Gerald's Column
by Gerald Fitton

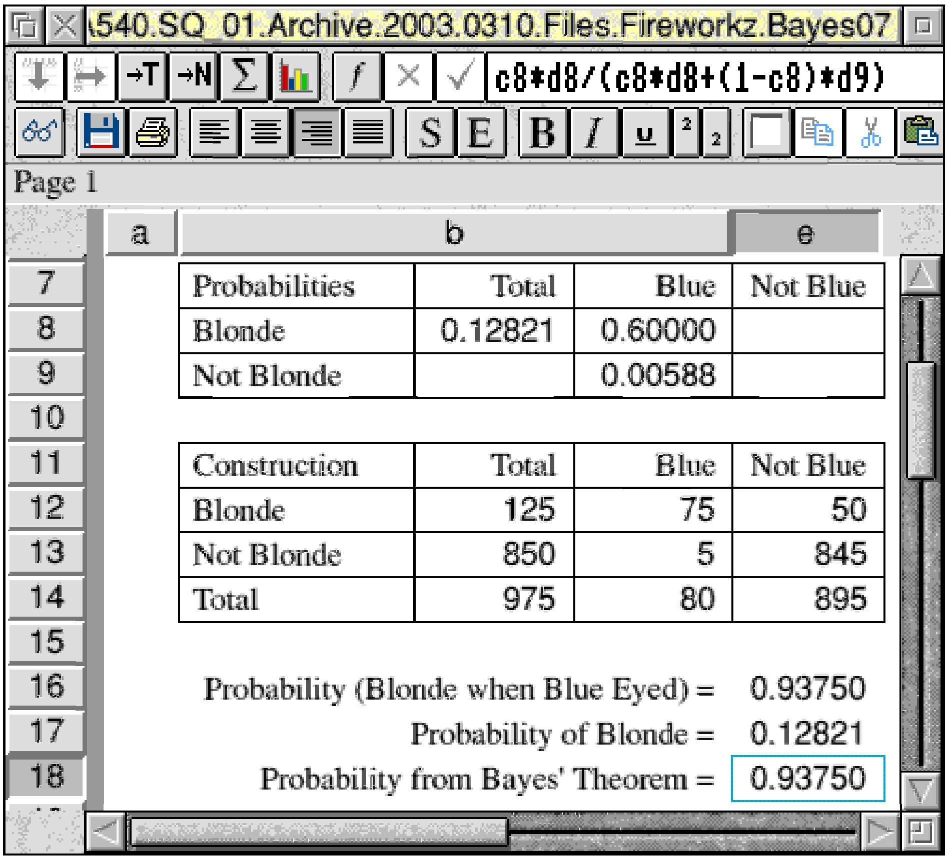
Last month I introduced you to Bayes' Theorem.
The formula which I gave you applies if there are only two Hypotheses, H_1 and H_2.

It is: $P(H_1|E_1) = P(H_1) * P(E_1|H_1) / (P(H_1) * P(E_1|H_1) + P(H_2) * P(E_1|H_2))$.

This month I shall describe how this formula can be extended to three Hypotheses.

Two Hypotheses

In my second example last month the two Hypotheses, H_1 and H_2, are that the person is Blonde (H_1) or Not Blonde (H_2). The Event (E_1) is that they have Blue Eyes.



Probabilities	Total	Blue	Not Blue
Blonde	0.12821	0.60000	
Not Blonde		0.00588	

Construction	Total	Blue	Not Blue
Blonde	125	75	50
Not Blonde	850	5	845
Total	975	80	895

Probability (Blonde when Blue Eyed) = 0.93750
Probability of Blonde = 0.12821
Probability from Bayes' Theorem = 0.93750

In words, $P(H_1)$ is spoken as: "The probability that the person is Blonde"; $P(E_1|H_1)$ is spoken as: "The probability that the person has Blue Eyes when you know that they are Blonde". Of course, what we want to know is $P(H_1|E_1)$, "The probability that the person is Blonde when we know they have Blue Eyes".

In the formula, the four probabilities on the right hand side of the equation are:

$P(H_1)$ = 0.12821 from cell c8 = 125/975
$P(H_2)$ = 1 − 0.12821 because $P(H_1) + P(H_2) = 1$
$P(E_1|H_1)$ = 0.60000 from cell d8 = 75/125
$P(E_1|H_2)$ = 0.00588 from cell d9 = 5/850

The formula is evaluated in cell e18 using Bayes' Theorem. You will see that it gives exactly the same answer as the 'long hand' method evaluated in e16—and so it should!

Let me remind you why these calculations are useful. If we want to select a Blonde and we choose a person at random then the probability of success is just under 13%. However, if we select only from those with Blue Eyes the probability of success rises to nearly 94%. As we gather information (Events happen to us) we improve the likelihood that our preferred Hypothesis will successfully predict the outcome of our chosen action(s).

Three Hypotheses

In my extension of this technique I will give you an example of Bayes' Theorem as it applies to situations where there are three Hypotheses. Furthermore I shall change the nature of the problem in a subtle way. Read on and see if you can spot the difference!

The formula for three Hypotheses is:

$$P(H_1|E_1) = P(H_1) * P(E_1|H_1) / (P(H_1) * P(E_1|H_1) + P(H_2) * P(E_1|H_2) + P(H_3) * P(E_1|H_3)).$$

The Missing Bus

I'd like to remind you of the story of the 'missing bus' which I first introduced in a series starting in the March 2002 Archive with the flippin' coin. In this article I shall show you how I used Bayes' Theorem as part of the solution to the problem of the 'missing bus'.

If the brief outline below is insufficient then I suggest that you refer back to the original articles (March to July 2002), particularly the July 2002 article, for a refresher.

I am waiting at a bus stop. I know that the buses are due every ten minutes. The time table says so. Even if I have just missed a bus (I don't know if I've just missed a bus or not) the maximum wait should be ten minutes.

I shall call this Hypothesis 2, H_2. I don't know if this Hypothesis is true—but it might be.

I wait ten minutes and there is no bus.

After ten minutes have passed it is quite apparent that the assumption that the buses run on time must be abandoned. We must guess again at the Probability Distribution Function.

It might be more realistic to give buses a bit of leeway. They never arrive spot on time but they might arrive within a few minutes of the scheduled time (most of the time). For simplicity let's consider the worst case scenario. I assume that the buses might start out on time but due to all sorts of dispersion effects (such as variable traffic), by the time they get to my bus stop, their arrival time is totally random.

Totally random in this hypothetical experiment still means that, on average, a bus turns up every ten minutes. In turn this means that for any and every minute the probability of a bus not turning up remains constant at 90%.

I shall call this Hypothesis 3, H_3. I don't know if this Hypothesis is true—but it might be.

Hypothesis 1

Those of you who are following this numbering of the Hypotheses will recall that the most salient feature of the Bayes' Theorem formula is that we must have a Hypotheses 1, H_1.

What is it? Read on.

Certainty

A probability of one corresponds to certainty and zero to impossible. Bayes' Theorem will work only if the sum of the three probabilities, $P(H_1) + P(H_2) + P(H_3) = 1$. This is a most important point to check before embarking upon an analysis. Secondly, each of the three hypotheses must be Mutually Exclusive. Only one of the three possible hypotheses can be true. Of course you can have a different 'Degree of Belief' about each of the three hypotheses; this is not the same as believing more than one can be true at the same time.

If you have two hypotheses which you believe could be true (let me remind you that you do not know which is true and that for some situations the truth is unknowable) then you can use the Two Hypothesis version of Bayes' Theorem. You can do so if and only if $P(H_1) + P(H_2) = 1$; the two hypotheses completely exhaust every possibility. The case of the missing bus does not fall into this category. It is a Three Hypotheses problem.

I didn't think of that

Up to now we have two possible Hypotheses, H_2 (bus every ten minutes) and H_3 (buses arrive at random times). This does not cover every contingency. Both H_2 and H_3 could be untrue! We must introduce another Hypothesis, H_1, in order to cover every eventually.

In my earlier articles I referred to this possibility, H_1, as "I didn't think of that".

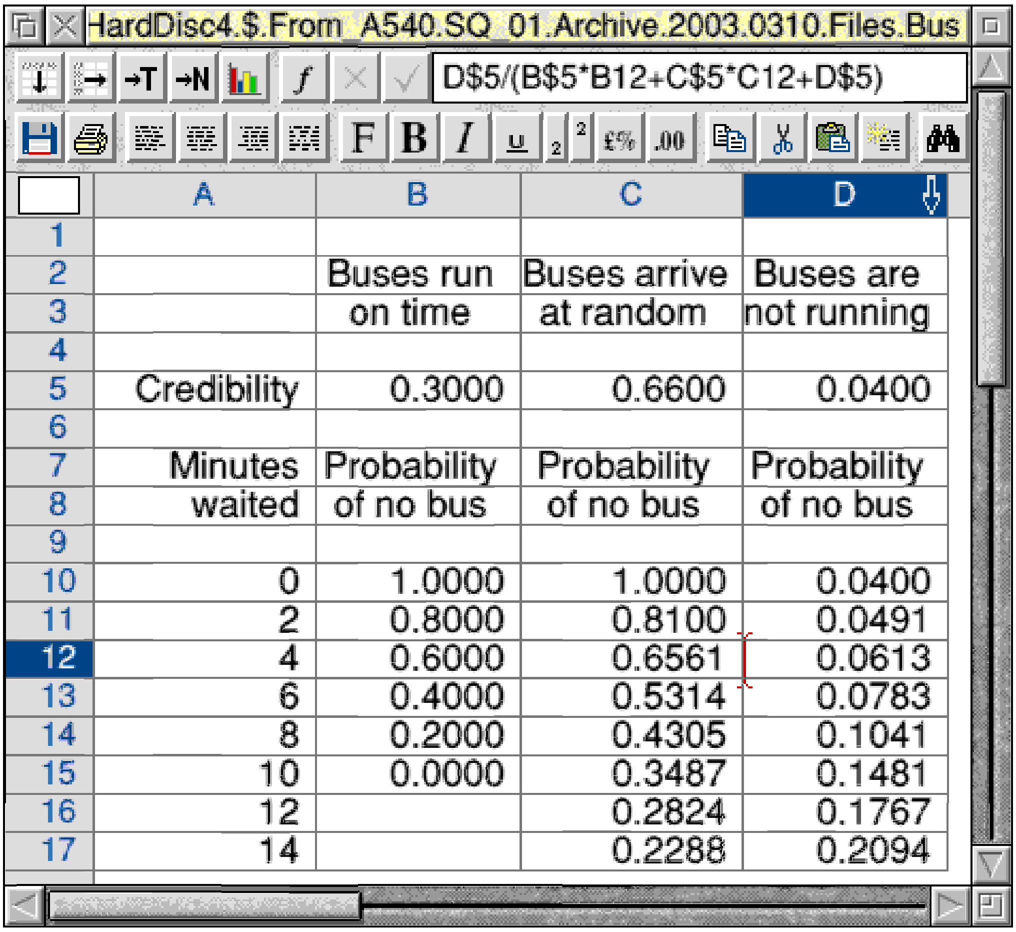
In mathematical terms we have to ensure that $P(H_1) + P(H_2) + P(H_3) = 1$ and it is the introduction of H_1which completes the 'set'. We calculate $P(H_1) = 1 - P(H_2) - P(H_3)$.

A Priori Probabilities

Now the time has come to look at the screenshot.

The three Hypotheses, H_2, H_3 and H_1 are used to calculate probabilities in the three columns B, C and D respectively. The Event which is relevant to row 12 is that no buses have appeared during a wait of four minutes.

In cell D12 you will see the formula which I have used to calculate the probability that H_1 ("Buses are not running") is correct. This formula is replicated down the column as the Event changes from 'No buses in four minutes' to 'No buses in 6 minutes' and you will in the screenshot values corresponding to 'No buses in 14 minutes'.



	A	B	C	D
1				
2		Buses run	Buses arrive	Buses are
3		on time	at random	not running
4				
5	Credibility	0.3000	0.6600	0.0400
6				
7	Minutes	Probability	Probability	Probability
8	waited	of no bus	of no bus	of no bus
9				
10	0	1.0000	1.0000	0.0400
11	2	0.8000	0.8100	0.0491
12	4	0.6000	0.6561	0.0613
13	6	0.4000	0.5314	0.0783
14	8	0.2000	0.4305	0.1041
15	10	0.0000	0.3487	0.1481
16	12		0.2824	0.1767
17	14		0.2288	0.2094

One subtle difference between this spreadsheet and the previous one (about Blondes with Blue Eyes) is that we do not have any measured values for $P(H_2)$ and $P(H_3)$; we have had to guess them. If you have looked at the earlier series of articles (or if you have a good memory) you will know that the probabilities in row 5 of the spreadsheet are those which Philosophers call the 'a priori' probabilities of the three Hypotheses, H_2, H_3 and H_1.

Let's have another look at them. My a priori 'Degree of Belief' that the buses will run near enough on time is $P(H_2) = 0.30$. I think there is a 30% chance that the buses will run every ten minutes. I think it more likely that they will arrive at random times but with a mean spacing of ten minutes. My a priori probability for this hypothesis is $P(H_3) = 0.66$.

I might be wrong. I must leave room for doubt however slight, that neither of these Hypotheses are true. I have introduced $P(H_1)$ with a value of 4% to take care of this possibility. The calculation is $P(H_1) = 1 - P(H_2) - P(H_3) = 1 - 0.30 - 0.66 = 0.04$.

Let me stress that these 'a priori' probabilities are chosen by guesswork or prejudice.

Bayes' Theorem

The formula in cell D12 is the Bayes' Theorem formula for three hypotheses. We are trying to find $P(H_1|E_1)$ and we know (or we can calculate) the values on the right hand side of the Bayes' Theorem formula for three hypotheses. Please refer back to that formula whilst reading the following explanation of the formula in cell D12.

If H_1 is true then $P(E_1|H_1)$ is exactly 1 (certainty).

If no buses are running (H_1 is true) then the numerator of the right hand side of the Bayes' Theorem equation is exactly $P(H_1)$ (our 'a priori' probability); this is D5.

That deals with the numerator.

Now to the denominator.

The first term in the denominator (of D12) is: $P(H_2) * P(E_1|H_2)$ = B$5 * B12.

$P(H_2)$ is the value in B5 and $P(E_1|H_2)$ is the value in B12 = 0.6000. The Probability Distribution Function (PDF) we have assumed is the Uniform Distribution Function and so the conditional probability, $P(E_1|H_2)$, of there being no bus whilst we are waiting, falls linearly until, after ten minutes, it falls to zero.

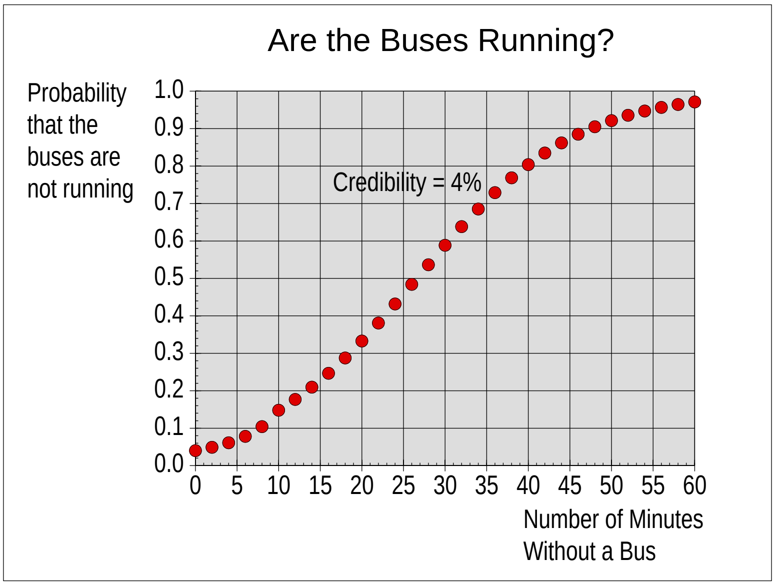
The second term in the denominator, $P(H_3) * P(E_1|H_3)$ = C$5 * C12.

If H_3 is true then $P(E_1|H_3)$ is 0.6561. This is calculated from the assumed Probability Distribution Function (PDF) that buses arrive randomly. The PDF is the Exponential Distribution Function with the factor 0.9 applied every minute. If H_3 is true then the conditional probability $P(E_1|H_3)$ of there being no bus in the first t minutes is (0.9 ^ t).

The Effect of Knowledge

As time passes the Events we experience increase our knowledge. This knowledge is still not 'absolute' (100% certain) but our 'Degree of Belief' (the extent of our uncertainty) changes and usually increases or falls substantially. Usually our 'a posteriori' probabilities, such $P(H_1|E_1)$, as are nearer to zero or one (impossibility or certainty) than, for example, $P(H_1)$, our 'a priori' probabilities. We become more certain as our experience grows.

In the case of the missing bus the calculation in D12 returns 0.0613. Let me spell out exactly what has happened to our 'Degree of Belief' in H_1 ("I didn't think of that!").



After 4 minutes without a bus, $P(H_1)$, the ‘a priori’ probability of 4% has increased to $P(H_1|E_1)$, the ‘a posteriori’ probability of 6%. It is the Event, E_1, (no bus after 4 minutes) which has driven up this percentage from $P(H_1) = 4\%$ to $P(H_1|E_1) = 6\%$.

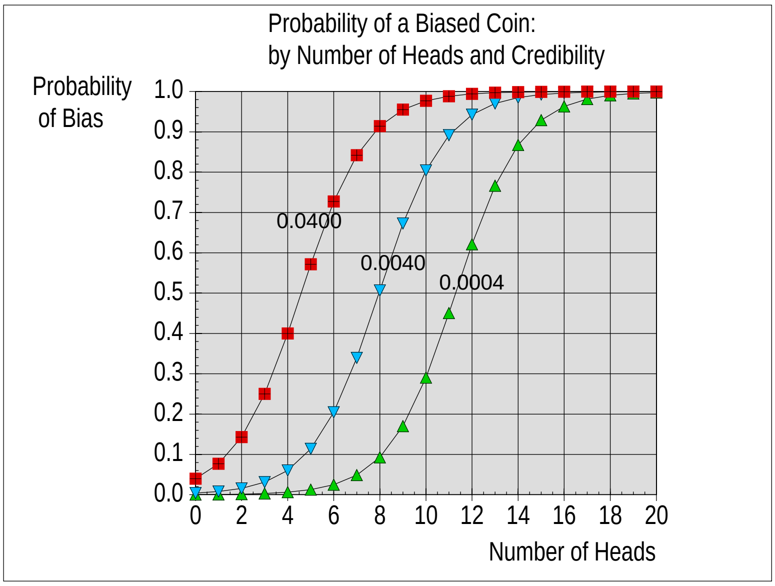
After about 27 minutes without a bus this probability, $P(H_1|E_1)$, has risen to about 50%. Probabilities have to add up to 100%. $P(H_1|E_1) + P(H_2|E_1) + P(H_3|E_1) = 1$. $P(H_1|E_1)$ is rising so the probability that H_2 or H_3 are ‘true’, $P(H_2|E_1) + P(H_3|E_1)$, falls, as I described in my July 2002 article. Both H_2 and H_3 become less believable as time passes.

When do people change their minds?

It is a mistake to think that people have very great confidence in buses or flippin’ coins.

I have characterised as H_2 and H_3 my hypotheses concerning the buses. I have called the ‘Degree of Belief’ that these might be false the “Credibility” and this I have quantified this as 4%. This 4% is the ‘a priori’ probability that both my two major theories are false.

After about 27 minutes the original ‘a priori’ probability of 4% has risen to an ‘a posteriori’ value of over 50%. If you run the spreadsheet with other values of ‘Credibility’ you will be able to discover how ‘Gullible’ you are (you wait longer than anyone else)!



From the correspondence I received about the flippin’ coin there are only a few of you who thought that five heads in a row was unusual enough to worry about. Five heads in a row corresponds to a Credibility of 4% similar to the Credibility which corresponds to a 27 minute wait for the missing bus. According to my straw poll for the bus, nobody would have waited more than half an hour before reverting to ‘Plan B’. The analysis of ‘a prori’ probability which I have condensed into these bus and coin graphs indicates to me that you have more ‘faith’ in coins being unbiased than buses being cancelled.

Let me rephrase this. Whilst your initial doubts about the coin (or me) are less than 4%, your initial doubts about the bus are higher than 4%.

In doing experiments such as the flippin' coin and the missing bus (which are designed to provide some estimate of 'a priori' probabilities) I have been surprised to discover how sceptical people are about many everyday things (such as bus timetables).

A high degree of scepticism is quantified by a high value of 'Credibility'. A value of 4% for the missing bus is a fairly typical value for many everyday things.

Conclusion

The calculations which allow these experiments to be analysed this way depend upon the rather mystical Bayes' Theorem. It is Bayes' Theorem which allows us to 'invert' not only measured probabilities but also probabilities calculated using guessed PDFs. This process of inversion is a tool which allows us to learn by 'converting' (guessed at) 'a priori' probabilities into more usable 'a posteriori' probabilities.

As we learn, these 'a posteriori' probabilities (usually) tend towards either zero or one (impossibility or certainty). Values of probability close to zero or one are much more usable than 'mediocre' probabilities.

Sometimes these 'a posteriori' do not tend towards zero or one. For example the probability that a truly unbiased coin will come up heads will tend towards 0.5 whatever assumptions ('a priori' probability) we assume. Deciding what action to take when the 'a posteriori' probability has converged on a 'mediocre' value will have to be the subject of another article—supported by spreadsheets of course!

I have used the word "learning" to describe the process whereby our insupportable 'a priori' probabilities are firmed up as much more usable 'a posteriori' probabilities. In philosophical jargon we change our 'Degree of Belief' about Hypotheses which may or may not be true as a result of our experiences.

I must emphasise that we can never know whether a Hypothesis is true or false; at best we will discard one Hypothesis with mediocre probabilities as 'useless' in favour of another which is more usable and more in accord with our measurements (or observations) than is our discarded Hypothesis.

The key to forcing probabilities to change to useful 'a posteriori' values is to make Events happen. On an individual basis this is the same as deliberately setting out to gain experience. Experience is usually a good thing because it enables us to develop a model in our minds of how the world works. In my personal view we should try to garner as much experience as we can—even if some of those experiences involve taking risks. Now and again we should do something which is a little 'dangerous' because the outcome of the 'dangerous' (but hopefully non fatal) experience is that our knowledge will increase. It is this knowledge which will enhance how useful we are (or could be) to others.

Communication

Please contact me by email (preferred) or by letter if you have any questions or comments.

Because of spam problems I shall shortly set up a Spam Filter which will reject (bounce) emails directed to <gerald@abacusline.demon.co.uk> unless you are on my personal 'white list'. Please email me at <Archive@abacusline.demon.co.uk> if your email relates to articles in the Archive magazine or at <GoldLine@abacusline.demon.co.uk> if your correspondence relates to PipeDream, Fireworkz or GoldLine or, if your correspondence relates to Living with Technology, then please use <LwT@abacusline.demon.co.uk>.