1016 Hanshaw Road

Ithaca, NY 14850   December 5, 2012

FUN WITH REVEREND BAYES – AN OVERVIEW

In the last three notes we looked at some interesting puzzles in probability theory, and Bayes theorem was mentioned briefly in a few cases. Here we will look at an overview of the Bayes viewpoint so that people can have a better understanding of the simple math associated with this fashionable appellation. We shall also tangentially try to address the more complex question of interpretations.

THREE BRIEF ISSUES

Let's dispense with three preliminary and perhaps (for the first two anyway) minor issues. Here we will consider the terms Bayes Theorem or Bayes Formula or Bayes Rule or Bayes equation (even Bayes Law) quite interchangeably. Further, we will reject the need to use the possessives Bayes' or Bayes's (often seen) as both awkward and unnecessary. Since mathematically it is more a simple algebraic formula or a rule than a theorem (indeed, the derivation is substantially its own proof) the term "Bayes" is more of a <u>label</u>. We don't find ourselves using the terminology "Fourier's Transform" for what is "just" an equation.

Secondly, Bayes was indeed a preacher born in about 1701 and who lived about 60 years, with his most famous paper published after his death. True also, his contemporary Richard Price, and later Laplace, likely deserve much credit for the current formulations and the popularity we now attribute mainly to Bayes. On the other hand, while very far from being flamboyant, Bayes was hardly a poor unknown "working out of his garage". Several books [1-3] give somewhat parallel descriptions of Bayes' life. Well worth knowing about.

The third point is that we hear of the conflict between a "frequentist" approach and a Bayesian approach to probability. The void is not, in many practical cases, as great as might be thought, and has little or nothing to do with whether or not the Bayes formula is used. Indeed, many statisticians will just write down a correct answer, just by looking at the occurrences of the various contents of the sample space with care, regardless of
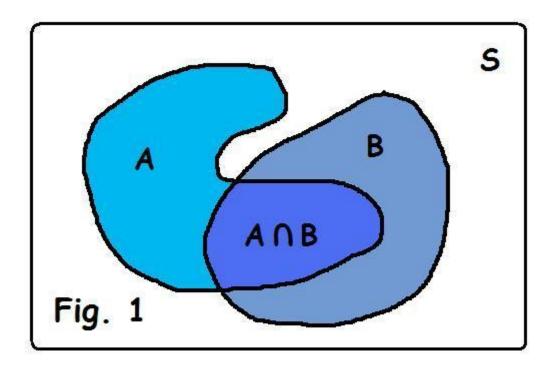
whether or not Bayes is even mentioned. We shall try to make the distinction between merely computing inverse probability (essentially as much a frequentist approach as Bayesian) and making a Bayesian inference (updating our certainty in light of new evidence).

CONDITIONAL PROBABILITY AND BAYES FORMULA

   The conditional probability that A is true given that B is true (has already occurred) is here (and in many places) denoted as $P(A|B)$ and spoken as the "probability of A given B".   It is my habit when confronted with an issue in probability and statistics to go back to my original textbook **Probability and Statistics** by Mosteller, Rourke, and Thomas (1961) [4] which I got for the educational TV program "Continental Classroom" (and latter in 1963 found to be a required text for Engineering 101 or 102 at Cornell).

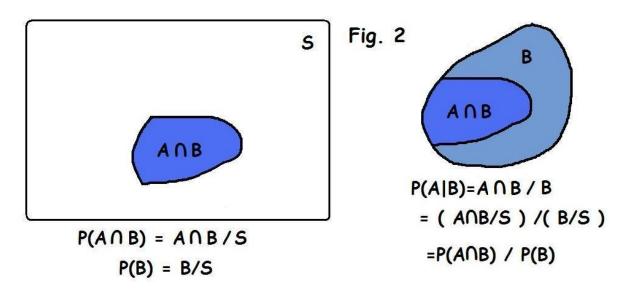   So how does one compute conditional probability?   The book says:

$$P(A|B) = P(A \cap B)/P(B) \tag{1}$$

Fig. 1

Fine – a formula – but do we remember what $P(A \cap B)$ is? What is A∩B for that matter? Okay – it's A and B, the upside down U is kind of like an A (for And) without the bar. Does equation (1) make sense.

In Fig. 1 we consider a sample space S which contains the set A and the set B (and more). There is some overlap set A ∩ B which belongs to both set A and set B. We can reasonably see that the probability of A (assuming areas are what is important) is $P(A) = A/S$, and $P(B) = B/S$. Likewise the probability $P(A \cap B)$ would be $(A \cap B)/S$. Note that when we are talking about $P(A|B)$ we now mean that B is given to us, not all of S. Thus the sample space is reduced from S to just B. B has happened – so we want to know how often A occurs in B. So $P(A|B) = (A \cap B)/B = P(A \cap B)/P(B)$. Note that when using the ratio of probabilities, in effect the S "cancels out" (see Fig. 2).

S

Fig. 2

B

A ∩ B

A ∩ B

P(A∩B) = A∩B /S

P(B) = B/S

P(A|B)=A ∩ B / B

= (A∩B/S) /(B/S)

=P(A∩B) / P(B)

In Fig. 2, on the left side we understand that $P(A \cap B) = (A \cap B)/S$. This is just the proportion of the total area that is common to sets A and B from Fig. 1, divided by the total area S. On the right side of Fig. 2, we now have the sample space reduced to just B (the given), so $P(A|B) = (A \cap B)/B$, the proportion of B taken up by A∩B. Now, dividing both top and bottom by S gives us $P(A|B)$ in terms of the original probabilities in all of S. We obtain equation (1).

So much for equation (1). This could be rearranged to give:

$$P(A \cap B) = P(A|B)P(B) \tag{2}$$

AN-392 (3)

This is not necessarily a very useful form of equation (1), which was about computing $P(A|B)$, not $P(A \cap B)$. The value of equation (2) is that we can write a completely equivalent form reversing the roles of A and B, giving a second expression for $P(A \cap B)$. Note that $A \cap B$ and $P(A \cap B)$, relative to S, must remain the same. Thus we have (isn't this neat!):

$$P(A \cap B) \;=\; P(B|A)P(A) \tag{3}$$

and equating (2) and (3) we have:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{4}$$

and this is often called the Bayes formula or indeed Bayes theorem. Remarkably remarkable!

   The formulas (the four equations above) are useful in computing conditional probabilities, but we seldom use them. We just write down the answers, having in mind some such arrangements as Fig. 1, or perhaps a tabulation of data (the enumerated contents of the sample space). So what is probably most useful about the Bayes formula (one of two things) is that it comes right out and suggests that you can consider calculating $P(B|A)$ knowing $P(A|B)$. That is sometimes called an inverse probability. We know the chances of A occurring given B. [This is like knowing the chance of getting a positive test (A) if you really have a disease (B).] The Bayes formula offers us the possibility of compute the chance of B being true if A has occurred. In the disease test example, the chance that you <u>really</u> have the disease if you have a positive test. Note that $P(A|B)$ and $P(B|A)$ are certainly not the same, and in fact are skewed by the ratio of $P(A)$ to $P(B)$. A further complication is that we may not know $P(A)$ and $P(B)$ exactly, or very well at all. Clearly we need to know the three terms on the right hand side, or be able to estimate them well, in order to compute $P(B|A)$ from equation (4).

   From equation (4) we see that the left side is computed knowing three terms on the right side. We have just assumed we at least know $P(A|B)$ well enough and want the inverse probability. What if we assume that $P(B/A)$ is not the inverse of $P(A/B)$ but rather an "update" of $P(B)$? That is, $P(B)$ was our original (prior) understanding of the probability of B but now, in light of new evidence A, we have a revised (posterior) understanding of the probability of B given that A has been established. Fig. 3 shows and contrasts the two ideas.

Equation (4) $P(B|A) = P(B) \, P(A|B) \, / \, P(A)$

Compute Inverse Probability

$$P(B|A) = \boxed{\frac{P(B)}{P(A)}} \, P(A|B)$$

B Given A   A Given B

Update "Certainty"

$$P(B|A) = \boxed{\frac{P(A|B)}{P(A)}} \, P(B)$$

Posterior   Prior

Fig. 3 Two Views of Equation (4)

It is hard to argue with anyone who contends that there is no difference between the two sides in Fig. 3. Indeed, for these to work we need to know all three terms on the right. We can however think of one term on the right as the actual "input" and the other two forming a ratio (in the boxes) that give the conversion factor. Also, the left side is an inversion process while the right side is an updating (refining) process. It is probably evident that not all the terms on the right side are necessarily available and/or completely accurate. In particular, we will note complications in computing $P(A)$ and in knowing $P(B)$.

KNOWING A PROBABILITY – FROM THE DATA

It might seem evident that knowing a probability ought to be simple. After all the probability of a head when flipping a coin is 1/2 and of drawing a Queen of Hearts from a deck of cards is 1/52. Perhaps even knowing the probability of catching the flu can be reasonably estimated. Possibly it is just a matter of observation. Say there are 14 boys in a class and 19 girls, so the probability that a child chosen at random is a boy is likely better chosen as 14/(14+19) = 14/33 = 0.4242 rather than just 1/2. Other classes may be different. So perhaps that works.

Further we know that the probability of two (independent) things both occurring is a product.   If we know that 73% of the students in our hypothetical class like math, in the class we know that 0.73 x 33 or 24 like math (and 27% or 9 don't like math). Here we have a "binary" in that a

	Boys	Girls	
Like Math	10	14	24
Don't like math	4	5	9
	14	19	33

Fig. 4

student is judged to like math or to not like math. The child's gender is another binary so we can form a 2x2 table as in Fig. 4, assuming we have assembled the necessary data. The probabilities must add to 1 for a binary categorization.

From the table, we can calculate everything we need. For example, the probability of a boy not liking math (we will denote "not" by the squiggle symbol "~") is $P(\sim\text{Math}|\text{Boy}) = 4/14$. The probability that a student not liking math is a boy is $P(\text{Boy}|\sim\text{Math}) = 4/9$. These obey the Bayes formula of course:

$$P(\sim\text{Math}|\text{Boy}) = P(\sim\text{Math})P(\text{Boy}|\sim\text{Math}) / P(\text{Boy})$$

$$= (9/33)(4/9) / (14/33) = 4/14 \tag{5}$$

This is as it should be, and everything is as it should be because we see all the data.

NO DATA TABULATION

There is no more famous example of the usefulness of the Bayes formula than the analysis of "false positives" in the test for a disease (see AN-390). In this case we have facts, but not a complete data table to work from.   Here we will change notation – substitute names for the terms in equation (4). This we do for an overall increase in clarity by eliminating the need for constant translation. But they are of course the same equation. Specifically the event B will become the actual disease denoted D, and the event A will be a <u>positive</u> test (true or false), denoted T.

$$P(D|T) = P(D) \, P(T|D) / P(T) \tag{6}$$

Thus we seek the probability of "really" having the disease D given a positive test for the disease T. This we obtain from the probability that the test will be positive when the disease is present, $P(T|D)$, which should be a high probability, approaching 1, which is what a test is for after all*. The "a priory" or "prior" probability of having the disease, knowing nothing else and not having the result of any test is $P(D)$, the "background" rate of the disease. So equation (6) is a revision of the chance of having the disease by the factor $P(T|D)/P(T)$.   We can look at the result in two ways. The most favorable is that if we have a positive test, the chance of having the disease is less (often much less) than we might have inferred from being given reliability scores. The second interpretation is that getting a positive test does (as it should) somewhat increase the chances we have a disease above the background $P(D)$.

*note at end

We looked at this example in some detail in AN-390, and noted there that the calculation of $P(T)$, which we were not given explicitly, involved a sum. This was because we wanted to know all positive tests, and we got the correct result by observing the data table. Writing a formula for $P(T)$ looks more complicated. Here it is:

$$P(T) = P(D)P(T \mid D) + P(\sim D)P(T \mid \sim D) \tag{7}$$

where

$$P(\sim D) = \left[1 - P(D) \right] \tag{7a}$$

Here we again use the squiggle symbol ($\sim$) to mean NOT, so $\sim D$ means NOT having the disease. Thus the total probability of having a positive test starts with the chance of having the disease times the chance that it is detected [first term of equation (7)]. The second term of equation (7) is the chance of not having the disease multiplied by the false positive rate, $P(T|\sim D)$. The probabilities of having or not having the disease must sum to 1, hence equation (7a). Here it is likely that:

$P(D)$ probability of disease (small)
$P(T|D)$ probability of true positive (large), test is fairly reliable
$P(\sim D)$ probability of not having the disease (large),
           uncommon disease
$P(T|\sim D)$ probability of false positive (small)

Thus of the two terms in equation (7) each is composed of a term that is likely to be small and one that is likely to be large (relative to 1). But we can't guess much about the sum of the two products ahead of time.

In our classroom example, suppose we had not had available the statistics on gender, but knew the conditional probabilities, and the overall vote that only 24 of 33 students liked math (9 did not). From this, analogous to equation (7) we can determine the probability of a boy $P(Boy)$ as:

$$P(Boy) = P(\sim Math)P(Boy|\sim Math) + [1 - P(\sim Math)]\ P(Boy|Math)$$

$$= P(\sim Math)P(Boy|\sim Math) + P(Math)P(Boy|Math)$$

$$= (9/33)(4/9) + (24/33)(10/24) = 14/33 \tag{8}$$

which is logical and in agreement with our tabulated data (which we were pretending we did not have).

It is expected that an event space S can be divided into more than just two partitions. Certainly we do have valid "binaries" such as boys/girls, likes-math/does-not-like-math, has-disease/does-not-have-disease. But we can well imagine such things as a student being indifferent to math, or having a test for a disease being three-way as positive, negative, or non-conclusive. So in general, the denominator of equation (4), the denominator of Bayes formula, which is what we are calculating in equation (7), need not just be the sum of two parts, but could be many parts. Still the total (unconditional) probabilities or all partitions must add to 1. (The denominator itself need not add to 1!) So we can often see equation (4), Bayes formula, written in a more general form: for the nth partition B_n:

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{\sum_i P(A|B_i)P(B_i)} \tag{9}$$

where:

$$P(A) = \sum_i P(A|B_i)P(B_i) \tag{9a}$$

*   *   *   *   *   *   *   *   *

We have looked at three probability puzzles: The Monty Hall (goat) puzzle (AN-389), the false-positive puzzle (AN-390), and the girl named Florida puzzle (AN-391) and here we have looked at the formulation of the Bayes equation. In AN-390 we did use the Bayes equation since the false positive example is used in many places as an example of the need for the Bayes rule. Here we will want to next see if the Bayes rule can be used for the other two puzzles.

*   *   *   *   *   *   *   *   *

BAYES AND THE GOATS

In AN-389 we looked at the classic Monty Hall puzzle and determined by two convincing methods that "switching" doubles your probability of winning the car from 1/3 to 2/3. Can we use Bayes rule in this case? Since we can always renumber (or re-letter) doors it will be sufficient here to consider that the contestant initially chooses door **a**, and that the host, knowing what's behind the doors, opens door **c** to show a goat. Should the contestant switch from door **a** to door **b**?   Yes.

AN-392 (8)

Bayes theorem is thought of as giving us the ability to re-evaluate our chances after we are given <u>new evidence</u>. Here we might suppose that the new evidence is merely the fact that door **c** is opened, now showing a goat. Thus we learn that the car is not there (~**c**). If this is the evidence, then applying Bayes formula we might think (wrongly):

$$P(a|\sim c) = P(a)P(\sim c|a) / P(\sim c)$$

$$= (1/3)(1)/ (2/3) = 1/2 \qquad\qquad (10a)$$

$$P(b|\sim c) = P(b)P(\sim c|b)/ P(\sim c)$$

$$= (1/3)(1) / (2/3) = 1/2 \qquad\qquad (10b)$$

Note here that $P(\sim c/a)$ or $P(\sim c/b)$ are both taken as 1 since if the car is behind a or b, it can't be behind **c**. But $P(\sim c)$ itself seems to be $1 - P(c) = 1 - 1/3 = 2/3$. $P(a)$ and $P(b)$ are both 1/3, the prior probabilities. Here both posterior probabilities are equal and equal to 1/2, what we <u>know to be the wrong answers!</u> This is the answer we would get if the contestant came to the show with the door already open showing a goat. We have left out of the <u>additional evidence</u> that door **c** was opened by the host who had complete knowledge of where the car was, and in response to a choice of door **a** by the contestant. Equations (10) are wrong.

Here are some revisions to consider [5]. First, $P(a)$ and $P(b)$ remain 1/3 as in equation (10). Secondly, if the car is really behind door **a**, then $P(\sim c|a) = 1/2$, since the contestant chose the <u>correct car door</u> and the host has the choice of <u>either</u> of the remaining wrong doors. Thirdly however, when we consider $P(\sim c|b)$ the situation is different. The host can't choose to open **a** (the contestant's choice) and he can't choose to open **b** because the car is there, so he must choose to open **c**. So now, $P(\sim c|b) = 1$. At last – a difference! And for that matter, $P(\sim c|c) = 0$: If the car <u>is</u> behind **c**, it can't also <u>not</u> be there.

Finally we complete the correct setup by computing the probability of $P(\sim c)$ which we originally thought was 2/3. Here we need to form the denominator in the manner or equation (9) as:

$$P(\sim c) = P(\sim c|a)P(a) + P(\sim c|b)P(b) + P(\sim c|c)P(c)$$

$$= (1/2)(1/3) + (1)(1/3) + (0)(1/3) = 1/2 \qquad\qquad (11)$$

So, tabulating: $P(a) = P(b) = 1/3$
$P(\sim c|a) = 1/2$
$P(\sim c|b) = 1$
$P(\sim c) = 1/2$

So we now put the corrected values into equation (10) and get instead:

$$P(a|\sim c) \ = \ P(a)P(\sim c|a) \ / \ P(\sim c)$$

$$= \ (1/3)(1/2) \ / \ (1/2) \ = \ 1/3 \qquad\qquad (12a)$$

$$P(b|\sim c) \ = \ P(b)P(\sim c|b) \ / \ P(\sim c)$$

$$= \ (1/3)(1) \ / \ (1/2) \ = \ 2/3 \qquad\qquad (12b)$$

Note that here we have two equations, two Bayes results because we wanted the probabilities for two different scenarios – switching of not switching. Since these are mutually exclusive, we actually needed just one. But isn't it nice to have two results which also give their correct counterpart? And this is the correct answer. It is twice as probable that the car is behind the door you did not initially choose.

BAYES AND FLORIDA

I am sometimes asked what the difference between application notes and our newsletter *Electronotes* is. The policy is not what anyone would call set in stone, but in general the newsletter is more about new stuff (news) and the ANs were about established, generally immediately useful material. There are many exceptions. But I have hesitated to put material I did not feel was "very solid" in the ANs. Thus forewarned that what follows may not be right, here is what I make of applying Bayes to the "Girl Named Florida" puzzle of AN-391. What I make is: It's too easy!

We are asked the probability that a two child family that has one girl named Florida has a second girl. Related problems might ask what the probability of two girls is (it's of course 1/4) and what the probability of two girls is if we know there is at least one girl (it's 1/3) and what the probability of two girls is if the first child is a girl (it's 1/2), as discussed in AN-391. These results seem solid and widely discussed. So, isn't the fact that the girl is named Florida irrelevant – it apparently would just tells us there is at least one girl – so the result should be a probability of 1/3 of there being two girls. But it seems to be 1/2, as discussed by Mlodinow and others, and as argued in AN-391 by showing the reduced sample space, and by computer simulation. So, with baited breath, what does Bayes say?

Simply set up, we seek the probability of a two girl family $P(TGF_{amily})$ given at least one girl who is named Florida (F). Bayes formula apparently gives us:

$$P(TGF_{amily}|F) \ = \ P(TGF_{amily}) \ P(F|TGF_{amily}) \ / \ P(F) \qquad\qquad (13)$$

First, the probability of a $\mathrm{TGF_{amily}}$ is just 1/4. So what is the probability of a girl named Florida? Well, it's very small, and most accounts of this seem to suggest that this is important. Of course it is of no particular importance that the girl's name is (specifically) Florida, except if it is necessary to argue that the name is very rare. However, is it necessary that the probability be <u>very</u> small, except for removing the case where a two girl family has two girls named Florida?   Mlodinow likes the value $P(F) = 0.000001$, but let's here just call it $P(F) = \varepsilon$.

So that leaves us with $P(F|\mathrm{TGF_{amily}})$. What is the probability of a family naming a girl Florida: well it's ε. Now, given the fact that they have two girls, two girls to name, it seems that now there are two chances of choosing Florida, and the probability becomes $P(F|\mathrm{TGF_{amily}}) = 2\varepsilon$, <u>at least for small enough ε</u>.   It reminds us of hitting the lottery. We have twice the chance with two tickets that we do with just one. If we were to examine the chance that a lottery player had won the prize given that he had bought multiple tickets, we would expect a larger number.

Well that gives us:

$$P(\mathrm{TGF_{amily}}|F) \;=\; (1/4)(2\varepsilon/\varepsilon) \;=\; 1/2 \tag{13a}$$

which in addition to being simple, is the right answer.   Is this luck?

$$* \qquad * \qquad * \qquad \text{ASIDE} \qquad * \qquad * \qquad * \qquad * \qquad * \qquad *$$

Similar but simpler derivations give us Bayes formula results for the three preliminary cases mentioned above:

$$P(\mathrm{TGF_{amily}}|\mathrm{SkyIsBlue}) \;=\; P(\mathrm{TGFF_{amily}})P(\mathrm{SkyIsBlue}|\mathrm{TGF_{amily}}) \,/\, P(\mathrm{SkyIsBlue})$$

$$= (1/4)(1)/(1) \;=\; 1/4 \tag{14a}$$

which is just $P(\mathrm{TGF_{amily}})$ since the sky is always blue (ask any child). Further:

$$P(\mathrm{TGF_{amily}} |\mathrm{AtLeastOneGirl}) = P(\mathrm{TGF_{amily}})P(\mathrm{AtLeastOneGirl} |\mathrm{TGF_{amily}})/P(\mathrm{AtLeastOneGirl})$$

$$= (1/4)(1)/(3/4) \;=\; 1/3 \tag{14b}$$

since a family that has two girls must have at least one, and the probability of at least one girl is 3 in 4 (not two boys). The remaining of our run-up cases is that of two girls given that the first child is a girl.

$$P(TGF_{amily}|FirstChildGirl) \ = \ P(TGF_{amily})P(FirstChildGirl|TGF_{amily})/P(FirstChildGirl)$$

$$= \ (1/4)(1)/(1/2) \ = \ 1/2 \qquad\qquad (14c)$$

*   *   *   *   *   *   *   *   *

Thus Bayes rule seems to work in four cases, the three lead-up cases (equations 14) and the Florida case. Further, both equations 13a and 14c have the same answer of 1/2, the required 2:1 ratio being $2\varepsilon/\varepsilon$ in the Florida case and $1/(0.5)$ in the first child girl case. This may be just a case of having one child well-defined while the gender of the second is still up to chance.

From equation (13a) we would suppose that the probability of being named Florida does not matter. Indeed, as long as ε is small, it doesn't seem to matter. In the more general case, this is only seen as an approximation. If ε is large enough, we have to consider that two girls named Florida is a factor to bring in. What we want is the probability of having a girl named Florida when we have two rather than one chance of choosing the name. This sort of probability problem seems to be easiest if we first find the probability that something does not happen. Thus, if the probability of naming a girl Florida is $P(F) = \varepsilon$, the probability of not naming her Florida is $P(\sim F) = (1 - \varepsilon)$. With the second girl we have the same $P(\sim F)$ so we multiply:

$$P(G1\sim F \text{ and } G2\sim F) \ = \ (1-\varepsilon)^2 \ = \ 1 \ - \ 2\varepsilon \ + \varepsilon^2 \qquad\qquad (15)$$

and the probability of at least one girl <u>being</u> named Florida is:

$$P(GF \text{ or } GF) = \ 1 - [1 \ - \ 2\varepsilon \ + \boldsymbol{\varepsilon^2}\,] \ = \ 2\varepsilon \ - \ \boldsymbol{\varepsilon^2} \qquad\qquad \boldsymbol{(}16)$$

which remains well approximated by 2ε for small ε.

But now we need to remember that the denominator of the Bayes formula needs to be a summation of partitions, as in equations (9).

$$P(TGF_{amily}|F) \ =$$

$$\frac{P(TGF_{amily})\ P(F|TGF_{amily})}{P(F|OGF_{amily})P(OGF_{amily})+P(F|TGF_{amily})P(TGF_{amily})} \qquad\qquad (17)$$

AN-392 (12)

where OGF_{amily} is obviously a one-girl family (Probability 1/2).

We just now need to plug in stuff we pretty much have already considered, but let's carefully tabulate:

$$P(OGF_{amily}) \;=\; 1/2 \tag{18a}$$

$$P(TGF_{amily}) \;=\; 1/4 \tag{18b}$$

$$P(F|OGF_{amily}) \;=\; \varepsilon \tag{18c}$$

$$P(F|TGF_{amily}) \;=\; 2\varepsilon \;-\; \varepsilon^2 \qquad\qquad \textbf{(Equation (16))} \tag{18d}$$

So: TGF_{amily}

$$P(TGF_{amily}|F) = \frac{\left(\frac{1}{4}\right)(2\varepsilon \;-\; \varepsilon^2)}{\varepsilon\left(\frac{1}{2}\right) + \;(2\varepsilon \;-\; \varepsilon^2)\left(\frac{1}{4}\right)} \tag{19}$$

which can be simplified to [6]:

$$P(TGF_{amily}|F) \;=\; (2-\varepsilon)/(4-\varepsilon) \tag{20}$$

which is indeed a simple result, and does become 1/2 as ε goes to 0. Wonder if it's right?

We saw in AN-391 that the simulation program, *florida.m*, verified that the value of 1/2 was correct. There we used ε=0.01 rather than 0.000001 just so that we had a moderate number of Florida outcomes. We set this probability and can change it to any value we want, with the notion of verifying equation 17. The results of running this for a range of ε is tabulated below.

The program *florida* was run for 10,000,000 trials for each value of ε. For each trial, a family of two was randomly constructed, with equal probabilities to boy/girl and with probability ε of a girl being named Florida (see program in AN-391 for the details). In the table, the first column is for the probability ε. The second two columns relate to occurrences, the fraction of occurrences out of 10,000,000, with the second column being the theoretical equation being tested, equation (16), and the third column the result of the simulation. Columns four and five are corresponding data relating to the conditional probability in question here. Thus column four is the theoretical equation (20), while the fifth column is the results from the simulation

ε	Theory $(1/2 - \varepsilon/4)\varepsilon$	Simulation $TGF_{am}Fl/N$	Theory $(2-\varepsilon)/(4-\varepsilon)$	Simulation $GF_{am}Fl / F_{am}Fl$
0.0001	0.0000500	0.0000530	0.5000	0.5048
0.001	0.000500	0.000497	0.4999	0.5008
0.01	0.00500	0.00500	0.4987	0.4975
0.02	0.0099	0.0099	0.4975	0.4985
0.05	0.0244	0.0243	0.4937	0.4937
0.1	0.0475	0.0476	0.4872	0.4875
0.2	0.0900	0.0901	0.4737	0.4738
0.4	0.1600	0.1601	0.4444	0.4441
0.6	0.2100	0.2102	0.4118	0.4121
0.8	0.2400	0.2402	0.3750	0.3753
1.0	0.2500	0.2501	0.3333	0.3330

Clearly we see outstanding agreement between theory and simulation here. Only for very small values of ε do we see much noise. This is because with very small probabilities of the name Florida, the integer counts are very small (even with 10,000,000 families), and the errors in ratios jitter about.

Overall, we have confidence that the value of 1/2 of the earlier analysis is correct, and a more general result (a range of probability) is also found.

FLORIDA LOOSE ENDS:

Here we want to eventually look at what Bayes has given us, but we need to clean up a few loose ends first. And, do the equations and the table of data tell us anything that provides insight.

First we need to address the difference between the Bayes equations as presented in equation (13) and equation (17), the difference being the denominators. The denominator should really be the expanded form of the sum, equation (17) through the

use of equations (9) and (9a). So we had a simple expression of $P(F) = \varepsilon$ and it is replaced by the more specific:

$$P(F) = \varepsilon\left(\frac{1}{2}\right) + (2\varepsilon - \varepsilon^2)\left(\frac{1}{4}\right) \qquad (21)$$

which is still really ε for small $\varepsilon \ll 1$. Indeed, it is only for quite unlikely large values of ε that $P(TGF_{amily}|F)$ departs from 1/2 and heads for the value of 1/3 as ε goes to 1. The difference is whether or not we take seriously the possibility that two girls in the same family can be named Florida, (considering just probable independent choices, cultural issues – traditions in naming children - aside).

As we let the probability of the name Florida increase from zero, or at least from a very small number, to a number approaching and finally reaching 1, the choice of the name Florida becomes <u>mandatory</u>. At the same time, the choice of any other name must disappear. For Fig. 5 below we reproduce a portion of the grid of the sample space from AN-391. This 3x3 grid had nine cells, but now GNF (Girl Not-Florida) is impossible, so only the four cells, inside the red frame, are possible. Now, only GF or B is possible. For the upper left cell (BB) this is not possible because we must have a girl Florida. We are left with only three possible cells, and two have one boy. Thus based on the sample space, only 1/3 of cases are favorable. Hence the 0.3333 in the data table. This is the same as the probability of two girls given one girl. The distinction between girl and girl-Florida is gone.
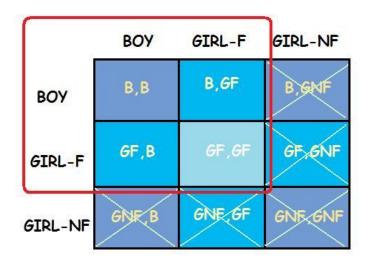

Fig. 5 Name Florida Becomes Mandatory

AN-392 (15)

As powerful as Bayes theorem is, it is quite possible to become confused. It should be apparent that in addition to just writing equations, we value simulation as a powerful verification technique, and the display of sample space as a grid or as a table provides very useful insight. In fact, when we encounter what is apparently an absurd conjecture about a changed probability, we really do need to look at the <u>changed sample space</u> to really see that no actual magic or slight-or-hand is involved.

WHAT DOES IT MEAN

There are a couple of points that can usefully be made here before we move on. First of all, in reading or writing this material, there is some sort of déjà vu because it is like putting different symbols in a familiar framework pattern. At times, we anticipate what is needed because it's like writing derivatives, "chain rules", gradients, that sort of thing. Putting in the right things is not necessarily easy, but it never seems "forced".

A second point is that we can't help being impressed by Bayes theorem in its simplicity and corresponding elegance, as well as its power to solve problems. While it can be viewed as simple inverse probability, this may remind us of other "transform" notions that are so useful in engineering (like Laplace transforms, for one example). Recall that one major reason for using transforms is that it is often much much easier to solve the problem in the "other domain". This is the case with Bayes theorem. We had, for example, no good idea of calculating $P(\text{TwoGirls}|\text{Florida})$. It seemed to be a joke. On the other hand, it was relatively easy to calculate $P(\text{Florida}|\text{TwoGirls})$. We needed $P(\text{Florida})$ and $P(\text{TwoGirls})$ as well, but these were easy enough.

Regardless of our level of comfort with using Bayes theorem and our notice of the fact that it sometimes makes solving a problem much much easier, we may well not be comfortable with the results. Clearly it <u>is impossible</u>, for example, that naming a girl Florida could make any difference in the gender makeup of a family. So it must <u>mean</u> something else. Bayes is about information and belief. It is about thinking of probability in terms of what we would consider "fair betting odds". So when we say something in terms of conditional probability, we are generally rearranging information and usually <u>changing the sample space</u>.

REFERENCES/NOTES   Regarding the first three references: I like these three books very much.   They all read easily, and contain a lot of interesting stories as one might expect from books that have relative long (or just long) subtitles. They all address Bayes as a person, and all give the false positive example, although McGrayne (in Appendix B) has a math error, and overall, almost no math.   None has nearly as much of the math as we have in this note.

[1] Leonard Mlodinow *The Drunkard's Walk – How Randomness Rules Our Lives*, (Pantheon 2008)

[2]   Sharon Bertsch McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, (Yale, 2011)

[3] Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't,* (Penguin, 2012)

[4] Fred Mosteller, Robert Rourke, and Robert Thomas, *Probability and Statistics* (Addison-Wesley1961)

[5]   After fighting with this problem for a while, I looked to the internet and found a very clear explanation (from 2005) of the application of Bayes to the Monty Hall problem at a website called "Devlin's Angle" by Keith Devlin. In my efforts, I had found two things with the original approach that might be the problem, but neither fixed it. Devlin took a similar approach - and fixed both problems. Head slapping time! Here is another Bayes problem to consider: What is the probability that both things are wrong given the non-success of trying each of two possible solutions individually!

[6] I first believed this equation because it agreed with my simulation. In truth, it was the case that I worked out what the answer had to be based on the "experimental" data. This is in line with my saying "Never Underestimate the Value of Knowing or Suspecting the Right Answer".   In fact, the answer I had as the theoretical value in my program (corrected for AN-391) was neither equation (19) or equation (20), but something equivalent in between. So when I went back to the internet to search to find if anyone had done this same calculation (always reassuring) to my surprise, I did find a simple answer, identical to what became equation (20). I didn't believe it, but the setup appeared equivalent to mine. So gosh – how blind we can be. One more step and my equation (19) result simplified to the one on the web, which is my equation (20). The web page is by Brian Blais of the "Professor Brian Blais' Blog" (From 2010):

http://bblais.blogspot.com/2010/01/there-once-was-girl-named-florida-aka.html

Curiously while he names his blog "Blais' Blog" he does use the term "Bayes Theorem" (no apostrophe).

We note that P(T|D), the probability of a positive test when a disease is present should approach 1, perhaps 0.9 or higher if the test is worth anything. Suppose we don't know this, so just assume it is 1. Then equation (6) would become just:

$$P(D|T) = \frac{P(D)}{P(T)}$$

That is, the probability of having the disease given a positive test is the background probability of the disease divided by the probability of a positive test. This probability gets small as P(D) gets smaller and/or P(T) gets larger (because of a large number of false positives). This is intuitive and does not require Bayes. If this alone tells you that there is only a 10% chance of your actually having the disease, rather than say 90%, that's the point. Either 10% or 9% (if the actual P(T,D) were used) really does not make much difference.

True, Bayes is hard to follow for the average person. As many point out, "natural frequencies", like cases of 1000 people like yourself, are much easier. The highly intuitive cases of just using the $P(D)/P(T)$ ratio above is simplest, and largely valid. It is a surprise that many health-care professionals may not be able to relate this, or even to understand it themselves. They must be wondering how it is, if the test is 90% valid, that the positive tests they themselves get back for their particular patients are usually wrong.