

DSE 210: Probability and statistics

Overview

The kinds of questions we'll study

- ▶ Design a spam filter.
- ▶ What fraction of San Diegans occasionally smoke pot?
- ▶ Categorize New York Times articles by their underlying topics.
- ▶ Two new malaria vaccines are under consideration. How can we determine which is better?
- ▶ We've obtained user ratings of many movies. Visualize them.
- ▶ A dating service asks each user to answer 200 multiple choice questions. Summarize each user's responses by a few numbers.

Intermediate-level questions

- ▶ **Regression**. How do you fit a line to a set of points?
- ▶ **Clustering**. Given a bunch of data points, partition them into groups that are distinct from each other.
- ▶ **Laws of large numbers**. A drunk starts off from a bar and at each time step, takes either a step to the right or a step to the left. Where will he be, approximately, after n time steps?
- ▶ **Hypothesis testing**. You are given two alternatives and wish to test which is better. Design an experiment to do this.
- ▶ **Dimensionality reduction**. Find the primary axes of variation in a data set.

Low-level questions

- ▶ If you toss a coin 10 times, what is the chance of getting heads every time?
- ▶ Throw 20 balls into 20 bins at random. What is the probability that at least one of the bins remains empty?
- ▶ If each cereal box contains one of k action figures, how many boxes do you need to buy, on average, before getting all the figures?
- ▶ What fraction of a bell curve lies at least one standard deviation away from the mean?
- ▶ Find a concise description of a data matrix.

Course outline

1. Probability basics
2. Fitting distributions to data
3. Regression, classification, embedding, and visualization
4. Sampling and hypothesis testing
5. Advanced probabilistic modeling

Sets and counting

DSE 210

Sets

$$A = \{a, b, c, \dots, z\} \quad |A| = 26$$

$$B = \{0, 1\} \quad |B| = 2$$

$$E = \{\text{all even integers}\} \quad |E| = \infty$$

$$S = \{x \in E : x \text{ is a multiple of } 3\}$$

$$I = [0, 1] = \{x : 0 \leq x \leq 1\}$$

In a set, the *order* of elements doesn't matter:

$$\{0, 1, 2\} = \{2, 0, 1\}$$

and there are no duplicates.

Tuples

Let $C = \{H, T\}$.

All pairs of elements from C :

$$\{(H, H), (H, T), (T, H), (T, T)\} = C \times C = C^2$$

All triples of elements of C :

$$\{(H, H, H), (H, H, T), (H, T, H), \dots\} = C \times C \times C = C^3$$

All sequences of k elements from C : denoted $C^k = C \times C \times \dots \times C$.

How many sequences of length k are there? $|C^k| = |C|^k = 2^k$.

In a sequence, the order of elements matters:

$$(H, T) \neq (T, H).$$

Let $A = \{a, b, c, \dots, z\}$.

How many sequences of length 2? 26^2

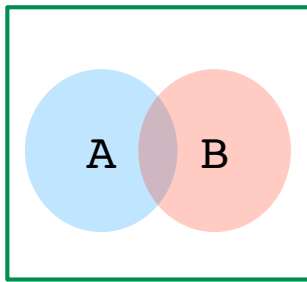
How many sequences of length 10? 26^{10}

How many sequences of length n ? 26^n

An alien language has an alphabet of size 10. Every sequence of ≤ 5 of these characters is a valid word. How many words are there in this language?

$$10^1 + 10^2 + 10^3 + 10^4 + 10^5 = 10 + 100 + 1000 + 10000 + 100000 = 111110.$$

Union and intersection



$A \cup B = \{\text{any element in } A \text{ or in } B \text{ or in both}\}$

$A \cap B = \{\text{any element in } A \text{ and in } B\}$

$M = \{2, 3, 5, 7, 11\}$ and $N = \{1, 3, 5, 7, 9\}$

$M \cup N = \{1, 2, 3, 5, 7, 9, 11\}$

$M \cap N = \{3, 5, 7\}$

$S = \{\text{all even integers}\}$ and $T = \{\text{all odd integers}\}$

$S \cup T = \{\text{all integers}\}$

$S \cap T = \emptyset$

Permutations

How many ways to order the three letters A, B, C ?

$ABC, ACB, BAC, BCA, CAB, CBA$

3 choices for the first, 2 choices for the second, 1 choice for the third

$3 \times 2 \times 1 = 6$. Call this $3!$

How many ways to order A, B, C, D, E ?

$5 \times 4 \times 3 \times 2 \times 1 = 5! = 120$

How many ways to place 6 men in a line-up?

$6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6! = 720$

How many possible outcomes of shuffling a deck of cards?

$52!$

General rule: The number of ways to order n distinct items is:

$$n! = n(n-1)(n-2) \cdots 1.$$

Combinations

An ice-cream parlor has flavors {chocolate, vanilla, strawberry, pecan}. You are allowed to pick two of them. How many options do you have?

CV, CS, CP, VS, VP, SP

In general, the number of ways to pick k items out of n is:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

For instance, $\binom{4}{2} = \frac{4 \cdot 3}{2!} = 6$.

How many ways to pick three ice-cream flavors?

$$\binom{4}{3} = 4$$

Pick any 4 of your favorite 100 songs. How many ways to do this?

$$\binom{100}{4} = \frac{100 \cdot 99 \cdot 98 \cdot 97}{4 \cdot 3 \cdot 2 \cdot 1}$$

Worksheet 1 — Sets and counting

1.
 - (a) Write down any set A of size 5.
 - (b) What is the formal notation for all sequences of three elements from A ?
 - (c) How many such sequences are there, exactly?
2. How many binary sequences of length 500 are there?
3. A and B are sets with $|A| = 3$ and $|B| = 4$.
 - (a) What is the largest size $A \cup B$ could possibly have?
 - (b) What is the smallest size $A \cup B$ could possibly have?
 - (c) Repeat for $A \cap B$.
4. A donkey, an ox, a goat, and a tiger need to cross a river. They have a boat that can only hold one animal, so they need to go one at a time. How many different orderings are there?
5. How many sequences of 5 English characters are there?
6. You have 10 good friends, and you want to choose 3 of them to accompany you on a trip. How many groups of three friends can you choose?
7. You have 10 different beer bottles, and you want to line 5 of them up on your mantelpiece. How many different arrangements can you make?

Probability spaces

DSE 210

Probability spaces

How to interpret a statement like:

The chance of getting a flush in a five-card poker hand is about 0.20%. (Flush = five of the same suit.)

The underlying **probability space** has two components:

1. The **sample space** (the space of outcomes).

In the example, $\Omega = \{\text{all possible five-card hands}\}$.

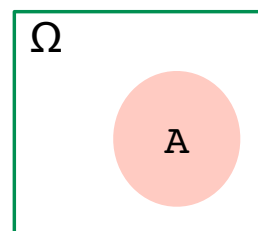
2. The **probabilities of outcomes**.

In the example, all hands are equally likely: probability $1/|\Omega|$.

Note: $\sum_{\omega \in \Omega} \Pr(\omega) = 1$.

Event of interest: the set of outcomes
 $A = \{\omega : \omega \text{ is a flush}\} \subset \Omega$.

$$\Pr(A) = \sum_{\omega \in A} \Pr(\omega) = \frac{|A|}{|\Omega|}$$



Examples

Roll a die. What is the chance of getting a number > 3 ?

Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Probabilities of outcomes: $\Pr(\omega) = \frac{1}{6}$.

Event of interest: $A = \{4, 5, 6\}$

$$\Pr(A) = \Pr(4) + \Pr(5) + \Pr(6) = \frac{1}{2}.$$

Roll three dice. What is the chance that their sum is 3?

Sample space

$$\begin{aligned}\Omega &= \{(1, 1, 1), (1, 1, 2), (1, 1, 3), \dots, (6, 6, 6)\} \\ &= \Omega_o \times \Omega_o \times \Omega_o\end{aligned}$$

where $\Omega_o = \{1, 2, 3, 4, 5, 6\}$.

Probabilities of outcomes:

$$\Pr(\omega) = \frac{1}{|\Omega|} = \frac{1}{216}$$

Event of interest: $A = \{(1, 1, 1)\}$. $\Pr(A) = \frac{1}{216}$.

Roll n dice.

Then $\Omega = \Omega_o \times \dots \times \Omega_o = \Omega_o^n$, where $\Omega_o = \{1, 2, 3, 4, 5, 6\}$.

What is $|\Omega|$? 6^n .

Probability of an outcome: $\Pr(\omega) = \frac{1}{6^n}$.

Socks in a drawer. A drawer has three blue socks and three red socks. You put your hand in and pull out two socks at random. What is the probability that they match?

Think of grabbing one sock first, then another.

$$\Omega = \{(B, B), (B, R), (R, B), (R, R)\} = \{B, R\}^2.$$

Probabilities:

$$\begin{aligned}\Pr((B, B)) &= \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{5} \\ \Pr((B, R)) &= \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10} \\ \Pr((R, B)) &= \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10} \\ \Pr((R, R)) &= \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{5}\end{aligned}$$

$$\text{Event of interest: } A = \{(B, B), (R, R)\}. \Pr(A) = \frac{2}{5}.$$

Socks in a drawer, cont'd. This time the drawer has three blue socks and two red socks. You put your hand in and pull out two socks at random. What is the probability that they match?

$$\text{Sample space, } \Omega = \{(B, B), (B, R), (R, B), (R, R)\} = \{B, R\}^2.$$

Different probabilities:

$$\begin{aligned}\Pr((B, B)) &= \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} \\ \Pr((B, R)) &= \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} \\ \Pr((R, B)) &= \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10} \\ \Pr((R, R)) &= \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}\end{aligned}$$

$$\text{Event of interest: } A = \{(B, B), (R, R)\}. \Pr(A) = \frac{2}{5}.$$

Shuffle a pack of cards.

Sample space $\Omega = \{\text{all possible orderings of 52 cards}\}$.

What is $|\Omega|$?

$$52! = 52 \cdot 51 \cdot 50 \cdot 49 \cdots 3 \cdot 2 \cdot 1$$

Toss a fair coin 10 times. What is the chance none are heads?

Sample space $\Omega = \{H, T\}^{10}$. It includes, for instance, $(H, T, H, T, H, T, H, T, H, T)$.

What is $|\Omega|$? $2^{10} = 1024$.

For any sequence of coin tosses $\omega \in \Omega$, we have $\Pr(\omega) = \frac{1}{1024}$.

Event of interest: $A = \{(T, T, T, T, T, T, T, T, T, T)\}$. $\Pr(A) = \frac{1}{1024}$.

What is the probability of exactly one head?

Event of interest: $A = \{\omega \in \Omega : \omega \text{ has exactly one } H\}$.

What is $|A|$? 10.

Each sequence in A can be specified by the location of the one H , and there are 10 choices for this.

What is $\Pr(A)$? $\frac{10}{1024}$.

Toss a fair coin 10 times. What is the chance of exactly two heads?

Again, sample space $\Omega = \{H, T\}^{10}$, with $|\Omega| = 2^{10} = 1024$.

For any sequence of coin tosses $\omega \in \Omega$, we have $\Pr(\omega) = \frac{1}{1024}$.

Event of interest: $A = \{\omega \in \Omega : \omega \text{ has exactly two } H\text{'s}\}$.

What is $|A|$? $\binom{10}{2} = \frac{10 \cdot 9}{2} = 45$.

Each sequence in A can be specified by the locations of the two H 's and there are $\binom{10}{2}$ choices for these locations.

What is $\Pr(A)$? $\frac{45}{1024}$.

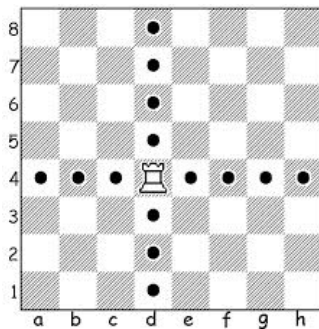
What is the probability of exactly k heads?

Event of interest: $A = \{\omega \in \Omega : \omega \text{ has exactly } k \text{ } H\text{'s}\}$.

What is $|A|$? $\binom{10}{k}$.

What is $\Pr(A)$? $\binom{10}{k}/1024$.

Rooks on a chessboard.



What is the maximum number of rooks you can place so that no rook is attacking any other? 8.

How many ways are there to place 8 rooks on the board, attacking or not? $\binom{64}{8}$

How many non-attacking placements of 8 rooks are there?

$$8 \cdot 7 \cdot 6 \cdot 5 \cdots = 8!$$

Randomly place 8 rooks on the board. What is the probability that it is a non-attacking placement?

$$\frac{8!}{\binom{64}{8}}.$$

Five-card poker. You are dealt 5 cards from a deck of 52.

Sample space $\Omega = \{\text{all possible hands}\}$.

Probabilities: each hand is equally likely, $\Pr(\omega) = 1/|\Omega|$.

What is $|\Omega|$? $\binom{52}{5}$.

What is the probability of getting a flush (five of the same suit)?

Event of interest: $F = \{\text{flush hands}\}$. Then $|F| = 4 \times \binom{13}{5}$

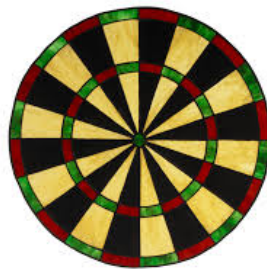
Therefore $\Pr(F) = |F|/|\Omega| = 4 \times \binom{13}{5} / \binom{52}{5}$.

What is the chance of a straight flush (flush, and in sequence)?

Let $S = \{\text{straight flush hands}\}$. Then $|S| = 4 \cdot 9 = 36$.

And $\Pr(S) = |S|/|\Omega|$.

Dartboard. A dartboard has radius 1 and its central bullseye has radius 0.1. You throw a dart and it lands at a random location on the board.



Sample space $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$.

All points are equally likely.

What is the probability of hitting the bullseye?

Event of interest: $B = \{(x, y) : x^2 + y^2 \leq (0.1)^2\}$.

$\Pr(B) = \text{area}(B)/\text{area}(\Omega) = 0.01$.

What is the probability of hitting the exact center? 0.

Birthday paradox. A room contains k people. What is the chance that they all have different birthdays?

Number the people $1, 2, \dots, k$.

Number the days of the year $1, 2, \dots, 365$.

Let $\omega = (\omega_1, \dots, \omega_k)$, where $\omega_i \in \{1, 2, \dots, 365\}$ is the birthday of person i . Thus $\Omega = \{1, 2, \dots, 365\}^k$.

What is $|\Omega|$? 365^k .

Event of interest: $A = \{(\omega_1, \dots, \omega_k) : \text{all } \omega_i \text{ different}\}$.

What is $|A|$? $365 \cdot 364 \cdot 363 \cdots (365 - k + 1)$.

Therefore,

$$\Pr(A) = \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k} = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - k + 1}{365}$$

For $k = 23$, this is less than $1/2$. In other words, **in a group of 23 random people, chances are some pair of them have a common birthday!**

Worksheet 2 — Probability spaces

1. Give a possible sample space Ω for each of the following experiments.
 - (a) An election decides between two candidates A and B .
 - (b) A two-sided coin is tossed.
 - (c) A student is asked for the month and day-of-week on which her birthday falls.
 - (d) A student is chosen at random from a class of ten students.
 - (e) You choose the color of your new car's exterior (choices: red, black, silver, green) and interior (choices: black, beige).
2. In each of the following situations, define the sample space Ω .
 - (a) A fair coin is tossed 200 times in a row.
 - (b) You count the number of people who enter a department store on a particular Sunday.
 - (c) You open up *Hamlet* and pick a word at random.
3. Let A , B , and C be events defined on a particular sample space Ω . Write expressions for the following combinations of events:
 - (a) All three events occur.
 - (b) At least one of the events occurs.
 - (c) A and B occur, but not C .
4. Consider a sample space $\Omega = \{a, b, c\}$ with probabilities $\Pr(a) = 1/2$ and $\Pr(b) = 1/3$.
 - (a) What is $\Pr(c)$?
 - (b) How many distinct events can be defined on this space?
 - (c) Find the probabilities of each of these possible events.
5. A fair coin is tossed three times in succession. Describe in words each of the following events on sample space $\{H, T\}^3$.
 - (a) $E_1 = \{HHH, HHT, HTH, HTT\}$
 - (b) $E_2 = \{HHH, TTT\}$
 - (c) $E_3 = \{HHT, HTH, THH\}$

What are the probabilities of each of these events?

6. Let A and B be events defined on a sample space Ω such that $\Pr(A \cap B) = 1/4$, $\Pr(A^c) = 1/3$, and $\Pr(B) = 1/2$. Here $A^c = \Omega \setminus A$ is the event that A *doesn't* happen. What is $\Pr(A \cup B)$?
7. A pair of dice are rolled. What is the probability that they show the same value?

8. In Morse code, each letter is formed by a succession of dashes and dots. For instance, the letter *S* is represented by three dots and the letter *O* is represented by three dashes. Suppose a child types a sequence of 9 dots/dashes at random (each position is equally likely to be a dot or a dash). What is the probability that it spells out *SOS*?
9. A die is loaded in such a way that the probability of each face turning up is proportional to the number of dots on that face (for instance, a six is three times as probable as a two). What is the probability of getting an even number in one throw?
10. A certain lottery has the following rules: you buy a ticket, choose 3 different numbers from 1 to 100, and write them on the ticket. The lottery has a box with 100 balls numbered 1 to 100. Three (different) balls are chosen. If any of the balls has one of the numbers you have chosen, you win. What is the probability of winning?
11. Five people of different heights are lined up against a wall in random order. What is the probability that they just happen to be in increasing order of height (left-to-right)?
12. Five people get on an elevator that stops at five floors. Assuming that each person has an equal probability of going to any one floor, find the probability that they all get off at different floors.
13. You are dealt five cards from a standard deck. What is the probability that the first four are aces and the fifth is a king?
14. A barrel contains 90 good apples and 10 rotten apples. If ten of the apples are chosen at random, what is the probability that they are all good?
15. Four women check their hats at a concert, but when each woman returns after the performance, she gets a hat chosen randomly from those remaining. What is the probability that each woman gets her own hat back?
16. Assume that whenever a child is born, it is equally likely to be a girl or boy, independent of any earlier children. What is the probability that a randomly-chosen family with six children has exactly three girls and three boys?
17. Snow White asks three of the seven dwarfs, chosen at random, to accompany her on a trip.
 - (a) What is the probability that Dopey is in this group?
 - (b) What is the probability that both Dopey and Sneezy are in the group?
 - (c) What is the probability that neither Dopey nor Sneezy are in the group?

Multiple events, conditioning, and independence

DSE 210

People's probability judgements

Experiment by Kahneman-Tversky. Subjects were told:

Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

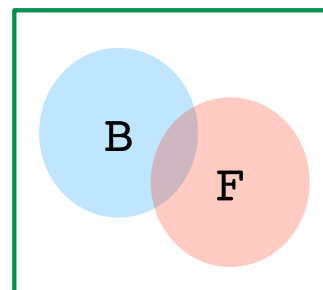
They were then asked to rank three possibilities:

- (a) Linda is active in the feminist movement.
- (b) Linda is a bank teller.
- (c) Linda is a bank teller and is active in the feminist movement.

Over 85% respondents chose (a) > (c) > (b).

But:

$\Pr(\text{bank teller, feminist}) \leq \Pr(\text{bank teller})$.

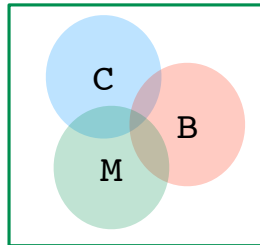


In a city, 60% of people have a car, 20% of people have a bike, and 10% of people have a motorcycle. Anyone without at least one of these walks to work. What is the minimum fraction of people who walk to work?

Let $\Omega = \{\text{people in the town}\}$. Let

$C = \{\text{has car}\}$, $B = \{\text{has bike}\}$, $M = \{\text{has motorcycle}\}$, $W = \{\text{walks}\}$.

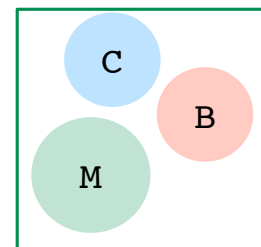
General picture:



$$\Pr(W) \geq 1 - \Pr(C \cup B \cup M)$$

$$\begin{aligned} \Pr(C \cup B \cup M) &\leq \Pr(C) + \Pr(B) + \Pr(M) \\ &= 0.6 + 0.2 + 0.1 = 0.9 \end{aligned}$$

and thus $\Pr(W) \geq 0.1$.



Complements and unions

The complement of an event.

Let Ω be a sample space and $E \subset \Omega$ an event.

Write E^c for the event that E does not occur, that is, $E^c = \Omega \setminus E$.

$$\Pr(E^c) = 1 - \Pr(E).$$

The union bound.

For any events E_1, \dots, E_k :

$$\Pr(E_1 \cup \dots \cup E_k) \leq \Pr(E_1) + \dots + \Pr(E_k).$$

This inequality is exact when the events are disjoint.

Coupon-collector problem

Each cereal box has one of k action figures. How many boxes do you need to buy so that you are likely to get all k figures?

Say we buy n boxes.

Let A_i be the event that the i th action figure is *not* obtained.

$$\begin{aligned}\Pr(A_i) &= \Pr(\text{not in 1st box}) \cdot \Pr(\text{not in 2nd box}) \cdots \Pr(\text{not in } n\text{th box}) \\ &= \left(1 - \frac{1}{k}\right)^n \leq e^{-n/k}\end{aligned}$$

By union bound, the probability of missing some figure is

$$\Pr(A_1 \cup \cdots \cup A_k) \leq \Pr(A_1) + \cdots + \Pr(A_k) \leq ke^{-n/k}.$$

Setting $n \geq k \ln 2k$ makes this $\leq 1/2$.

Therefore: enough to buy $O(k \log k)$ cereal boxes.

Conditional probability

You meet a stranger at a bar. What is the chance he votes Republican?

Just use the average for your town: 0.5, say.

Now suppose you find out he plays tennis.

Sample space $\Omega = \{\text{all people in your town}\}$

Two events of interest:

$R = \{\text{votes Republican}\}$

$T = \{\text{plays tennis}\}$

What is $\Pr(R|T)$?

Formula for conditional probability:

$$\Pr(R|T) = \frac{\Pr(R \cap T)}{\Pr(T)} = \frac{\# \text{ people who vote Republican and play tennis}}{\# \text{ people who play tennis}}.$$

Pregnancy test.

The following data is obtained on a pregnancy test:

$\Omega = \{\text{women who use the test}\}$

$P = \{\text{women using the test who are actually pregnant}\}$

$T = \{\text{women for whom the test comes out positive}\}$

Suppose $T \subset P$ and $\Pr(P) = 0.4$ and $\Pr(T) = 0.3$.

Suppose the test comes out positive. What is the chance of pregnancy?
Exactly 1.

Suppose the test comes out negative. What is the chance of pregnancy?

$$\Pr(P|T^c) = \frac{\Pr(P \cap T^c)}{\Pr(T^c)} = \frac{\Pr(P) - \Pr(T)}{1 - \Pr(T)} = \frac{0.1}{0.7} = \frac{1}{7}.$$

Rolls of a die.

You roll a die twice. What is the probability that the sum is ≥ 10 :

If the first roll is 6?

$$\Pr(\text{sum} \geq 10 | \text{first} = 6) = \Pr(\text{second} \geq 4) = \frac{1}{2}.$$

If the first roll is ≥ 3 ?

$$\begin{aligned} \Pr(\text{sum} \geq 10 | \text{first} \geq 3) &= \frac{\Pr(\text{sum} \geq 10, \text{first} \geq 3)}{\Pr(\text{first} \geq 3)} \\ &= \frac{\Pr(\{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\})}{2/3} = \frac{1}{4}. \end{aligned}$$

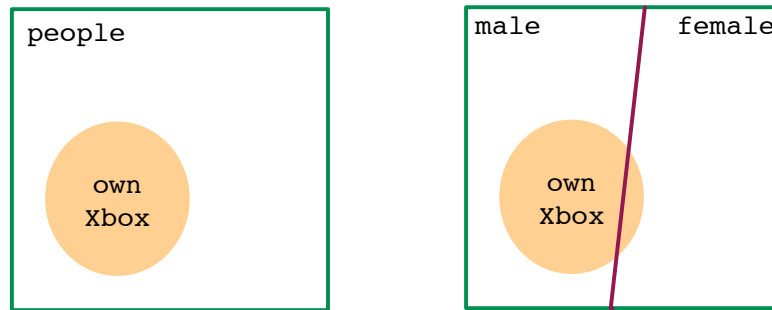
If the first roll is < 6 ?

$$\begin{aligned} \Pr(\text{sum} \geq 10 | \text{first} < 6) &= \frac{\Pr(\text{sum} \geq 10, \text{first} < 6)}{\Pr(\text{first} < 6)} \\ &= \frac{\Pr(\{(5, 5), (5, 6), (4, 6)\})}{5/6} = \frac{1}{10}. \end{aligned}$$

Summation rule

Breaking down a probability into disjoint pieces.

Example: what fraction of people own an Xbox?



$$\begin{aligned}\Pr(\text{own Xbox}) \\ &= \Pr(\text{own Xbox, male}) + \Pr(\text{own Xbox, female}) \\ &= \Pr(\text{own Xbox}|\text{male})\Pr(\text{male}) + \Pr(\text{own Xbox}|\text{female})\Pr(\text{female})\end{aligned}$$

Suppose events A_1, \dots, A_k are disjoint and $A_1 \cup \dots \cup A_k = \Omega$: that is, one of these events must occur. Then for any other event E ,

$$\begin{aligned}\Pr(E) &= \Pr(E, A_1) + \Pr(E, A_2) + \dots + \Pr(E, A_k) \\ &= \Pr(E|A_1)\Pr(A_1) + \Pr(E|A_2)\Pr(A_2) + \dots + \Pr(E|A_k)\Pr(A_k)\end{aligned}$$

The Monty Hall game

Three doors: one has a treasure chest behind it and the other two have goats. You pick a door and indicate it to Monty. He opens one of the other two doors to reveal a goat. Now, should you stick to your initial choice, or switch to the other unopened door?

You should switch.

First argument:

$$\begin{aligned}\Pr(\text{initial choice has treasure}) &= 1/3 \\ \text{No matter what Monty does, he can't change this fact. So} \\ \Pr(\text{other unopened door has treasure}) &= 2/3\end{aligned}$$

Second argument:

$$\begin{aligned}\Pr(\text{treasure in other door}) \\ &= \Pr(\text{treasure in other door}|\text{initial choice correct})\Pr(\text{initial choice correct}) + \\ &\quad \Pr(\text{treasure in other door}|\text{initial choice wrong})\Pr(\text{initial choice wrong}) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.\end{aligned}$$

Sex bias in graduate admissions

In 1969, there were 12673 applicants for graduate study at Berkeley. 44% of the male applicants were accepted, and 35% of the female applicants.

Define:

- ▶ $\Omega = \{\text{all applicants}\}$
- ▶ $M = \{\text{male applicants}\}$
- ▶ What is M^c ? $M^c = \{\text{female applicants}\}$
- ▶ $A = \{\text{accepted applicants}\}$

What do the percentages 44% and 35% correspond to?

$$\Pr(A|M) = 0.44 \text{ and } \Pr(A|M^c) = 0.35.$$

The administration found, however, that in every department, the accept rate for female applicants was at least as high as the accept rate for male applicants. How could this be?

Bayes' rule

Pearl: You wake up in the middle of the night to the shrill sound of your burglar alarm. What is the chance that a burglary has been attempted?

The facts:

- ▶ There is a 95% chance that an attempted burglary will trigger the alarm.

$$\Pr(\text{alarm}|\text{burglary}) = 0.95$$

- ▶ There is a 1% chance of a false alarm.

$$\Pr(\text{alarm}|\text{no burglary}) = 0.01$$

- ▶ Based on local crime statistics, there is a 1-in-10,000 chance that a given house will be burglarized on a given night.

$$\Pr(\text{burglary}) = 10^{-4}$$

We need to compute

$$\Pr(\text{burglary}|\text{alarm}) = \frac{\Pr(\text{burglary, alarm})}{\Pr(\text{alarm})} = \frac{\Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary})}{\Pr(\text{alarm})}$$

We need to compute

$$\Pr(\text{burglary}|\text{alarm}) = \frac{\Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary})}{\Pr(\text{alarm})}$$

Now,

$$\Pr(\text{alarm}) = \Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary}) + \Pr(\text{alarm}|\text{no burglary})\Pr(\text{no burglary})$$

Therefore,

$$\Pr(\text{burglary}|\text{alarm}) = \frac{0.95 \times 10^{-4}}{0.95 \times 10^{-4} + 0.01 \times (1 - 10^{-4})} \approx 0.00941$$

The alarm increases one's belief in a burglary hundredfold, from 1/10000 to roughly 1/100.

Bayes' rule:

$$\Pr(H|E) = \frac{\Pr(E|H)}{\Pr(E)}\Pr(H).$$

The three prisoners

Three prisoners – A , B , C – are in a jail one night and one of them (they don't know whom) will be declared guilty and executed in the morning. Racked by worry, prisoner A calls the prison guard and begs to be told whether he is the unlucky one. The guard is not allowed to tell him – but he can say only that B will be declared innocent. Now A thinks to himself, “previously my chance of being executed was $1/3$, and now, because of an innocuous inquiry, it seems to have gone up to $1/2$. How can this be?”

Analyze using these events:

G_A = the event that A will be declared guilty

I_B = the event that the guard, when prompted, will declare B innocent

$$\Pr(G_A|I_B) = \frac{\Pr(I_B|G_A)\Pr(G_A)}{\Pr(I_B)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

Independence

Two events A, B are **independent** if the probability of B occurring is the same whether or not A occurs.

Example: toss two coins.

$A = \{\text{first coin is heads}\}$

$B = \{\text{second coin is heads}\}$

Formally, we say A, B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

The independence of A and B implies:

- ▶ $\Pr(A|B) = \Pr(A)$
- ▶ $\Pr(B|A) = \Pr(B)$
- ▶ $\Pr(A|B^c) = \Pr(A)$

Examples: independent or not?

1. You have two children.

$A = \{\text{first child is a boy}\}$, $B = \{\text{second child is a girl}\}$.

Independent.

2. You throw two dice.

$A = \{\text{first is a six}\}$, $B = \{\text{sum} > 10\}$.

Not independent.

3. You get dealt two cards at random from a deck of 52.

$A = \{\text{first is a heart}\}$, $B = \{\text{second is a club}\}$.

Not independent: $\Pr(A) = \Pr(B) = 1/4$, but

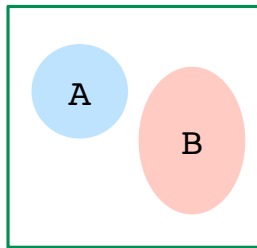
$$\Pr(A \cap B) = \frac{1}{4} \cdot \frac{13}{51} > \Pr(A)\Pr(B).$$

4. You are dealt two cards.

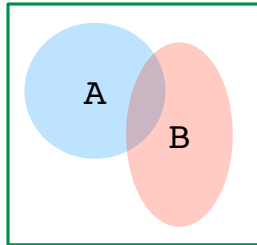
$A = \{\text{first is a heart}\}$, $B = \{\text{second is a 10}\}$.

Independent: $\Pr(A) = 1/4$, $\Pr(B) = 1/13$, and

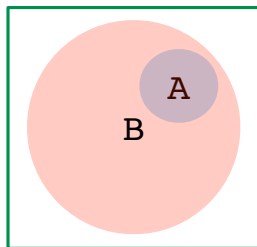
$$\Pr(A \cap B) = \frac{12}{52} \cdot \frac{4}{51} + \frac{1}{52} \cdot \frac{3}{51} = \frac{1}{52} = \Pr(A)\Pr(B).$$



Not independent



Possibly independent



Not independent

Worksheet 3 — Multiple events, conditioning, and independence

1. A coin is tossed three times. What is the probability that there are exactly two heads, given that:
 - (a) the first outcome is a head?
 - (b) the first outcome is a tail?
 - (c) the first two outcomes are both heads?
 - (d) the first two outcomes are both tails?
 - (e) the first outcome is a head and the third outcome is a tail?
2. A student must choose exactly two of the following three electives: art, French, or mathematics. The probability that he chooses art is $5/8$, the probability he chooses French is $5/8$, and the probability that he chooses both art and French is $1/4$.
 - (a) What is the probability that he chooses mathematics?
 - (b) What is the probability that he chooses either art or French?
3. For a bill to come before the president of the United States, it must be passed by both the House of Representatives and the Senate. Assume that, of the bills presented to the two bodies, 60% pass the House, 80% pass the Senate, and 90% pass at least one of the two. Calculate the probability that the next bill presented to the two groups will come before the president.
4. In a fierce battle, not less than 70% of the soldiers lost one eye, not less than 75% lost one ear, not less than 80% lost one hand, and not less than 85% lost one leg. What is the minimal possible percentage of those who simultaneously lost one ear, one eye, one hand, and one leg?
5. A card is drawn at random from a standard deck. What is the probability that:
 - (a) it is a heart, given that it is red?
 - (b) it is higher than a ten, given that it is a heart (interpret J, Q, K, A as having numeric value 11, 12, 13, 14)?
 - (c) it is a jack, given that it is higher than a 10?
6. If $\Pr(B^c) = 1/4$ and $\Pr(A|B) = 1/2$, what is $\Pr(A \cap B)$?
7. A die is rolled twice. What is the probability that the sum of the two rolls is > 7 , given that:
 - (a) the first roll is a 4?
 - (b) the first roll is a 1?
 - (c) the first roll is > 3 ?
 - (d) the first roll is < 5 ?
8. Two cards are drawn successively from a deck of 52 cards.

- (a) Find the probability that the second card is equal in rank to the first card. (Rank is defined according to the following ordering: $2, 3, \dots, 10, J, Q, K, A$. The suit is irrelevant.)
- (b) Find the probability that the second card is higher in rank than the first card.
9. A particular car manufacturer has three factories F_1, F_2, F_3 making 25%, 35%, and 40%, respectively, of its cars. Of their output, 5%, 4%, and 2%, respectively, are defective. A car is chosen at random from the manufacturer's supply.
- (a) What is the probability that the car is defective?
- (b) Given that it is defective, what is the probability that it came from factory F_1 ?
10. Suppose that there are equal numbers of men and women in the world, and that 5% of men are colorblind whereas only 1% of women are colorblind. A person is chosen at random and found to be colorblind. What is the probability that the person is male?
11. A doctor assumes that his patients has one of the three diseases d_1, d_2 , or d_3 , each with probability $1/3$. He carries out a test that will be positive with probability 0.8 if the patient has d_1 , with probability 0.6 if the patient has d_2 , and with probability 0.4 if the patient has d_3 .
- (a) What is the probability that the test will be positive?
- (b) Suppose that the outcome of the test is positive. What probabilities should the doctor now assign to the three possible diseases?
12. One coin in a collection of 65 coins has two heads; the rest of the coins are fair. If a coin, chosen at random from the lot and then tossed, turns up heads six times in a row, what is the probability that it is the two-headed coin?
13. A scientist discovers a fossil fragment that he believes is either some kind of tiger (with probability $1/3$) or mammoth (with probability $2/3$). To shed further light on this question, he conducts a test which has the property that for tigers, it will come out positive with probability $5/6$ whereas for mammoths it will come out positive with probability just $1/3$. Suppose the test comes out negative. What is the probability, given the outcome of the test, that the fossil comes from a tiger?
14. Sherlock Holmes finds paw prints at the scene of a murder, and thinks that they are either from a dog, with probability $3/4$, or from a small bear, with probability $1/4$. He then discovers some unusual scratches on a nearby tree. The probability that a dog would produce these scratches is $1/10$, while the probability that a bear would is $3/5$. What is the probability, given the presence of scratches, that the animal is a bear?
15. A coin is tossed three times. Consider the following five events:
- A : Heads on the first toss
 - B : Tails on the second toss
 - C : Heads on the third toss
 - D : All three outcomes the same
 - E : Exactly one head
- (a) Which of the following pairs of events are independent?
- (1) A, B
 - (2) A, D

- (3) A, E
- (4) D, E
- (b) Which of the following triples of events are independent?
 - (1) A, B, C
 - (2) A, B, D
 - (3) C, D, E
- 16. You randomly shuffle a standard deck and deal two cards. Which of the following pairs of events are independent?
 - (1) $A = \{\text{first card is a heart}\}, B = \{\text{second card is a heart}\}$
 - (2) $A = \{\text{first card is a heart}\}, B = \{\text{first card is a 10}\}$
 - (3) $A = \{\text{first card is a 10}\}, B = \{\text{second card is a 9}\}$
 - (4) $A = \{\text{first card is a heart}\}, B = \{\text{second card is a 10}\}$
- 17. A student applies to UCLA and UCSD. He estimates that he has a probability of 0.5 of being accepted at UCLA and a probability of 0.3 of being accepted at UCSD. He further estimates that the probability that he will be accepted by both is 0.2.
 - (a) What is the probability that he is accepted at UCSD if he is accepted at UCLA?
 - (b) Is the event “accepted at UCLA” independent of the event “accepted at UCSD”?