# Conditional Probability - Part 3
## by Gerald Fitton

Once is happenstance
Twice is coincidence
Three times is enemy action
Four times ... and you don't need to be a Statistician

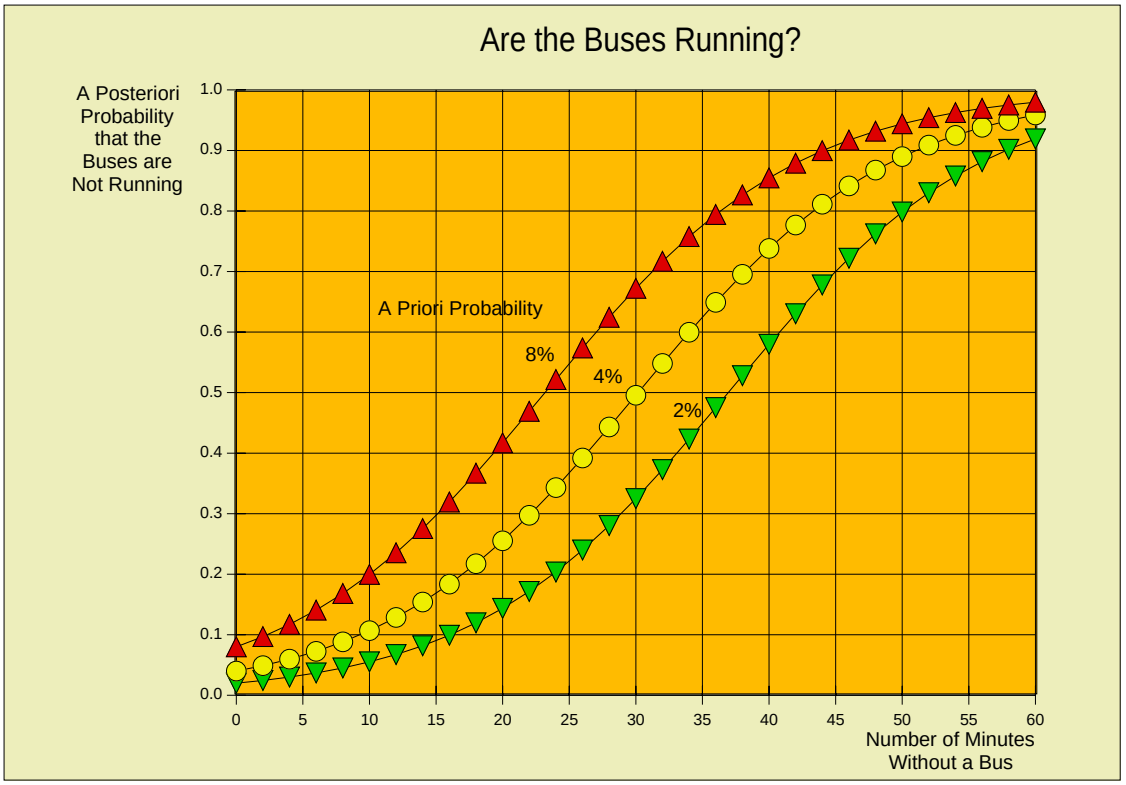
My acknowledgements to the James Bond movie, "Goldfinger" for the first three lines.

I have seen the words "enemy action" replaced with "a conspiracy". The fourth line was added by one of my students in an article which she wrote for me as part of a college project about Bayesian Epistemology. In her essay she makes the point that, when making decisions (about an action we might take) we should not rely blindly on logic nor on reason (nor even on the Scientific Method) when these 'tools' conflict with the evidence of our experience. We must learn from our experience and put aside blind logic and reason.

Even if we are 'totally convinced' that the buses must be running, "It said so on the local radio and in the paper", we would be foolish to continue to have faith in those sources after waiting a long time for the non existent bus to appear. It is much wiser to believe in the cumulative evidence of our own personal experiences and bravely reject our original hypothesis, however reasonable and logical it seemed originally.

### Where were we?

This is the graph we are trying to produce. The three lines on the graph are the result of applying Bayes' Theorem (not Hypothesis Testing) to the "Will the Bus Come?" problem.



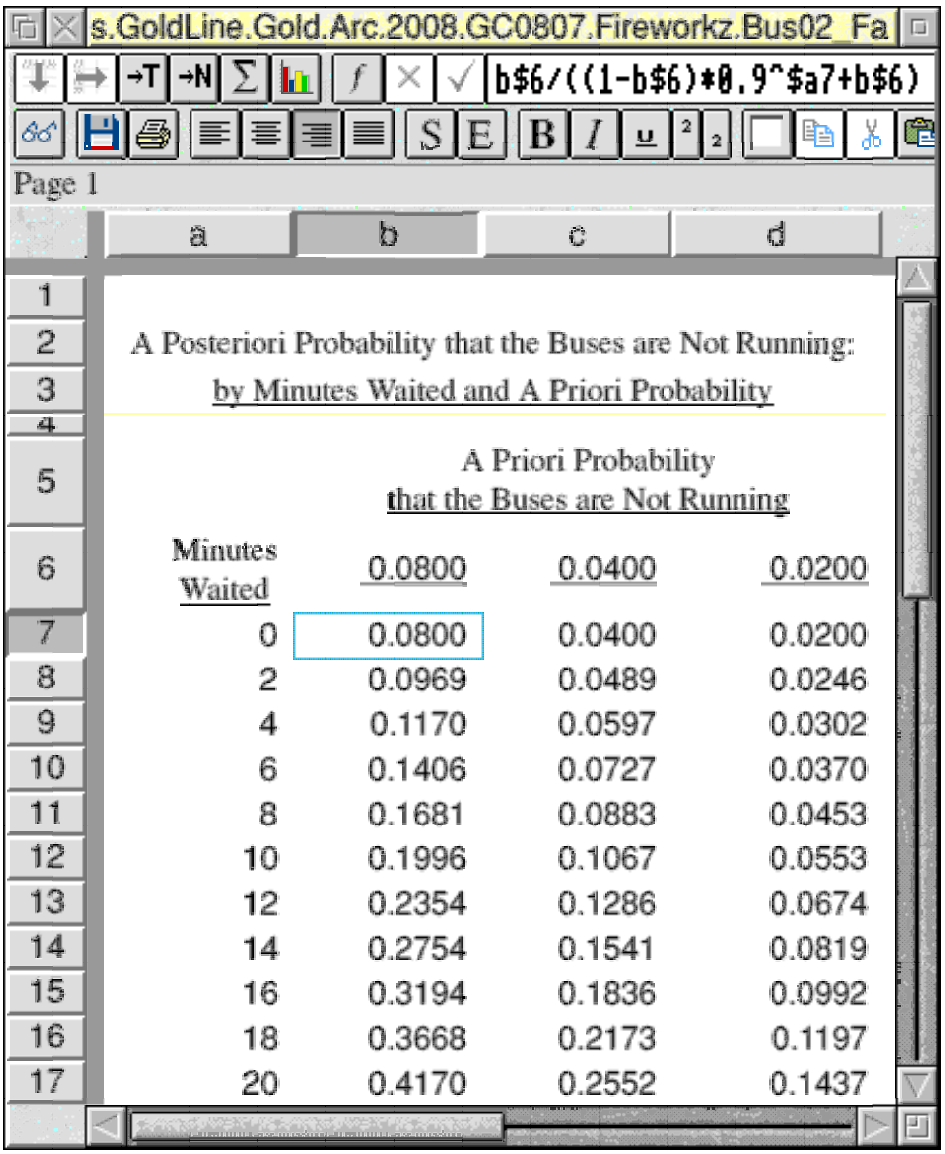
## More than one Probability

The thing which worries people most about this graph is that after waiting, say, 30 minutes without a bus, there appears to be three different probabilities ('A Posteriori Probabilities') that the buses are not running. These probabilities seem to depend not only on the number of minutes waited (this seems reasonable) but also on the seemingly arbitrary value of the 'A Priori Probability'. How can a probability depend on an arbitrary value? We shall see!

I have chosen three values, 2%, 4% and 8%, for this 'A Priori Probability' but, in reality, it can have any value between 0% and 100% - but it can't have exactly 0% nor 100%!

## The Spreadsheet

The screenshot below shows of part of the spreadsheet. In the screenshot the minutes waited range from 0 to 20, but the sheet continues (out of shot) to the 60 minutes plotted on the chart. The 'A Priori Probabilities', 2%, 4% and 8% are written as 0.0200, 0.0400 and 0.0800 respectively in row 6 across the columns b, c and d.



The formula in b7 is b$6/((1-b$6)*0.9^$a7+b$6). This formula is replicated both down and to the right through the whole sheet by marking the block from b7d37 and then clicking on the down and right pointing arrows (top left). The $ symbols in the formula 'lock' the column or row immediately following the $ symbol. For example in d17, b$6 has changed to d$6 and the $a7 has changed to $a17.

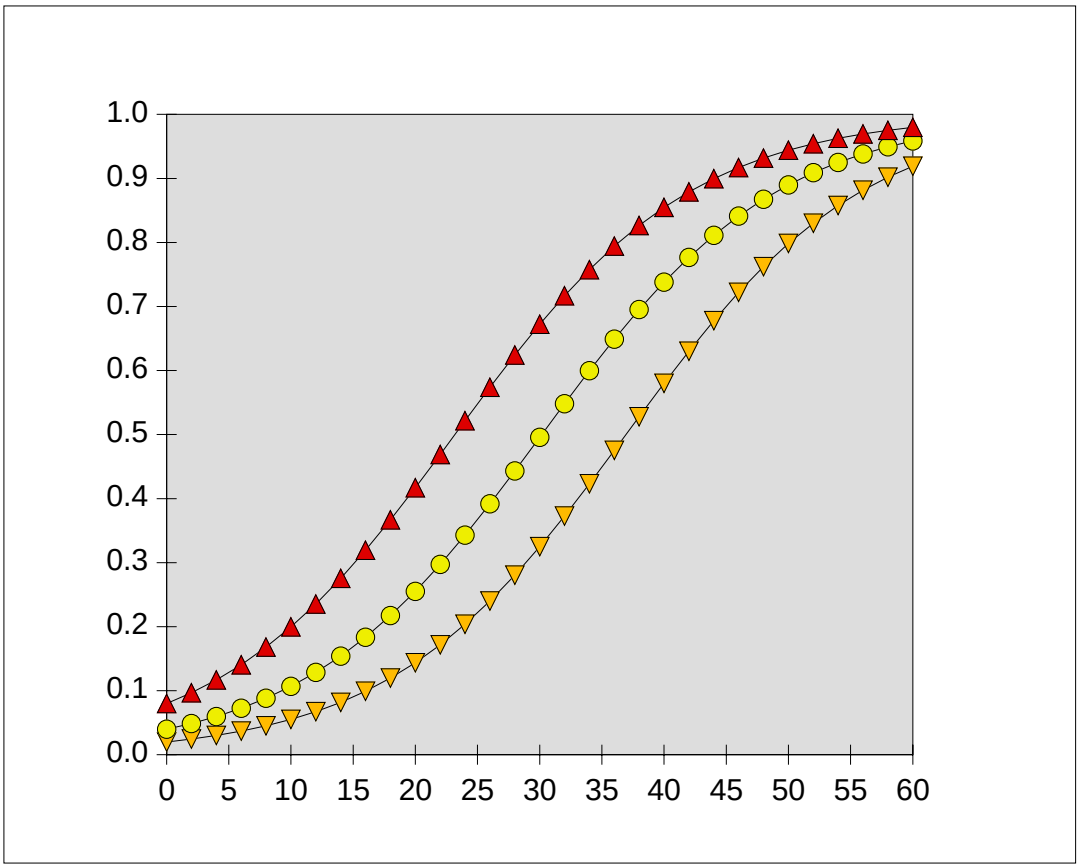
The formula used in b7 and throughout the sheet to find P is derived from Bayes' Theorem:

$$P = \frac{1}{(1 + r)}$$

In this formula P is the 'A Posteriori Probability' that the buses are not running and r is the ratio of two probabilities. This ratio varies as the evidence comes in; r starts at a high value (P is small) and, as the minutes inexorably roll by, r falls and P rises. This ratio, r, falls by a factor of 0.9 every minute. The 0.9 factor can be expressed as "There is a 90% probability that during the last minute, a bus would have turned up - but it didn't!"

## Drawing the graph

Last month I described in detail how to produce this graph. Start by marking the block b7d37, then click on the Create Chart icon, select the type of chart you want and then, when you click on OK, Fireworkz creates the chart shown in the drawfile below.



This is a much simpler process than using PipeDream. However, there are two drawbacks. The chart is drawn at the bottom of the block of data in the same file as the data. Worse in my opinion is that there is no facility for adding text to this chart from within Fireworkz.

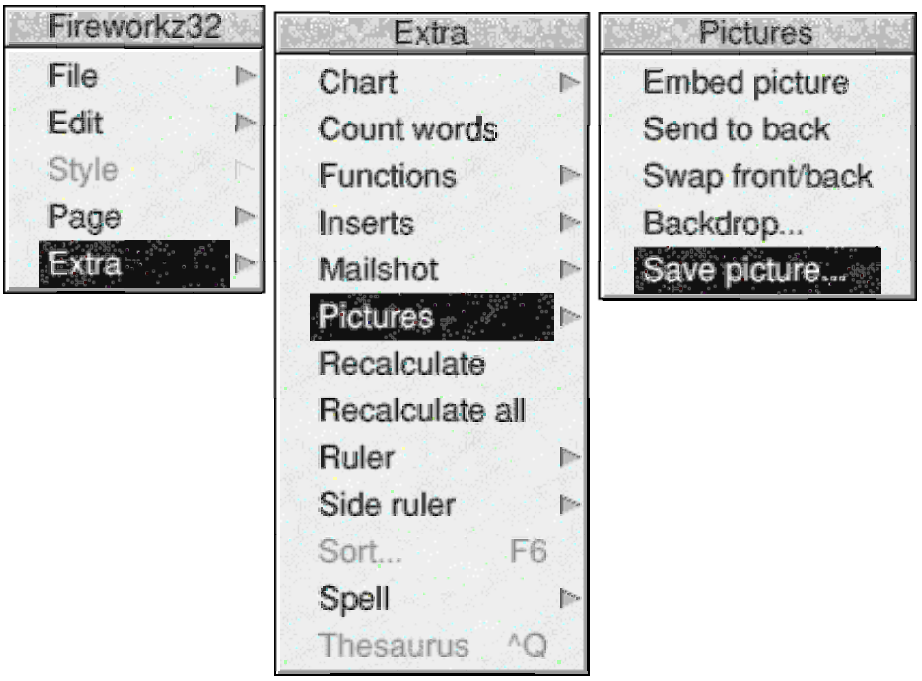
## Adding text

The only way I know of for adding text to a Fireworkz chart is it save it as a DrawFile and then add the text using an application such as !Draw, !Vector or, my favourite. !DrawPlus.

Having created the chart in Fireworkz, click on it once to select it. The chart will be surrounded by a red rectangular box. Do not double click on the chart. When you double click on a chart in Fireworkz you enter the Chart Editing Menu. You are in editing mode if there is a double rectangular box surrounding the Fireworkz chart.

With the single rectangular box surrounding the chart, click the <menu> button and the Fireworkz32 pop up menu shown below.

Run the pointer through Extra - Pictures - Save picture.




Give the file a suitable name and then drag the icon to an open directory.

From there you have to use the facilities of your chosen application, !Draw, !Vector, !DrawPlus, etc. to add and position the text exactly where and how you want it.

I have added a title for the chart, labels for the axes and labels for the three lines.

The result of my efforts is shown in the picture at the beginning of this article. The drawfile which is created in !Draw etc can be loaded back into any Fireworkz document but, unlike PipeDream, the chart is not ‘live’.

## PipeDream or Fireworkz

If I want to draw a very simple chart quickly then I might choose Fireworkz because I do not have to use the "Add to chart" sequence. However, if I want a live chart with text then I use PipeDream.

Sometimes what you want to do is write a mainly textual report which includes in it both an illustrative table (the numerical data), and an illustrative chart (a graph of the data). The main part of the document is the text; the table and the chart are there only to illustrate the points which you are making in the text of your report.

For an exercise like this I would always choose Fireworkz.

Sometimes what I want to do is to create a graph from a table of data and then observe visually the changes in the graph which results from changing parameters. For that purpose I find PipeDream much better.

Let me give use as my example the "Are the Buses Running?" chart.

I wanted to see visually the effect of changing the 'A Priori Probability' so, in PipeDream, I created a rather basic chart and then I varied the parameter (the 'A Priori Probability') to see what effect it had on the shape of the curve. Eventually I found three values which were suitable for me to illustrate the way in which the 'A Posteriori Probability' varied with time over a period of an hour for the three different 'A Priori Probabilities'.

Having decided that 2%, 4% and then 8% would demonstrate the way in which my confidence in the possible arrival of a bus gradually ebbed as time went by, I saved the DrawFile and included it in my mainly textual report where it served to illustrate this example of what is called Bayesian Epistemology. I might use Fireworkz, Impression or Ovation Pro for my printed article, but I would have done all the 'research' using PipeDream rather than Fireworkz.

## Finding "r"

Let me say right away that $r = r_0 * r_1 * r_2 * r_3 * r_4 * r_5 * ...$

The "..." means that this series can be extended indefinitely.

Each of the values of r is a ratio of probabilities. In the case of a Bayesian Spam Filter each of the values of $r_n$ are different; there is one value of $r_n$ for each word in the email. In the case of the bus which didn't arrive all these values of $r_n$ are 0.9. If the bus doesn't arrive in the first minute then the value of $r_1$ is 0.9. If it doesn't arrive in the second minute then the value of $r_2$ is 0.9 ... and so on.

I have said that $r_n$ is the ratio of two probabilities; both of these probabilities are the probability that a bus will NOT arrive during the nth minute (the minute under consideration). How can we have two different probabilities for the same event? That is a good question which I shall try to answer in the next few paragraphs.

In my simple ‘Are the buses running?” example we have two hypotheses. The first hypothesis is that the buses are not running; the second hypothesis is that the buses are running and arrive randomly at an average time interval of ten minutes.

Although in this example of the ‘missing bus’ I have only two hypotheses, Bayes’ Theorem can be extended to cover many hypotheses - but that’s another story for another day.

How can we have two different probabilities for one event? The answer is that we can have two probabilities for the same event when we have two different theories (hypotheses) to consider. I have said that there is no such thing as an absolute probability. Your estimate of a probability depends not only on how much knowledge you have but (believe me when I say this next bit!) it depends upon what you believe originally! By this enigmatic statement I mean that your assessment of a probability depends not only upon the data you collect but also which hypothesis you favour before you start collecting data!

For the ‘missing bus’ problem we have two hypotheses. Our first is that he buses are not running and our second is that they are running at approximately ten minute intervals.

First we calculate the probability that NO bus will arrive when the buses are NOT running. I’m sure you will agree that if the buses are not running at all then the probability of a bus NOT arriving is 100% (or 1.00 if you prefer).

Then we calculate the probability that NO bus will arrive when the buses are running at approximately ten minute intervals. The probability of a bus arriving in the nth minute is 1 in 10 (10% or 0.1); the probability of NO bus arriving is 90% (or 0.9 if you prefer).

The ratio, $r_n$ , is found by dividing the 0.9 of the second hypothesis by the 1.0 of the first hypothesis giving us 0.9. For every minute the value of $r_n$ is the same; it is 0.9. The exception is the zeroth minute or, if you follow my notation, $r_0$ .

When we use the method called Hypothesis Testing we have only one hypothesis. After collecting enough data we may be able to reject our chosen hypothesis as being improbable. When we use the Bayesian method we can have as many hypotheses as we can think of and, using a suitable spreadsheet, we can test them all at once and then, when we need to make a decision, we can select the one which seems most likely at the time!

Back to $r = r_0*r_1*r_2*r_3*r_4*r_5*...$

We have every value of $r_n$ (except $r_0$ ) for the ‘missing bus’ problem; they are all 0.9. This first value, $r_0$ , is something which seems to baffle even the expert. It is this first value of r which makes the Bayesian method what it is and distinguishes it from Hypothesis Testing.

## What is the value of $r_0$ ?

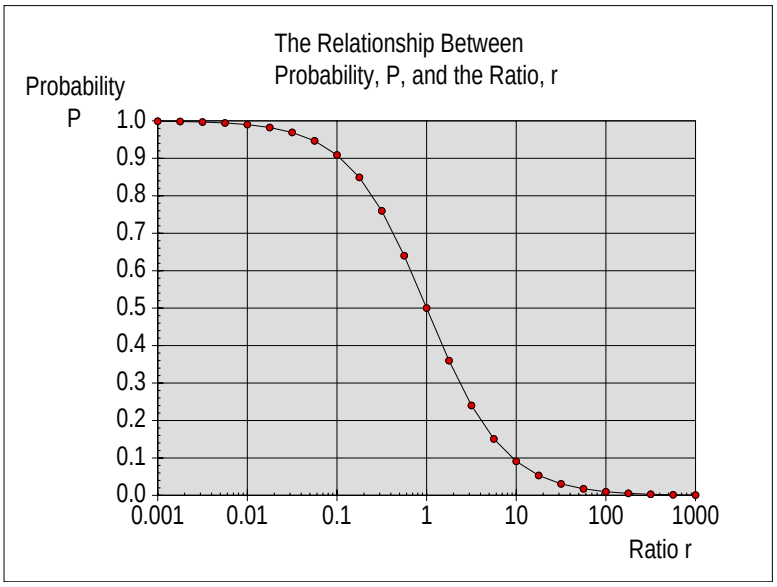
So, for our missing bus problem we have $r = r_0*(0.9)^n$ where n is the number of minutes we have waited without success. This is the formula I have used in the spreadsheet.

The key to understanding why there are three lines on this Bayesian Graph and only one on a Hypothesis Testing Graph is to understand where the $r_0$ comes from. I invite you to tell me the answer to this question.

## The relationship between P and r

Next month I shall discuss this. Also I shall use this relationship to describe how a Bayesian Spam filter works. We need to assign a value of r, actually we use log(r), to every word in the email. The Bayesian Formula can be used to calculate the probability that the email is Spam based on the accumulated evidence of the log(r) values for every word.

Here is a graph (created using PipeDream) of the relationship between P and r.




You will see that when r is smaller than 0.01 the probability is close to certainty. When r is larger than 100 the probability of the hypothesis being true is negligible. For values of r between 0.1 and 10 we might regard the chances of the hypothesis being true as 'Uncertain'. When the probable truth of the hypothesis is uncertain then a Scientist, given time and financial resources, would try again. A Manager (and we are all Managers of our own lives) has to take a chance!

Even the decision to do nothing is a decision - making decisions is what Managers do!

In life, now and again, we need to take a chance rather than do nothing!