Bayes' Theorem

Bayes' Theorem (Short form)

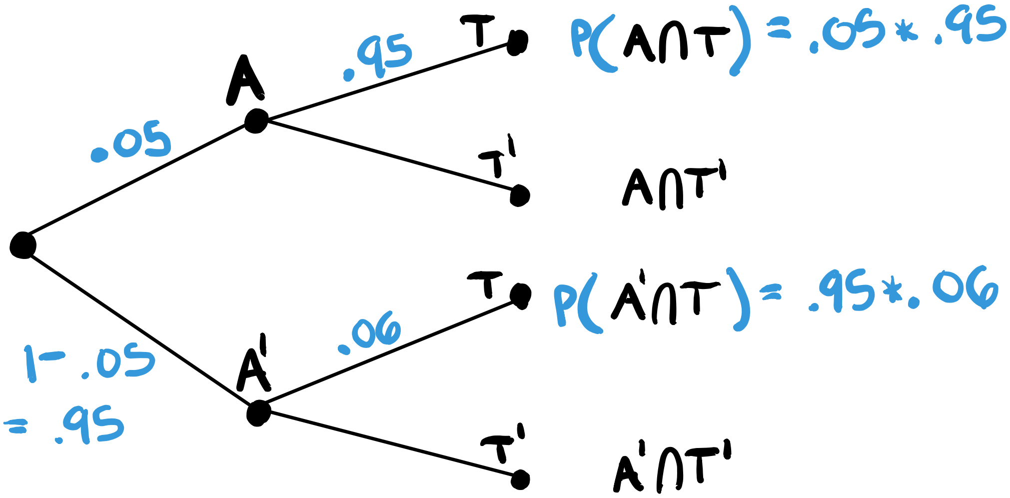
A college sports league has found that 5% of its players use an anabolic steroid. The sports league would like to contract with a testing firm. One firm claims that it correctly identifies users 95% of the time and that non users only test positive 6% of the time. To evaluate the efficacy of their tests you wonder how likely someone that tests positive is actually a user.

$P(A) = .05$ $\quad P(T \mid A) = .95$ $\quad P(T \mid A') = .06$

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

$P(A) = .05 \quad P(T|A) = .95 \quad P(T|A') = .06$

$$P(A|T) = \frac{P(A \cap T)}{P(T)} = \frac{0.05 * .95}{0.05 * .95 + .95 * .06}$$



$$P(T) = P(A \cap T) + P(A' \cap T)$$
$$= .05 * .95 + .95 * .06$$

$$P(A|T) = \frac{P(A \cap T)}{P(T)} = \frac{0.05 * .95}{0.05 * .95 + .95 * .06}$$

$$= 0.\overline{45}$$

Less than half the people that test positive are users meaning this might be a very poor diagnostic

The process we went through is encapsulated in Bayes' Theorem, which can be stated as

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

ex $P(L) = .005$, $P(T'|L) = .01$, $P(T|L') = .2$, find $P(L|T)$

$$P(L|T) \overset{BT}{=} \frac{P(L)P(T|L)}{P(L)P(T|L) + P(L')P(T|L')}$$

$P(L')$: $P(L) + P(L') = 1$, $.005 + P(L') = 1$

$P(L') = .995$

$P(T|L)$: $P(T|L) + P(T'|L) = 1$

$P(T|L) + .01 = 1$

$P(T|L) = .99$

$$P(L|T) = \frac{0.005 * 0.99}{0.005 * 0.99 + 0.995 * 0.2} \approx 0.024$$

Expanded Bayes' Theorem

The shortened Bayes' Theorem tells us that

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

and this denominator comes from

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

This can be done because B and B' <u>partition</u> the sample space. Meaning

(i) $B \cap B' = \emptyset$

(ii) $B \cup B' = S$

For different problems we can use different partitions. If we had 3 events T_1, T_2, T_3 that were all mutually exclusive and $T_1 \cup T_2 \cup T_3 = S$, then we could say

$$P(T_1|A) = \frac{P(T_1)P(A|T_1)}{P(T_1)P(A|T_1) + P(T_2)P(A|T_2) + P(T_3)P(A|T_3)}$$

In general, if $E_1, \ldots, E_n$ partition the sample space, then

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{P(E_1)P(A|E_1) + \cdots + P(E_n)P(A|E_n)}$$

this can be called the expanded Bayes' theorem.

2 Word Problems

On average, every 100 years the town of Castle Rock experiences a major earthquake. There is thought to be a 5% chance that this town's natural rock formation would collapse during a major earthquake. Due to regular erosion, it is also believed that there is a 0.07% chance that this formation would collapse in any given year. What is the probability that there is a major earthquake in the same year that the formation collapses?

$$P(E \mid C) = ?$$

$$P(E) = 1/100 = 0.01$$

$$P(C \mid E) = .05$$

$$P(C) = .0007$$

$$P(E|C) = \frac{P(E \cap C)}{P(C)} = \boxed{\frac{P(E)P(C|E)}{P(C)}} = \frac{P(E)P(C|E)}{P(E)P(C|E) + P(E')P(C|E')}$$

$$P(E|C) = \frac{0.01 * 0.05}{0.0007} = 0.714$$

71.4% chance of an earthquake in the year of collapse

A dog food company is doing research on single dog owners and found that 27% of single dog owners had a small dog, 39% had a medium sized dog, and that 34% had a large dog. They also found that, by dog size, the probability the owner bought food from them was

74%	80%	60%
small	medium	large.

They want to know for s.d.o.'s that buy their food if it's more likely for a small dog or a large dog.

$P(S) = .27 \quad P(M) = .39 \quad P(L) = .34$

$P(F|S) = .74 \quad P(F|M) = .8 \quad P(F|L) = .6$

We want to find and compare

$P(S|F) \qquad P(L|F)$

$$P(S|F) = \frac{P(S)P(F|S)}{P(S)P(F|S) + P(M)P(F|M) + P(L)P(F|L)}$$

$$= \frac{.27 * .74}{.27 * .74 + .39 * .8 + .34 * .6} = 0.279$$

$$P(L|F) = \frac{.34 * .6}{.27 * .74 + .39 * .8 + .34 * .6} = 0.285$$