Journal of the California Mathematics Project
Volume **6**, 2013

Counting on Bayes' Theorem, or "Back to the Future"

Angelo Segalla, Angelo.Segalla@csulb.edu
and Yonghee Kim-Park, Yonghee.KimPark@csulb.edu

Dear Sir, I now send you an essay found among the papers of our deceased friend Mr. Bayes, and which, in my opinion has great merit and well deserves to be preserved.
A letter to John Bowlton, dated December 23, 1763 from Richard Price.

Abstract. The main thrust of this article is that Bayes' Theorem becomes plausible for high school AP Statistics classes since students are generally uneasy with conditional probability. A "Back to the Future" metaphor demonstrates the cleverness of the theorem, which never fails to surprise students. A simple example, followed by visuals that are quasi-proofs will hopefully enhance the conceptual basis of this important theorem. Finally, suggestions of more sociologically significant examples illustrate the theorem's statistical power.

1. Introduction

Bayes' (1701-1761) work "An Essay towards solving a Problem in the Doctrine of Chances" was published posthumously in 1763 by his friend Richard Price (see quotation). Also, Joseph-Louis Lagrange (1736-1813, born in Turin, Italy as Giuseppe Luigi Lagrancia), unaware of Bayes' publication, worked on the same topic and extended the theory in an essay of his own in 1774.

Bayes' Theorem, from a counting point of view, is a sensible approach for introducing probabilities in high school: first as fractions with decimals, then as fractions with a common denominator, and finally with problems whose probabilities are proper fractions with (usually) different denominators. This approach provides a bonus for high school students. It is a conceptual framework for understanding fractions through a systematic review of their properties (an often neglected skill in school mathematics).

We have found that Bayes' Theorem works nicely in high school AP Statistics classes. Our approach also reinforce students' facility with those rational numbers

©2013 Author

that are ordinarily called "proper fractions" (with different denominators), that is, fractions we get from working out probability problems, then as fractions with the same (common) denominator, then as decimals, and finally as percents. For example, the probability at least one "heads" on a flip of three honest coins is 7/8 or 0.875 or 87.5%.

Most introductory statistics textbooks have a section on Bayes' Theorem, usually following a discussion of conditional probability, and illustrate by examples of how *a priori* information about a compound event (an earlier event in a tandem of a sequence of two events) will change the calculation of the probability of a set of outcomes, often dramatically.

2. A Simple Experiment

Consider two bowls, Bowl A, and Bowl B. Bowl A contains 10 red marbles and 30 green marbles for a total of 40 marbles. Bowl B contains 20 red marbles and 20 green marbles also for a total of 40 marbles. Grand total: 80 marbles. All marbles are indistinguishable by touch. The list below summarizes the experiment.

STEP 1: Choose one of the two bowls at random, say by flipping a fair coin; the identity of the chosen bowl is not revealed.
STEP 2: Also randomly, choose a single marble from the given bowl.
STEP 3: The marble is green. This we know.
STEP 4: Find the probability that the green marble came from Bowl A. That is that Bowl A was the one chosen in STEP 1.

This problem, starting with the concluding event and recovering the antecedent event, forces us to "think backwards" to put it informally. Therefore, our subtitle "Back to the Future" as in the movie. Intuition would suggest that the marble most likely came from Bowl A since that bowl contains more green marbles than Bowl B. What percent of the time would our intuition be correct?

Symbolically, let A be the event of picking Bowl A, B be the event of picking Bowl B, G the event of picking a green marble, and R the event of picking a red marble.

In this example choose Bowl A or Bowl B with equal probability. So here $Pr(A) = \frac{1}{2} = 0.5 = 50\%$ and $Pr(B) = \frac{1}{2} = 0.5 = 50\%$. Let $Pr(G)$ and $Pr(R)$ be the probability of picking a green marble or red marble respectively (whether it be from Bowl A or B. We seek the probability that we picked Bowl A, given that the marble we have is green, G. Symbolically, we seek $Pr(A|G)$.

Now, reversing A and G (correcting the order), $Pr(G|A)$ represents the probability of picking a green marble given that we actually chose Bowl A. Since in our case there is an equal number of marbles in each bowl (a special condition to be sure), the combined number of marbles can be used to find the total probability of each color marble. This and other information are summarized in Table 1 where it is easily seen that $Pr(G|A) = \frac{30}{40} = \frac{3}{4} = 0.75 = 75\%$.

Table 1. Summary of probability using fractions, decimals, and percents.

	Bowl A	Bowl B	Total
Red marbles	$10/40 = 0.25 = 25\%$	$20/40 = 0.50 = 50\%$	$30/80 = 0.375 = 37.5\%$
Green marbles	$30/40 = 0.75 = 75\%$	$20/40 = 0.50 = 50\%$	$50/80 = 0.625 = 62.5\%$
Totals	$40/40 = 1.00 = 100\%$	$40/40 = 1.00 = 100\%$	$80/80 = 1.000 = 100\%$

Taking some liberty with probability theory, to be in the cell where "$30/40 = 0.75 = 75\%$" we must have chosen Bowl A, which occurs with $Pr(A) = 0.5$, and, since the ball is green, the probability that that green ball came from the entire collection of green balls is $(0.75)/0.625$. Putting all this together (again, informally) the probability of having chosen Bowl A given that the marble is green is $((0.75)(0.5))/0.625 = 0.6 = 60\%$.

3. Bayes' Theorem

Bayes' Theorem (Rule) tells us how to compute the *a posteriori* probabilities when we know information ahead of time, the antecedent. The probability of a hypothesis X in light of a piece of new evidence, Y, is

$$Pr(X|Y) = \frac{Pr(Y|X)Pr(X)}{Pr(Y)}$$

In our example:

$$Pr(A|G) = \frac{Pr(G|A)Pr(A)}{Pr(G)} = \frac{(0.75)(0.5)}{0.625} = 0.60 = 60\%$$

Why is this so? In Figure 1 we use Venn diagrams to describe our experiment visually. We take the interior of the rectangle to be the sample space S and its area to be equal to one. The next four diagrams illustrate the probability (area) that an event in sets X, Y, $X \cup Y$, and $X \cap Y$, respectively, occurs: $Pr(X)$, $Pr(Y)$, $Pr(X \cup Y)$, and $Pr(X \cup Y)$.

The last two diagrams are pertinent to our example. They demonstrate that the area of the intersection of X and Y, $(X \cap Y)$, does not change whether we are given that event X has occurred or event Y has occurred. This is key to proving Bayes' Theorem. This fact and some assumptions about conditional probability (which can be defined and proved) illustrate the answers to the questions we ask (and others we can ask) about our experiment. That is, the last two pictures demonstrate what happens if we are told, *a priori*, that "X has happened" so that all the points in the sample set S except for those in set X, including those in set Y are now excluded! Set X is now the new (total) sample space. Similarly for "Y has happened."

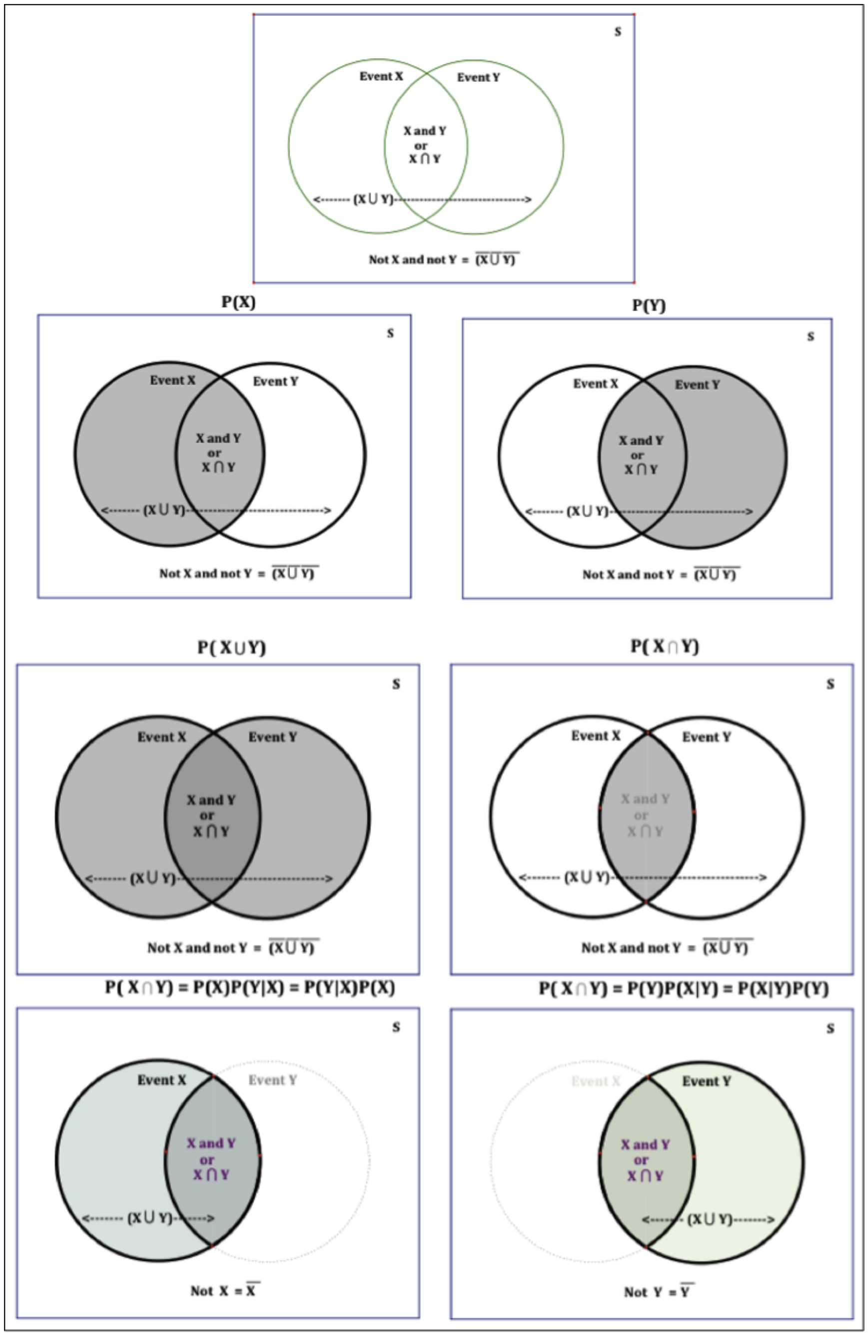

FIGURE 1. Venn diagrams for the sample space S, and sets X, Y, $X \cup Y$, and $X \cap Y$.

4. Proof of the Theorem

Now note in the diagrams that the area (probability) of $(X \cap Y)$ does not change if we are told that X has happened or Y has happened.

This hints that we can write the following if X happened

$$Pr(Y|X) = \frac{Pr(X \cap Y)}{Pr(X)}$$

and if Y happened,

$$Pr(X|Y) = \frac{Pr(Y \cap X)}{Pr(Y)}$$

But

$$Pr(X \cap Y) = Pr(Y \cap X)$$

So

$$Pr(Y|X)Pr(X) = Pr(Y \cap X) = Pr(X \cap Y) = Pr(X|Y)Pr(Y)$$

And

$$Pr(Y|X) = \frac{Pr(X|Y)Pr(Y)}{Pr(X)}$$

With the proper exchange of letters, the solution to the problem of the marbles is:

$$Pr(A|G) = \frac{Pr(G|A)Pr(A)}{Pr(G)} = \frac{(0.75)(0.5)}{0.625} = 0.60 = 60\%$$

And this backs up our intuition – but not overwhelmingly!

5. Trees

Yet another visually appealing method for simple probability problems, and simple Bayes' Theorem problems, that we found to work well in high school classes is illustrated in Figure 2 using a tree diagram. The answer to the question in our experiment (60%) is in the tree diagram, but needs to be interpreted. We leave this interpretation open for the reader with the hint that $37.5 + 25 = 62.5$.

Also note a tree diagram can elicit interesting questions to be posed by the teacher. For example, the sum of all the percents at the right of the tree must be $100\% = 1$, the probability of the entire sample space S.

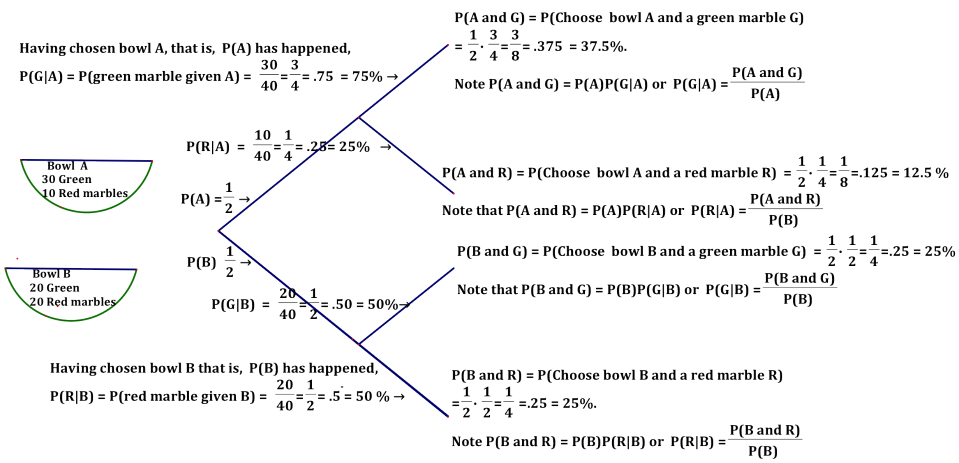

FIGURE 2. Tree probabilities of fractions, decimals, and percents.

6. Conclusion

The power of Bayes' Theorem can be appreciated even further in sociological examples that involve drug testing and profiling. Examples that challenge our intuition abound in the literature. Surprise can be the name of the game, so to speak, in statistics. Consider how on the evening of a presidential Election Day television networks project the winner with less than 1% of the votes counted!

Bayes' Theorem can play an important part in high school students' AP Statistics. Paired with the pedagogical emphasis on the different ways we can express rational numbers, the topic might also be used in other parts of the high school mathematics curriculum.

References

Barnard, G. A. (1958). Studies in the history of probability and statistics: Ix. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 293–295. (Biographical remarks.)

Stigler, S. M. (1982). Thomas Bayes' bayesian inference. *Journal of the Royal Statistical Society, Series A*, *145*, 250–258. (Stigler argues for a revised interpretation of the essay – recommended.)

Stigler, S. M. (1983). Who discovered Bayes' Theorem? *The American Statistician*, *37*(4), 290–296.

Stigler, S. M. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*(3), 359–378.