THE 50:50 LIFELINE ON "WHO WANTS TO BE A MILLIONAIRE?"

Diane L. Evans
Mathematics Department
Rose-Hulman Institute of Technology
5500 Wabash Avenue
Terre Haute, Indiana 47803
diane.evans@rose-hulman.edu

Isaac Weintraub, Student
Rose-Hulman Institute of Technology
Campus Mailbox 1494
5500 Wabash Avenue
Terre Haute, Indiana 47803

Abstract: The Monty Hall problem can be a useful exercise in statistics and probability classes. Besides the lively interaction of students proposing various arguments (for or against) the answer, some students begin to consider problems and games where the Monty Hall dilemma may be hidden beneath the surface. Since *Who Wants to Be a Millionaire?* continues to be a popular game show, one problem raised in class is "How does the probability of choosing the correct answer differ depending on the elimination method used by the 50:50 lifeline?" This question is of particular interest to students since probabilistic choices determine a contestant's winnings. This paper will examine the probability calculations of two possible elimination methods.

1. INTRODUCTION

In our Introductory Engineering Statistics and Introductory Probability classes, in which all students have a calculus background, we never pass up the opportunity to liven up the classes with the exasperating Monty Hall problem when we discuss probability. (See the appendix for an overview of the Monty Hall problem and its solution.) When students first encounter this problem's solution, i.e., the contestant should switch doors to improve his probability of winning the valuable prize, a wild flurry of arguments, challenges, and questions abound. Comparable game show situations are debated where switching answers would or would not be beneficial for improving contestants' odds of winning. One game show that is familiar to today's college age students is *Who Wants to Be a Millionaire?*, or *Millionaire* for short. One student's question about this game show and switching answers drew particular interest from the class: "After the 50:50 lifeline eliminates two of the incorrect answers from the contestant's four choices, should the contestant switch answers to increase his probability of selecting the correct one?" Assuming that the contestant is merely guessing at the correct answer, assigns equally likely probabilities to the four choices, and loudly announces his guess before the host has two incorrect choices eliminated (but never the contestant's if it is incorrect), then the answer is "yes." The probability of selecting the correct answer by switching choices increases from $\frac{1}{4}$ to $\frac{3}{4}$. But if we remove these contrived assumptions, which also allows the contestant to assign asymmetric probabilities to the four choices, then this answer changes from "yes" to the more typical answer relayed in a statistics course, "It depends."

The "it depends" answer prompted another student to retort, "Assumptions or not, the 50:50 lifeline is useless anyway if the contestant is stuck between two answers that seem correct. The game's computer elimination of the two incorrect answers is not random; the computer always leaves the two answers that a contestant would most likely choose." Whether this statement is true or not is secondary to the fact that it leads to a more interesting question of how conditional probabilities for selecting the correct answer differ depending on *how* incorrect answers are eliminated by the 50:50 lifeline, ***randomly*** or ***non-randomly***. Calculating these probabilities for various examples became the topic of the rest of the day's class and subsequently we considered calculating these probabilities for general probability assignments. Specifically, this paper will address how the 50:50 lifeline elimination method in *Millionaire*, random versus non-random, affects the probability of the correct answer being chosen by the contestant after the elimination of two incorrect answers.

2. *MILLIONAIRE* AND THE 50:50 LIFELINE

The game *Millionaire* is played with one contestant attempting to answer a sequence of up to fifteen multiple choice questions in a quest to win increasing amounts of money from $100 to $1,000,000. The more questions answered correctly by the contestant, the more prize money the contestant earns. The questions increase in difficulty as the contestant proceeds, and four possible answers are provided for each question. The contestant must choose the most correct answer from the four possible choices. The contestant is aided by three lifelines, which include "asking the audience," "phoning a friend," and "50:50." The lifelines can be used once each during the game to assist the contestant in choosing the correct answer to a question.

The information about the 50:50 lifeline as provided on the official U.S. *Millionaire* website [1] is:

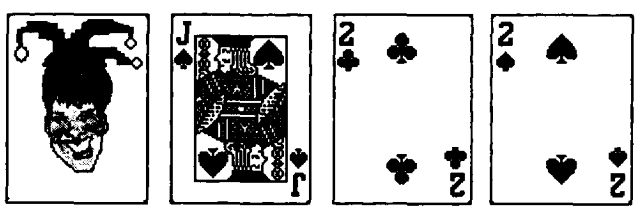
> *"**50:50** - The Main Game Contestant asks the computer to eliminate two of the possible answer choices, leaving the Main Game Contestant a choice of two answers [one of which is the correct answer] from which to select. The Main Game Contestant then has the choice of selecting an answer, quitting the game or using another lifeline, if available."*

The site itself does not explain how the two possible answer choices are eliminated, though the host of *Millionaire* informs the contestant and audience that the elimination of two incorrect answers is random. Some *Millionaire* fans, as did our students, speculated that the elimination is not random, and the two remaining answers are the correct one and the "next best answer." Following is a description of the two methods used to eliminate incorrect answers and the probabilities associated with determining the correct answer once the 50:50 lifeline is applied under each method.

3. PROBABILITY CALCULATIONS

Let C_A, C_B, C_C, and C_D represent the events that the correct answer out of the four possibilities is *A*, *B*, *C*, or *D*, respectively. Let p_A, p_B, p_C, and p_D represent the probabilities that the contestant assigns to events C_A, C_B, C_C, and C_D, respectively, when all four answers are available to him. By the rules of the game, we know $C_i \cap C_j = \varnothing$ for all $i \neq j$, where *i*, $j \in \{A,B,C,D\}$ and $p_A + p_B + p_C + p_D = 1$. Without loss of generality, let's assume the correct answer to a contestant's question is *A*, and the "next best answer" is *B*. Here the terminology "next best answer" means the answer that would receive the highest proportion of votes from a large simple random sample of the U.S. population after disregarding the proportion of votes for the true correct answer. Example 1 displays a possible *Millionaire* question where *A* is the answer and *B* is the "next best answer."

Example 1. *Millionaire* question: "Which playing card is the author of this paper posing as a Joker?"



A **B** **C** **D**

Solution: The correct answer is *A*, though you may have been fooled by the "next best answer" *B*, the Jack. Even if you have never seen the above picture, choices *C* and *D* are clearly incorrect. If you've handled or used playing cards, then you would most likely know that the "J" in choice *B* represents a Jack, not Joker. Although *A* and *B* are the only possible correct answers, *B* is the "next best answer" to *A* because of most peoples' knowledge of playing cards.

Recall the problem we are considering in this paper is how the contestant's probability of selecting the correct answer differs depending on the elimination method used by the 50:50 lifeline, which I'm calling random versus non-random. Assuming that the true correct answer is A and the next best answer is B, then using either elimination method, A will not be eliminated by the 50:50 lifeline because A is the correct answer. But, if the elimination method is non-random, i.e., the game's computer leaves the best two answers, then answer B will remain after the 50:50 lifeline is applied. In the random elimination case, the computer randomly chooses two of the three incorrect answers to eliminate. So, the contestant will be left with answers A and B, A and C, or A and D, after the 50:50 lifeline is applied. What are the contestant's updated probabilities for selecting answer A given that A and B remain in the non-random case and answers A and B, A and C, or A and D remain in the random case after the 50:50 lifeline is applied?

If the elimination method of two incorrect answers is **random**, then two of the three incorrect answers are randomly eliminated and the correct answer remains. So, in our case, the correct answer A remains and the next best answer B may remain, though A and C or A and D may remain since the elimination of two of B, C, and D is truly random. Let p_r represent the probability that the contestant will choose answer A given that A and B, C, or D remain as choices after the 50:50 lifeline is applied. Note that p_r and p_{nr}, defined below, are personal probabilities calculated *a priori*. Then p_r can be calculated as:

$$p_r = \sum_{i=B,C,D} P\left(C_A \mid \text{Answers } A \text{ and } i \text{ remain after 50:50}\right) \times$$

$$P\left(\text{Answers } A \text{ and } i \text{ remain after 50:50}\right) =$$

$$\frac{p_A}{p_A + p_B} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_C} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_D} \cdot \frac{1}{3}.$$

Now if the elimination method of two incorrect answers is **non-random**, then answers C and D will be eliminated by design. Let p_{nr} represent the probability that the contestant will choose answer A given that A and B remain as choices after the 50:50 lifeline non-randomly removes answers C and D. Then p_{nr} can be calculated as:

$$p_{nr} = \sum_{i=B,C,D} P\left(C_A \mid \text{Answers } A \text{ and } i \text{ remain after 50:50}\right) \times$$

$$P\left(\text{Answers } A \text{ and } i \text{ remain after 50:50}\right) =$$

$$\frac{p_A}{p_A + p_B} \cdot 1 + \frac{p_A}{p_A + p_C} \cdot 0 + \frac{p_A}{p_A + p_D} \cdot 0 = \frac{p_A}{p_A + p_B}.$$

The probabilities for selecting answer A, given that A and one of B, C, or D remain as answers after the 50:50 lifeline is applied, is

$p_r = \frac{p_A}{p_A + p_B} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_C} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_D} \cdot \frac{1}{3}$ for the random elimination method and $p_{nr} = \frac{p_A}{p_A + p_B}$ for the non-random elimination method. Are there any values for p_A, p_B, p_C, and p_D in which the updated probability for selecting the correct answer *A* for the non-random method p_{nr} (where the best two answers remain) is greater than or equal to the probability for the random method p_r? That is, can the non-random method of the 50:50 lifeline prove useful for the contestant in selecting the correct answer? By this we mean, can the non-random method provide an updated probability p_{nr} that is greater than p_r, which gives the contestant more motivation for staying with answer *A* after the elimination of two incorrect answers? Certainly there are fans of *Millionaire* who don't think so!

Let me be clear in saying that I'm not trying to determine when answer *A* becomes the more likely choice than answer *B* after the 50:50 elimination, but rather how the conditional probabilities of selecting *A* differ after the two 50:50 elimination methods are applied. In fact, if the original probability values for p_A and p_B are set by the contestant such that $p_A > p_B$, then choice *A* will remain more likely than choice *B* after the 50:50 non-random elimination method, and also after the random 50:50 elimination method (since $p_A > p_B$ implies that $\frac{p_A}{p_A + j} > \frac{p_B}{p_B + j}$ for $0 < j < 1$).

Let's first consider four special cases in computing the probabilities p_{nr} and p_r when $p_A = 0$, $p_B = 0$, or both, or when $p_A > 0$, $p_B > 0$, and $p_B = p_C = p_D$.

Case 1. If $p_A = 0$, $p_B > 0$, then $p_r = 0$ and p_{nr} is zero or undefined. But in the random elimination case, the contestant may be given a chance to "start over" if the 50:50 lifeline leaves one of *C* or *D* with $p_C = 0$ or $p_D = 0$.

Case 2. If $p_A > 0$, $p_B = 0$, then $p_r \le p_{nr} = 1$.

Case 3. If $p_A = 0$ and $p_B = 0$, then p_r and p_{nr} are undefined. In the non-random elimination case, the contestant will be given a chance to start over since the 50:50 lifeline will leave answers *A* and *B*. In the random elimination case, the contestant may be given the same chance to start over if the remaining choice besides *A* is one that has been assigned zero probability. Otherwise, if p_C or p_D remains with a nonzero probability, then the contestant will not choose answer A.

Case 4. If $p_A > 0$, $p_B > 0$, and $p_B = p_C = p_D$, then

$p_r = \frac{p_A}{p_A + p_B} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_C} \cdot \frac{1}{3} + \frac{p_A}{p_A + p_D} \cdot \frac{1}{3}$ is equal to $p_{nr} = \frac{p_A}{p_A + p_B}$.

This is a case in which neither elimination method produces a more probable final answer of A after the 50:50 lifeline is used.

By examining the formulas for p_r and p_{nr}, we can also conclude that if p_B is greater than both p_C and p_D, then $p_r > p_{nr}$. In the same manner, if p_B is less than both p_C and p_D, then $p_r < p_{nr}$. So, the non-random elimination method provides a greater updated probability for selecting A if the contestant originally assigns p_B less than both p_C and p_D. But there are also instances when the non-random elimination method is better when p_B is less than just one of p_C or p_D. Example 2 illustrates this situation.

Example 2. Suppose we are given the same *Millionaire* question as in Example 1; we need to determine which of four playing cards the author of this paper is posing as a Joker. Answer A is the correct answer, while B is the Jack of spades, C is the two of clubs, and D is the two of spades. Let's place various probabilities on A through D being the correct answer and see how they affect the conditional probabilities p_r and p_{nr}.

Solution: Let's suppose the contestant is fairly certain that the correct answer is A, and gives answer B less probability than both answers C and D because the contestant recalls that a "J" represents Jack. Perhaps he also thinks this is a trick question and that for some reason the answer is the number two playing card, but he can't decide which suit is more likely correct. For example, p_A = 0.750, p_B = 0.050, p_C = p_D = 0.100. If the 50:50 lifeline is used, then

$$p_r = \frac{1}{3}\left(\frac{0.750}{0.750+0.050} + \frac{0.750}{0.750+0.100} + \frac{0.750}{0.750+0.100}\right) = \frac{245}{272} \approx 0.901,$$

$$\text{and } p_{nr} = \frac{0.750}{0.750+0.050} = \frac{15}{16} \approx 0.938.$$

As explained above, or by logical reasoning, p_{nr} exceeds p_r if less probability is given to answer B than both answers C and D. (Or, try p_C=0.080 and p_D=0.120 for non-equally likely probabilities for answers C and D.)

Now suppose that the contestant gives a higher probability to answer B than just one of answers C and D. Although it makes little sense why a contestant would do this with the playing card example (unless he thinks a club is a Joker), let $p_A = p_B = 0.200$, $p_C = 0.500$, and $p_D = 0.100$. Then after the 50:50 lifeline is used and A and B remain as the best two answers, we have:

$$p_r = \frac{1}{3}\left(\frac{0.200}{0.200+0.200}+\frac{0.200}{0.200+0.500}+\frac{0.200}{0.200+0.100}\right)=\frac{61}{126}\approx 0.484,$$

$$\text{and } p_{nr} = \frac{0.200}{0.200+0.200}=\frac{1}{2}.$$

But, if the contestant had put a slightly higher probability on answer C, $p_C = 0.550$, while taking more probability away from answer D, $p_D = 0.050$, then the outcomes for p_r and p_{nr} would be reversed since

$$p_r = \frac{1}{3}\left(\frac{0.200}{0.200+0.200}+\frac{0.200}{0.200+0.550}+\frac{0.200}{0.200+0.050}\right)=\frac{53}{90}\approx 0.589.$$

While the mathematics easily confirms the new ranking of probability values, this reversal is non-intuitive and requires additional sleuthing to determine why.

In order to determine the complete solution region of probability values for p_A, p_B, p_C, and p_D when the non-random elimination method provides a greater updated probability for selecting A than the random elimination method, the following arguments, algebra, and geometry are introduced. The rest of the analysis of the methods will assume $p_A > 0$, $p_B > 0$; and $p_C > 0$ or $p_D > 0$, or both.

For p_{nr} to be greater than or equal to p_r, we need:

$$\frac{p_A}{p_A+p_B} \ge \frac{p_A}{p_A+p_B}\cdot\frac{1}{3}+\frac{p_A}{p_A+p_C}\cdot\frac{1}{3}+\frac{p_A}{p_A+p_D}\cdot\frac{1}{3},$$

which simplifies to: $\dfrac{2}{p_A+p_B} \ge \dfrac{1}{p_A+p_C}+\dfrac{1}{p_A+p_D}$.

Let region R be the interior of the triangle in the p_A, p_B plane with vertices $(0,0)$, $(1,0)$, and $\left(0,\frac{1}{3}\right)$. Let T be the triangle with vertices at $(0,0)$, $(1,0)$, and $(0,1)$. Both the region R and triangle T are displayed in Figure 1. The non-random elimination method provides a greater updated probability for selecting A than the random elimination method when (p_A, p_B) is selected on, or within the region R and certain restrictions are placed on p_C and p_D.

Theorem 1 with proof provides this explanation for $p_A, p_B > 0$.

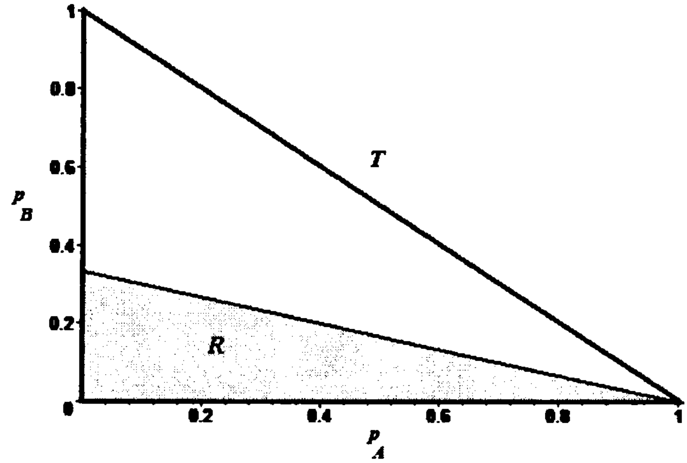


Figure 1. The region R and the triangle T used in Theorem 1.

***Theorem* 1.** Let $p_A > 0$, $p_B > 0$; and $p_C > 0$ or $p_D > 0$, or both, with the restriction $p_A + p_B + p_C + p_D = 1$. If, in addition, the inequality

$$\frac{2}{p_A + p_B} \geq \frac{1}{p_A + p_C} + \frac{1}{p_A + p_D} \quad \textbf{(1)}$$

is satisfied and p_C and p_D are chosen as $p_C = \frac{1 - p_A - p_B}{2} + t$ and $p_D = \frac{1 - p_A - p_B}{2} - t$, where $|t| \leq \frac{\sqrt{(1 + p_A - p_B)(1 - p_A - 3p_B)}}{2}$, then (p_A, p_B) is within R or on its diagonal boundary $p_B = \frac{1 - p_A}{3}$.

Proof: The restriction $p_A + p_B + p_C + p_D = 1$ with $p_C > 0$ or $p_D > 0$ implies that $p_A + p_B < 1$. Since p_A and p_B are both positive, then the point (p_A, p_B) is within triangle T or on its top diagonal boundary $p_B = 1 - p_A$.

Now since $p_C + p_D = 1 - p_A - p_B$, then $0 \leq p_C \leq 1 - p_A - p_B$ since at least one of p_C and p_D is greater than 0. Thus,

$$-\frac{1 - p_A - p_B}{2} \leq p_C - \frac{1 - p_A - p_B}{2} \leq 1 - p_A - p_B - \frac{1 - p_A - p_B}{2}, \text{ or}$$

$$\left| p_C - \frac{1 - p_A - p_B}{2} \right| \leq \frac{1 - p_A - p_B}{2}.$$

Define the variable t as $t = p_C - \frac{1 - p_A - p_B}{2}$, and so $|t| \leq \frac{1 - p_A - p_B}{2}$.

Since $p_D = 1 - p_A - p_B - p_C$, then in terms of t, $p_D = \frac{1 - p_A - p_B}{2} - t$.

Inequality (1) can now be rewritten as:

$$\frac{2}{p_A + p_B} \geq \frac{1}{p_A + \frac{1 - p_A - p_B}{2} + t} + \frac{1}{p_A + \frac{1 - p_A - p_B}{2} - t}$$
$$= \frac{1}{\frac{1}{2}(1 + p_A - p_B) + t} + \frac{1}{\frac{1}{2}(1 + p_A - p_B) - t}.$$

Additional algebra to the above yields:

$$\frac{1}{p_A + p_B} \geq \frac{1}{1 + p_A - p_B + 2t} + \frac{1}{1 + p_A - p_B - 2t} = \frac{2(1 + p_A - p_B)}{(1 + p_A - p_B)^2 - 4t^2}.$$

Since the denominator on the right hand side of the inequality is just the product of $p_A + p_C$ and $p_A + p_D$, which are both positive, then we can cross multiply the expression above and rewrite the inequality as:

$$(1 + p_A - p_B)^2 - 4t^2 \geq 2(1 + p_A - p_B)(p_A + p_B), \text{ which simplifies to}$$
$$(1 + p_A - p_B)[(1 + p_A - p_B) - 2(p_A + p_B)] \geq 4t^2, \text{ or}$$
$$(1 + p_A - p_B)(1 - p_A - 3p_B) \geq 4t^2.$$

The factor $1 + p_A - p_B$ is positive, which means the factor $1 - p_A - 3p_B$ must be greater than or equal to 0. For this to happen, we require $p_B \leq \frac{1 - p_A}{3}$. This means that (p_A, p_B) is a point in R or on its diagonal boundary $p_B = \frac{1 - p_A}{3}$. But this last inequality

$$(1 + p_A - p_B)(1 - p_A - 3p_B) \geq 4t^2$$

shows us that we need to place a tighter bound on t, previously $|t| \leq \frac{1 - p_A - p_B}{2}$, if we want p_A, p_B, p_C, and p_D to satisfy (1). For (1) to hold true, we need to choose t so that $|t| \leq \frac{\sqrt{(1 + p_A - p_B)(1 - p_A - 3p_B)}}{2}$. ■

4. CONCLUSION

Considering the outcome of this theorem and the four special cases, we can conclude that when the correct answer to a contestant's question is A, and the next best answer is B (where A and B can be replaced by the appropriate choices for any question), the non-random elimination can provide an equal or greater updated probability for selecting A than the random elimination method. The theorem and four cases provided us with a region R, displayed in Figure 1, in which this is true. Thus, any value (p_A, p_B) on the boundary of or within R, where its top diagonal is defined by $p_B = \frac{1-p_A}{3}$, with $p_C = \frac{1-p_A-p_B}{2}+t$ and $p_D = \frac{1-p_A-p_B}{2}-t$, where $|t| \le \frac{\sqrt{(1+p_A-p_B)(1-p_A-3p_B)}}{2}$, will give the contestant a greater updated probability for selecting A after the 50:50 elimination if the elimination method is non-random (computer leaves the best two answers), rather than random. So, if the computer automatically leaves the best two answers after the 50:50 lifeline on *Millionaire*, we have determined the situations in which this could be beneficial to the contestant in selecting the correct answer.

An interesting extension to the problem would be to determine what the initial probability values for p_A, p_B, p_C, and p_D have to be to make the updated probabilities (after the 50:50 elimination method) for p_r and p_{nr} greater than 50%. That is, when can the 50:50 elimination "overturn" (yield a greater than 50% updated probability for answer A) a contestant's original answer? Is this more likely to happen given one of the two elimination methods?

REFERENCE

1. *Who Wants to Be a Millionaire?* official rules page. Retrieved on July 27, 2005 from http://abc.go.com/primetime/millionaire/abouttheshow/rules2.htmlH.

APPENDIX: THE MONTY HALL PROBLEM

The classic Monty Hall problem is from the popular 1960's and 1970's game show *Let's Make a Deal.* The host, Monty Hall, gives a contestant a choice of three closed doors concealing prizes. Behind one of the doors is a valuable prize (e.g., cash, car, vacation), while the other two contain funny gag prizes or "zonks" (e.g., can of tuna fish, giant stuffed animal, a goat). Before the contestant's door is opened, Monty Hall (who knows which doors hide "zonks") opens one of the two remaining doors to unveil a "zonk." Monty then offers the contestant a chance to switch doors before revealing the prizes behind them. The contestant's dilemma is whether or not to switch doors in order to maximize the chances of winning the valuable prize.

The solution to the dilemma is that the contestant doubles the probability of winning the valuable prize if the contestant switches doors. Below are the details of this solution using Bayes Theorem.

Suppose you're the contestant on the show, and you must choose either Door 1, 2, or 3 as the correct door that conceals the valuable prize. The prize has an equally likely chance of being behind any one of these doors. Without loss of generality, let's say you choose Door 1. Then Monty counters with opening Door 2 and revealing a goat ("zonk"). Now recall, Monty knows which two doors conceal goats, and he can and will always open a door with a goat behind it. That is, assuming that you have chosen Door 1 and the valuable prize is behind:

- Door 2, then he'll open Door 3.
- Door 3, then he'll open Door 2.
- Door 1, then he'll either open Door 2 or 3.

Upon the opening of Door 2 revealing the goat, you should choose to switch doors (to Door 3) in order to double your likelihood of winning the valuable prize.

Let P_i represent the event that the valuable prize is behind Door i, where $i = 1,2,3$. Let MH_j represent the event that Monty Hall opens Door j, where j = 2 or 3. The probability that you win the valuable prize given that you switch doors after Monty reveals the goat behind Door 2 is the conditional probability $P(P_3 \mid MH_2)$. Its value is:

$$P(P_3 \mid MH_2) = \frac{P(P_3 \cap MH_2)}{P(MH_2)} = \frac{P(MH_2 \mid P_3) \cdot P(P_3)}{P(MH_2)}$$

$$= \frac{P(MH_2 \mid P_3) \cdot P(P_3)}{P(MH_2 \mid P_1) \cdot P(P_1) + P(MH_2 \mid P_2) \cdot P(P_2) + P(MH_2 \mid P_3) \cdot P(P_3)}$$

$$= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3},$$

which is double your original probability ($\frac{1}{3}$) of winning the prize.