

# Bayes' For Beginners

Chris Westbury

Department of Psychology, P220 Biological Sciences Bldg., University of Alberta,

Edmonton, AB, T6G 2E9, Canada.

E-mail: [chrisw@ualberta.ca](mailto:chrisw@ualberta.ca)

This manuscript was written for my 4<sup>th</sup> year Psychometrics class. It is released under a Creative Commons license: Attribution-NonCommercial-ShareAlike 2.5. You are free to copy, distribute, display, and (God forbid!) perform the work under the following conditions:

- By Attribution. You must attribute the work to the author.
- Noncommercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

\* For any reuse or distribution, you must make clear to others the license terms of this work.

\* Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code that can be obtained from:

<http://creativecommons.org/licenses/by-nc-sa/2.5/legalcode>

Acknowledgements– Thanks to Gail Moroschan for feedback on an earlier version of this manuscript.

### Abstract

Bayes' Rule is a simple way of calculating conditional probabilities. Although a great deal has been written about the relevance of Bayes' Rule in clinical settings, it is difficult to find a single article that is both mathematically comprehensive and easily accessible to students and professionals with clinical backgrounds. This article tries to fill that void, by laying out the nature and implications of Bayes' Rule in a way that requires little or no background in probability theory. It builds on Meehl & Rosen's classic (1955) paper, *Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores*, by laying out algebraic proofs that they simply allude to, and by providing extremely simple and intuitively accessible examples of the concepts that they simply assumed their reader understood.

Bayes' Rule is a way of calculating conditional probabilities. Although it is quite simple in its conception, Baye's Rule can be fiendishly difficult for beginners to understand and apply. In part this is because it forces us to confront and overcome strong biases in our natural way of thinking and in part it is because it is not easy to be specific about exactly where Bayes' Rule will apply, or how it may apply in any particular case. The purpose of this paper is to present and explore the simplest forms of Bayes' Rule, and to understand how it may be used in practical reasoning, especially in clinical settings.

A great deal has been written about the importance of conditional probability in diagnostic situations. However, there are, so far as I know, no papers that are both comprehensive and simple. Most writing on the topic, particularly in probability textbooks, assumes too much knowledge of probability for diagnosticians, losing the clinical reader by alluding to simple proofs without giving them. Many introductory psychometrics textbooks err on the other side, either ignoring conditional probability altogether, or by considering it in such a cursory manner that the reader has little chance to understand what it is and why it matters. This paper is intended to fill the perceived void between simplicity and thoroughness. The exposition provided here assumes only the most basic understanding of non-conditional probability, and provides both concrete examples and simple algebraic proofs of some implications of Bayes' Rule that are clinically relevant. It may be reasonably considered an interpretive guide to perhaps the best paper on Bayes' Rule for the clinician, Meehl & Rosen's classic (1955) paper, *Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores*. Most students of psychology find this paper very difficult to understand, in part

because the authors do make mathematical claims without providing any detailed explanation of where they came from. The present paper frames Meehl & Rosen's claims with a more basic introduction than they give, and fills in some simple proofs that they only allude to.

The first section consists of a general introduction to understanding conditional probabilities. The second section introduces Bayes' Rule itself, in an historical and mathematical setting. The third section lays out some implications of Bayes' Rule that follow as a direct result of its definition.

### *Conditional Probabilities*

Conditional probabilities are those probabilities whose value depends on the value of another probability. Such probabilities are ubiquitous. For example, we may wish to calculate the probability that a particular patient has a disease, given the presence of a particular set of symptoms. The probability of disease may be more or less close to certain, depending on the nature and number of symptoms. Or we may wish to calculate the probability that a given hypothesis is true, given a diverse set of evidence (say, results from several experiments) for and against it. Hypothesis testing is just one way of assigning weight to belief. Conditional probabilities also come into play when we wish to decide how much confidence we wish to assign to a given belief.

A very simple example of conditional probability will elucidate its nature. Consider a 6/49 lottery, in which players are invited to choose 6 out of 49 numbers, and win a jackpot if their six numbers are chosen. The probability that any particular six numbers will be chosen is 1 in  $(49 \times 48 \times 47 \times 46 \times 45 \times 44)$ , or 1/10,496,787,840. These

clearly are not very good odds: if you entered a 6/49 lottery every day from your eighteenth to your eightieth birthday, you would still only have about one chance out of 464,000 of winning the lottery.

To understand conditional probability, consider the question: How likely is that you would win the jackpot in a 6/49 ticket if you didn't have a ticket? It should be obvious that the answer is zero— you certainly could not win if you didn't even have a ticket. So the probability of winning a 6/49 lottery is really a conditional probability, where your odds of winning are conditional on the number of tickets you have purchased. If you have zero tickets, then you have no chance of winning. With one ticket, you have 1/10,496,787,840 chances to win. With two tickets, your odds will be twice as good, and you will have 2/10,496,787,840 chances of winning.

We symbolize conditionality by using a vertical slash ' | ', which can be read as 'given'. Then the odds of winning the 6/49 with one ticket could be expressed as  $p(\text{Winning} \mid \text{One ticket})$ . There are many 'keywords' in a problem's definition that may (but need not necessarily) suggest that you are dealing with a problem of conditional probability. Phrases like 'given', 'if', 'with the constraint that', 'assuming that', 'under the assumption that' and so on all suggest that there may be a conditional clause in the problem.

One thing that sometimes confuses students of probability is the fact that *all* probability problems are really conditional. Consider the simple probability question: 'What is the probability of getting a head with a coin toss?' The question implicitly assumes that the coin is fair (that is, that heads and tails are equally probable), and should really be phrased 'What is the probability of getting a head with a coin toss, *given that the*

*coin is fair?* Non-conditional probability problems conceal their conditional clause in the background assumptions that either explicitly or implicitly limit the domain in which the probability calculation is supposed to apply.

This observation sheds light on what conditionality actually does. A condition always serves exactly this role: to limit the domain in which the 'non-conditional' portion of the question is supposed to apply. When you are asked 'What is the probability of getting a head with a coin toss?' you are supposed to understand that we are only considering fair coins. When you are asked 'What is the probability that you have disease X, given that you have symptom Y?', you are supposed to understand that now the probability calculation only applies to those people who do have symptom Y. An appropriate way of thinking about conditional probability is to understand that a conditional limits the number and kind of cases you are supposed to consider. You can think of the vertical slash as meaning something like 'ignoring everything to which the following constraint does not apply'. So 'What is the probability of getting a head with a coin toss, given that the coin is fair?' means 'What is the probability of getting a head with a coin toss, ignoring every coin to which the following statement does not apply: The coin is fair'.

Bayes' Rule and other methods of solving conditional probability questions are simply mathematical means of limiting the domain across which a calculation is being computed. To see that this is so, consider the following simple question:

Three tall and two short men went on a picnic with four tall and four short woman. What is  $P(\text{Tall} \mid \text{Female})$ , the probability that a person is tall, given that the person is female?

The solution to this problem may be immediately obvious, but it is worth working through a few ways of solving it. These are all formally the same, though they may appear to be different.

The first way is just to turn the question into a very simple non-conditional question that we know how to solve. Following the discussion above, the question can be re-phrased to say 'What is the probability that a person is tall, ignoring everyone who is not a woman?' If we ignore the men, we have a really simple question, viz. 'Four tall and four short woman went on a picnic? What is the probability that a woman who went on the picnic was tall?' This is simple (that is, non-conditional) probability. Like any simple probability question, it can be solved by dividing the number of ways the outcome of interest ('being tall') can happen by the number of ways any outcome in the domain ('being a woman') can happen. So:  $4 \text{ tall women} / (4 \text{ tall woman} + 4 \text{ short woman}) = 0.5$  probability that a person on the picnic was tall, given that she was a woman.

A formally identical way of solving the same problem can be seen by drawing a 2 x 2 table such as the following:

GENDER / HEIGHT	FEMALE	MALE
TALL	4	3
SHORT	4	2

The condition 'Given that she was a female' means that we can simply ignore the rightmost column of this box, the males, and act as if the question about the probability of being tall only applied to the leftmost column, the woman.

Here comes the tricky part. This diagram makes clear what the question is asking: What is the ratio of people who are both tall and female (top left cell) to people who are female (sum of left column)? We can re-state this and solve the problem in a third way by asking: What is the ratio of the probability that a person is both female and tall to the probability that a person is female? To see why, consider the concrete example again. There were thirteen people on the picnic. Since 4 were tall females, the probability of being a tall female is  $4/13$ . Since 8 were females, the probability of being female was  $8/13$ . The ratio of people who were both tall and female to people who were female is therefore  $4/13 / 8/13$ , or  $4/8$ , or 50%. The reason this may seem 'tricky' is that here we consider the domain as a whole— all people who went on the picnic— and then take the ratio of two probabilities in that domain.

If you understand this third method of calculating the conditional probability, then you will understand Bayes' Rule. Bayes Rule is a way to 'automatically' pick out this very same ratio: the ratio of the probability of being in the cell of interest (in this case, the cell tall and female picnickers) to the probability of being in the sub-domain of interest that is specified by the conditional clause (in this case, woman, a subset of all the people who went on the picnic).

Before we look at how the math works, let's introduce the rule itself.

### *Bayes' Rule*

Bayes' Rule is very often referred to Bayes' Theorem, but it is not really a theorem, and should more properly be referred to as Bayes' Rule (Hacking, 2001). In either case, it is so-called because it was first stated (in a different form than we consider



here) by Reverend Thomas Bayes in his 'Essay towards solving a problem in the doctrine of chances', which was published in the *Philosophical Transactions of the Royal Society of London* in 1764. Bayes was a minister interested in probability and stated a form of his famous rule in the context of solving a somewhat complex problem involving billiard balls that need not concern us here.

Bayes' Rule has many analogous forms of varying degrees of apparent complexity. This paper concerns itself almost entirely with the simplest form, which covers the cases in which two sets of mutually exclusive possibilities A and B are considered, and where the total probability in each set is 1. At the end of the paper we will briefly examine how this most simple case is just a specific case of a more general form of Bayes' Rule. The simplest case covers many diagnostic situations, in which the patient either has or does not have a disease (possibility set A) and either has or does not have a set of symptoms (possibility set B). For such cases, Bayes' Rule can be used to calculate  $P(A | B)$ , the probability that the patient has the disease given the symptom set. Bayes' Rule says that:

$$P(A | B) = P(B | A) P(A) / P(B)$$

$P(A)$  is called *the marginal or prior probability of A*, since it is the probability of A prior to having any information about B. Similarly, the term  $P(B)$  is *the marginal or prior probability of B*. Because it does depend on having information about B, the term  $P(A | B)$  is called *the posterior probability of A given B*. The term  $P(B | A)$  is called *the likelihood function for B given A*.

In the third solution to the example above, we solve for the probability of being female, given that you are tall, by considering the ratio of those who were tall and female to those who were female:

$$P(\text{Tall} \mid \text{Female}) = P(\text{Tall} \ \& \ \text{Female}) / P(\text{Female})$$

This suggests that Bayes' Rule can also be stated in the following form:

$$P(A \mid B) = P(A \ \& \ B) / P(B)$$

From this it should be evident, by equating the numerators of the two equations above, that:

$$P(A \ \& \ B) = P(B \mid A) P(A)$$

This is true by the definition of '&'. Let us try to understand why this is so, by again considering the three tall and two short men went on a picnic with four tall and four short woman. We have already convinced ourselves that  $P(\text{Female} \ \& \ \text{tall})$  is  $4/13$ , because there are 4 people in the cell of interest and thirteen people in the problem's domain. Let's see how the definition agrees with this answer. The definition above says that  $P(\text{Female} \ \& \ \text{Tall}) = P(\text{Tall} \mid \text{Female})P(\text{Female})$ .  $P(\text{Tall} \mid \text{Female})$ , the probability of a picnicker being tall given that she is female, is  $4/8$ .  $P(\text{Female})$  is  $8/13$ , because eight of the thirteen people on the picnic are females.  $4/8$  multiplied by  $8/13$  is  $4/13$ .

Note that it is equally correct to write that:

$$P(A \ \& \ B) = P(A \mid B) P(B)$$

In other words:

$$P(B \mid A)P(A) = P(A \mid B) P(B)$$

Let's see why using the same example. Now we will see that  $P(\text{Female} \ \& \ \text{Tall}) = P(\text{Female} \mid \text{Tall})P(\text{Tall})$ .  $P(\text{Female} \mid \text{Tall})$ , the probability of a picnicker being female

given that he or she is tall, is  $4/7$ , because there are four tall females and seven tall people altogether.  $P(\text{Tall})$  is  $7/13$ , because seven of the thirteen people on the picnic are tall.  $4/7$  multiplied by  $7/13$  is  $4/13$ .

If you go back and look at the  $2 \times 2$  table above, you should be able to understand why these two calculations of  $P(A \& B)$  must be the same. The first calculation picks out the cell of tall females *by column*. The second picks it out *by row*. It doesn't matter if you concern yourself with females who are tall or tall people who are females— in the end you must get to the same answer if you want to know about people who are both tall and female. A tall female person is also a female tall person.

So now we have

$$P(A | B) = P(B | A)P(A)/P(B) = P(A | B)P(B)/P(B)$$

Although either form will give the same answer, the first form is the 'canonical' form of Bayes' Rule, for a reason that should be obvious: because the second form contains the same element on the right,  $P(A | B)$ , as the left element that we are trying to calculate. If we already know  $P(A | B)$ , then we don't need to compute it. If we don't know it, then it will not help us to include it in the equation we will use to calculate it.

Bayes' Rule can be easily derived from the definition of  $P(A | B)$ , in the following manner:

- |                                       |                            |
|---------------------------------------|----------------------------|
| 1.) $P(A   B) = P(A \& B) / P(B)$     | [ By definition ]          |
| 2.) $P(B   A) = P(A \& B) / P(A)$     | [ By definition ]          |
| 3.) $P(B   A) P(A) = P(A \& B)$       | [ Multiply 2.) by $P(A)$ ] |
| 4.) $P(A   B) P(B) = P(B   A) P(A)$   | [ Substitute 1.) in 3.) ]  |
| 5.) $P(A   B) = P(B   A) P(A) / P(B)$ | [ Bayes' Rule ]            |

It might seem at first glance that Bayes' Rule cannot be a very helpful rule, because it says that to solve a conditional probability  $P(A | B)$  you have to know another

conditional probability  $P(B | A)$ . However, Reverend Bayes' insight was that in many cases the second possibility is knowable when the first is not. In diagnostic cases where we are trying to calculate  $P(\text{Disease} | \text{Symptom})$  we often know  $P(\text{Symptom} | \text{Disease})$ , the probability that you have the symptom given the disease, because this data has been collected from previous confirmed cases. In scientific cases where we want to know  $P(\text{Hypothesis} | \text{Result})$ , the probability that a hypothesis is true given some relevant result, we may know  $P(\text{Result} | \text{Hypothesis})$ , the probability that we would obtain that result given that the hypothesis is true- this is often statistically calculable, as when we have a p-value.

### *Implications of Bayes Rule*

Bayes' Rule is very simple. However, its implications are often unexpected. Many studies have shown that people of all kinds- even those who are trained in probability theory- tend to be very poor at estimating conditional probabilities. It seems to be kind of innate incompetence in our species. As a result, people are often surprised by what Bayes' Rule tells them.

Let us consider an example given in Meehl & Rosen (1955), from which much of the discussion in this section is drawn.

A particular disorder has a *base rate* occurrence of 1/1000 people. A test to detect this disease has a *false positive rate* of 5%- that is, 5% of the time that it says a person has the disease, it is mistaken. Assume that the false negative rate is 0%- the test correctly diagnoses every person who does have the disease. What is

the chance that a randomly selected person with a positive result actually has the disease?

When this question was posed to Harvard University medical students, about half said that the answer was 95%, presumably because the test has a 5% false positive rate. The average response was 56%. Only 16% gave the correct answer, which can be computed with Bayes' Rule in the following manner:

$$\begin{aligned}
 \text{Let: } P(A) &= \text{Probability of having the disease} = 0.001 \\
 P(B) &= \text{Probability of positive test} \\
 &= \text{Sum of probabilities of all independent ways to get a positive test} \\
 &= \text{Probability of true positive} + \text{probability of false positive} \\
 &= (\text{True positive base rate} \times \text{Percent correctly identified}) + (\text{Negative} \\
 &\text{Base Rate} \times \text{Percent incorrectly identified}) \\
 &= (0.001 \times 1) + (0.999 \times 0.05) \\
 &= 0.051 \\
 P(B | A) &= \text{Probability of positive test given disease} = 1
 \end{aligned}$$

$$\begin{aligned}
 \text{Then: } P(A | B) &= P(B | A) P(A) / P(B) \\
 &= (1 \times 0.001) / (0.051) \\
 &= 0.02, \text{ or } 2\%
 \end{aligned}$$

Although the test is highly accurate, it in fact gives a correct positive result just 2% of the time. How can this be? The answer (and the importance of Bayes' Rule in diagnostic situations) lies in the highly skewed base rates of the disease. Since so few people actually have the disease, the probability of a true positive test result is very small. It is swamped by the probability of a false positive result, which is fifty times larger than the probability of a true positive result.

You can concretely understand how the false positive rate swamps the true positive rate by considering a population of 10,000 people who are given the test. Just 1/1000th or 10 of those people will actually have the disease and therefore a true positive test result. However, 5% of the remaining 9990 people, or 500 people, will have a false

positive test result. So the probability that a person has the disease given that they have a positive test result is  $10/510$ , or 2%.

Many cases are subtle. Consider another case cited by Meehl & Rosen (1955). This involved a test to detect psychological adjustment in soldiers. The authors of the instrument validated their test by giving it to 415 soldiers known to be well-adjusted, and 89 soldiers known to be mal-adjusted. The test correctly diagnosed 55% of the mal-adjusted soldiers as mal-adjusted, and incorrectly diagnosed only 19% of the adjusted soldiers. Since the true positive rate (55%) is much higher than the false positive rate (19%), the authors believed their test was good. However, they failed to take into account base rates. Meehl & Rosen did not know  $P(\text{Maladjusted})$ , the probability that a randomly-selected soldier was maladjusted, but they guessed that it might be as high as 5%. With this estimate, we can use Bayes' Rule as follows:

$$\begin{aligned}
 &\text{Let } P(M) = \text{Probability of being maladjusted} = 0.05, \text{ by assumption} \\
 &\text{Let } P(D) = \text{Probability of being diagnosed as being maladjusted.} \\
 &\quad = \text{Probability of true positive} + \text{probability of false positive} \\
 &\quad = (\text{True positive base rate} \times \text{Percent correctly identified}) + (\text{Negative} \\
 &\quad \text{Base Rate} \times \text{Percent incorrectly identified}) \\
 &\quad = (0.55 \times 0.05) + (0.95 \times 0.19) \\
 &\quad = 0.208 \\
 &P(D | M) = \text{Probability of being diagnosed, given maladjustment.} \\
 &\quad = 0.55, \text{ as found by the authors.} \\
 &P(M | D) = \text{Probability of maladjustment given diagnosis as maladjusted} \\
 &\quad = P(D | M)P(M)/P(D) \quad [ \text{Bayes' Rule} ] \\
 &\quad = (0.55)(0.05)/0.208 \\
 &\quad = 0.13 \text{ or } 13\%
 \end{aligned}$$

When base rates are taken into account, the test's true positive rate is just 13%, not 55% as claimed. The test is still better than guessing that everyone is maladjusted. With that strategy 5% of positive diagnoses would be correct. However, note that the test's diagnosis of maladjustment is much more likely to be wrong (87% probability) than right

(13% probability).

Of course we prefer to make diagnoses that are more likely to be right than wrong. We can state this desire more formally by saying that we want the fraction of the population that is diagnosed correctly to be greater than the fraction of the population that is diagnosed incorrectly. Mathematically this leads to a useful conclusion in the following manner:

$$\begin{aligned} \text{Fraction diagnosed correctly} &> \text{Fraction diagnosed incorrectly} \\ \text{Fraction diagnosed incorrectly} / \text{Fraction diagnosed correctly} &< 1 \end{aligned}$$

Let D = Diseased and S = Selected ('~' means 'not')

$$\begin{aligned} P(D \ \& \ \sim S) / P(D \ \& \ S) &< 1 && \text{[ Substitute symbols ]} \\ P(D \mid \sim S)P(\sim S) / P(D \mid S) P(S) &< 1 && \text{[ By definition of '\&' ]} \\ P(D \mid \sim S) / P(D \mid S) P(S) &< 1 / P(\sim S) && \text{[ Divide by } P(\sim S) \text{ ]} \\ P(D \mid \sim S) / P(D \mid S) &< P(S) / P(\sim S) && \text{[ Multiply by } P(S) \text{ ]} \end{aligned}$$

In English this can be expressed as:

$$\text{False positive rate} / \text{True positive rate} < \text{Positive base rate} / \text{Negative base rate}$$

We need the ratio of positive to negative base rates to be greater than the ratio of the false positive rate to the true positive rate, if we want to be more likely to be right than wrong.

This can be a handy heuristic because it allows us to calculate the minimum proportion of the population we are working with that needs to be diseased in order for our diagnostic methods to be useful. In the example above, the ratio of false positive to true positive rates is 0.19 / 0.55 or 0.34. This means that the test can only be useful— in the sense of having a positive diagnosis that is more likely to be true than false— when it is used in settings in which the ratio of the maladjusted people (positive base rate) to the number of people who are not maladjusted (negative base rate) is at least 0.34.

Again we can consider another example from Meehl & Rosen (1955). Imagine that you have a test that correctly identifies 80% of brain-damaged patients, but also

misidentifies 15% of non-brain-damaged people. The calculation above says that this test will only be reliable if the ratio of brain-damaged to non-brain-damaged people is greater than  $0.15 / 0.80$ , or about 0.19. If we are using the test in a setting which has a lower ratio of brain damaged people, we will run in to the problem described above, in which we find that the base rates have made it more likely that we are wrong than right when we make a diagnosis.

Note that the requirement given by this heuristic does not mean that the true population base rate must be that high— it is sufficient for the base rate of the subpopulation to which the test is exposed to be high enough. If the test is used in settings (such a mental clinic to which front-line physicians refer) that have ‘higher concentration’ of maladjusted subjects than the general population as a result on non-random sampling of that population, then the test may be useful in that setting, even though it would not be reliable if subjects were *randomly* selected from the population as a whole.

This ability to skew true diagnosis rates in a favorable direction by pre-selecting subjects has important implications. In most of the examples we have considered so far, we have assumed low base rates. The implications of a conditional clause, such as a the probability of that a person has a disease given a positive tests results, become more severe as the base rates moves away from 0.5. The further the base rate is from 50/50, the further it takes the posterior probability  $P(A | B)$  from the simple ‘hit rate’, given by taking the ratio of the true positive rate to the positive diagnoses rate (the sum of the true and false positive rate).

Mathematically, we can see this by expanding the canonical form of Bayes' Rule



given above, just as we did with the example of the maladjusted soldiers above:

Let  $P(C)$  = Probability of belonging to the diagnostic category

Let  $TP$  = True positive rate =  $P(C \text{ \& } \text{Diagnosed})$

Let  $FP$  = False positive rate =  $P(\sim C \text{ \& } \text{Diagnosed})$

Let  $B$  = Base rate of the diagnostic category

Let  $P(D)$  = Probability of being diagnosed as being maladjusted.

= Probability of true positive + probability of false positive

= (True positive base rate x Percent correctly identified) + (Negative Base Rate x Percent incorrectly identified)

=  $(B * TP) + ((1 - B) * FP)$

$P(C | D)$  = Probability of belonging to the category given diagnosis

=  $P(D | C)P(C) / P(D)$  [ Bayes' Rule ]

=  $(TP * B) / (B * TP) + ((1 - B) * FP)$  [ Substitute  $P(D)$  ]

=  $(TP * 0.5) / (0.5 * TP) + (0.5 * FP)$  [ Let the base rate  $B = 0.5$  ]

=  $TP / TP + FP$  [ Divide by 0.5 ]

This is a degenerate case of Bayes' Rule, since the conditional collapses to the simple unconditional probability that is given by the ratio of the probability of getting diagnosed correctly to the probability of getting diagnosed at all, whether correctly or not. One way of understanding what is happening in this case is to note that the true and false positive rates are sampling equally from the population. When this is so, we don't need to bother to 'weight' their respective contributions to the conditional probability of belonging to the category given a diagnosis.

A concrete example may make this interpretation more clear. Consider the conditional probability of having blue eyes, given that you are female. Since eye color is not a sex-linked character, the conditional is the same for both those who are in the group of interest (females) and those who are not (males). You may be able to intuit in this case that the conditional is therefore irrelevant: that is, the probability of being blue-eyed given that you are female is just the same as the probability of being blue-eyed.

This degenerate case of exactly equal base rates with and without the character of

interest may occur only rarely, but the general principle illustrated by this case is of wider relevance for the reason note above: the further the positive and negative base rates are from being equal, the greater the difference between the conditional probability that depends on that base rate and the simple probability given by the ratio of the probability of getting diagnosed correctly to the probability of getting diagnosed at all (that is, the ratio of the true positives to the sum of the true and false positives).

Intuitively, this makes sense for the following reasons. Insofar as a disease is less common, it becomes more likely that a larger portion of the positives are false positives, as in the case considered above that bamboozled so many of the Harvard medical students. By the same token, insofar as a disease is more common, it becomes more likely that many of the negative diagnoses are false. At some point as base rates increase, they may come to exceed the ability of the test to identify them, rendering the test worse than guessing, as discussed above.

Bayes' Rule may be easily generalized to incorporate multiple pieces of evidence bearing on a single belief, hypothesis, or diagnosis, or to incorporate multiple pieces of evidence bearing on multiple beliefs, hypotheses, or diagnoses.

The simplest way to 'extend' Bayes' Rule is to note that the posterior probability may depend on more than one piece of evidence. This is not an extension at all, since we noted at the beginning what was given in a conditional may be a set of evidence rather than a single piece. However, it is worth emphasizing this point, since so many of the examples considered in this paper have treated the conditional as a single piece of evidence. Given a belief, hypothesis, or diagnosis  $H$ , and a single relevant piece of evidence  $E1$ , we have seen how to compute some new probability  $P(H | E1)$ . If we get a

new piece of relevant evidence  $E_2$ , that is independent from  $E_1$ , we could as easily calculate  $P(H | E_2)$  for the same  $H$ . However, that calculation would not take into account the fact that we already attached a certain level of probability to  $H$  because of the prior evidence  $A$ . To get that, we need to calculate  $P(H | E_1 \& E_2)$ .

For example, imagining trying to guess a single card from a deck. If you know it is red, then you have  $P(\text{Guess} | \text{Red}) = 1/26$ , because there are 26 red cards in a deck. If you know it is a face card, you have  $P(\text{Guess} | \text{Face}) = 4/13$ , because there are four face cards per suit of 13 cards. If you know it is both a face card and red, you need to calculate  $P(\text{Guess} | (\text{Face} \& \text{Red})) = 1/8$ , because there are eight cards that are both red and a face card.

A slightly more complex way of generalizing Bayes' Rule comes about when there is more than one competing hypothesis, diagnosis, or possibility to be considered. In that case, evidence brought to bear in favor of any single hypothesis needs to be considered in the context of the domain of all other competing hypotheses. In fact the simple forms of Bayes' Rule we have considered in this paper does exactly this. We have seen that  $P(H | E) = P(E | H) P(H) / P(E)$ , where  $H$  is some hypothesis, diagnosis, or possibility, and  $E$  is some evidence bearing on it. We have also seen in several examples that the denominator  $P(E)$ — to be concrete, the probability of getting a positive diagnosis— can be expanded into sum of (the true positive rate \* the positive base rate) and (the false positive rate \* the negative base rate). The two elements in this sum are just two different hypotheses about where a positive diagnosis could have come from: it could either have come from a mistaken diagnosis or a true diagnosis. If there was also a possibility of a deliberately fraudulent diagnosis, we would have to add that in to our calculation of the

probability of getting a positive diagnosis, as a third term in  $P(E)$ .

The generalization of Bayes' Rule to handle any number of competing hypotheses simply makes explicit that the denominator in Bayes' Rule is the domain of possible kinds of evidence that could explain  $H$ - or said another way, the domain of possible ways the evidence under consideration could come about. The generalized expression is:

$$P(H_n | E) = P(E | H_n)P(H_n) / \sum [P(E | H_{n-1}) P(H_{n-1})]$$

$H_n$  is a current hypothesis, and  $E$  is, as ever, some new piece of evidence, such as a diagnostic sign. The denominator, as above in the specific cases we have considered, is simply the sum of all ways the diagnostic sign might occur, howsoever that may be.

### *Conclusion*

The goal of this paper has been to introduce conditional probability in general, and Bayes' Rule in specific, in a manner that is both comprehensive and accessible. If I have succeeded, then  $P(\text{Understanding Bayes' Rule} | \text{Reading this paper})$  will be high. I hope that it is.

Hacking, I. (2001). *An introduction to probability and deductive logic*. Cambridge, England: Cambridge University Press.

Meehl, P. & Rosen A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52(3):194-216. As reprinted in: Meehl, P. (1977). *Psychodiagnosis: Selected Papers*. New York, USA: W.W. Norton & Sons.