Intuition Test

An olympic gymnast fails a doping test that is known to be 95 percent accurate. How likely is it that he is really guilty?

You cant really answer it without knowing some other things:

- How many gymnasts are actually using illegal performance enhancing drugs?
- How many total gymnasts are there (those who use the drug and those who don't)

Intuition Test

A rare disease X affects 1,000 out of 1,000,000 individuals. It has no obvious external symptoms, and susceptibility can't be inferred from medical or family history. It strikes at random.

A test for X is 99% accurate — the test correctly identifies the presence of X 99% of the time that it is present, and correctly identifies the absence of X 99% of the time that is not present.

Being a hypocondriac, I have myself tested for X, and it comes back positive. What is the probability that I have X?

A: more than 90%
B: around 60%
C: around 30%
D: less than 10%

Bayes' Theorem

Example: Two bags contain red and white marbles:

$$\underbrace{\boxed{3R \quad 5W}}_{B_1} \qquad\qquad \underbrace{\boxed{7R \quad 4W}}_{B_2}$$

You pick a bag at random (each bag is equally likely to be picked), and select a marble. What is the probability that the marble is Red?

$$P(R) = \frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{7}{11} \approx 0.5057$$

Q: The marble you picked turns out to be Red. What is the probability that it came from Bag 1?

Bayes' Theorem

Q: The marble you picked turns out to be Red. What is the probability that it came from Bag 1?

$$\underbrace{\boxed{3R \quad 5W}}_{B_1} \qquad\qquad \underbrace{\boxed{7R \quad 4W}}_{B_2}$$

We are looking for $P(B_1|R)$. The events B_1 (picking Bag 1) and B_2 (picking Bag 2) **partition the sample space** into two pieces. Event R can be written as:

$$P(R) = \underbrace{P(B_1)P(R|B_1)}_{P(B_1 \cap R)} + \underbrace{P(B_2)P(R|B_2)}_{P(B_2 \cap R)}$$

$$P(B_1|R) = \frac{P(B_1 \cap R)}{P(R)} = \frac{\frac{1}{2} \cdot \frac{3}{8}}{\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{7}{11}} \approx 0.3708$$

Bayes' Theorem

This was an example of using **Bayes' Theorem**. It allows us to calculate **reverse conditional probabilities** — i.e., how we calculate $P(B|A)$ when we know $P(A|B)$.

Suppose B_1 and B_2 are mutually exclusive events that **partition our sample space**, and E is any event. Notice

$$P(E) = \underbrace{P(B_1)P(E|B_1)}_{P(E\cap B_1)} + \underbrace{P(B_2)P(E|B_2)}_{P(E\cap B_2)}$$

Then if E occured, the chance that B_1 occured is

$$\boxed{\begin{aligned} P(B_1|E) = \frac{P(B_1 \cap E)}{P(E)} &= \frac{P(B_1 \cap E)}{P(E \cap B_1) + P(E \cap B_2)} \\ &= \frac{P(B_1)P(E|B_1)}{P(B_1)P(E|B_1) + P(B_2)P(E|B_2)} \end{aligned}}$$

Factory example

Example: A factory has two machines (A and B), both producing touch screens. 40% of production is from Machine A and 60% from Machine B. 10% of the screens produced by Machine A are defective and 5% from Machine B are defective.

If I randomly choose a touch screen produced in the factory, then there is a 40% probability that it came from Machine A.

I test the randomly chosen screen, and find that it is defective. What is the probability that it came from Machine A? Greater or less than 40%?

$$P(A) = 0.4, \quad P(B) = 0.6, \quad P(D|A) = 0.10, \quad P(D|B) = 0.05$$

Factory example

The given information is listen below. Notice that A and B partition our sample space in two pieces.

$$P(A) = 0.4, \quad P(B) = 0.6, \quad P(D|A) = 0.10, \quad P(D|B) = 0.05$$

We want $P(A|D)$, a reverse probability, so we use Bayes' Theorem.

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{(0.4)(0.10)}{\underbrace{(0.4)(0.10)}_{P(D \cap A)} + \underbrace{(0.6)(0.05)}_{P(D \cap A)}} \approx 0.5714$$

so there is a 57% chance that the defective screen came from machine A. This is larger than the 40% chance that any randomly chosen screen comes from A.

A More General Bayes' Theorem

If $B_1, B_2, \ldots, B_n$ are (mutually exclusive) events that partition our sample space, and E is any event, then

$$P(E_1|F) = \frac{P(E_1 \cap F)}{P(F)} =$$

$$\frac{P(E_1 \cap F)}{P(E_1 \cap F) + P(E_2 \cap F) + \cdots + P(E_n \cap F)} =$$

$$\frac{P(E_1)P(F|E_1)}{P(E_1)P(F|E_1) + P(E_2)P(F|E_2) + \cdots + P(E_n)P(F|E_n)}$$

A More General Bayes' Theorem

Example: A pile of 8 playing cards has 4 aces, 2 kings and 2 queens. A second pile of 8 playing cards has 1 ace, 4 kings and 3 queens.

First Pile	Second Pile
$\boxed{4A \quad 2K \quad 2Q}$	$\boxed{1A \quad 4K \quad 3Q}$

You conduct the following experiment

- randomly choose a card from the first pile and place it in the second pile
- shuffle the second pile, and you randomly choose a card from the second pile

Q: If the card drawn from the second deck was an ace, what is the probability that the first card was also an ace?

A More General Bayes' Theorem

$\boxed{4A \;\; 2K \;\; 2Q}$
First Pile

$\boxed{1A \;\; 4K \;\; 3Q}$
Second Pile

For the first card, let **A** be the event that you draw an ace, **K** the event that you draw a king and **Q** be the event that you draw a queen.

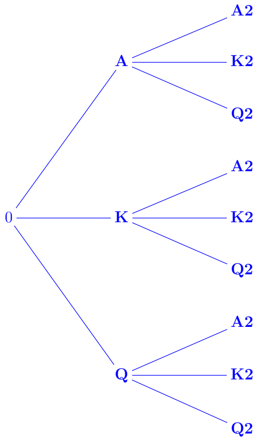
For the second card, let **A2** be the event that you draw an ace, **K2** the event that you draw a king and **Q2** be the event that you draw a queen.

We are looking to find $P(A|A2)$.

$$P(A|A2) = \frac{P(A \cap A_2)}{P(A2)} = \frac{\frac{4}{8} \cdot \frac{2}{9}}{\underbrace{\frac{4}{8} \cdot \frac{2}{9}}_{A \cap A_2} + \underbrace{\frac{2}{8} \cdot \frac{1}{9}}_{K \cap A_2} + \underbrace{\frac{2}{8} \cdot \frac{1}{9}}_{Q \cap A_2}} \approx 0.667$$

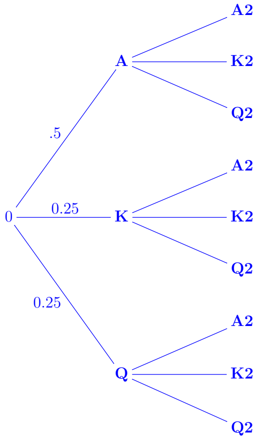
A More General Bayes' Theorem

The next few slides show a tree diagram method of solving the problem.



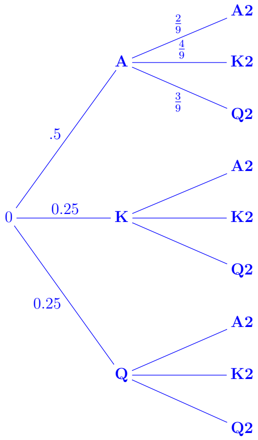
A More General Bayes' Theorem

In the first round there are $4 + 2 + 2 = 8$ cards so the probabilities in the first round are



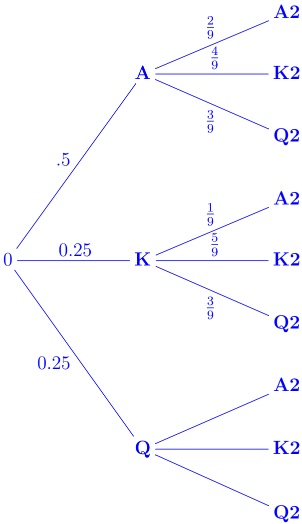
A More General Bayes' Theorem

In the second round there are $1 + 4 + 3 + 1 = 9$ cards and the probabilities are different at the various nodes. If you draw an ace in round 1 the cards are 2 aces, 4 kings and 3 queens so we get



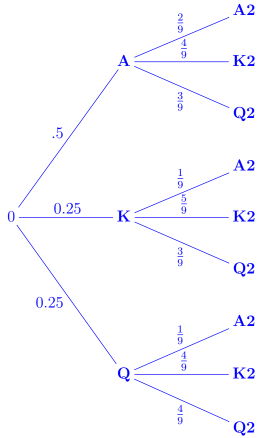
A More General Bayes' Theorem

If you draw a king in round 1 the cards are 1 ace, 5 kings and 3 queens so we get

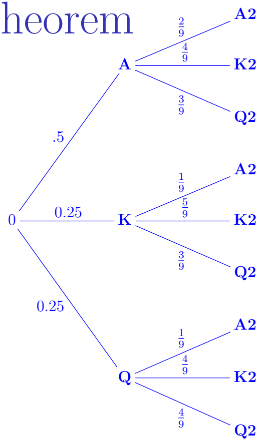


A More General Bayes' Theorem

If you draw a queen in round 1 the cards are 1 ace, 4 kings and 4 queens so we get



A More General Bayes' Theorem



The question asks for

$$P(A|A2) = \frac{P(A \cap A2)}{P(A2)} = \frac{\frac{1}{2} \cdot \frac{2}{9}}{\frac{1}{2} \cdot \frac{2}{9} + \frac{1}{4} \cdot \frac{1}{9} + \frac{1}{4} \cdot \frac{1}{9}} = \frac{2}{3}$$

Predictive Value Of Diagnostic Tests

Example: A rare disease X affects 1,000 out of 1,000,000 individuals. It has no obvious external symptoms, and susceptibility can't be inferred from medical or family history. It strikes at random.

A test for X is 99% accurate — the test correctly identifies the presence of X 99% of the time that it is present, and correctly identifies the absence of X 99% of the time that is not present.

Being a hypocondriac, I have myself tested for X, and it comes back positive. What is the probability that I have X?

In other words, what is $P(X|Pos)$? 0.0991 or 9.9% chance

Predictive Value Of Diagnostic Tests

Example: In a certain country, 40% of the residents have condition X, and the test for X is 95% accurate. Suppose a random resident of Country C tests positive for X. What is the probability that the person actually has X?

$$
\begin{aligned}
P(X|Pos) &= \frac{P(X \cap Pos)}{P(Pos)} \\
&= \frac{(0.4)(0.95)}{(0.4)(0.95) + (0.6)(0.05)} \approx 93\%
\end{aligned}
$$

Predictive Value Of Diagnostic Tests

Example: A test for Lyme disease is 60% accurate when a person has the disease and 99% accurate when a person does not have the disease. In a certain country, 0.01% of the population has Lyme disease. If a randomly chosen person tests positive, what is the probability that the person actually has the disease?

$$\begin{aligned} P(L|Pos) &= \frac{P(L \cap Pos)}{P(Pos)} \\ &= \frac{(0.0001)(0.60)}{(0.0001)(0.60) + (0.9999)(0.01)} \\ &\approx 0.0059 \text{ or } 0.6\% \end{aligned}$$

The Papanicolaou Smear

The Pap smear is a screening procedure used to detect cervical cancer. Out of women with this cancer, 16% are *false negatives*:

$$P(TN|C) = 0.16 \quad \text{and} \quad P(TP|C) = 0.84$$

For women without this cancer, there are about 19% *false positives*:

$$P(TP|C') = 0.19 \quad \text{and} \quad P(TN|C') = 0.81$$

In the U.S., there are about 8 women in 100,000 who have this cancer:

$$P(C) = 0.00008 \quad \text{and} \quad P(C') = 0.99992$$

The Papanicolaou Smear

Q: Given that a patient tests positive, what is the probability that she actually has cervical cancer?

- $P(TN|C) = 0.16$ and $P(TP|C) = 0.84$
- $P(TP|C') = 0.19$ and $P(TN|C') = 0.81$
- $P(C) = 0.00008$ and $P(C') = 0.99992$

$$\begin{aligned} P(C|TP) &= \frac{P(C \text{ and } TP)}{P(TP)} \\ &= \frac{(0.00008)(0.84)}{(0.00008)(0.84) + (0.99992)(0.19)} \\ &\approx 0.000354 \end{aligned}$$

What this means: for every million positive Pap smears, only about 354 represent true cases of cervical cancer!

Legal Cases

A crime has been committed and the only evidence is a blood spatter that could only have come from the perpetrator. The chance of a random individual having the same blood type as that of the spatter is 10%. Joe has been arrested and charged. The trial goes as follows:

Prosecutor: Since there is only a 10% chance that Joe's blood would match, there is a 90% chance that Joe did it.

Defence Lawyer: There are two hundred people in the neighborhood who could have done the crime. Twenty of them (10% of 200) will have the same blood type as the sample. Hence the chances that Joe did it are $\frac{1}{20} = 5\%$ so there is a 95% chance that Joe is innocent.

Legal Cases

Reverend Thomas Bayes: You're all nuts!

Consider the events:

- I = *Joe is innocent*
- G = *Joe is guilty* (the complement of I)
- M = *blood type is a match*

If $P(I) = x$ (unknown), then $P(G) = 1 - x$ and

$$P(I|M) = \frac{0.1 \cdot x}{0.1 \cdot x + 1 \cdot (1 - x)} = \frac{0.1x}{1 - 0.9x}$$

If $P(I) = x$ then $P(I|M) = \dfrac{0.1x}{1 - 0.9x}$.

If you are initially certain that $P(G) = 1$, then $x = 0$, and after seeing the evidence you are still certain: $P(I|M) = 0$.

If you are initially certain that $P(I) = 1$, then $x = 1$ and after seeing the evidence you are still certain: $P(I|M) = 1$.

If you initially think that $P(G) = 60\%$, then $x = 0.4$ and after seeing the evidence, $P(I|M) = 0.0625$.

If you suspect that the cops searched a blood-type database until they came up with a name in the neighborhood (20 total with that blood type), then you might initially think that $x = P(I) = \frac{19}{20} = 95\%$ (assuming only 1 of the 20 is guilty). Now after seeing the evidence, Bayes suggests revising to $P(I|M) = 0.66$.

Notice: The evidence lowers the perception of innocence.

Some resources

Here's an article on the predictive value of diagnostic tests:

`Doctors flunk quiz on screening-test math`

If you are more legally inclined, here is a discussion of Bayes Theorem as it applies to criminal trials.