CSCI 3022 Intro to Data Science
Bayes and pdfs

Opening Examples: Suppose we have a standard 52 card playing deck, and are being dealt 5 cards at random.

1. What is the probability of being dealt a flush (all 5 cards share a suit)?
2. What is the probability of being dealt all 4 kings?

Opening Example Sol'n

Example: What is the probability of being dealt a flush in poker (five cards)?

Opening Example Sol'n

Example: What is the probability of being dealt a flush in poker (five cards)?

Solution: Two ways

1. Count all possible selections of five cards - $C(52,5)$ - then count all possible selections of flushes: $C(13,5)$ for the values on the flush and $C(4,1)$ for the possible suits. Then
$$P(\textit{flush}) = \frac{C(13,5)C(4,1)}{C(52,5)}$$

2. Do things *conditionally*:
$$P(\textit{all 5 cards same suit})$$
$$= P(\textit{cards 1-4 match suit AND card 5 matches that suit})$$
$$= P(\textit{cards 1-4 match suit})P(\textit{ card 5 matches that suit GIVEN cards 1-4 match suit})$$
$$= \cdots = \frac{52}{52}\frac{12}{51}\frac{11}{50}\frac{10}{49}\frac{9}{48}$$

Second Opening Example

What is the probability of being dealt all 4 kings in poker (five cards)?

Second Opening Example

What is the probability of being dealt all 4 kings in poker (five cards)?
The 52 card deck has 48 "N" non-Kings and 4 "Ki" Kings. We are interested in 5 possible outcomes: that we are dealt NKiKiKiKi, KiNKiKiKi, KiKiNKiKi, KiKiKiNKi, or KiKiKiKiN. It turns out that these each have the same probability:

$$\begin{aligned}
P(\{NKiKiKiKi\}) &= P(\#5 = N|KiKiKiKi) \cdot P(KiKiKiKi) \\
&= \frac{48}{48} \cdot P(KiKiKiKi) \\
&= \frac{48}{48} \cdot P(\#4 = K|KiKiKi) \cdot P(KiKiKi) \ldots \\
&= \frac{48}{48} \cdot \frac{1}{49} \cdot P(KiKiKi) \ldots \\
&\ldots \\
&= \frac{48}{48} \cdot \frac{1}{49} \cdot \frac{2}{50} \cdot \frac{3}{51} \cdot \frac{4}{52}
\end{aligned}$$

Announcements and To-Dos

Announcements:

1. HW 2 for Monday.
2. Another nb day this Friday!

Last time we learned:

1. Basics of Probability in review.

To do:

Last Time...

A few big takeaways from our second lecture on probability.

- If all the outcomes in Ω are *equally likely*, calculating probabilities collapses down to *counting outcomes*, where $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{\# of ways A can happen}}{\text{\# of total outcomes in sample space}}$
- *Conditional Probability*: $P(A|B) = \frac{P(A \cap B)}{P(B)}$,
- *Multiplication Rule*: $P(A \cap B) = P(A|B)P(B)$

Independence, formally

Definition: Two events A and B are said to be *independent* if $P(A|B) = P(A)$.

This definition, combined with the product rule give us three equivalent tests for independence:

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

Independence, in detail!

We don't have to stop at two sets. Sometimes we have lots of outcomes we want to be unrelated.

Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if for every $k = 2, 3, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

In other words, for any selection of A mutually independent event, the probability of their intersection is equal to the product of their individual probabilities.

Independence, in detail!

Why does this matter? Consider the following
Example: Flip a fair coin twice, and define

1. A : "heads on flip 1"
2. B : "heads on flip 2"
3. C : "same outcomes on both flips"

What are $P(A), P(B), P(C), P(A|B), P(A|C), P(B|C)$?
What about $P(A \cap B \cap C?)$

Independence, in detail!

Why does this matter? Consider the following
Example: Flip a fair coin twice, and define

1. A : "heads on flip 1"
2. B : "heads on flip 2"
3. C : "same outcomes on both flips"

What are $P(A), P(B), P(C), P(A|B), P(A|C), P(B|C)$?
What about $P(A \cap B \cap C?)$
Any *pair* of A, B, C looks independent, since
$P(A) = P(B) = P(C) = P(A|B) = P(A|C) = P(B|C) = 1/2.$
However, $P(A \cap B \cap C?) = P(\{HH\}) = 1/4$ which is not the same as the triple product
$P(A)P(B)P(C) = \frac{1}{8}$.
Ultimately, event C is determined by the combination of A and B.

The Law of Total Probability

Example: Suppose I have a couple of bags of marbles. The first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles.

Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, and then grab a marble from that bag, what is the probability that it is black?

The Law of Total Probability

Example: Suppose I have a couple of bags of marbles. The first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles.
Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, and then grab a marble from that bag, what is the probability that it is black?

There are two 'ways' we get a black marble: from bag 1 or from bag 2. We just have to add both up!

$$\begin{aligned} P(\textbf{B}) &= P(\textbf{B from 1}) + P(\textbf{B from 2}) \\ &= P(\textbf{B} \cap \mathbf{1}) + P(\textbf{B} \cap \mathbf{2}) \\ &= P(\textbf{B}|\mathbf{1})P(\mathbf{1}) + P(\textbf{B}|\mathbf{2})P(\mathbf{2}) \\ &= \frac{4}{10} \cdot \frac{1}{2} + \frac{7}{10} \cdot \frac{1}{2} \\ &= \frac{11}{20} \end{aligned}$$

The Law of Total Probability

Example: As before, the first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles.
But what if bag 1 is made of a much nicer material to touch, so I am twice as likely to randomly select bag 1 from between the 2 bags?

The Law of Total Probability

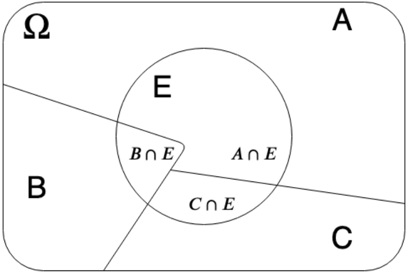
Example: As before, the first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles.
But what if bag 1 is made of a much nicer material to touch, so I am twice as likely to randomly select bag 1 from between the 2 bags?

Same solution!:

$$\begin{aligned}
P(\textbf{B}) &= P(\textbf{B from 1}) + P(\textbf{B from 2}) \\
&= P(\textbf{B} \cap \mathbf{1}) + P(\textbf{B} \cap \mathbf{2}) \\
&= P(\textbf{B}|\mathbf{1})P(\mathbf{1}) + P(\textbf{B}|\mathbf{2})P(\mathbf{2}) \\
&= \frac{4}{10} \cdot \frac{2}{3} + \frac{7}{10} \cdot \frac{1}{3} \\
&= \frac{1}{2}
\end{aligned}$$

The Law of Total Probability

Definition: A *Partition* of Ω is a set of disjoint events $E_1, E_2, \ldots E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$. Given such a partition, any event A can be decomposed into:



$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k)$$

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)$$

Independence

Recall: Independence.
Example: In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. One student is selected at random. Let S represent the a senior is chosen, and M represent that a student taking math is chosen.

If the randomly chosen student is a senior, then what is the probability that they are taking math?

Are these events independent?

Independence

Recall: Independence.
Example: In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. One student is selected at random. Let S represent the a senior is chosen, and M represent that a student taking math is chosen.

If the randomly chosen student is a senior, then what is the probability that they are taking math?
$P(S) = 250/1200$; $P(M) = 150/1200$; $P(M \cap S) = 40/1200/$
So, $P(M|S) = P(M \cap S)/P(S) = 250/1200 = 4/25$.

Are these events independent?
Does $P(M|S) = P(M)$? *No.*

Bayes' Theorem

The formula for $P(M|S)$ on the prior example is an example of *Bayes' Theorem*.

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

The proof follows directly from the multiplication rule, that

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

Bayes' theorem is most important mathematical way to describe *how much new information matters.*

$P(A)$ is called the *prior* information about A, and $P(A|B)$ is the *posterior* (post-data!) information about A.

Bayes' Theorem

Example 1:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

Bayes' Theorem

Example 1:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?
We know:

$$P(1) = .7; P(2) = .2; P(3) = .1; P(S|1) = .01; P(S|2) = .02; P(S|3) = .05;$$

and by LTP

$$P(S) = P(S|1)P(1) + P(S|2)P(2) + P(S|3)P(3)$$

$$P(S) = .007 + .004 + .005 = .018$$

Bayes' Theorem

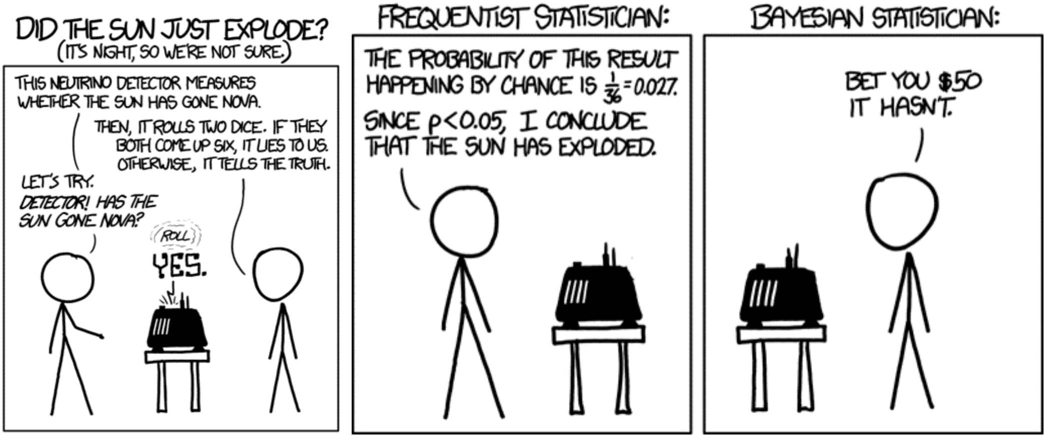
Example 2:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. Say she selects a message at random and it is spam. What is the probability that the message came from account #1?

Bayes' Theorem

Example 2:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. Say she selects a message at random and it is spam. What is the probability that the message came from account #1?
Now we use Bayes'!

$$P(1|S) = \frac{P(S|1)P(1)}{P(S)}$$

$$P(1|S) = \frac{.007}{.018} = \frac{7}{18}$$


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Probability Wrapup

- If all outcomes are *equally likely*, we can just count outcomes: $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{\# of ways A can happen}}{\text{\# of total outcomes in sample space}}$
- *Conditional Probability*: $P(A|B) = \frac{P(A \cap B)}{P(B)}$,
- *Multiplication Rule*: $P(A \cap B) = P(A|B)P(B)$
- The following are equivalent: Two events A and B are said to be *independent*; $P(A|B) = P(A)$; $P(B|A) = P(B)$; $P(A \cap B) = P(A)P(B)$.
- *Law of Total Probability*: Given disjoint $E_1, E_2, \ldots E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$, for any A: $P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)$
- *Bayes:* $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$

Random Variables

Definition: *Random Variable*
A *random variable* is a (measurable) function that maps elements or events in the sample space Ω to the real numbers $a_1, a_2, \ldots$ (or, more generally, to a measurable space...whatever that is!)
Example: Consider rolling two dice. The *Sample Space* is the full list of outcomes $\{\omega_1, \omega_2\}$.
But what if we only care about summing the two dice? We could skip the sample space and just count the *random variable:*

$X :=$ the sum of the two dice.

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

If X is discrete, the pdf provides answers to questions like ________________. It is also called a probability mass function (pmf).

If X is continuous, then ____________ $=0$ for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

If X is discrete, the pdf provides answers to questions like $\underline{f(x) = P(X = x)}$. It is also called a probability mass function (pmf).

If X is continuous, then $\underline{P(X = x)=0}$ for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:
"What is the probability that X takes on a value between a and b?"

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

2. (For discrete distributions:)

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1. $$f(x) = P(X = x) \geq 0 \qquad \forall x \text{ (with events in } \Omega)$$

2. (For discrete distributions:)
$$\sum_{x \in \Omega} f(x) = \sum_{x \in \Omega} P(X = x) = 1$$

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Making a pdf

Recall; last time our opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space:
- Associated r.v. possible values or *support:*
- pdf $P(X = x) =$ probability of x tails before a heads:

Making a pdf

Recall; last time our opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space: $\{H, TH, TTH, TTTH, \dots\}$
- Associated r.v. possible values or *support:* $\{0, 1, 2, 3, \dots\}$
- pdf $P(X = x) =$ probability of x tails before a heads:

$$P(X = x) = P(\{T \dots TH\}) = P(\{T\})^x P(\{H\}) = (1-p)^x \cdot p$$

So we report $f(x) = (1-p)^x \cdot p$

Daily Recap

Today we learned

1. Bayes Review... and pdfs and cdfs!

Moving forward:

- nb day Friday!
- Monday: HW 2

Next time in lecture:

- We start giving special and common pdfs names!