The Odds Are – My Random Walk Through Probability Theory

P. Reany

May 9, 2021

Abstract

In this introductory paper we look at certain problems in probability theory. I gleaned these problems from the Internet. I regard my further explication of their work as my personal notes (that augments the online presentations) that I can share with others. Special thanks go to Elliot Nicholson and to Tom Leighton (Instructor at MIT), and others for their wonderful online presentations on probability theory.

1 Introduction

It's not the purpose of this paper to systematically present the rudiments of probability theory, but, rather, to use the basic theory to solve some interesting problems, especially to reveal what this author thinks are the best heuristics to use in the effort. So, the reader should already know basic probability theory up to and including conditional probability.

A small reminder on conditional probability is this: Let A and B be events, with $\Pr(B) \neq 0$. Then the probability of A being true given that B is true is expressed symbolically as $\Pr(A|B)$. Writing it in terms of previously defined expressions (within probability theory), we have that

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \,. \tag{1}$$

The symbol '$\cap$' stands for logical 'and'. Another common symbol for it is '&'.

Let's begin by looking at the problem of screening a population for some disease.[1] We shall not be concerned here with how one determines the accuracy of a test or how the knowledge of the incidence of a particular disease is known in a general population. These number will simply be provided to us.

2 Problem 1: Disease Screening

In a certain population the incidence of a certain disease is .001. If the test for this disease is 99% accurate, what is the probability that a person who

[1] Elliot Nicholson at https://www.youtube.com/watch?v=m66UfZafZdE.

tests positive for the disease actually has the disease? Symbolically, that's $\Pr(\mathrm{D}^+ | \mathrm{T}^+)$.

Before calculating any probabilities, let's look carefully at Figure 1. We'll be solving this problem by calculating the probability entries that go in the column with circles 1 and 2. They are subtotals from their respective rows, and together they should add to 1.00. Similarly, the bottom row with the circles 3 and 4 are subtotals from their respective columns, and they should also add to 1.00. We'll be filling in these values as we go, starting with the easiest ones to solve for.

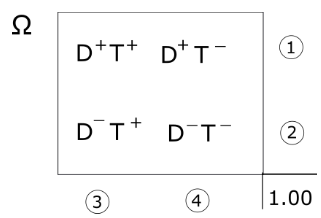

Figure 1. The rectangle represents the sample space Ω, containing the four mutually exclusively events of interest. 'D^+' stands for 'has the disease', 'D^-' stands for 'doesn't have the disease', 'T^+' stands for 'has tested positive', 'T^-' stands for 'has tested negative'. Each of the four outcomes is a shorthand for logical 'and'; for example, '$\mathrm{D}^+\mathrm{T}^+$' stands for $\mathrm{D}^+ \cap \mathrm{T}^+$ or $\mathrm{D}^+ \,\&\, \mathrm{T}^+$.

Let's begin by asking ourselves what is meant by the claim that the *incidence of a certain disease is* .001. It means that any given person being screened has a 0.001 probability of having the disease. $\mathrm{D}^+/\mathrm{D}^-$ are our boolean variables to indicate the presence/absense of the disease.

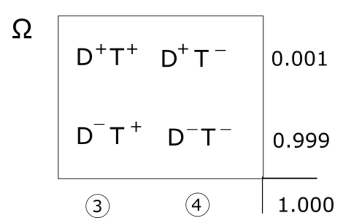

Figure 2. We've replaced circle 1 by .001 because the probability of having the disease, whether you test positive or negative, is .001. Similarly, we've replaced circle 2 by .999 because the probability of not having the disease, whether you test positive for it or not, is .999.

Therefore,

$$\Pr(\mathrm{D}^+) = .001 \quad \text{and, of course,} \quad \Pr(\mathrm{D}^-) = .999\,. \tag{2}$$

So, what is meant by the claim that the test for this disease is 99% accurate? It means two things: First, that if the testee actually has the disease, he or she has a .99 probability of testing positive, or, symbolically:

$$\Pr(T^+ | D^+) = .99 \,. \tag{3}$$

Second, that if the testee actually does not have the disease, he or she has a .99 probability of testing negative, or, symbolically:

$$\Pr(T^- | D^-) = .99 \,. \tag{4}$$

Viewed another way, this level of testing accuracy means that the testee has only a 1% chance of getting a false positive or a 1% chance of getting false negative.

We now have enough information already to get two event probabilities in the sample space:

$$\Pr(T^+ \cap D^+) = \Pr(T^+ | D^+)\Pr(D^+) = (.99)(.001) = .00099 \,, \tag{5a}$$

$$\Pr(T^- \cap D^-) = \Pr(T^- | D^-)\Pr(D^-) = (.99)(.999) = .98901 \,. \tag{5b}$$

But $T^+ \cap D^+ = D^+T^+$ and $T^- \cap D^- = D^-T^-$. The updated graphic appears in Figure 3.

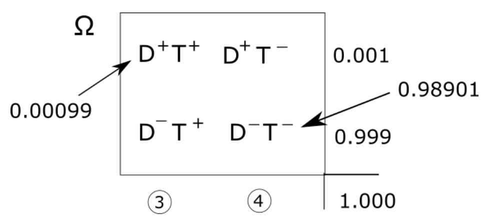

Figure 3. We've added probabilities that are easily calculated from the given information.

We now have enough information to calculate $\Pr(D^+ \cap T^-)$ and $\Pr(D^- \cap T^+)$ by forming the equations

$$\Pr(D^+ \cap T^+) + \Pr(D^+ \cap T^-) = .001 \,, \tag{6a}$$

$$\Pr(D^- \cap T^+) + \Pr(D^- \cap T^-) = .999 \,. \tag{6b}$$

Or, expressing these in terms of the entries in the sample space:

$$\Pr(D^+T^+) + \Pr(D^+T^-) = .001 \,, \tag{7a}$$

$$\Pr(D^-T^+) + \Pr(D^-T^-) = .999 \,. \tag{7b}$$

We obtain the solutions

$$\Pr(D^+T^-) = 0.00001 \quad \text{and} \quad \Pr(D^-T^+) = 0.00999 \,. \tag{8}$$

We can now update the probabilities for the remaining two events and then calculate the subtotals for circles 3 and 4.

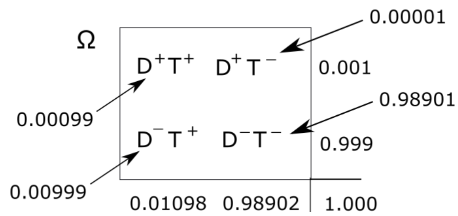

Figure 4. We've calculated the subtotals from the columns and placed them in the bottom row, and they add to 1.000.

Now we can calculate what we were orignally asked to find, namely,

$$\Pr(D^+ | T^+) = \frac{\Pr(D^+ \cap T^+)}{\Pr(T^+)} = \frac{\Pr(D^+T^+)}{\Pr(T^+)} = \frac{.00099}{.01098} \approx 0.09\,. \tag{9}$$

This is one of those unintuitive results of probability theory. Even with such a highly reliable test used in this screening, getting a positive result still means that you have less than 10% chance of having the disease.

3 Problem 2: The Monty Hall Problem

Here I'm working the so-called Monty Hall Problem similar to how Elliot Nicholson did it.[2] I assume the reader knows the basic format of the Monty Hall Show known as *Let's Make a Deal*. Anyway, the basic situation for this intriguing problem is this: A contestant is shown three doors, behind each of which is hidden either a car or a goat. There is one car and two goats. On the assumption the contestant wants the car more than a goat, we presume, without loss of generality, that the contestant chooses Door # 1.

But before the prize behind Door #1 is revealed, Monte makes the offer that he will reveal the prize behind one of the other doors and then let the contestant choose to stay with his or her original choice or swap the original door with what's behind the other closed door, and, of course, Monte is going to choose a door to open with a goat behind it. Our job is to use probability theory to decide which strategy is better to win: To stay with the original choice or to swap?

[2] https://www.youtube.com/watch?v=6bqEcuUsQZk&index=16&list=PLAvgI3H-gclbyWnE9WrSq68_HRW9rOZUw.

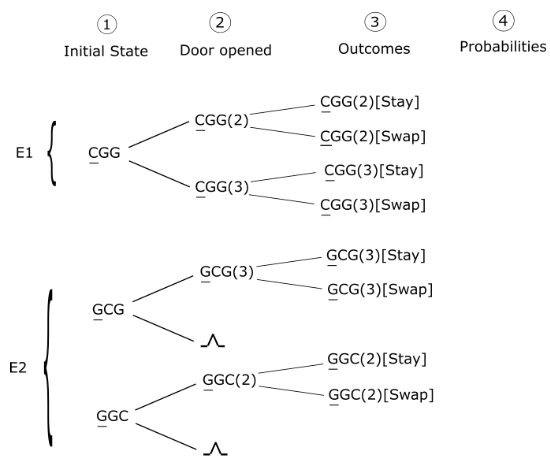

Figure 5. Step-by-step formation of all outcomes and two events.

The first time I encountered this problem, my intuition was that it should make no difference which door to choose of the remaining two closed doors. This seems to be the typical intuitive feeling of most people who encounter this problem. But on second thought, whichever door one initially chooses, he or she has 2 to 1 'odds' of picking a goat rather than the car. Therefore, maybe swapping doors is the reasonable thing to do.

We will now prove this rigorously. My plan of attack against this problem is to

1. systematically produce all the possible outcomes/events of this contest,
2. assign a reasonable probability to each outcome,
3. use the formulas of conditional probability to find the result I need.

Step 1: Systematically produce all the possible outcomes of this contest.

Listing the door numbers 1,2,3 from left to right, the basic possibilities of the order of the hidden prizes are the following

$$\underline{C}GG \qquad \underline{G}CG \qquad \underline{G}GC$$

where the underbar indicates the contestant's choice (Door #1).

In Figure 5, we see the step-by-step construction of the outcomes of this contest (experiment). The E's on the left are the events: E1 = set of all outcomes with the car behind Door # 1, E2 = set of all outcomes with a goat behind Door #1. Clearly, P(E1) = 1/3, and P(E2) = 2/3.

In column 1, we see the initial state of the game, in which the possible configurations of the car and goats are shown, along with the contestant's choice of Door #1. In column 2, we add the information about which door Monte opened. (The meaning of the shark tooth (_/_) is to represent a null pointer.)

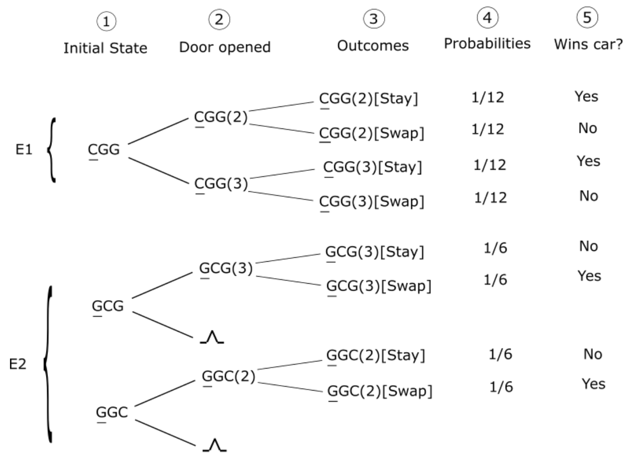

Figure 6. We've listed all possible outcomes and their consequences.

Now, if there was a goat behind Door #1, Monte has only one choice of doors to open – hence then null pointer. In Column 3 we see all eight outcomes. And now it's time to assign probabilities to each outcome.

Step 2: Assign a reasonable probability to each outcome.

Since $\Pr(\text{E1}) = 1/3$, the sum of the probabilities in E1 in Column 3 must add to 1/3, I see no reason not to assign all these probabilities equally, making each one to have probability 1/12. And, since $\Pr(\text{E2}) = 2/3$, the sum of the probabilities in E1 must add to 1/3, I see no reason not to assign all these probabilities equally, making each one to have probability 1/6.

Step 3: Use the formulas of conditional probability to find the result I need.

What we need to know is the probability of winning if we swap compared to if we stay on the original choice. Since the two choices are mutually exclusive and collectively exhaustive, their probabilities must add to one. We can read-off the probabilities we need directly from Figure 6.

$$\Pr(\text{Win} \,|\, \text{Swap}) = \frac{\Pr(\text{Win} \cap \text{Swap})}{\Pr(\text{Swap})} = \frac{\frac{1}{6} + \frac{1}{6}}{\frac{1}{12} + \frac{1}{12} + \frac{1}{6} + \frac{1}{6}} = \frac{2}{3}\,. \tag{10}$$

Therefore, by swapping, the odds of winning go from 1:2 (staying with original choice) to 2:1. The player should swap.

4 Problem 3: Derivation of the Hypergeometric Distribution

Our problem here is to construct the probability mass function for the hypergeometric distribution. First, let's describe how this distribution occurs. Say we have N objects split into two kinds: K blues chips and $N-K$ red chips. We will repeatedly draw one chip at random at a time without replacement. After having drawn out n chips, what is the probability that k of those chips are blue? (We assume that $k \leq n$ and $n < K$ and $n < N-K$.) Show that, for random variable X,

$$\Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \tag{11}$$

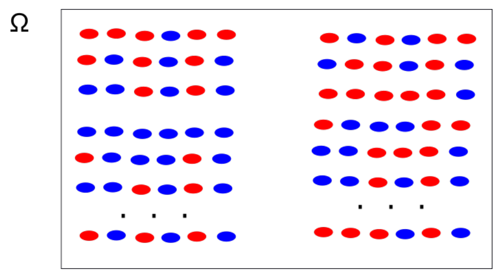

Figure 7. Each row of n chips is a possible outcome of the sample space. We define E_k as the event that the outcomes have exactly k blue chips.

Solution. We begin our analysis by imagining each chip to be distinguishable from all others by marking each of them with a unique number from 1 to N. Next, we imagine having a row of n bowls to place the n chips in. The total number of ways we can place n chips from among N distinguishable chips (irrespective of their color) is P_n^N, where P stands for permutations.

$$\mathrm{P}_T^S \equiv S(S-1)\cdots(S-(T-1)) = \frac{S!}{T!}. \tag{12}$$

Now consider a generic draw of k blus chips (all distinguishable) and $n-k$ red chips (all distinguishable) and we will first randomly choose k little bowls out of an ordered row of n little white bowls to place the distinguishable blue chips. Having made this choice, we'll replace those bowls by blue bowls. Now, into the remaining $n-k$ white bowls, we'll 1) replace them with red bowls and then 2) place into them the $n-k$ distinguishable red chips.

Fixing this configuration of blue and red bowls, how many strings of blue chips can go into the blue bowls? In terms of permutations, it's P_k^K. And for each of these, how many strings of distinguishable red chips can go into the remaining $n-k$ red bowls? That's P_{n-k}^{N-K}. Therefore, for this one confiuration of little bowls, their product is the number of ways of placing P_k^K distinguishable blue chips strings of lenght k intermixed with P_{n-k}^{N-K} distinguishable red chip strings.

The next question is, How many ways are there to choose k bowls out of the n little bowls to take as places to put blue chips? The answer is $\binom{n}{k}$. Therefore, the number of ways of placing P_k^K distinguishable blue chips into a row of P_{n-k}^{N-K} distinguisble red chips is

$$\text{No. of Ways} = \mathrm{P}_k^K \, \mathrm{P}_{n-k}^{N-K} \binom{n}{k} . \tag{13}$$

Now, to get the probability $\Pr(X = k)$, we need to divide this last result by P_n^N:

$$\Pr(X = k) = \frac{\mathrm{P}_k^K \, \mathrm{P}_{n-k}^{N-K} \binom{n}{k}}{\mathrm{P}_n^N} . \tag{14}$$

Now, to put this into the form given in (12), we use that

$$\mathrm{P}_q^p = \binom{p}{q} q! \, . \tag{15}$$

So,

$$\Pr(X = k) = \frac{\binom{K}{k} k! \binom{N-K}{n-k} (n-k)! \binom{n}{k}}{\binom{N}{n} n!} . \tag{16}$$

But since

$$\frac{k!(n-k)! \binom{n}{k}}{n!} = 1 \, , \tag{17}$$

then (17) becomes

$$\Pr(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} . \tag{18}$$

QED

5 Problem 4: Sneaky Problem Using the Hypergeometric Formula

Michigan's Classic Lotto 47[3]

In Michigan's Classic Lotto 47 Lottery, a player must choose 6 numbers between 1 and 47, inclusive. Six balls numbered from 1 and 47 are then randomly selected from an urn. The random variable X represents the number of matching numbers.

(a) What is the probability of matching 3 numbers?
(b) What is the probability of matching 4 numbers?
(c) What is the probability of matching 5 numbers?
(d) What is the probability of matching 6 numbers?
(e) A winning ticket is one in which the player matches 3, 4, 5, or 6 numbers. What is the probability of purchasing a winning ticket? Would it be unusual to purchase a winning ticket?
(f) What is the mean and standard deviation of the random variable X? For a randomly selected ticket, how many numbers do you expect to match?

Solution to (a) and (c) only:

What makes this problem interesting to me is that I need to massage it so that I can apply the Hypergeometric Formula. The first thing I want to do is to change the order of events. So, without altering the probabilities involved, the winning numbers are first selected and hiddden from the public. There are 6 of them, and they represent the fixed 'successes' among the total. Second, the players may now choose a sample size of number from the set of numbers 1 to 47. So, the act of choosing 6 numbers is equivalent to a random pick because the winning numbers are not known to the public.

Therefore, we have these parameters: $N = 47$, $K = 6$, $n = 6$. The value of k depends on which question we're trying to answer. For (a), $k = 3$. Repeating the PMF, we have

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}\,. \tag{19}$$

Then,

$$P(X = 3) = \frac{\binom{6}{3}\binom{47-6}{6-3}}{\binom{47}{6}} \approx 0.01986\,. \tag{20}$$

[3]This problem was found at http://staffwww.fullcoll.edu/kchang/MATH%20120 %20Primavera%202012/6.4_HypergeometricProbability.pdf, p. 6–6, problem 17.

For (b), $k = 5$,

$$P(X = 5) = \frac{\binom{6}{5}\binom{47-6}{6-5}}{\binom{47}{6}} \approx 0.00002291\,. \qquad (21)$$

6 Problem 5: Random Walk on Gambler's Ruin

Here I present my notes to the YouTube presentation[4] given by Professor Tom Leighton on the Gambler's Ruin (Mathematics for Computer Science, Fall 2010: tawkaw OpenCourseWare, Lecture 25).

We will model this situation as a gambler who goes to the casino to play roulette. You can play for a dollar a play in this version. That means that on each play you will either win a dollar or lose a dollar. The game continues until you win m dollars more than you started with, say a \$1000. There are 36 numbered slots, 1–36, half are red, half are black, and there are two more slots: a 0 and a 00. To model the random-walk situation, you can bet on either red or on black.

The chance of winning is $p = 18/38 \approx 0.473$, and the chance of losing is $1 - p = 20/38$. Surprisingly, it is virtually imposible to win a hundred dollars before losing your \$1000. The outcome on each bet is independent of all previous bet outcomes. If $p = 1/2$ the random walk is said to be unbiased.

We start with n dollars. The lower boundary to this play is \$0. That is, if we go 'broke' playing this game, the game is over. The upper boundary is $n + m$ dollars and we declare (arbitrarily) that we have won and the game is over. (We should remember that the current state, tracked by the variable n, is always a new 'starting state'.)

Definitions:

1) Introduce our goal parameter $T = n + m$.

2) W^* is the event that our goal parameter T is reached before we go broke.

3) D is the number of dollars we start with.

4) The variable n is the number of dollars in the current state.

5) Let X_n be the probability of winning if we start with n dollars, or,

$$X_n = \Pr(W^*|D = n)\,. \qquad (22)$$

Theorem:

$$X_n = \begin{cases} 0 \text{ if } n = 0, \\ 1 \text{ if } n = T, \\ pX_{n+1} + (1-p)X_{n-1} \text{ if } 0 < n < T\,. \end{cases} \qquad (23)$$

[4] https://www.youtube.com/watch?v=56iFMY8QW2k.

Proof:

$$X_0 = \Pr(W^*|D=0) = 0\,, \tag{24a}$$
$$X_T = \Pr(W^*|D=T) = 1\,. \tag{24b}$$

We will consider (24a) and (24b) as our boundary conditions, leaving (22) for for when $0 < n < T$.

Now, define E = 'event wins a dollar on 1st bet' and $\bar{E}$ = 'event loses a dollar on 1st bet'. So, using the Law of Total Probability with virtual emplacement of $E \vee \bar{E}$, we get that[5]

$$\begin{aligned}
X_n &= \Pr(W^*|D=n) \\
&= \Pr(W^* \wedge (E \vee \bar{E})|D=n) \\
&= \Pr(W^* \wedge E \,|\, D=n) + \Pr(W^* \wedge \bar{E} \,|\, D=n) \\
&= \Pr(E \,|\, D=n)\Pr(W^*|E \wedge D=n) \\
&\qquad + \Pr(\bar{E} \,|\, D=n)\Pr(W^*|\bar{E} \wedge D=n)\,.
\end{aligned} \tag{25}$$

See the footnote to fill in some of the details that were left out.[6]

Now, since the probability of winning or losing is independent of the state we're in, $\Pr(E \,|\, D=n) = p$ and $\Pr(\bar{E} \,|\, D=n) = 1-p$, and that

$$E \wedge (D=n) \implies D=n+1 \quad \text{and} \quad \bar{E} \wedge (D=n) \implies D=n-1\,, \tag{27}$$

therefore

$$\begin{aligned}
X_n &= p\Pr(W^*|D=n+1) + (1-p)\Pr(W^*|D=n-1) \\
&= pX_{n+1} + (1-p)X_{n-1}\,.
\end{aligned} \tag{28}$$

Or,

$$pX_{n+1} - X_n + (1-p)X_{n-1} = 0 \quad \text{with} \quad X_0 = 0, \quad X_T = 1\,. \tag{29}$$

This is a linear, homogeneous recurrence equation. Its characteristic equation is:

$$pr^2 - r + (1-p) = 0\,, \tag{30}$$

[5] We will use the Distributive Law from logic: $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$.

[6] Here's how we arrived at $\Pr(W^* \wedge E \,|\, D=n) = \Pr(E \,|\, D=n)\Pr(W^*|E \wedge D=n)$: Applying (1),

$$\begin{aligned}
\Pr(W^* \wedge E \,|\, D=n) &= \frac{\Pr(W^* \wedge E \wedge (D=n)}{\Pr(D=n)} \\
&= \frac{\Pr(W^* \wedge E \wedge (D=n))\Pr(W^* \wedge (D=n))}{\Pr(W^* \wedge (D=n))\Pr(D=n)} \\
&= \Pr(W^*|E \wedge (D=n))\Pr(E \,|\, (D=n))\,.
\end{aligned} \tag{26}$$

with roots 1 and $\dfrac{1-p}{p}$. If $p \neq \frac{1}{2}$, then, by the method of undetermined coefficients,

$$X_n = A\Big(\frac{1-p}{p}\Big)^n + B(1)^n = A\Big(\frac{1-p}{p}\Big)^n + B\,. \tag{31}$$

Using the boundary conditions, we get

$$A = \frac{1}{\Big(\frac{1-p}{p}\Big)^T - 1}\,, \qquad B = \frac{-1}{\Big(\frac{1-p}{p}\Big)^T - 1}\,. \tag{32}$$

Hence,

$$X_n = \frac{\Big(\frac{1-p}{p}\Big)^n - 1}{\Big(\frac{1-p}{p}\Big)^T - 1}\,. \tag{33}$$

Since $p < 1/2$,[7] then $\dfrac{1-p}{p} > 1$, and since for $T > n$,

$$\Big(\frac{1-p}{p}\Big)^T > \Big(\frac{1-p}{p}\Big)^n\,. \tag{34}$$

Therefore,

$$X_n = \frac{\Big(\frac{1-p}{p}\Big)^n - 1}{\Big(\frac{1-p}{p}\Big)^T - 1} < \frac{\Big(\frac{1-p}{p}\Big)^n}{\Big(\frac{1-p}{p}\Big)^T} = \Big(\frac{p}{1-p}\Big)^{T-n}\,, \tag{35}$$

where we have used that, for $a < b$, $\dfrac{a}{b} < \dfrac{a+1}{b+1}$. From which we have that

$$X_n < (\frac{p}{1-p})^m\,. \tag{36}$$

Let's use this equation for our roulette problem, with $m = 100$ and $p = 9/19$ and $\frac{1-p}{p} = \frac{9}{10}$ then

$$X_n = \Pr(\text{winning \$100 (before losing your stake)}) < \Big(\frac{9}{10}\Big)^{100} < 1/37,648\,. \tag{37}$$

[7]Of course, it's biased towards the House.

7 Problem 6: Computers and RAM

[8] At a technology consulting firm with x computers, all of which are laptops or desktops, 30% are laptops; if 80% of the total number of computers have more than 1GB of RAM and 10% of the computers with less than 1 GB of RAM are laptops (and no computers have exactly 1GB of RAM), approximately what percent of desktops have more than 1GB of RAM?
A) 75%
B) 60%
C) 52%
D) 40%
E) 45%

The correct answer is reported to be A).

Solution

So, where to begin? For the moment let's ignore the details given us in the form of numbers, percentages, etc, and just concern ourselves with the objects presented and their characteristics: There are two kinds of objects, namely, laptops (LTs) and desktops (DTs). Both of these objects share the property of having RAM, characterized in two different types: the type > 1GB and the type < 1 GB. Well, two objects with two characteristics – that sounds like we should create a foursquare to represent the data given and to extract the conclusions we need. We'll start with Figure 8, which already contains the given information.

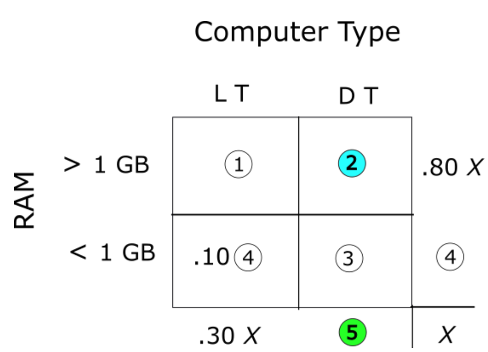

Figure 8. We start the process of constructing the foursquare by first noting that the bottom row and the rightmost column must each separately add to X, the total number of computers. The blue and green colors highlight the squares whose values we're looking for.

Now, we're asked to find the percentage of desktops that have more that 1 GB of RAM. We can get at this number by taking the ratio of the number in location

[8]Found at https://www.beatthegmat.com/650-800-question-t68610.html .

Circle 2 and dividing it by the number in location Circle 5, because we are interested in the ratio compared to the total number of desktops, not the total number of computers altogether.

Circle 5 is equal to $.70X$ because the bottom row must add up to X. Circle 4 is $.20X$ because its column must add up to X. Therefore .10 x (Circle 4) = $.20X$.

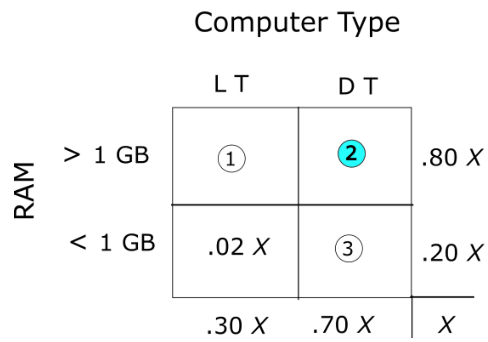

Figure 9. We are told that 10% of the computers with less than 1 GB of RAM are laptops. With this information we can calculate what should go into Circle 3.

The value of Circle 3 in Figure 9 is $.18X$ because its row must add to $.20X$. Therefore, the value of Circle 2 is $.52X$ because its column must add to $.70X$.

Finally, the ratio of $.52X$ to $.70X$ is .7429. When we round off this to a whole number percentage, we get about 75%.

8 Problem 7: A GMAT Problem

[9] The events A and B are independent. The probability that event A occurs is 0.6, and the probability that at least one of the events A or B occurs is 0.94. What is the probability that event B occurs?
(A) 0.34
(B) 0.65
(C) 0.72
(D) 0.76
(E) 0.85

Solution:

[9] Found at http://www.beatthegmat.com/mba/2011/04/23/gmat-sample-problem-probability -problem-solving-2.

Of course we will use a similar solution strategy to the last problem, employing a probability foursquare, only this time we will use more variables than circles. See Figure 10.

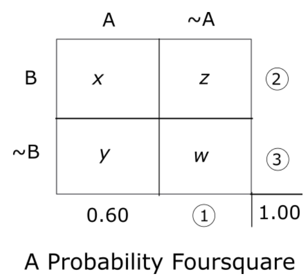

Figure 10. Since we were given that the probability of event A occurring is .60, I already entered that value at the bottom of column ‘A’. Naturally, the value to be entered at Circle 1 is 0.40.

We were also given that $\Pr(A \cup B) = 0.94$. If I hadn't gone to the trouble to construct this probability foursquare to aid in the solution, I would probaby just employ the Inclusion-Exclusion Principle to expand $\Pr(A \cup B)$. But since I have the foursquare in use, I'll instead write

$$\Pr(A \cup B) = x + y + z = .94\,. \tag{38a}$$

But we also have that

$$x + y = .60\,. \tag{38b}$$

Therefore

$$z = .34\,. \tag{39}$$

And that implies that $w = .06$.

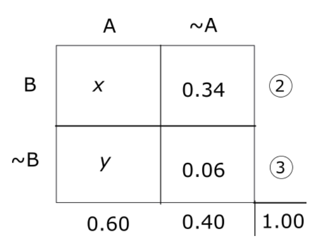

Figure 11. We are proceeding well. We've been asked to determine the probability of B, which is the value that belongs in Circle 2.

There's one last piece of given information: Events A and B are independent, or

$$\Pr(A \cap B) = \Pr(A)\Pr(B)\,. \tag{40}$$

From Figure 11, we can write

$$x = (.60)\,\Pr(B)\,, \tag{41a}$$

and

$$x + .34 = \Pr(B)\,. \tag{41b}$$

The solutions to these are

$$x = .51 \quad \text{and} \quad \Pr(B) = .85\,. \tag{42}$$

We'll go ahead and complete the foursquare as a double-check on our result for the probability of B:

	A	~A	
B	0.51	0.34	0.85
~B	0.09	0.06	0.15
	0.60	0.40	1.00

Figure 12. The foursquare is complete and everything adds up properly.

9 Problem 8: Find the PMF for the Geometric Distribution

I am again following a presention by Elliot Nicholson on Youtube.[10] The Geometric Distribution relies on a knowledge of the Bernoulli Distribution, which maps on a given trial the success state to the formal probability space element 1, and failure is mapped to 0.

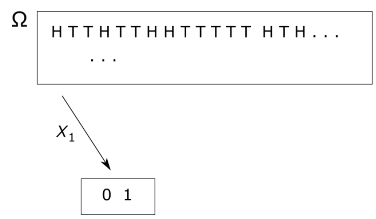

Figure 13. Sample space Ω containing all possible infinitely long strings of H's and T's. We also show X_1, the first of many random variables, which we'll use to break down the complicated probability into simpler Bernoulli factors.

The Geometric Distribution[11] seeks to answer the question: What is the probability that i failures preceed the first success. We'll model this in the form of the case of flipping a coin and calling heads success and tails failure. Each flip (trial) is independent of all others. The probability of heads we'll call p and the probability of tails we'll call $q = 1 - p$.

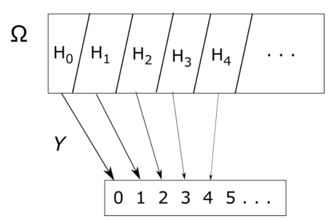

Figure 14. Sample space Ω is partitioned into groups H_i according as the the number i of tails that preceed the first head. The random variable Y maps each partition onto the number i of the nonnegative integer set.

We'll introduce as many random variables X_j as we need, each one maps (on the jth flip) a tail to 0 and a head to 1. In notation: $X_j \sim \mathrm{Bern}(p)$.

[10] https://www.youtube.com/watch?v=prSf31ttKhI.

[11] The description 'Geometric' comes from the fact that to verify that the PMF derived really does add to 1 when adding over all probabilities, we get a simple geometric series to sum.

$$
\begin{aligned}
P(H_i) &= P(\text{tail}_1,\ \text{tail}_2,\ ...,\text{tail}_i,\ \text{head, head or tail, ...}) \\
&= P(X_1 = 0, X_2 = 0, ..., X_i = 0, X_{i+1} = 1, \ldots) \\
&= P(X_1 = 0) \cdots P(X_i = 0) P(X_{i+1} = 1) P(X_{i+2} = 1 \vee X_{i+2} = 0) \cdots \\
&= q^i p \cdot 1 \cdots = q^i p \qquad (43)
\end{aligned}
$$

Thus, PMF=$P(Y = i) = q^i p$. However, to prove that $P(Y = i)$ is a PMF, we have a consistency requirement to prove, namely:

$$\sum_{i=0}^{\infty} P(Y = i) = 1\,. \qquad (44)$$

Thus, if we sum over all possible outcomes, we get

$$\sum_{i=0}^{\infty} P(Y = i) = \sum_{i=0}^{\infty} q^i p = p \sum_{i=0}^{\infty} q^i = \frac{p}{1-q} = 1\,, \qquad (45)$$

as was necessary, and this confirms that the sum of all probabilities for all simple events is unity.

10 Problem 9: Find the PMF for the Negative Binomial Distribution

I am again following a presention by Elliot Nicholson on Youtube.[12] The Negative Binomial Distribution is a generalization of the geometric distribution. In the latter case, we want to know the probability mass function of getting a success after i failures. In the present case, we want to know the PMF (probability mass function) of getting r successes after getting i failures.

We'll model the problem as flipping a coin, in which getting r heads (the successes) after getting i tails (the failures).

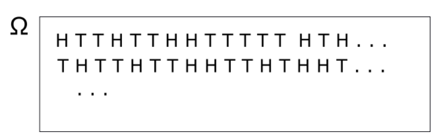

Figure 15. Sample space Ω containing all possible infinitely long strings of H's and T's..

Let p be the probability of getting a head on a single trial. Let r be the number of heads needed to stop flipping the coin. Let X be a random variable that maps Ω to the nonnegative integers.

[12]https://www.youtube.com/watch?v=LxeIXPToJl8&index=44&list=PLAvgI3H-gclbyWnE9WrSq68_HRW9rOZUw.

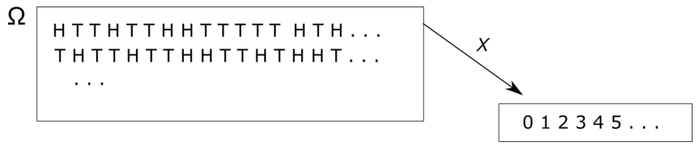

Figure 16. Random variable X maps Ω to the nonnegative integers.

Let $s \in \Omega$, then

$$X \mapsto \#\text{ tails preceding the } r\text{th head.} \tag{46}$$

A nice way to proceed here is to break up the problem into r number of subproblems. Let X_j be the random variable mapping the number of T's between the $(j-1)$st tail and the jth head. It's not hard to see that each X_j is distributed as Geom(p).

Now we write X as

$$X = X_1 + X_2 + \cdots + X_r\,. \tag{47}$$

Therefore,

$$\Pr(X = i) = \Pr(X_1 + X_2 + \ldots + X_r = i)\,. \tag{48}$$

Now, the i tails are situated between the r heads. Let's give these tails in each partition a symbol, namely i_k. Then

$$i_1 + i_2 + \ldots + i_r = i\,. \tag{49}$$

Then the individual random variables X_j are independent, individually distributed as geometric distributions of parameter p, or Geom(p). In other words, $\Pr(X_j = i_j) \sim \text{Geom}(p)$. Therefore, $\Pr(X_j = i_j) = pq^{i_j}$.

$$\begin{aligned}\Pr(X_1 + X_2 + \cdots + X_r = i) &= \sum_{i_1+i_2+\ldots+i_r=i} \Pr(X_1 = i_1)\Pr(X_2 = i_2)\cdots\Pr(X_r = i_r)\\ &= \sum_{i_1+i_2+\ldots+i_r=i} (pq^{i_1})(pq^{i_2})\cdots(pq^{i_r})\\ &= \sum_{i_1+i_2+\ldots+i_r=i} p^r q^i\,. \end{aligned} \tag{50}$$

Now, each term in the above expansion will have the same factor $p^r q^i$. So, how many ways can we distribute i objects into r partitions? The answer is $\binom{i+r-1}{r-1}$ ways. Therefore,

$$\text{PMF} = \binom{i+r-1}{r-1} p^r q^i\,. \tag{51}$$

To establish the expectation value of this distribution, we use the theorem that the expectation value of a sum of random variables is the sum of the expectation values, or,

$$
\begin{aligned}
E(X) &= E(X_1 + X_2 + \ldots + X_r = i) \\
&= E(X_1) + E(X_2) + \ldots + E(X_r) \\
&= \sum_{k=1}^{r} \frac{q}{p} = \frac{qr}{p} \,. \qquad (52)
\end{aligned}
$$