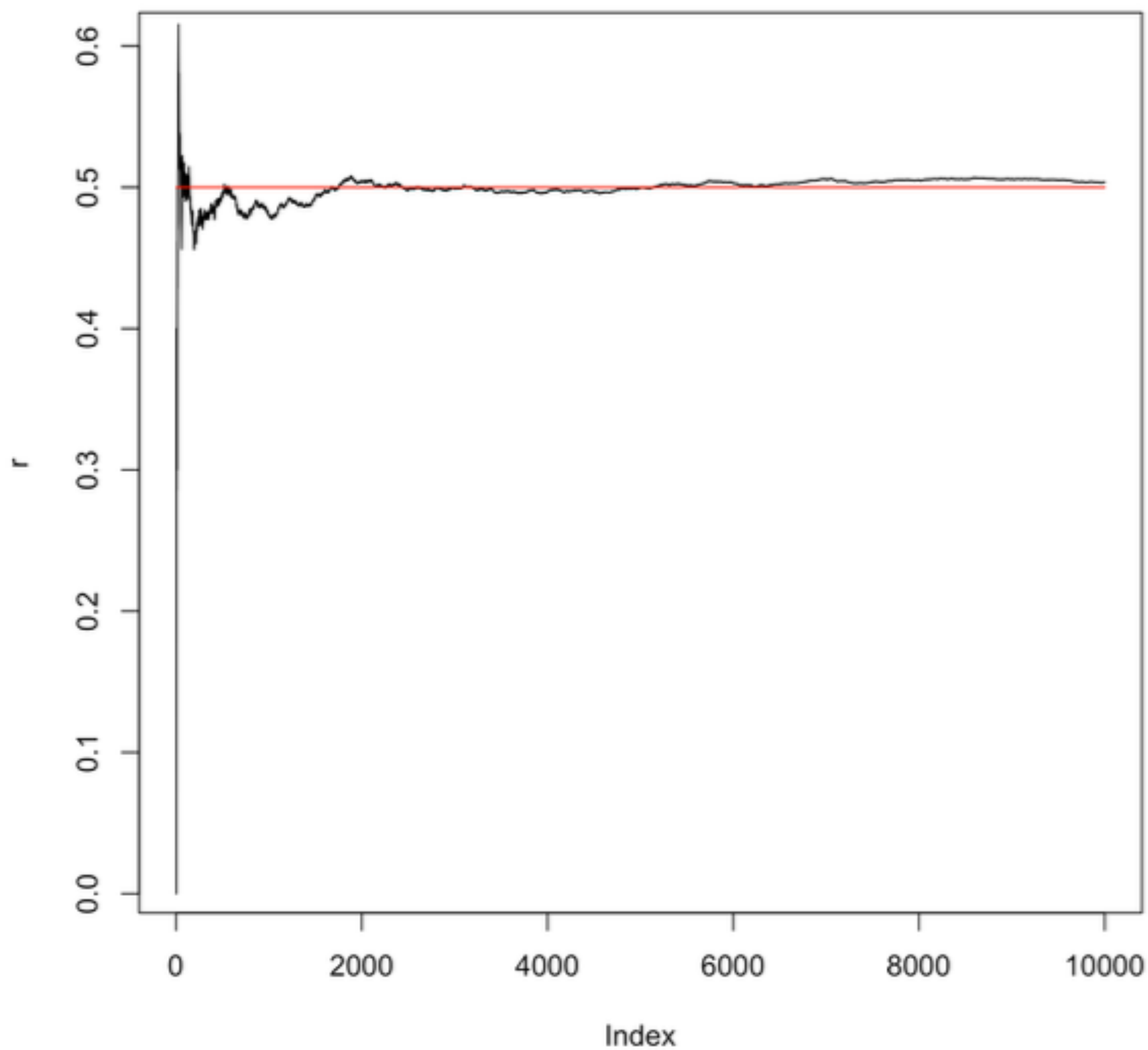
- Trial: Each draw of a card
- Outcome: There are 52 cards, 4 suits (clubs, spades, diamonds, hearts), each having 13 cards (one ace, numbers 2-10, one jack, one queen, one king)
- Probability: P(individual card)=1/52
- Sample Space:

  One draw: $S = \{13\clubsuit, 13\spadesuit, 13\diamondsuit, 13\heartsuit\}$
- Independence: The outcome of one card draw does not affect the outcome of the next card draw.

Law of Large Numbers

- The Law of Large Numbers says that the long-run relative frequency of repeated independent events gets closer and closer to the true relative frequency as the number of trials increases. After many, may repetitions, the observed frequency will be close to the true frequency.
    - If a coin is tossed many times, the overall percentage of heads should settle down to about 50% as the number of tosses increases.
- This is why we define probability as the long run relative-frequency of an event.
- The common (mis)understanding of the law of large numbers is that random phenomena are supposed to compensate for whatever happened in the past; this is just not true (also called gambler's fallacy or law of averages).

Law of Large Numbers



Law of Large Numbers

- When tossing a fair coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss?

H H H H H H H H H H

- The probability is still 0.5 or there is still a 50% chance that another head will come up on the next toss.
- Prob(H on 11th toss) = Prob(T on 11th toss) = 0.5
- The coin is <u>not</u> due for a tails.

Probability Properties

- Probabilities must be between 0 and 1 (or 0% to 100%), inclusive. It may be expressed as a fraction, a decimal, or a precent.
- A probability of 0 indicates impossibility.
- A probability of 1 indicated certainty.

$$0 \leq \text{Prob(an event)} \leq 1$$

Note: If you calculate a probability and get a negative number or a number greater than 1, there is a mistake!

Probability Properties

- The probability of the set of all possible outcomes of a trial must be equal to 1.

$$\text{Prob(sample space)} = 1$$

- Coin toss: Prob(H) + Prob(T) = 0.5 +0.5 = 1
- Cards: Prob(face card) + Prob(non-face card)

  = 12/52 + 40/52 = 52/52 = 1

Probability Rules

- **Not** P(not A) means A does not happen
- **And** P(A and B) = P(both A and B happen)
- **Or** P(A or B) = P(A happens, or B happens, or both happen)

Example

		politics					
		conservative	liberal	moderate	very conservative	very liberal	row summary
death penalty	favor	13	37	59	2	2	113
	no opinion	4	24	28	2	1	59
	oppose	7	25	48	3	6	89
	strongly favor	6	7	18	7	1	39
	strongly oppose	1	14	10	0	7	32
	column summary	31	107	163	14	17	332

Let’s select one of these 332 people at random.

A = Event person is liberal

B = Event person favors the death penalty

Example

		politics					
		conservative	liberal	moderate	very conservative	very liberal	row summary
death penalty	favor	13	37	59	2	2	113
	no opinion	4	24	28	2	1	59
	oppose	7	25	48	3	6	89
	strongly favor	6	7	18	7	1	39
	strongly oppose	1	14	10	0	7	32
	column summary	31	107	163	14	17	332

P(A) = P(liberal) = 107/332 = 32%

P(B) = P(favors death penalty) = 113/332 = 34%

P(A and B) = P(liberal and favors death penalty)

= 37/332 = 11%

Example

counted twice

		politics					
		conservative	liberal	moderate	very conservative	very liberal	row summary
death penalty	favor	13	37	59	2	2	113
	no opinion	4	24	28	2	1	59
	oppose	7	25	48	3	6	89
	strongly favor	6	7	18	7	1	39
	strongly oppose	1	14	10	0	7	32
	column summary	31	107	163	14	17	332

P(A or B) = P(liberal or favors death penalty)

$= P(A) + P(B) - P(A \text{ and } B)$

$= (107/332) + (113/332) - (37/332) = 183/332$

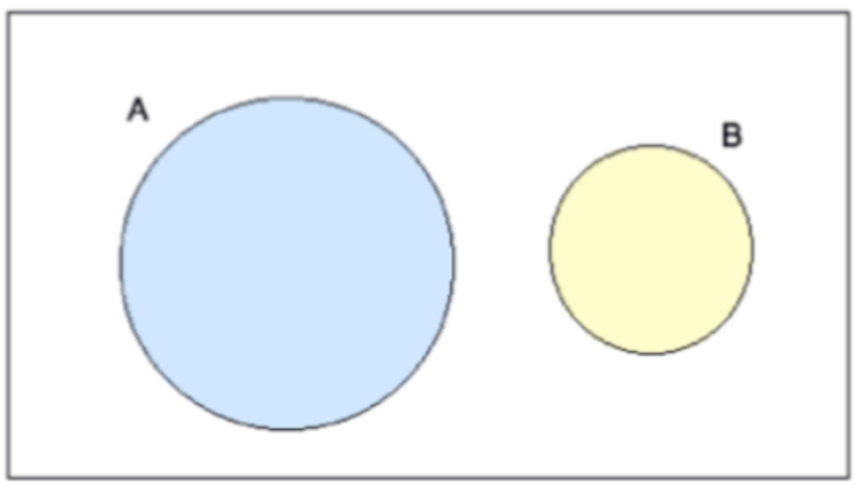
$= 55\%$

Disjoint Events

- Events that have no outcomes in common (and, thus, cannot occur together) are called disjoint (or mutually exclusive).

  P(A and B) = 0

  - The outcome of a coin toss cannot be a head and a tail.
  - A student cannot fail and pass a class.
  - A card drawn from a deck cannot be an ace and a queen.



Example

Clicker!

		politics					
		conservative	liberal	moderate	very conservative	very liberal	row summary
death penalty	favor	13	37	59	2	2	113
	no opinion	4	24	28	2	1	59
	oppose	7	25	48	3	6	89
	strongly favor	6	7	18	7	1	39
	strongly oppose	1	14	10	0	7	32
	column summary	31	107	163	14	17	332

What's the probability the person selected is both a liberal and a conservative?

A. 37/332

B. 13/332

C. 0

Example

		politics					
		conservative	liberal	moderate	very conservative	very liberal	row summary
death penalty	favor	13	37	59	2	2	113
	no opinion	4	24	28	2	1	59
	oppose	7	25	48	3	6	89
	strongly favor	6	7	18	7	1	39
	strongly oppose	1	14	10	0	7	32
	column summary	31	107	163	14	17	332

What's the probability the person selected is both a liberal and a conservative?

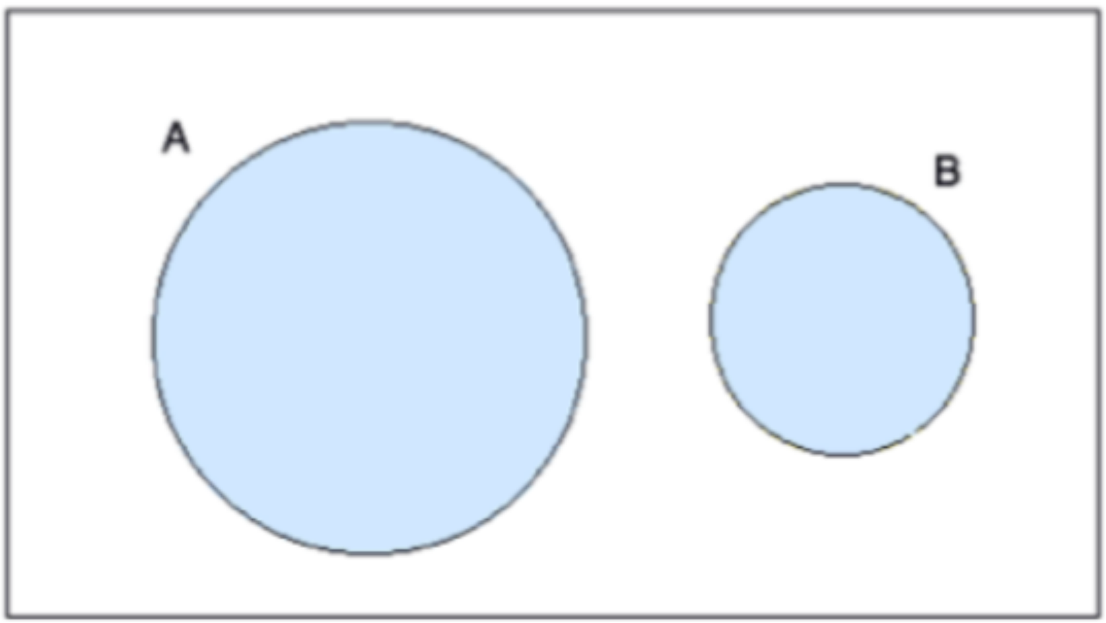
P(liberal and conservative) = 0

These two events are disjoint or mutually exclusive.

Addition Rule

- For two disjoint events A and B, the probability that one OR the other occurs is the sum of probabilities of the two events.

Prob(A OR B) = Prob(A) + Prob(B)



Note: This is also called a union, denoted by $\cup$.

Example

Roll a fair, six-sided die.

- What is the probability that the die shows an odd number OR a number greater than 3 on top?

Example

- What is the probability that the die shows an odd number OR a number greater than 3 on top?

The probability of getting an odd number (1, 3, 5) is 3/6. The probability of getting a number greater than 3 (4, 5, 6) is 3/6. The probability of getting an odd number and greater than 3 (just 5) is 1/6.

P(odd OR greater than 3) = P(odd) + P(greater than 3) - P(odd AND greater than 3)

= 3/6 + 3/6 - 1/6

= 5/6

Example

Clicker!

Roll a fair, six-sided die.

- What is the probability that the die shows an odd number OR the number 2 on top?

  A. 3/6 + 1/6 - 1/6

  B. 3/6 + 1/6 - 0

  C. 3/6 + 3/6 - 0

  D. 3/6 + 3/6 - 1/6

Example

- What is the probability that the die shows an odd number OR the number 2 on top?

The probability of getting an odd number (1, 3, 5) is 3/6. The probability of getting the number 2 is 1/6. The probability of getting an odd number and the number 2 is 0 (mutually exclusive).

P(odd OR the number 2) = P(odd) + P(the number 2) - P(odd AND the number 2)

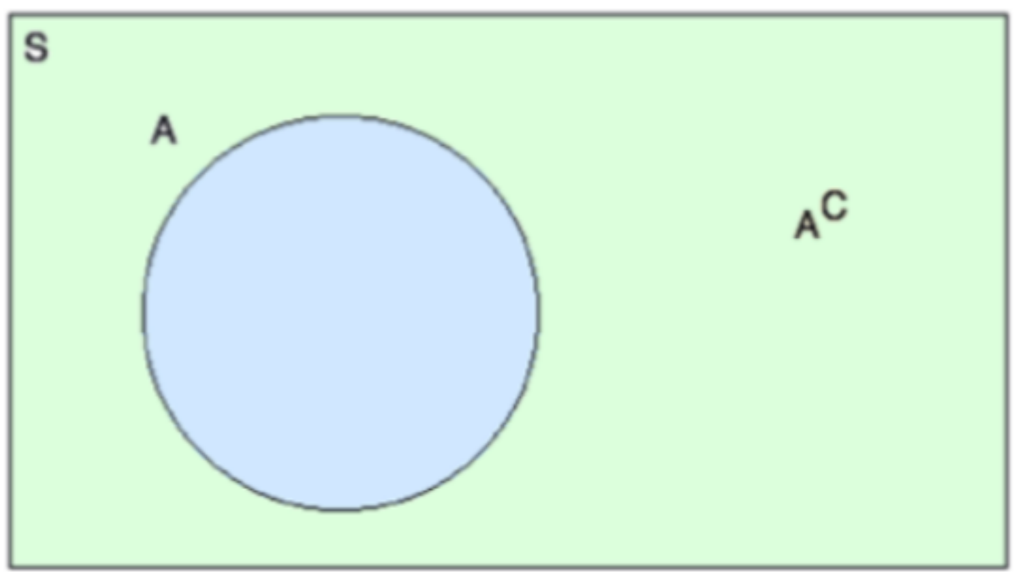
= 3/6 + 1/6 - 0

= 4/6

Complement Rule

- The set of outcomes that are not in the event A is called the complement of A denoted A^C.
- The probability of an event not occurring is 1 minus the probability that it does occur.

$$\text{Prob}\left(A^C\right) = 1 - \text{Prob}\left(A\right)$$



Example

Clicker!

A gumball machine contains gumballs of five different colors: 36 red, 44 white, 15 blue, 20 green and 5 orange. The machine dispenser randomly selects one gumball.

- What is the probability that the gumball selected is green?

  A. 20/120

  B. 20/80

  C. 40/120

Example

A gumball machine contains gumballs of five different colors: 35 red, 45 white, 15 blue, 20 green and 5 orange. The machine dispenser randomly selects one gumball.

- What is the probability that the gumball selected is green?

There are a total of 120 gumballs, 20 of them are green.

P(green gumball) = 20/120 = 1/6

Example

Clicker!

A gumball machine contains gumballs of five different colors: 36 red, 44 white, 15 blue, 20 green and 5 orange. The machine dispenser randomly selects one gumball.

- What is the probability that the gumball selected is not green?

  A. 20/120

  B. 5/6

  C. 1/6

Example

A gumball machine contains gumballs of five different colors: 35 red, 45 white, 15 blue, 20 green and 5 orange. The machine dispenser randomly selects one gumball.

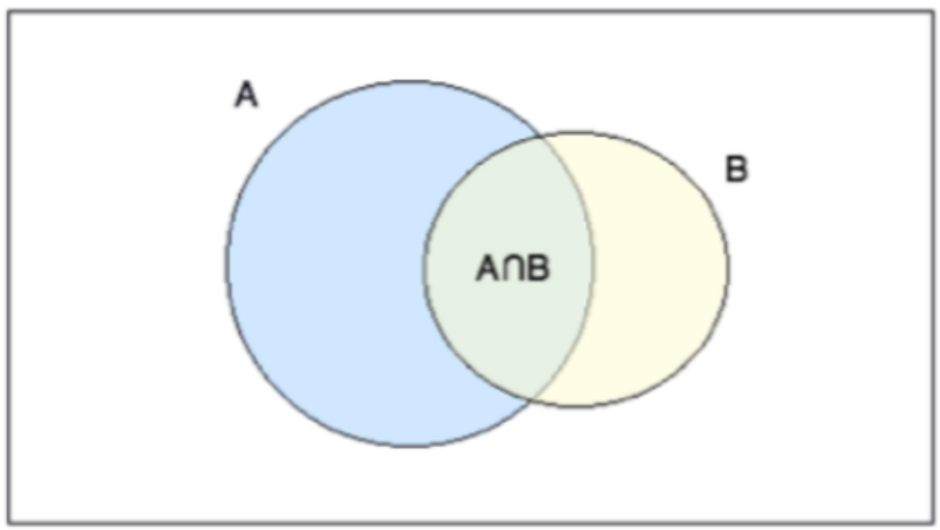
- What is the probability that the gumball selected is not green?

P(not green) = 1- 20/120 = 1-1/6 = 5/6

Multiplication Rule

- Sometimes we want to find the probabilities for a sequence of events.
- For two independent events A and B, the probability that both A AND B occur is the product of the probabilities of the two events.

Prob(A AND B) = Prob(A) x Prob(B)



Note: This is also called the intersection, denoted by $\cap$.

Example

Clicker!

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability she gets the *first* question right?

  A. 1/4

  B. 1/5

  C. 3/5

  D. 3/4

Example

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability she gets the *first* question right?

Prob(1st question right) = 1/4 = 0.25

Example

Clicker!

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability she gets the *2nd* question wrong?

  A. 1/4

  B. 1/5

  C. 3/5

  D. 3/4

Example

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability she gets the *2nd* question wrong?

Prob(2nd question wrong) = 3/4 = 0.75

Example

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability that the first question she gets right is the *5th* question?

Example

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability that the first question she gets right is the *5th* question?

Prob(Wrong, Wrong, Wrong, Wrong, Right)

$= 0.75 \times 0.75 \times 0.75 \times 0.75 \times 0.25$

$= 0.0791$

Example

Clicker!

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability that she gets all the questions right?

  A. 0.75 x 0.75 x 0.75 x 0.75 x 0.75

  B. 0.25 x 0.25 x 0.25 x 0.25 x 0.25

Example

In a multiple choice exam there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all, and decided to randomly guess the answers.

- What is the probability that she gets all the questions right?

  Prob(Right, Right, Right, Right, Right)

  $= 0.25 \times 0.25 \times 0.25 \times 0.25 \times 0.25$

  $= 0.00098$

Example

- What is the probability that she gets at least one question right?

If there are 5 questions, the possible number of questions she gets right are:

$$S = \{0, 1, 2, 3, 4, 5\}$$

We are interested in instances where she gets at least 1 question right:

$$S = \{0, 1, 2, 3, 4, 5\}$$

So we can divide up the sample space into two categories:

$$S = \{0, \text{at least } 1\}$$

Example

Since the probability of the sample space must add up to 1:

Prob(at least 1 right) = 1 - Prob(none right)

= 1- Prob(all wrong)

$$= 1 - (0.75 \times 0.75 \times 0.75 \times 0.75 \times 0.75)$$

$$= 1 - 0.75^5$$

$$= 1 - 0.2373$$

$$= 0.7627$$

Independent vs Disjoint

- Can two events be independent and disjoint?

  Remember: Independence means that the outcome of one trial doesn't influence the outcome of another. Disjoint means that two events can't happen at the same time.

Disjoint events are NOT independent

Example: Flipping heads or tails are two disjoint events. If we know that the outcome of a coin toss is heads, then we know that it is not tails. So whether or not the outcome is tails depends on whether or not the outcome is heads.

Independent vs Disjoint

Non disjoint events may or may not be independent

Two events can happen at the same time but still have no effect on each other.

Example: Let’s say we pick two random people in the class who scored 100% on the midterm. They could be complete strangers whose performance had nothing to do with each other, or they could’ve been close friends who studied together a lot.

Associations in Categorical Variables

- By looking at the table below, is there an association between martial status and having a college education?
- If there is an association, the probability that a randomly selected college-educated person is married would be different from the probability that a person with less than a college education is married.

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	70	10	28	125
High school	68	240	59	30	397
College or higher	27	98	15	3	143
Total	112	408	84	61	665

Conditional Probabilities

- What is the probability that a person is college-educated AND is married?
- What is the probability that a college-educated person is married?
- Is there a difference between these two questions?

Conditional Probabilities

- What's the difference?
  - What is the probability that a person is college-educated AND is married?

    We are looking at everyone in the sample.

  - What is the probability that a college-educated person is married?

    We are looking at only the people with college degrees.

Conditional Probabilities

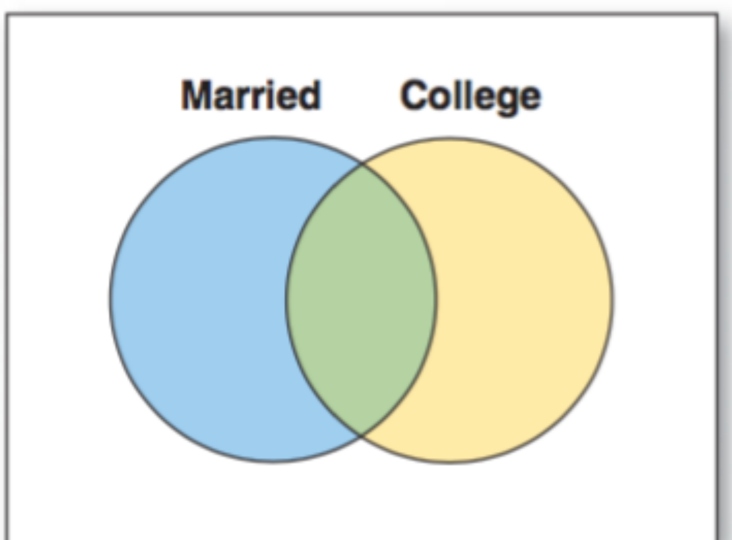
- Conditional probabilities are probabilities where we focus on just one group of objects and imagine taking a random sample from that group alone.

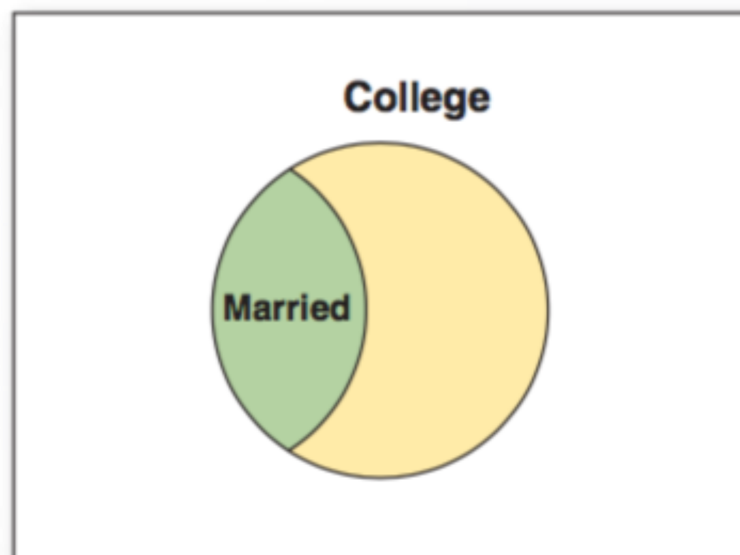
$$P(A|B) = \frac{P(A \text{ AND } B)}{P(B)}$$

- The line between A and B denotes “given”. The probabilities that A occurs given that we know B occurs.

Conditional Probabilities

- What is the probability that a college-educated person is married?
- Another way to phrase the question is “What is the probability that a person is married given that the person has a college degree?”





Conditional Probabilities

- "What is the probability that a person is married given that the person has a college degree?"

$$P(\text{person is married} \mid \text{person has college degree}) = \frac{98/665}{143/665} = \frac{98}{143}$$

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	70	10	28	125
High school	68	240	59	30	397
College or higher	27	98	15	3	143
Total	112	408	84	61	665

Example

Clicker!

- If a person is randomly chosen from this sample, what is the probability that the person owns a dog given that the person is a male?

	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

Answer:

A. 42/100
B. 52/100
C. 42/52
D. 0

Example

- If a person is randomly chosen from this sample, what is the probability that the person owns a dog given that the person is a male?

$$P(\text{dog owner} \mid \text{male}) = \frac{P(\text{dog owner AND male})}{P(\text{male})}$$

	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

$$= \frac{42/100}{52/100} = \frac{42}{52}$$

Example

Clicker!

In a class where everyone's native language is English, 70% of students speak Spanish as a second language, 45% speak French as a second language, and 20% speak both. Assume that there are no students who speak another second language. What is the probability that a randomly selected student speaks French given that they speak Spanish?

Answer:

A. .45/.70
B. .20/.70
C. .20/.45
D. 0

Example

In a class where everyone's native language is English, 70% of students speak Spanish as a second language, 45% speak French as a second language, and 20% speak both. Assume that there are no students who speak another second language. What is the probability that a randomly selected student speaks French given that they speak Spanish?

$$\mathrm{P}(\text{French} \mid \text{Spanish}) = \frac{\mathrm{P}(\text{French AND Spanish})}{\mathrm{P}(\text{Spanish})}$$

$$= \frac{0.20}{0.70} \approx 0.29$$

Probability Rules

You can use the following formula when calculating conditional probabilities:

$$P(A \mid B) = \frac{P(A \text{ AND } B)}{P(B)}$$

Another way of finding AND probabilities:

$$P(A \text{ AND } B) = P(A) \times P(B \mid A)$$

$$P(A \text{ AND } B) = P(B) \times P(A \mid B)$$

It doesn't matter which event is called A and which is B.

Flipping the Condition

Some common mistakes with conditional probabilities that you should be careful of:

$$P(B|A) \neq P(A|B)$$

$$P(B|A) \neq \frac{1}{P(A|B)}$$

Clicker Test

Clicker!

- Freebie! Click in choice A!

Independence

- Independent events are variables or events that are not associated.
- Two events are independent if knowledge that one event has happened tells you nothing about whether or not the other event has happened.

A and B are independent if the following occur:

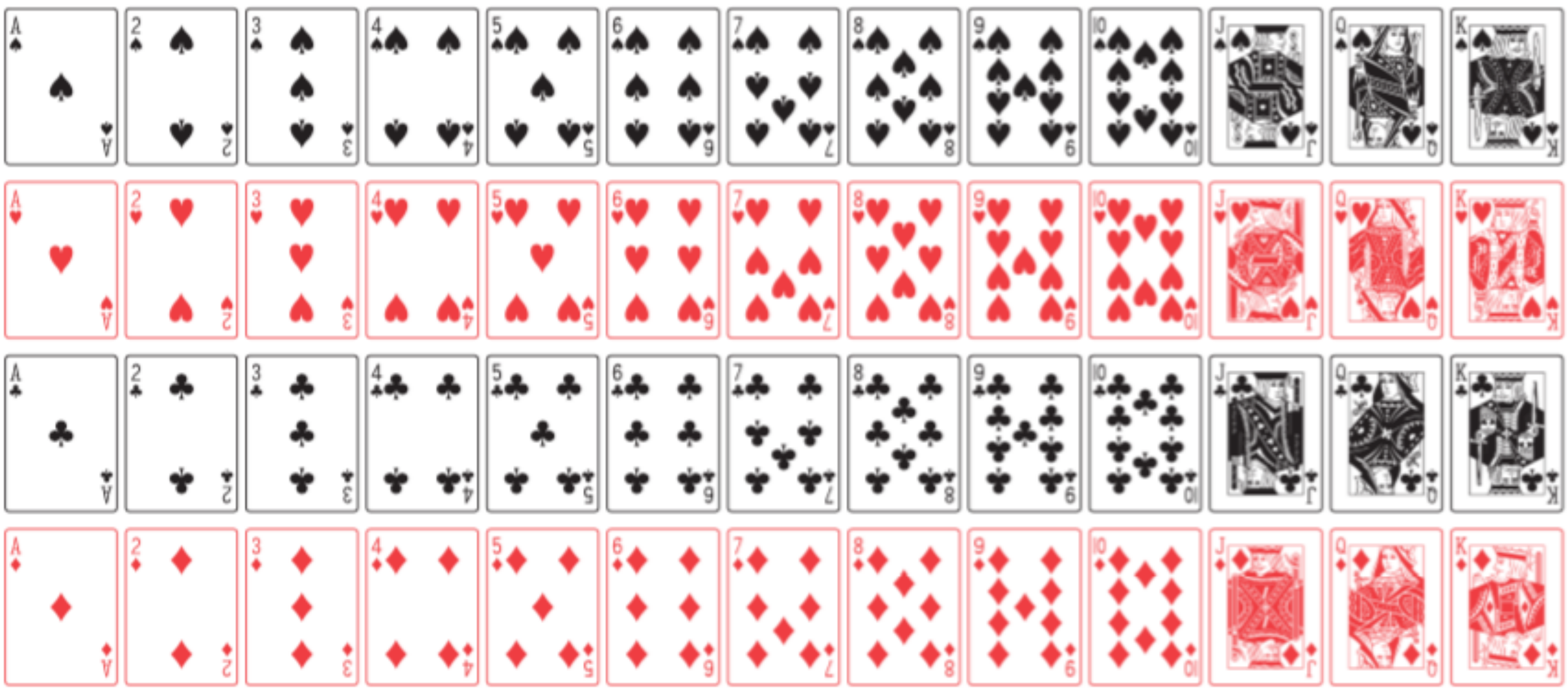
$$P(A) \times P(B) = P(A \text{ and } B)$$

$$P(B|A) = P(B)$$

$$P(A|B) = P(A)$$

Example

Suppose a deck of cards is shuffled and one card is dealt facedown on the table. Are the events “card is a diamond” and “card is red” independent?



Example

Suppose a deck of cards is shuffled and one card is dealt facedown on the table. Are the events “card is a diamond” and “card is red” independent?

Compare P(diamond) to P(diamond|red).

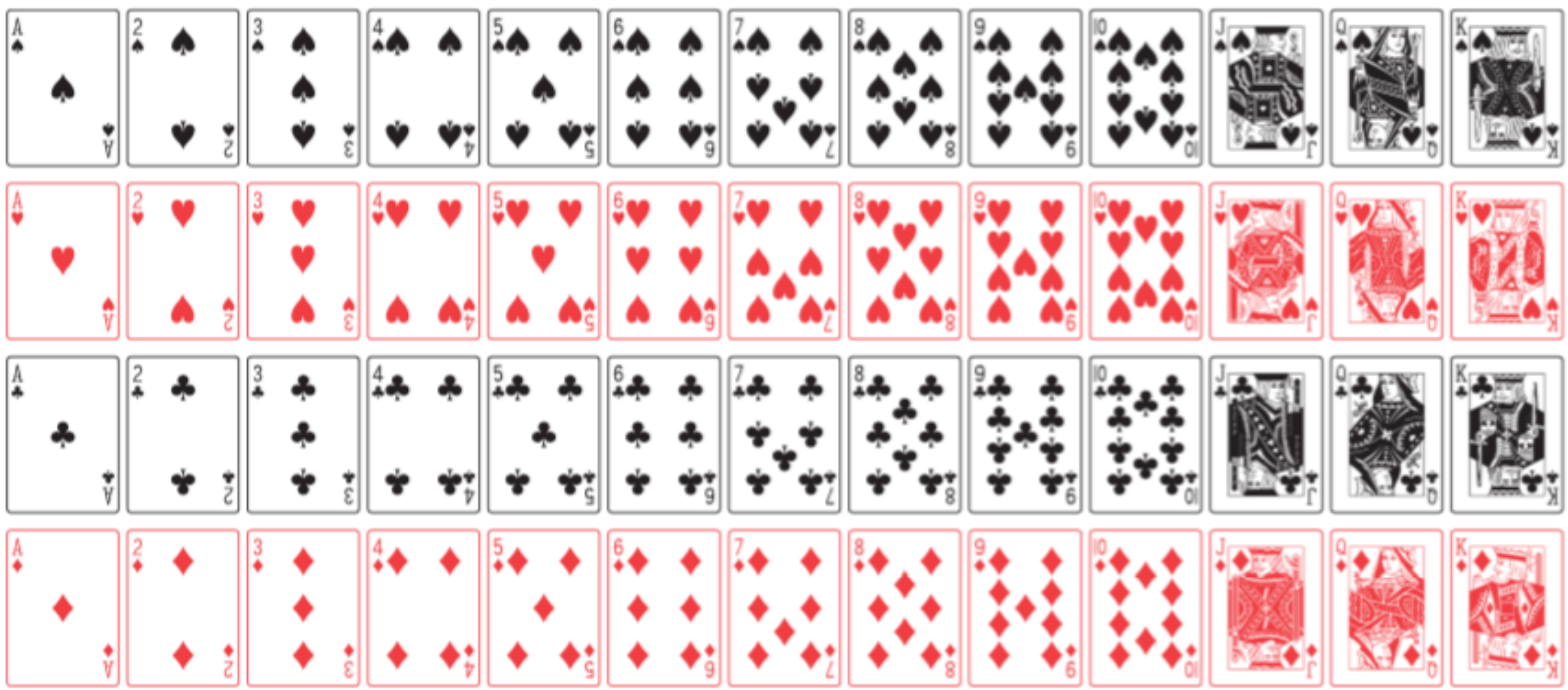
$$P(\text{diamond}) = \frac{13}{52} = \frac{1}{4}$$

$$P(\text{diamond}|\text{red}) = \frac{13/52}{26/52} = \frac{13}{26} = \frac{1}{2}$$

The probabilities are not equal so the two events are associated. They are not independent.

Example

Suppose a deck of cards is shuffled and one card is dealt facedown on the table. Are the events “card is a diamond” and “card is a queen” independent?



Example

Suppose a deck of cards is shuffled and one card is dealt facedown on the table. Are the events “card is a diamond” and “card is a queen” independent?

Compare P(diamond) to P(diamond|queen).

$$\mathrm{P}(\text{diamond}) = \frac{13}{52} = \frac{1}{4}$$

$$\mathrm{P}(\text{diamond} \mid \text{queen}) = \frac{1/52}{4/52} = \frac{1}{4}$$

The probabilities are equal so the two events are independent. You can also compare P(queen) to P(queen|diamond) to get the same result.

Multiplication Rule

- If A and B are independent,

$$P(A \text{ AND } B) = P(A) \times P(B)$$

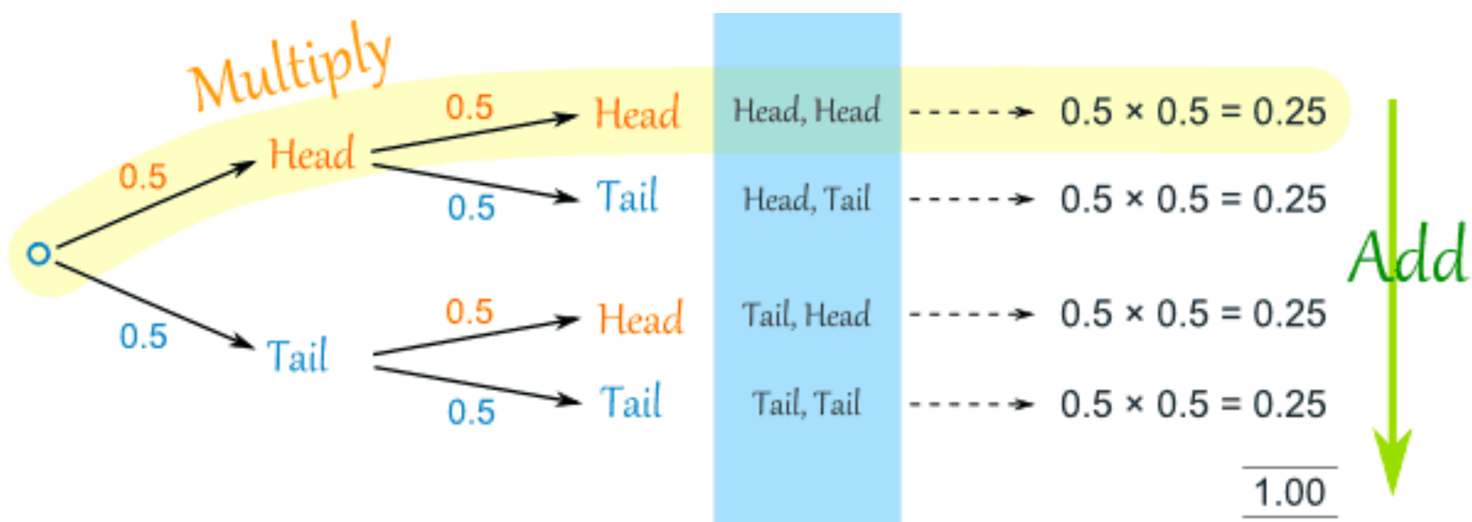
- If A and B are not independent,

$$P(A \text{ AND } B) = P(A|B) \times P(B)$$

Note: Assume independence only if the question specifies it or if you have checked for it yourself.

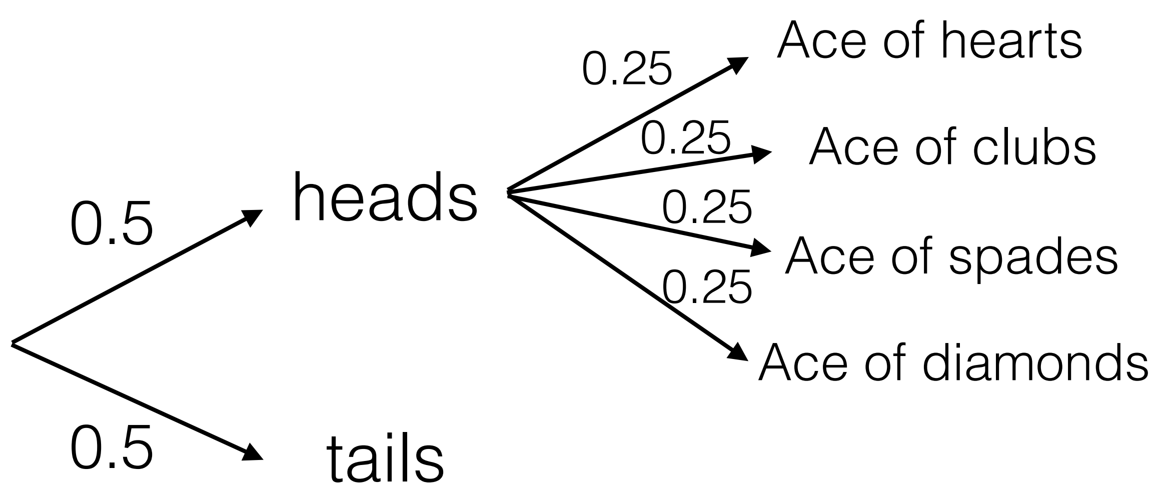
The Tree Diagram

- Tree diagrams are helpful in finding conditional probabilities.
- Tree diagrams can be used to represent problems in which events occur in a sequence.
- Tree diagrams show all the possible outcomes.



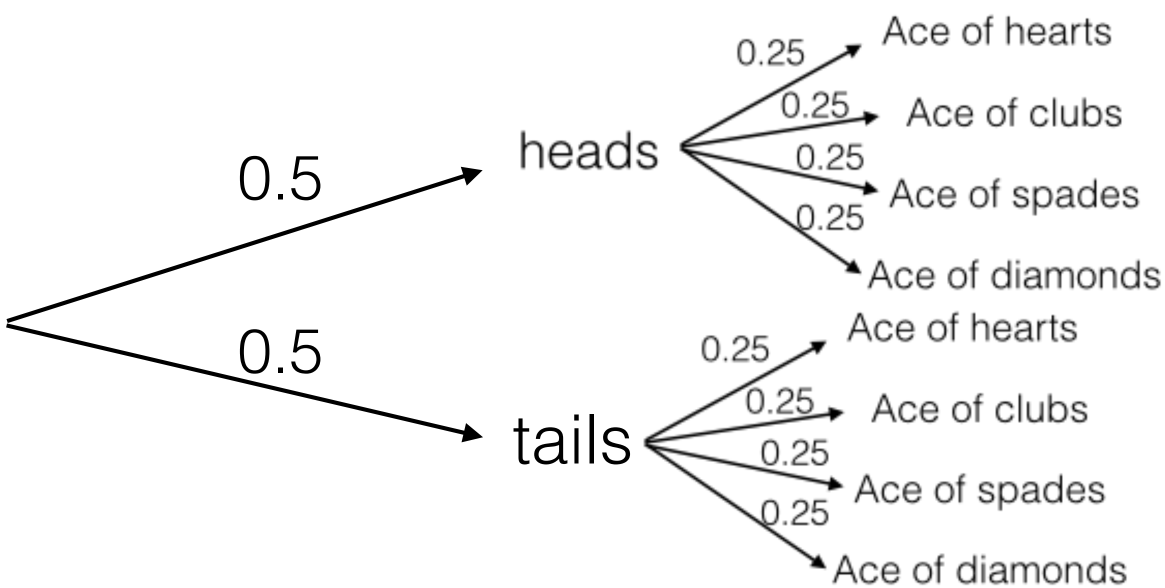
Example

Suppose 4 aces (heart, spade, diamond, club) are removed from a deck of cards. A coin is tossed and one of the aces is chosen. What is the probability of getting a heads on the coin and the ace of hearts?



Example

Suppose 4 aces (heart, spade, diamond, club) are removed from a deck of cards. A coin is tossed and one of the aces is chosen. What is the probability of getting a heads on the coin and the ace of hearts?



P(Heads AND Ace of Hearts)

= P(Heads) x P(Ace of Hearts|Heads) =0.5 x 0.25 = 0.125

Example

Clicker!

Suppose of your sample, 1% of people actually have cat allergies. A cat allergy test is correct for 80% of the people who actually have it. The test yields a 10% false positive result (test will say “yes” to 10% of those who don’t have it.)

What is the probability that someone is allergic to cats and tests positive?

Answer:

A. .01 x .80
B. .10 x .80
C. .01 x .10
D. 0

Example

What is the probability that someone is allergic to cats and tests positive?

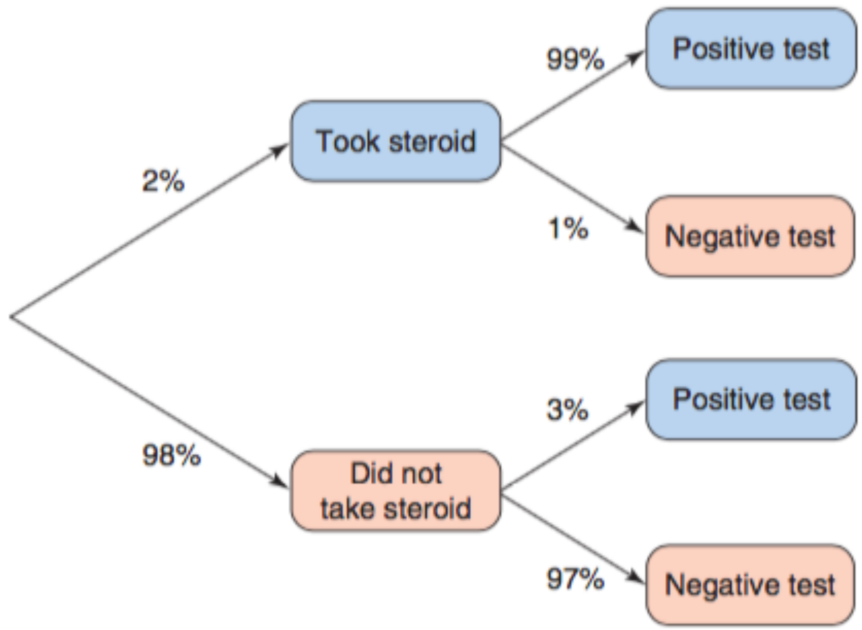
P(Allergic AND Yes) = P(Allergic) x P(Yes|Allergic) =0.01 x 0.80 = 0.008 or 0.8%

Example

Suppose you randomly sample cyclists to participate in a drug test to test for steroids. Let's imagine that 2% of the cyclists have taken an illegal steroid. Assume that if they took the drug, there is a 99% chance that the test will return a "positive" reading. Even if the cyclist did not take a steroid, there is the 3% chance that the test will be positive for drugs.

What is the probability that a randomly chosen cyclist will be steroid free and will still test positive for steroids?

Example



What is the probability that a randomly chosen cyclist will be steroid free and will still test positive for steroids?

$$
\begin{aligned}
\text{P(did not take steroid AND positive result)} &= \text{P(did not take steroid)} \times \\
&\quad \text{P (positive result} \mid \text{did not take steroid)} \\
&= 0.98 \times 0.03 \\
&= 0.0294
\end{aligned}
$$