This is a draft; please send me corrections and/or suggestions. Also, this is not a carbon-copy of section, but is meant to provide a reference if your notes are shy of detail.

Supermarket Optimization

Safeway is trying to maximize sales of its store-brand breakfast cereals. It can sell cereal from General Mills, Post, Kellogg, or its own brand. There is room for three brands in the cereal aisle. If order is not important, how many ways can Safeway-brand cereal be sold next to other brands?

There is room for three brands of cereal in the cereal aisle; since Safeway-brand must be one, the store managers should choose two brands from the remaining three to fill the shelves. The number of possible sales configurations is then

$$_3C_2 = \binom{3}{2} = \frac{3!}{2!(3-2!)} = 3.$$

Let S represent Safeway, G represent General Mills, K represent Kellogg, and T represent Post. Safeway finds three stores, one offering each of the desired lineups, and runs the following experiment: 100 customers go shopping, and it records the number of sales of various brands of cereal. The results are as follows:

	S	**G**	**K**	**T**
SGK	5	10	6	0
SGT	4	7	0	3
SKT	6	0	7	9

Which treatment maximizes the probability of purchasing Safeway-brand cereal? Which treatment maximizes the probability of purchasing Safeway-brand cereal, among the customers buying cereal?

Given 100 customers in each treatment, we know that there is a probability of $\frac{100}{300} = \frac{1}{3}$ that a customer receives a particular cereal lineup. We can then compute

$$\mathrm{P(S|SGK)} = \frac{\mathrm{P(S \cap SGK)}}{\mathrm{P(SGK)}} = \frac{5/300}{100/300} = 0.05,$$
$$\mathrm{P(S|SGT)} = \frac{\mathrm{P(S \cap SGT)}}{\mathrm{P(SGT)}} = \frac{4/300}{100/300} = 0.04,$$
$$\mathrm{P(S|SKT)} = \frac{\mathrm{P(S \cap SKT)}}{\mathrm{P(SKT)}} = \frac{6/300}{100/300} = 0.06.$$

So putting out Safeway-brand cereal with Kellogg- and Post-brand cereals maximizes the probability that an individual customer buys cereal.

To compute the latter quantities, we again use the definition of conditional probability,

$$\mathrm{P(S|SGK, cereal)} = \frac{\mathrm{P(S \cap SGK \cap cereal)}}{\mathrm{P(SGK \cap cereal)}} = \frac{5/300}{21/300} = \frac{5}{21} < \frac{6}{21},$$
$$\mathrm{P(S|SGT, cereal)} = \frac{\mathrm{P(S \cap SGT \cap cereal)}}{\mathrm{P(SGT \cap cereal)}} = \frac{4/300}{14/300} = \frac{4}{14} = \frac{6}{21},$$
$$\mathrm{P(S|SKT, cereal)} = \frac{\mathrm{P(S \cap SKT \cap cereal)}}{\mathrm{P(SKT \cap cereal)}} = \frac{6/300}{22/300} = \frac{6}{22} < \frac{6}{21}.$$

Then we can see that the offering Safeway-, General Mills-, and Post-brand cereals maximizes the purchasing of Safeway-brand cereals by cereal-buyers.

Does it look like our simple model of the world — "only the cereals offered matters" — correctly captures what is going on here? Why or why not? Ignoring this, is there a reason that a store manager might want to sell more Safeway-brand cereal to cereal buyers, rather than simply more Safeway-brand cereal overall? These are not "Econ 41 topics," but better-capture the spirit in which actual econometrics is applied – we have the tools to discuss them, so why not.

Bayes' Law

As we have seen in class, the definition of conditional probability tells us that

$$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B),$$
$$\mathrm{P}(A \cap B) = \mathrm{P}(B|A)\mathrm{P}(A).$$

Putting these two facts together, we obtain a result known as *Bayes' Law* (alterately, *Bayes' Theorem*),

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

While Bayes' Law is simple to derive from the definition of conditional probability, it provides a useful result that allows us to state one (unknown) conditional probability in terms of a (known) conditional probability and the probabilities of the sets involved.

A True Friend

You have a friend who tells the truth 90% of the time; there is a 1% chance that Bigfoot exists. You're hanging out with your friend when he mentions that he saw Bigfoot off the 405. Is he more likely telling the truth or lying?[1]

For useful shorthand, we will consider the events {claims} and {saw} to represent that your friend claims he saw Bigfoot, and that he actually saw Bigfoot (respectively). The question asks us to consider P(saw|claims). By Bayes' Law, we know

$$\mathrm{P}(\text{saw}|\text{claims}) = \frac{\mathrm{P}(\text{claims}|\text{saw})\mathrm{P}(\text{saw})}{\mathrm{P}(\text{claims})}.$$

The two values in the numerator are parameters to the question — $\mathrm{P}(\text{claims}|\text{saw}) = 0.9$, and $\mathrm{P}(\text{saw}) = 0.01$ — but the denominator must be calculated. Applying the law of total probability, we know

$$\begin{aligned}\mathrm{P}(\text{claims}) &= \mathrm{P}(\text{claims} \cap \text{saw}) + \mathrm{P}(\text{claims} \cap \text{saw}')\\ &= \mathrm{P}(\text{claims}|\text{saw})\mathrm{P}(\text{saw}) + \mathrm{P}(\text{claims}|\text{saw}')\mathrm{P}(\text{saw}')\\ &= 0.9(0.01) + 0.1(0.99).\end{aligned}$$

The probability that your friend is telling the truth is then

$$\mathrm{P}(\text{saw}|\text{claims}) = \frac{0.9(0.01)}{0.9(0.01) + 0.1(0.99)} = \frac{1}{12}.$$

Using the same logic, how "trustworthy" must your friend be in order for you to believe him?

[1] We will assume that your friend tells you everything that has ever happened to him, and that if Bigfoot exists he has seen it. This is to clarify a slight confusion between Bigfoot's existence and whether or not your friend saw it, or that he saw it but didn't bother to tell you.

Now suppose that there the probability that your friend tells the truth is t. The formula above says that

$$\begin{aligned}
\mathrm{P}(\text{saw}|\text{claims}) &= \frac{\mathrm{P}(\text{claims}|\text{saw})\mathrm{P}(\text{saw})}{\mathrm{P}(\text{claims})} \\
&= \frac{\mathrm{P}(\text{claims}|\text{saw})\mathrm{P}(\text{saw})}{\mathrm{P}(\text{claims}|\text{saw})\mathrm{P}(\text{saw}) + \mathrm{P}(\text{claims}|\text{saw}')\mathrm{P}(\text{saw}')} \\
&= \frac{0.01t}{0.01t + 0.99(1-t)} \\
&= \frac{t}{99 - 98t}.
\end{aligned}$$

For you to believe your friend, there should be a 50% chance he is telling the truth. So we want

$$\begin{aligned}
& & \mathrm{P}(\text{saw}|\text{claims}) &\geq \frac{1}{2} \\
&\iff & \frac{t}{99-98t} &\geq \frac{1}{2} \\
&\iff & 2t &\geq 99 - 98t \\
&\iff & 100t &\geq 99 \\
&\iff & t &\geq 0.99.
\end{aligned}$$

So in order for you to believe your friend, there should be a 99% chance that he tells the truth. That this is exactly the probability that Bigfoot does not exist is a mathematical coincidence, resulting from the simple (binary) nature of the problem and the symmetric (identical regardless of whether or not Bigfoot has been seen) probability of truth-telling.

A Medical Test

Ponies[2] *is a terminal illness which can be cured if detected early. 5% of the general population has* Ponies*. Fortunately, there is a test for* Ponies*, which has no false negatives but has a 10% false-positive rate. You are tested for* Ponies *and the result is positive; what is the probability that you have contracted* Ponies*?*

For ease of notation, let {Ponies} and {positive} denote the events that you have Ponies and that you test positive for Ponies, respectively. We want to know

$$\mathrm{P}(\textsc{Ponies}|\text{positive}).$$

Using Bayes' Law, we know that

$$\mathrm{P}(\textsc{Ponies}|\text{positive}) = \frac{\mathrm{P}(\text{positive}|\textsc{Ponies})\mathrm{P}(\textsc{Ponies})}{\mathrm{P}(\text{positive})};$$

applying the law of total probability, we have

$$\mathrm{P}(\textsc{Ponies}|\text{positive}) = \frac{\mathrm{P}(\text{positive}|\textsc{Ponies})\mathrm{P}(\textsc{Ponies})}{\mathrm{P}(\text{positive}|\textsc{Ponies})\mathrm{P}(\textsc{Ponies}) + \mathrm{P}(\text{positive}|\textsc{Ponies}')\mathrm{P}(\textsc{Ponies}')}.$$

Putting this all together, we see

$$\mathrm{P}(\textsc{Ponies}|\text{positive}) = \frac{1(0.05)}{1(0.05) + 0.1(0.95)} = \frac{1}{2.9} \approx 0.3448.$$

[2]Traditionally, this question is stated in terms of breast cancer. So as to not give off the impression that these numbers have anything to do with actual cancer rates, we'll be doing this question as an analogy. Specifically, I am not a doctor and you should not read too much into this.

So given that you have tested positive, there is a 34.5% chance that you have contracted PONIES.

Since treatment for PONIES *is both expensive and painful, it is not advisable to undergo therapy without being somewhat sure that you have contracted the illness. You have the option to take the test again; the probability of two consecutive false positives is* 1%. *If you test positive for* PONIES *in both tests, what is the probability that you have* PONIES*?*

The solution method here differs little from the above. We see

$$\begin{aligned}
\text{P}(\text{PONIES}|\text{positive x 2}) &= \frac{\text{P}(\text{positive x 2}|\text{PONIES})\text{P}(\text{PONIES})}{\text{P}(\text{positive x 2})} \\
&= \frac{1(0.05)}{\text{P}(\text{positive x 2}|\text{PONIES})\text{P}(\text{PONIES}) + \text{P}(\text{positive x 2}|\text{PONIES}')\text{P}(\text{PONIES}')} \\
&= \frac{0.05}{1(0.05) + (0.01)(0.95)} \\
&= \frac{1}{1.19} \approx 0.8403.
\end{aligned}$$

Consider this: what does it mean that the probability of two consecutive false positives is equal to the probability of a single false positive, squared? Why may or may not this be a reasonable statement? If there is a more intuitive value, is it greater than or less than 1%*? The lecture on independence of events may provide some insight.*

The Pepsi Challenge

You are sampling a series of 10 *soft drinks;* 8 *are Coca-Cola and* 2 *are Pepsi. If the sequence is randomized, what is the probability that the fourth soda you sample is a Pepsi?*

Since writing the initial solution, I've realized there is a much simpler way (proof that TAs also have to work out answers, I suppose). Read the first solution for an exercise in conditional probability and combinatorics, and the second for an exercise in pure combinatorics.

First solution

Using the law of total probability, we can see

$$\begin{aligned}
\text{P}(\text{fourth is Pepsi}) =&\text{P}(\text{fourth is Pepsi} \cap \text{none of first three is Pepsi}) \\
&+ \text{P}(\text{fourth is Pepsi} \cap \text{one of first three is Pepsi}) \\
&+ \text{P}(\text{fourth is Pepsi} \cap \text{more than one of first three is Pepsi}).
\end{aligned}$$

Since there are only two Pepsis in the lineup, the probability that the fourth is Pepsi while more than one of the first three is Pepsi is 0, so the rightmost probability falls out. Using the definition of conditional probability, we know that

$$\begin{aligned}
&\text{P}(\text{fourth is Pepsi} \cap \text{none of first three is Pepsi}) + \text{P}(\text{fourth is Pepsi} \cap \text{one of first three is Pepsi}) \\
&= \text{P}(\text{fourth is Pepsi}|\text{none of first three is Pepsi})\text{P}(\text{none of first three is Pepsi}) \\
&\quad + \text{P}(\text{fourth is Pepsi}|\text{one of first three is Pepsi})\text{P}(\text{one of first three is Pepsi}).
\end{aligned}$$

We can now substitute in for these quantities:

$$\begin{aligned}
\text{P}(\text{fourth is Pepsi}|\text{none of first three is Pepsi}) &= \frac{2}{7}, \\
\text{P}(\text{fourth is Pepsi}|\text{one of first three is Pepsi}) &= \frac{1}{7}, \\
\text{P}(\text{none of first three is Pepsi}) &= \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} = \frac{7}{15}.
\end{aligned}$$

The last necessary quantitiy is slightly more complicated. For just one of the first three sodas sampled to be Pepsi, we may have CCP, CPC, or PCC (where C is Coca-Cola and P is Pepsi). These probabilities are

$$\begin{aligned} \mathrm{P(CCP)} &= \frac{8}{10} \cdot \frac{7}{9} \cdot \frac{2}{8} = \frac{7}{45}, \\ \mathrm{P(CPC)} &= \frac{8}{10} \cdot \frac{2}{9} \cdot \frac{7}{8} = \frac{7}{45}, \\ \mathrm{P(PCC)} &= \frac{2}{10} \cdot \frac{8}{9} \cdot \frac{7}{8} = \frac{7}{45}. \end{aligned}$$

So the probability that one of the first three sodas sampled is Pepsi is $3(\frac{7}{45}) = \frac{7}{15}$ — that this is exactly the probability that none of the first three sodas sampled is Pepsi is mere coincidence and not a general principle (consider what happens if we change "fourth" in the statement of the problem to "fifth").

We are now set to finally compute,

$$\mathrm{P(fourth\ is\ Pepsi)} = \frac{2}{7} \cdot \frac{7}{15} + \frac{1}{7} \cdot \frac{7}{15} = \frac{3}{15} = \frac{1}{5}.$$

Second solution

We begin by computing the total number ways which the sodas may be ordered. Since we don't care about particular cans, only whether a given can is Pepsi or Coca-Cola, this is equivalent to finding the two Pepsi cans from the lineup of ten; rather, it is the number of ways we can *choose* the locations of the two Pepsi cans from the lineup of ten. This number is

$$\binom{10}{2} = \frac{10!}{2!(10-2)!} = 45.$$

How many ways can a Pepsi be the fourth soda? It would be painfully long to enumerate all 45 cases (although in other problems this may be a worthwhile approach), so let's decompose the solution. If we know that the fourth can is a Pepsi, there are nine locations for the "other" can of Pepsi. The number of places this can can occur in the sequence is then equivalent to the number of ways we can pick one location from nine available, or

$$\binom{9}{1} = \frac{9!}{1!(9-1)!} = 9.$$

The probability that the fourth can is Pepsi can be computed from the definition of probability: the number of ways that a Pepsi can be the fourth can divided by the total number of arrangements of the ten cans (by brand). This is

$$\mathrm{P(fourth\ is\ Pepsi)} = \frac{9}{45} = \frac{1}{5}.$$

This answer matches the "Bayesian" approach, validation of both answers.

What would have happened if we had looked at permutations of 1 and 2 cans (from 9 and 10 locations, respectively), instead of ways to choose the cans[3]? By what factor is the naïve application of this approach off? Does this make sense, and what should be considered to fix it? What if we'd simply looked at permutations of 9 and 10 cans? Combinatorial logic is meant to be — and is — robust to these considerations.

[3] This result should hold even if we care about the locations of particular cans, not just their brands; but remember that there are 2 Pepsis which may occupy the fourth spot.