**Puzzling Probabilities of Probability Puzzles**

Kicab Castaneda-Mendez, kicabcm@yahoo.com, 878 Fearrington Post, Pittsboro, NC 27312

Puzzles are used to teach probability in an entertaining way. Often, they are nuanced or counter-intuitive for the purpose of a deeper understanding of probability theory. Solutions can be derived using axiomatic probability theory. But for counter-intuitive answers, relative frequency probability theory and empirical frequency probability theory are used to convince others of and confirm the answers, whether in class or online discussion of the answers to these types of puzzles.

Unfortunately, numerous answers are incorrect and misteach probability theory when the puzzle does not include the required information or the puzzle is about past events. The following dialogue between a student (S) and teacher (T) serve as a model for the classroom on how to teach probability theory using these puzzles in activities. This results in 12 lessons to correct these types of puzzle statements, derive correct answers, and avoid creating faulty puzzles.

#### **Probability Distribution Requirement**

Teacher: We are going to learn about probability theory using probability puzzles. You may have come across probability puzzles online. Let’s start with this puzzle: What’s the probability of getting heads with a toss of a coin?

Student: Easy—1/2.

T: No. It’s 0. The coin has tails on both sides.

S: But you didn’t say that.

T: I also didn’t say it was a head-tail sided coin with each equally likely. I could have said 1 is the probability because both sides are heads.

S: How is the puzzle solver supposed to know?

Lesson 1: A probability distribution (sample space and their probabilities) is required to calculate a probability.

Lesson 2: If the puzzle is to have one correct answer, it must state which events are random and provide their probability distribution.

T: Apply Lessons 1 and 2 to these two versions of the Same Birthday Problem where both have the answer n = 23:

Scenario 1. “How many people [n] must be gathered together in a room, before you can be certain that there is a greater than 50/50 chance that at least two of them have the same birthday?” (https://gpuzzles.com/mind-teasers/chances-of-having-same-birthday-problem)

Scenario 2: “An entertaining example is to determine the probability that in a randomly selected group of n people at least two have the same birthday. If one assumes for simplicity that a year contains 365 days and that each day is equally likely to be the birthday of a randomly selected person, then in a group of n people there are possible combinations of birthdays. The simplest solution is to determine the probability of no matching birthdays and then subtract this probability from 1.” (https://www.britannica.com/science/probability-theory/The-birthday-problem)

S: Scenario 1 doesn’t say if any events are random and doesn’t include any probability distribution. So, there are really many answers. Does leap year change the answer? A gathering of people could be all the babies born on January 1 at one maternity ward. Isn’t S1 mixing the gathering of people even if random with the random selection of days of the year?

T: Yes, it is. This puzzle is often erroneously used with any group of people.

S: I like Scenario 2. It provides all the information to get the correct answer.

T: How about the Four Balls in a Bowl Problem? (https://www.riddlesandanswers.com/v/234771/this-is-a-famous-paradox-probability-riddle-which-has-caused-a-great-deal-of-argument-and-disbelief):

Four balls are placed in a bowl. One is Green, one is Black and the other two are Yellow. The bowl is shaken and someone draws two balls from the bowl. He looks at the two balls and announces that at least one of them is Yellow. What are the chances that the other ball he has drawn is also Yellow? Puzzle answer: 1/5.

S: There is no probability distribution included so there are really many possible answers each conditional on what is assumed.

T: Unfortunately, this violation of Lessons 1-2 occurs often. as we shall see.

#### **Past Event Probability**

T: One principle of probability puzzles is that there is only one correct answer that applies to everyone. Let’s do an exercise. This is a head-tail sided coin with each side equally likely. Close your eyes while I toss the coin. Now what’s the probability the outcome is heads?

S: This time you gave the probability distribution, so it’s 1/2.

T: Let’s reverse roles. Here’s a pencil and paper. I will close my eyes as you toss the coin 10 times and record the outcome and the probability of heads. Done? Now I don’t know the outcome, so I should answer as you did when you didn’t know: 1/2 for each toss. True?

S: Well, no. The probabilities were six 1s and four 0s; none was 1/2.

T: The relative frequency across the five results is 6/10. Why isn’t that approximately the probability of heads?

S: No, because you asked about each toss. Besides, I know what the outcomes were and those probabilities apply to everyone if there is only one correct probability.

T: Exactly. The correct probability is based on the fact of the outcome, not whether one knows the outcome. When you didn’t know the outcome, the probability of heads wasn’t 1/2 but 0 or 1.

Lesson 3: For any outcome A whether known or not and past event E, the probability A was/is the outcome of E is 0 or 1. To know which requires knowing the outcome.

Analyze the Second Girl Child Puzzle (https://www.briddles.com/2011/07/probability-of-second-girl-child.html).

A married couple…(has) two kids, one of them is a girl. Assume safely that the probability of each gender is 1/2. What is the probability that the other kid is also a girl? Puzzle answer: 1/3

S: Finally, a probability distribution. But this doesn’t make sense. The parents know if they have two girls. So, they know which of 0 or 1 is the correct probability—for them and everyone.

T: Good. What about the Three Prisoners Problem (Gardner 1959, “Mathematical Games: Problems involving questions of probability and ambiguity,” *Scientific American*, 201, 4, 174–182)?

The governor selects one of three prisoners to be pardoned and tells the warden who it is. The warden informs three prisoners but won’t say who it is pardoned. Since the warden won’t say who is pardoned, Prisoner A privately asks him to tell him which prisoner B or C is to be executed. The warden says B. Prisoner A thinks his probability of being pardoned increased from 1/3 to 1/2.

S: The same thing. The governor and warden know who was pardoned, so the correct probability for everyone is 0 or 1—they know which it is. But what if no one knows the outcome?

T: Let’s see using the Two-Heads Puzzle:

You have two head-tail sided coins with each side equally likely. The probability of getting two heads by tossing the coins consecutively is P(2 heads) = P(heads 1)\*P(heads 2) = 1/2\*1/2 = 1/4. After the first coin is tossed what is the probability of getting two heads? Puzzle Answer: 1/4.

S: I would not have answered 1/4. If the first toss is tails, then you can’t get two heads so the probability is 0. And if the first toss is heads, then only if the second toss is heads can you get two heads. That probability is 1/2. So, 0 or 1/2 is the probability. What am I doing wrong?

T: Nothing. The reason I put 1/4 as the answer is because when I asked you what was the probability the outcome of a coin toss *was* heads when you didn’t know the result, you said 1/2. If that were true, then even after the first toss, you should have said it’s still 1/4 for two heads. How would you show someone the probability 1/4 is wrong?

S: I’d say I’ll toss the first coin so that you can’t see the outcome. Then, toss the second coin multiple times and record the results like I just did. We’d repeat that sequence until both heads and tails have occurred on the first toss. Then for each set we’d calculate the frequency of two heads. But they should quickly see if the first toss is tails, they don’t need to toss the second coin. If the first toss is heads, they’ll see that heads will occur on the second toss about half the time. To me it’s obvious the probabilities are 0 or 1/2 just by looking at the outcome of the first toss and knowing the probability distribution of the second toss.

T: Note that when you don’t know the outcome, there are two choices for the probability. But all the puzzles incorrectly provide only one answer. Secondly, if the two choices are 0 and 1/2 but P(heads 2) = 1/2, what is the probability that heads was the outcome of the first toss?

S: If I replace P(heads 2) by 1/2 in the equation, I get the probability of heads 1 is either 0 when tails or 1 when heads. So, this is another way of showing Lesson 3.

T: And, you have confirmed Lesson 3 using three forms of probability: axiomatic probability, relative frequency probability, and empirical frequency probability. Let’s look at each.

*Axiomatic Probability*

There is a simple theorem of axiomatic probability that states P(A|A) = 1 and its corollary P(A|not A) = 0. That is what you used in your reasoning: the probability of heads 1 given heads 1 occurred is 1 and the probability of heads 1 given tails 1 occurred is 0, even not knowing the outcome.

The beauty of axiomatic probability is that none of its axioms, lemmas, theorems, or corollaries depends on anyone’s knowledge of the outcome or the probability distribution.

Lesson 4: Knowledge or lack of knowledge of an outcome is irrelevant to its probability using axiomatic probability theory.

T: Here is the famous Monty Hall Problem (Selvin 1975a, “A problem in probability (letter to the editor),” *The American Statistician*, 29, 1, 67):

A car is placed behind one of three doors. The contestant selects a door. Monty opens one of the other two doors showing no car behind it. To win the car, Monty offers the contestant the option to stay with the original selection or switch to the other unopened door. The contestant asks “Should I stay or switch?” Puzzle answer: The probability of winning by switching is 2/3.

How did Selvin (August 1975b, "On the Monty Hall problem (letter to the editor)," *The American Statistician,* 29, 3, 134) get his 2/3 probability?

S: While no probability distributions are given, he uses 1/3 as the probability the car is behind a specific door and 1/3 as the probability that the contestant chooses a specific door. But those are the probabilities *before* those events occur. *After* the car is placed behind a door and the contestant chooses a door, the correct probabilities are 0 or 1. That makes the probability of winning by switching 0 or 1. Oh, and one more thing—Monty knows which it is.

Lesson 5: Once an event has occurred and its outcome is unknown, do not use the probability distribution before it occurred to calculate probabilities of outcomes after it has occurred.

*Relative Frequency Probability*

T: Right. Next, you used relative frequency probability to recognize that the sample space changes after the event occurs. The sample space for Two Heads Puzzle before the first toss is {head 1, tail 1}. But you reasoned by separating those outcomes. After the toss, there are two sample spaces each with only one outcome: {head 1} and {tail 1}. After the toss, there is only case and it is either favorable, probability 1, or it is unfavorable, probability 0.

Lesson 6: After an event occurs, its possible sample spaces for its outcome are single outcomes, each favorable or unfavorable.

Use this to analyze Four Balls in a Bowl Problem’s 1/5 probability and why is it wrong.

S: The puzzle’s reasoning is that the original sample space containing a yellow ball is {Green-Yellow, Black-Yellow, Yellow-Green, Yellow-Black, Yellow-Yellow}, for all the first ball-second ball combinations. Since Yellow-Yellow occurs once in five, the probability is 1/5. This is wrong. As we did with the Two Heads Puzzle, once the balls are drawn there are five sample spaces each with one combination: {Green-Yellow}, {Black-Yellow}, {Yellow-Green}, {Yellow-Black}, and {Yellow-Yellow}. The first four sample spaces are unfavorable so their correct probability is 0 while the last sample space is favorable with a correct probability of 1.

T: Exactly.

Lesson 7: Past event sample spaces consist of only one case. The outcome is either favorable (probability 1) or it is unfavorable (probability 0).

*Empirical Frequency Probability*

T: Thirdly, you recognized that observing the first toss outcome confirms the outcome. You used empirical frequency probability by repeating the Two Heads Puzzle 10 times but correctly determined the probability for *each* toss rather than combine them as I had asked.

S: Then that should be true for all these puzzles even if I, the reader, don’t know the outcome.

T: That’s right. Apply that to Bertrand’s Box Paradox (https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095501915):

“There are three boxes, one with two gold coins, one with one gold and one silver, one with two silver coins. A coin drawn at random is gold. What is the probability that the other coin in the same box is gold? Puzzle answer: The coin came either from the first box, or the second; there is no reason for preferring either, so the chance should be a half. But the coin picked was either the first in the first box, or the second in the first box, or the one gold coin in the second box. Two of these three possibilities give it that the other coin in the same box is gold, so the probability is 2/3.”

S: Another probability puzzle without a probability distribution. The coin drawn is from a box that *has been* selected. So, there are three sample spaces with one outcome in each and not the original sample space with three outcomes. I can show this each time I draw a coin by looking at the other coin in the box I selected and see that it is either gold, probability 1, or not gold, probability 0. But never 2/3.

Lesson 8: Simulations or experimental runs of an event that has occurred, requires repeating the same outcome—not all possible outcomes as if none had occurred.

T: And just as you said about the two-heads puzzle, looking at the other coin confirms which of 0 or 1 is correct. In Monty Hall Problem, Monty opens a door knowing that the car is not behind it. Was the car behind the door before it was opened?

S: Of course, not. Once the car is placed behind a door it doesn’t move.

Lesson 9: Revealing the outcome of a past event does not change the outcome or its probability before it was revealed.

T: We have seen that in each case these puzzles ask about the probability of a specific outcome of an event that has occurred. But probabilities other than 0 or 1 are calculated as if the event hasn’t occurred. Why do they do this?

S: I imagine they do what I did when you first asked me about the coin toss outcome. I said it was 1/2 because I didn’t know the outcome. Readers don’t know the outcome in puzzles that do not include it. So, they answer from that perspective—even when someone else knows the outcome. But when someone else knows the outcome, their probability is correct for everyone. There can’t be two correct probabilities, as that contradicts there is only one answer. Besides, the puzzles never say the probability is *just for the reader* or *anyone who doesn’t know the outcome*.

Also, it sure seems contradictory that these puzzles say the answer based on the fact of the outcome is wrong while the one based on ignorance of the fact is right.

T: Great. One way to show this is to switch roles in a simulation. Assume you, the reader, are the dealer in Snake-Eyes Problem (https://mindyourdecisions.com/blog/2016/05/15/snake-eyes-probability-riddle-sunday-puzzle):

With eyes closed, you roll a pair of dice at a casino table. The dealer, seeing the result, tells you one of the dice is a one. What is the probability of snake-eyes (two ones)? Puzzle answer: 1/11.

S: As the dealer, I will see whether the outcome is snake-eyes or not. So, I will know which of 0 or 1 is the correct probability, not 1/11. Since there is only one correct probability, it is the probability for the roller too. And it is the probability for the reader in the original puzzle.

T: Because these puzzles claim to be counter-intuitive, their answers are often supported by using simulations, a form of empirical frequency. Three Cards Riddle (https://www.youtube.com/watch?v=1IFM1ngwEb0) uses simulations:

“I have three cards; one is blue on both sides, another is red on both sides, and the last is red on one side and blue on the other. I put the cards in bag and randomly grab a card. Without looking at the card, I place the card on a table. The side showing is blue. What is the probability the other side of the card is also blue?” Puzzle answer: 1/3.

The simulation consists of 10,000 times randomly selecting each of the cards with each equally likely, displaying one side, and noting what color the other side is. Then calculating the frequency both sides are blue across all cases with blue on one side. Is this correct?

S: No. When we did the Two Heads Puzzle, all the probabilities I wrote were 0s and 1s. Each was the empirical frequency probability of two heads, not the proportion across all 10 results.

PT. Probability is defined as the extent to which an event is likely *to* *occur*. How does that apply to these puzzles?

S: They are all about events that *have* *occurred*.

#### **Real-World Application with Incorrect Probabilities**

T: That’s correct. Let’s look at the real world. Have you seen the televised “World Series of Poker, Texas Hold ‘Em”?

S: Yes. I like it because it shows probabilities. A regular 52-card deck is shuffled. Each player is dealt two cards (pocket or hole) from the top. The top card in the remaining deck is discarded and the next three community cards are shown (flop). Top card is again discarded and a fourth community card is shown (turn). Top card is discarded and a fifth community card is shown (river). Players make their best hands with any combination of 5 cards from the hole and community cards. Bets can be made before and after any cards are dealt.

The TV screen shows the probability of each player having the winning hand based on all the cards each player has and whatever community cards are revealed. As the community cards are revealed, these probabilities change until all the community cards are shown. At that point, one player’s probability of having the winning hand is 1, and 0 for every other player.

T: Is a probability distribution provided?

S: I forgot about that. No, there isn’t one. So, it’s probabilities are based on the assumption that all cards are equally likely.

Lesson 10: Real-world problems do not come with probability distributions so they must be assumed. There are multiple probability answers each conditional on the assumed probability distribution.

T: Incorrect teachings applied to real-world events lead to misapplying probabilities (e.g., World Series of Poker, Texas Hold ‘Em) and not understanding that probability distributions are models that *must* *be* *assumed* when applying probability and statistics to real-world events.

Aren’t all the cards whether known or not, dealt *after* the one shuffle?

S: Yes. I see the problem. All the probabilities displayed, except 0 or 1 are wrong since the only random event, shuffling the cards, has occurred. At that point, the players’ cards, the discarded cards, and the community cards are fixed. The first 2n cards go to the n players. The 2n+1 card is discarded, the 2n+2, 2n+3, and 2n+4 cards are the flop. The 2n+5 card is discarded and 2n+6 card is the turn. The 2n+7 card is discarded and the 2n+8 card is the river. By Lesson 9, revealing the community cards did not change what they are. If all player and community cards are used to calculate the probabilities, it will be obvious the final probabilities 0 and 1 are the correct probabilities throughout the game soon as the shuffling is done. Of course, the players won’t know but the audience will know who has the best hand.

T: Since the players don’t know all the cards that will be displayed, what will they do?

S: They will guess what is likely to occur. But their probabilities of guessing will be incorrect since they don’t even see the other players’ cards.

#### **Guessing the Outcome of a Past Event**

T: Perhaps these puzzles are really asking what is the probability of correctly guessing the unknown outcome that has occurred. Consider the High-Low Problem (Fletcher 2021, “Probability puzzle: Higher or lower?,” *Significance*, 10):

“You’re a contestant and you’ve been dealt the following three cards: a Jack of Hearts and two mystery cards. Aces are high. So, is the next card higher or lower? We can inform our guess by calculating the probability of the next card being lower… 36/51 ≈ 70.6%. Your task now is to guess whether the third card is higher or lower than the second card… what is the probability that both your guesses are correct?”

Is 70.6% the probability of the next card being lower? Or, of correctly guessing?

S: Don’t you still need a probability distribution for guessing? This doesn’t even have a probability distribution for selecting cards.

T: Yes, you do need one.

S: Since the second and third cards have already been dealt, their probabilities of being higher or lower are each either 0 or 1. The 70.6% might be the probability of guessing correctly, but it doesn’t equal 0 or 1, the probability of being higher or lower. Besides, if someone decided to toss a fair coin to guess then their probability of guessing correctly would be 50%, not 70.6%.

Lesson 11: The probability of guessing what outcome occurred is not the same as the probability that a specific outcome occurred.

T: These puzzles are supposedly counter-intuitive. They do this by asking for the probability of a specific outcome of an event *after* occurred and then answering with the probability of that outcome *before* it occurred. As these puzzles are stated, the correct probability is 0 or 1—which is revealed upon revealing the outcome. Incidentally, for past events no before-the-outcome probability distribution is required nor does the event have to be random.

Can we make the classical puzzle answers correct?

S: Don’t you have to make sure the key random event has not occurred when the probability of its outcome is calculated?

T: Yes. Sometimes that’s easy. In Texas Hold ‘Em, what is the random event?

S: It’s shuffling the cards. Oh, I get it. If they simply reshuffle the remaining cards *before* the discard and flop, *before* the discard and turn, and *before* the discard and river, the screen probabilities would be correct throughout the game. And, we keep the suspense of the game.

T: How do we make 1/3 the correct probability for prisoner A in Three Prisoners Problem?

S: Let’s see. The pardoning is the random event so it has to occur *after* prisoner A determines his probability. What if we change the order so the warden tells prisoner A that one of the three prisoners *will* be pardoned rather than *has been* pardoned and prisoner A *will* be told which of the other two prisoners will be executed? Given that information and prior to any pardoning, prisoner A’s probability of being pardoned stays at 1/3 while the other prisoner who will not be executed has a 2/3 probability of being pardoned. Is that right?

T: Perfect. And now you can see why it is 2/3: it could be either of the Prisoner B or C, whose combined probabilities is 1/3 + 1/3 = 2/3.

Either inserting a random event or reordering can make the classical answers correct.

Lesson 12: For probabilities other than 0 or 1 of a random event’s outcome, the event must not have occurred.

Your homework is to rewrite the other puzzles so their puzzle answers are correct, including stating what events are random and their probability distribution when not provided. How can you convince others that 0 or 1 are the correct probabilities for outcomes of past events?

S: Doing the puzzles with the person who doubts the 0 or 1 probabilities are correct be the person who knows what the outcome is, seems to be an easy way to convince them. I will also use the Two Heads Problem to show the 12 lessons.

T: Good. We need to reduce the proliferation of puzzles asking about the probability of an event that has occurred and answering with the probability before the event occurs and of not including a probability distribution for the event of interest (e.g., the links below).

#### **Further Reading**

https://puzzlefry.com/top-10-interesting-and-popular-probability-puzzles/.

https://towardsdatascience.com/five-tricky-statistics-and-probability-riddles-that-90-of-people-fail-77db1eda2e15.

https://mathcircle.berkeley.edu/sites/default/files/handouts/2020/BMC\_Int%20II\_Probability.pdf