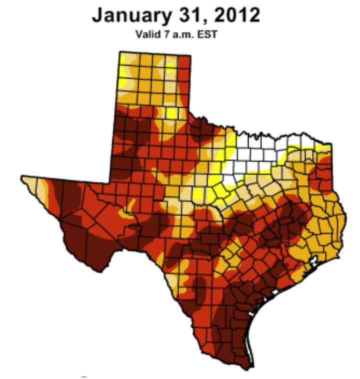
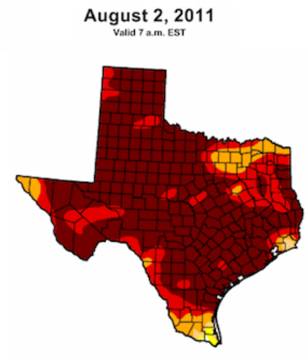
Conditional Probability

CE 311S

INTRODUCTION

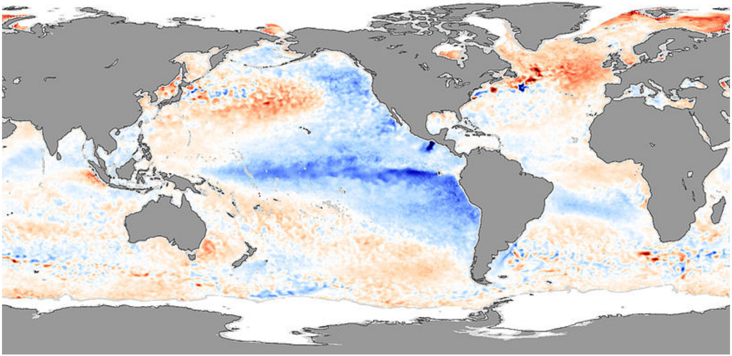
Sometimes, two events are related in that knowing one has happened influences the probability that the other has happened.





What caused the 2011 drought in Texas?

The El Niño/La Niña-Southern Oscillation is a climate pattern which occurs in the Pacific Ocean.



La Niña events tend to correspond with drier weather in Texas.

Consider the following (fictional) historical record from the last 100 years

	La Niña years	No La Niña
Drought years	20	25
No drought	10	45

In general, what is the probability there will be a drought? If this year is a La Niña year, what is the probability there will be a drought?

In general, what is the probability there will be a drought?

	La Niña years	No La Niña
Drought years	20	25
No drought	10	45

From the data, there have been 45 drought years out of 100 on record, so the probability is 0.45

If this year is a La Niña year, what is the probability there will be a drought?

	La Niña years	~~No La Niña~~
Drought years	**20**	~~25~~
No drought	**10**	~~45~~

Only looking at the La Niña years, 20 out of 30 have been drought years, so the probability is 2/3.

Can we generalize the logic we used?

	La Niña years	~~No La Niña~~
Drought years	**20**	~~25~~
No drought	**10**	~~45~~

Essentially, we did the following calculation:

$$\frac{20}{30} = \frac{20/100}{30/100} = \frac{P(\text{drought} \cap \text{La Niña})}{\text{La Niña}}$$

This is called the **conditional probability** of a drought, given that this is a La Niña year, which we write as

$$P(\text{drought} \mid \text{La Niña})$$

For any two events A and B, the conditional probability of A given B is

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

if $P(B) > 0$.

We can rearrange this expression to get

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

which is sometimes called the **multiplication rule**.

EXAMPLE

Example

A recent survey shows that 50% of computer users primarily use a Gmail account, 40% primarily use a Yahoo! account, and 10% primarily use an AOL account. 10% of the Gmail users' computers are infected with a virus, compared to 40% of the Yahoo! users' computers and 90% of the AOL-ers'.

1. What is the probability that a random person uses a Gmail account and has a virus?
2. What is the probability that a random person has a virus?
3. If a random person has a virus, what is the probability they use AOL?

Solution by Formula

What is the probability that a random person uses a Gmail account and has a virus?

Define the following events:

G	The person in question primarily uses Gmail
Y	The person in question primarily uses Yahoo!
A	The person in question primarily uses AOL
V	The person in question has a virus on their computer
V^c	The person in question dose not have a virus on their computer

$$P(G \cap V) = P(V \mid G)P(G) = 0.1 \times 0.5 = 0.05$$

Solution by Formula

What is the probability that a random person has a virus?

Because either G or Y, or A occurs (and because they are disjoint), $P(V) = P(V \cap G) + P(V \cap Y) + P(V \cap A)$. (Why?)

We already found $P(V \cap G) = 0.05$; in the same way $P(V \cap Y) = 0.16$ and $P(V \cap A) = 0.09$.

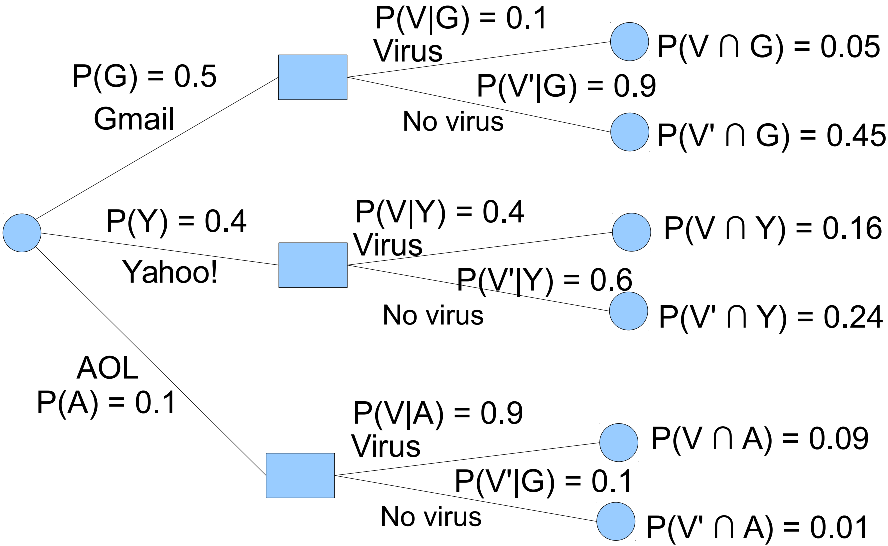
Therefore $P(V) = 0.05 + 0.16 + 0.09 = 0.3$

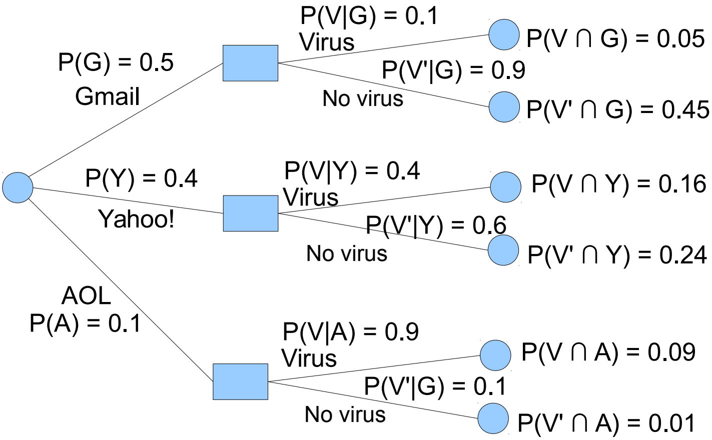
Solution by Formula

If a random person has a virus, what is the probability they use AOL?

$$P(A \mid V) = \frac{P(A \cap V)}{P(V)} = \frac{0.09}{0.3} = 0.3$$

Tree diagrams offer another solution method.





- Could I have built the tree in a different order (with V at the "top" level?)
- What is the probability that a random citizen has a virus?

LAW OF TOTAL PROBABILITY

The most difficult question from the previous example was "what is the probability that a random person has a virus?"

We answered this question by saying that if a person has a virus, then *either* they are a Gmail user with a virus *or* a Yahoo! user with a virus *or* an AOL-er with a virus (but not more than one of these!), so we can just add those probabilities.

This logic is generalized in the **law of total probability**.

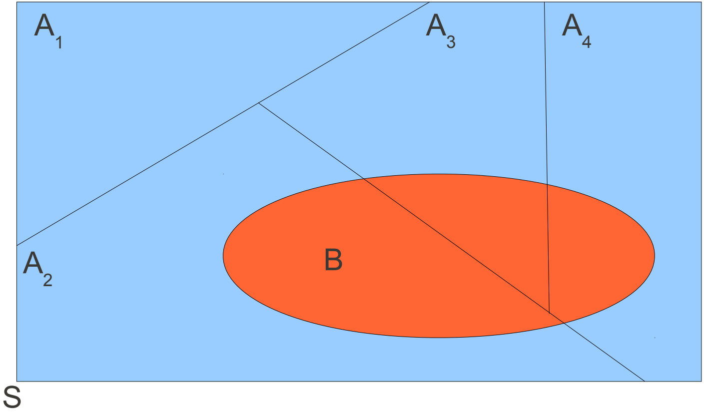
Law of Total Probability

If $B_1, \ldots, B_k$ are a partition (mutually exclusive and exhaustive events), then the probability of any event A is given by

$$\begin{aligned} P(A) &= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A \mid B_i)P(B_i) \end{aligned}$$

Proof Sketch

$P(A \mid B_i)P(B_i) = P(A \cap B_i)$ by the multiplication rule



Since the B_i are mutually exclusive and exhaustive, $P(A) = \sum_i P(A \cap B_i)$ by Axiom 3

BAYES’ THEOREM

Bayes' Theorem

Let $B_1, B_2, \ldots, B_k$ be a partition. Then for any event A with $P(A) > 0$,

$$P(B_j \mid A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

for $j = 1, \ldots, k$

Bayes' Theorem lets us flip the condition around and calculate $P(B_j \mid A)$ if we know $P(A \mid B_i)$ and $P(B_i)$ for all i.

The law of total probability and Bayes' theorem formalize the logic we used to find out the second and third answers in the computer virus example.

The TSA is considering automated facial recognition software to identify terrorists and suspicious-acting people.

Is this technology a good idea?

Let's give the system the benefit of the doubt (by many orders of magnitude):

1. If a passenger is a terrorist, they are correctly identified **with probability 1**.
2. If a passenger is not a terrorist, they are correctly identified **with probability 0.999**.

There are roughly 600,000,000 air passengers each year in the United States. (Very generously) assuming that 100 of them are terrorists, **what is the probability that somebody flagged by this system is a terrorist?**

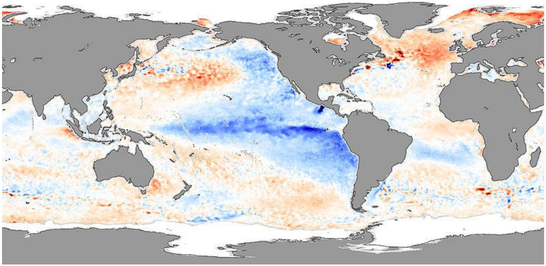
- Experiment: A passenger is analyzed by the software.
- Events: the passenger is a terrorist (T); the passenger is flagged by the system (F)

By Bayes' Theorem

$$\begin{aligned} P(T|F) &= \frac{P(F|T)P(T)}{P(F|T)P(T) + P(F|T')P(T')} \\ &= \frac{1 \times \frac{100}{600{,}000{,}000}}{1 \times \frac{100}{600{,}000{,}000} + 0.001 \times \frac{599{,}999{,}900}{600{,}000{,}000}} \\ &\approx 0.00017 \end{aligned}$$

(Remember this was done with extremely optimistic assumptions. Using more realistic numbers, the probability is about 0.00000001)

INDEPENDENCE



The idea of conditional probability is that **two events may be related, so knowing that one occurs influences the likelihood of the other.** In this class, we develop the opposite line of thought: what does it mean for two events to be completely unrelated?

Two events A and B are **independent** if $P(A \mid B) = P(A)$ and **dependent** otherwise.

Is this definition "asymmetric?"

Example

If I flip a coin twice, let A and B be the events "heads on the first flip" and "heads on the second flip."

The four possible outcomes are $\{HH, HT, TH, TT\}$. By counting, $P(A) = 2/4 = 1/2$.

If we know that event B happens, the only possible outcomes are $\{HH, TH\}$ and $P(A \mid B) = 1/2$.

Thus $P(A) = P(A \mid B)$, so A and B are independent.

Example

Let A and B be mutually exclusive events with positive probability. Are they independent?

MULTIPLICATION RULE

Events A and B are independent iff $P(A \cap B) = P(A)P(B)$.

Because these two criteria are equivalent, we could have used $P(A \cap B) = P(A)P(B)$ as our **definition** of independence and then derived $P(A \mid B) = P(A)$ from this definition.

Example

If I flip a coin twice, let A and B be the events "heads on the first flip" and "heads on the second flip."

Because A and B are independent, the probability $P(A \cap B) = P(A)P(B) = 1/4$.

One way to describe the multiplication rule is that "independent probabilities multiply" when you want to find the probability that both events occur.

The following result can be shown to be true from the mathematical definition of independence:

A and B independent $\iff$ A^c and B independent $\iff$ A and B^c independent $\iff$ A^c and B^c independent

Example

If a roll a six-sided die twice, what is the probability that at least one of the rolls is a six?

It is easier to find the probability that this *doesn't* happen, then subtract from 1. Let A be the event "the first roll is a six" and B be the event "the second roll is a six."

From a Venn diagram, we see that $A \cup B$ (the event we are interested in) is $(A^c \cap B^c)'$.

A and B are independent, so
$P(A^c \cap B^c) = P(A^c)P(B^c) = (5/6)(5/6) = 25/36$. Thus
$P(A \cup B) = 1 - 25/36 = 11/36$.

What would independence look like on a tree diagram?

EXAMPLE

Seventy percent of all vehicles examined at a certain emissions inspection station pass the inspection. Assuming that successive vehicles pass or fail independently of one another, calculate the probabilities of the following events:

1. All of the next three vehicles inspected pass
2. At least one of the next three inspected fails
3. Exactly one of the next three inspected passes
4. At most one of the next three inspected passes
5. Given that at least one of the next three vehicles passes inspection, what is the probability that all three pass?

CONDITIONAL INDEPENDENCE

Definition

Events A and B are **conditionally independent given** C if $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$.

The idea is that once we know that C occurred, knowing that A occurs tells us nothing more about B, and vice versa.

Example

Assume that I can either take the 5 or 7 bus home. Are the events "the 5 is late" and "the 7 is late" independent?

No, because whatever made the 5 late could also make the 7 late (bad traffic downtown, etc.).

However, they could be conditionally independent given the event "bad traffic downtown." Once I know that there is congestion downtown, whether the 5 is late or not may not have any bearing on whether the 7 is late.

SOME REAL-WORLD TAKEAWAYS

Conditional probabilities are a common source of confusion — both for students and in real-world applications.

- Confusing the condition: "most murderers are men" is NOT the same as "most men are murderers." $P(A \mid B) \neq P(B \mid A)$
- Confusing conditioning with "and," $P(A \mid B)$ vs. $P(A \cap B)$.

The prosecutor's fallacy

In 1998, a woman was accused of killing her two children (both died a few months old). An expert witness testified that the odds of a child dying from SIDS is about 1 in 8500, so the odds of *two* children dying from SIDS in the same family was 1 in 73 million.

The prosecution used this low probability to argue that she was almost certainly guilty of murder.

What things are wrong with these arguments?

Where should airplanes be reinforced?

In World War II, the Army commissioned a statistician to observe where planes returning from missions had been shot. The plan was to provide additional reinforcement to planes in areas with many bullet holes.

What's wrong with this?

Data dredging

The odds of a false positive in a DNA test are 1 in 10,000.

A match is found when examining a bank of 20,000 samples.

How strong is the evidence against the match?