On The Confusion in Some Popular Probability Problems

Nikunj C. Oza
March, 1993

Abstract. In this paper, we will look at three probability problems that have caused widespread disagreement and much confusion. We will look at some false solutions to these problems, the reasons why they are incorrect, and the correct solutions. We will also find a common pattern in the three problems which probably accounts for these false solutions.

1. Introduction. In a series of articles entitled “Ask Marilyn” in *Parade Magazine* [**1**], an interesting probability problem was discussed. The problem, often called the “Let’s Make a Deal” Problem, was stated in the first article [**1a**]: “Suppose you’re on a game show and given a choice of three doors. Behind one is a car; behind the other two are goats. You pick Door No. 1, and the host, who knows what’s behind them, opens No. 3, which has a goat. He then asks if you want to pick No. 2. Should you switch?” Marilyn vos Savant, who is in the “Guinness Book of World Records Hall of Fame” for “Highest IQ,” and author of the “Ask Marilyn” articles, replied [**1a**], “Yes, you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance.” In a subsequent *Parade Magazine* article (see [**1b**]), there were letters from three Ph.D.’s insisting that Marilyn’s answer was incorrect. Two of the three people claimed that the probability of winning the car no matter which door you pick is 1/2.

In a class called “Principles of Applied Mathematics I” held at the

Massachusetts Institute of Technology (M.I.T.) in the fall of 1991, Professor James G. Propp asked his students a question similar to the game show problem: "You know that a family has two children and that one of them is a boy. What is the probability that the other child is a boy?" The result was an argument that was, at times, heated. Just as in the *Parade Magazine* articles, the two sides in the argument advocated two different answers: 1/2 and 1/3.

Martin Gardner sent a probability problem to Marilyn vos Savant. It was printed in an "Ask Marilyn" article in *Parade Magazine* [**3**]:

> The Greens and the Blacks are playing bridge. After a deal, Mr. Brown, an onlooker, asks Mrs. Black: "Do you have an ace in your hand?" She nods. There is a certain probability that her hand holds at least one other ace.
>
> After the next deal, he asks her: "Do you have the ace of spades?" She nods. Again, there is a certain probability that her hand holds at least one other ace.
>
> Which probability is greater? Or are they both the same?

Marilyn vos Savant, in the same article, gave an answer that is, at first, counterintuitive: "Mrs. Black's second hand--the one with the ace of spades--is more likely to have another ace..." The initially intuitive answer, which many of my friends gave, is that both hands have equal probability of having at least one additional ace.

All three of the above problems have two main things in common: All three have answers that are, at first, counterintuitive, and all three have frequently given wrong answers. In Sections 2 through 4, we will discuss these three problems individually. In Section 2, we will discuss the "Let's Make a Deal" Problem in the form that Marilyn dealt with it, by giving a frequently given wrong solution, explaining why it is wrong, and giving the correct answer. (In an article [**4**] entitled "Let's Make a Deal: The

Player's Dilemma," the authors point out that Marilyn vos Savant actually solved a slightly different problem from the one she gave in the first *Parade Magazine* article. In this paper, however, we will only look at the problem Marilyn actually solved.) In Section 3, we will discuss the "Boy/Girl" Problem. In Section 4, we will look at the Bridge Problem. The Bridge Problem actually has a twist to it that makes it more difficult than the other two problems, but one error that many people make when they attempt to solve it is quite similar to the error they make when attempting to solve the other two problems. All three problems have a common trap that leads to incorrect solutions: incorrect specification of the sample space. In the appendix, we will briefly discuss the notions of conditional probability and sample spaces as they relate to the problems given in Sections 2 through 4 of this paper. More detailed explanations can be found in *A First Course in Probability* by Sheldon Ross [**2**].

2. The "Let's Make a Deal" Problem. In an article [**4**] called "Let's Make a Deal: The Player's Dilemma," the authors distinguish the problem that Marilyn vos Savant gave in her *Parade Magazine* article (the "conditional problem") from the one Marilyn vos Savant actually solved, which they call "the unconditional problem." The unconditional problem is the following: "You will be offered the choice of three doors, and after you choose the host will open a different door, revealing a goat. What is the probability that you win if your strategy is to switch?" In the *Parade Magazine* article, Marilyn vos Savant specified which door the contestant picks and which door the host chooses to open. In the "unconditional problem," no such specification is given. We will consider only the latter "unconditional problem" because it is closely related to the other two problems that we will discuss in this paper.

> Here is an incorrect solution:
>
> The sample space as of the beginning of the problem is {AGG, GAG, GGA} where AGG, for example, means that there is an Auto behind Door 1, a Goat behind Door 2, and a Goat behind Door 3. However, since the game show host opens a door that has a goat behind it, the original sample point indicating that a car is behind the door that the host opened is no longer a part of the sample space (for example, if the host opens Door 3 revealing a goat, then GGA would no longer be part of the sample space since GGA indicates that an auto is behind Door 3). The reduced sample space is now {AGG, GAG}.The car could be behind one of the two remaining doors. Therefore, the probability of winning by opening either door is 1/2. “Should you switch?” is, therefore, an unnecessary question, since it does not matter which door the person picks.

The above answer is incorrect because the sample space is incorrectly specified. For convenience, we will assume that the contestant picks Door 1, but our explanation can easily be generalized to the case where the contestant picks any door. When the contestant picks Door 1, the correct sample space is {AGG2, AGG3, GAG3, GGA2} where, for example, AGG3 means that there is an Auto behind Door 1, Goat behind Door 2, Goat behind Door 3, and the game show host opens Door 3. We want the probability that the car is not behind Door 1. In other words, we want the probability that the contestant should switch. We will first calculate the probability that the car is behind Door 1. This probability subtracted from one will give the probability that the car is not behind Door 1. Conditioning on the door that the game show host opens, we obtain

P{car is behind Door 1} =

P{(car is behind Door 1) | (host opens Door 2)}*P{host opens Door 2} +

P{(car is behind Door 1) | (host opens Door 3)}*P{host opens Door 3}.

We will assume that, when the car is behind Door 1, the host is as likely to open Door 2 as Door 3. We can rewrite the above expression in terms of the sample points given above as well as our assumption:

$$\frac{P\{AGG2\}}{P\{AGG2\} + P\{GGA2\}}\left(\frac{1}{2}\right) + \frac{P\{AGG3\}}{P\{AGG3\} + P\{GAG3\}}\left(\frac{1}{2}\right)$$

where

P{AGG3} = P{auto is behind Door 1}*P{host opens Door 3} = (1/3)*(1/2)
= 1/6,

P{AGG2} = P{auto is behind Door 1}*P{host opens Door 2} = (1/3)*(1/2)
= 1/6,

P{GGA2} = P{auto is behind Door 3}*P{host opens Door 2} = (1/3)*1
= 1/3,

P{GAG3} = P{auto is behind Door 2}*P{host opens Door 3} = (1/3)*1
= 1/3.

Now we can calculate the probability that we wanted:

$$\frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}}\left(\frac{1}{2}\right)+\frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}}\left(\frac{1}{2}\right)=\frac{1}{3}.$$

Note that 1/3 is the probability that the car is behind Door 1. The probability that the car is not behind Door 1 is (1 - 1/3) = 2/3. Hence the probability that the contestant wins by switching is 2/3. Note that the reduced sample space given in the incorrect solution, {AGG, GAG}, is incorrect because the sample point AGG includes both AGG2 (the host opens Door 2) and AGG3 (the host opens Door 3). The correct reduced sample space is {AGG3, GAG3} if the host opens Door 3. Equivalently, if the host opens Door 2, then the correct reduced sample space would be {AGG2, GGA2} and not {AGG, GGA} as the person writing the incorrect solution would probably have assumed.

There is a simpler way to think about the problem that makes the above answer more intuitive. There is a 1/3 chance that the contestant picks the door that has a car behind it, hence there is a 1/3 chance that the contestant should stay with his original choice. Because the only other possibility is that the contestant switches to win the car, winning by switching must have probability 1 - 1/3 = 2/3.

3. The “Boy/Girl” Problem. The following is the “Boy/Girl” Problem: “You know that a family has two children and that one of them is a boy. What is the probability that the other child is a boy?” Just as in the “Let’s Make a Deal” Problem, most incorrect solutions to the “Boy/Girl” Problem are the result of incorrect specification of the sample space.

The following solution is incorrect:

> You are told that one child is a boy. There are obviously only two possibilities for the other child: boy and girl. Each possibility has an equal probability of occurring; therefore, the probability that the second child is a boy is 1/2.

This statement's author correctly assumes that the original sample space, {(b,b),(b,g),(g,g)}, is reduced to {(b,b),(b,g)} where, for example, (b,g) means that one child is a boy and the other child is a girl. However, the author incorrectly assumes that "Each possibility [(b,b), (b,g)] has an equal probability of occurring." The possibility (b,g) is twice as likely to occur as the possibility (b,b): recall that (b,g) means that either the first child is a boy and the second child is a girl or the first child is a girl and the second child is a boy. Many people, because they do not carefully read the problem, assume that the first child is known to be a boy and reason that the second child is as likely to be a girl as it is to be a boy. In effect, they fail to realize that (b,g) contains two possibilities. If one child is a boy, then the other child is twice as likely to be a girl as it is to be a boy. Therefore, the probability that the other child is a boy given that one child is a boy is 1/3.

4. The Bridge Problem. As in both of the previous problems, improper specification of the sample space causes some of the errors. In addition, in the Bridge Problem, errors also result from an underestimation of the complexity of the problem. From now on, we will use P_{ace} and P_{spade} to signify the two probabilities of the Bridge Problem. Here are the definitions of P_{ace} and P_{spade}.

$$P_{ace} = P\{\text{hand has at least one more ace} \mid \text{hand has at least one ace}\};$$

$$P_{spade} = P\{\text{hand has at least one more ace} \mid \text{hand includes the ace of spades}\}.$$

Here is an incorrect idea:

> There is nothing special about a bridge hand having the ace of spades. A person playing bridge is as likely to have the ace of spades as to have any of the other three aces. Therefore, the definitions of P_{ace} and P_{spade} are equivalent and the two probabilities must be the same.

People who make the above statement simply do not realize the complexity of this problem.

The solution that Marilyn gave in *Parade Magazine* [**3**] also has a problem with it:

> ...There are fewer opportunities to get a particular ace than there are to get any ace at all. But each of these groups of opportunities contains an equal number of “golden opportunities” to get more aces. Therefore, the smaller group provides the greater chance of success.

Marilyn specified the reduced sample space correctly (the sample space of bridge hands including the ace of spades is smaller than the sample space of hands containing at least one ace); however, her statement “...each of these [sample spaces] contains an equal number of “golden opportunities” to get more aces” is a naive statement as we will show below. We will now analyze this problem more deeply in order to solve it. First, note that a bridge hand having at least one more ace is the same as a hand having at least 2 aces. From now on, A_n denotes the bridge hand having at least n aces (of course, $0 <= n <= 4$) and A_S denotes the bridge hand including the ace of spades. Now, the equations for P_{ace} and P_{spade} can be rewritten as follows:

$$P_{ace} = P\{A_2 | A_1\} = \frac{P\{A_2 A_1\}}{P\{A_1\}} = \frac{P\{A_2\}}{P\{A_1\}};$$

$$P_{spade} = P\{A_2 | A_s\} = \frac{P\{A_2 A_s\}}{P\{A_s\}}.$$

As we will show below, $P\{A_s\} < P\{A_1\}$, supporting Marilyn's description of the reduced sample space; however, $P\{A_2 A_s\} < P\{A_2\}$, which shows that her statement about "equal number of "golden opportunities" to get more aces" does not apply here. We can calculate each of the probabilities in the above expressions separately. Here are the calculations:

$$P\{A_2\} = \frac{\binom{4}{2}\binom{48}{11} + \binom{4}{3}\binom{48}{10} + \binom{4}{4}\binom{48}{9}}{\binom{52}{13}};$$

$$P\{A_1\} = \frac{\binom{4}{1}\binom{48}{12} + \binom{4}{2}\binom{48}{11} + \binom{4}{3}\binom{48}{10} + \binom{4}{4}\binom{48}{9}}{\binom{52}{13}};$$

$$P\{A_2 A_s\} = \frac{\binom{3}{1}\binom{48}{11} + \binom{3}{2}\binom{48}{10} + \binom{3}{3}\binom{48}{9}}{\binom{52}{13}};$$

$$P\{A_S\} = \frac{\binom{3}{0}\binom{48}{12} + \binom{3}{1}\binom{48}{11} + \binom{3}{2}\binom{48}{10} + \binom{3}{3}\binom{48}{9}}{\binom{52}{13}}.$$

The ideas behind the above calculations are relatively simple, and a complete explanation would take a large amount of space, so such an explanation is not included here. After substituting the above calculations in the last equations for P_{ace} and P_{spade}, we get the following results: $P_{ace} = 0.369637$ and $P_{spade} = 0.561152$. As we can see, $P_{spade} > P_{ace}$. Indeed, as we noticed for the other two problems, the answer for the Bridge Problem is counterintuitive when first examined.

APPENDIX: SAMPLE SPACE AND CONDITIONAL PROBABILITY

A *sample space* is a set consisting of all outcomes of an event or group of events. For example, in the “Boy/Girl” Problem, the event was having two children. The three possible outcomes of this event are (b,b),(b,g), and (g,g), where the first and second letters in every pair specify the genders of the two children. The set containing those three possible outcomes, {(b,b),(b,g),(g,g)}, is called the sample space.

Conditional probability is the probability that one event occurs conditional upon the occurrence of some other event. If E and F are two events, then P{E | F} (read, “the probability of E given F”) is the probability that E occurs given that F has occurred. An important formula used in calculating conditional probabilities is the following:

$$P\{E \mid F\} = \frac{P\{EF\}}{P\{F\}}$$

if P{F} > 0. This formula states that P{E | F} is the answer to the question: Out of the events in which F occurs, in what proportion does E also occur? For this reason, people often calculate conditional probabilities by using the idea of a reduced sample space: the original sample space is reduced to include only those sample points in which F occurs. Given the new sample space, the proportion of events in which E occurs is P{E | F}.

We will use the “Boy/Girl” Problem again as an example. In that problem, we wanted to calculate P{(other child is a boy) | (one child is a boy)}. We can calculate this probability by using either of the two methods given above: the method of conditional probabilities or the method involving reduced sample spaces. Using the first method, we get

$$P\{\text{(other child is a boy)} \mid \text{(one child is a boy)}\} = \frac{P\{\text{one child and other child are boys}\}}{P\{\text{one child is a boy}\}}.$$

In terms of the original sample space, this probability is

$$\frac{P\{(b,b)\}}{P\{(b,b),(b,g),(g,b)\}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Using the second method given above, we can reduce the original sample space to only those points where one child is a boy: {(b,b),(b,g)}. Out of these, there is only one sample point where the other child is a boy and the other sample point actually contains two possibilities within it (the two possible orderings of the boy and girl), therefore, the probability that the other child is a boy is 1/3.

In Section 2, we use the notion of *conditioning* to calculate the probability that the car is behind Door 1. Conditioning is a technique involving use of the following formula known as Bayes' formula:

$$P\{E\} = P\{E \mid F\} P\{F\} + P\{E \mid F^c\} P\{F^c\}.$$

Conditioning allows us to calculate probabilities based upon whether or not some other event has occurred. This is significant because there are problems in which it is difficult to compute the probability of an event directly, but it is relatively easy to calculate the probability once we know whether some other event has occurred or not. For example, in the "Let's Make a Deal" Problem, once we knew which door the host opened, calculating the probability that the car was behind Door 1 was simple.

REFERENCES

[1a] Marilyn vos Savant, “Ask Marilyn,” *Parade Magazine*, September 9, 1990, p.15.

[1b] Marilyn vos Savant, “Ask Marilyn,” *Parade Magazine*, December 2, 1990, p. 25.

[2] Sheldon Ross, *A First Course in Probability*, Third edition, New York; MacMillan Publishing Company, 1988.

[3] Marilyn vos Savant, “Ask Marilyn,” *Parade Magazine*, September 8, 1991, p. 7.

[4] J.P. Morgan, et al, “Let’s Make a Deal,” *The American Statistician*, November 1991, pp. 284-287.