

Preface

Probability problems fascinate just about everyone, despite wildly varying backgrounds that span the sophistication spectrum. Mathematicians love them because probability theory is so beautiful, one of the gems of mathematics. Physicists love them because probability often is the key to answering many of their technical questions. And everybody loves probability problems simply because they can be a lot of fun. A probability problem can simultaneously be so easy to state that anybody with a brain can understand the question, while also being so puzzling that even the experts are stumped (at least for a while).

Here's an example of what I mean by that. Suppose we have an urn with 100 red balls and 100 black balls in it. Suppose further that we select balls, one after the other, *with replacement*—that is, after selecting a ball we put it back in the urn. Then,

- (a) What's the probability the first ball is red? The answer is, of course, $1/2$.
- (b) What's the probability the second ball is red? Ho, hum, you say, it's $1/2$, too, of course!

Now, suppose we select balls *without* replacement, that is, after drawing a ball we discard it. Then,

- (c) What's the probability the first ball is red? My gosh, you say, this is getting boring—it's clearly still $1/2$ because it doesn't matter what we do with first ball after we select it.
- (d) What's the probability the second ball is red? Okay, you shout, at last a nonobvious question! *Now*, you surely say, the answer depends on what the first ball was because if the first ball was red, then there would be just 99 red balls in the urn for the second drawing, while if the first ball was black, then there would still be 100 red balls available but only 99 black balls.

xvi Preface

This does seem to be more complicated than were the earlier cases. What we *can* write is the following, using what are called *conditional* probabilities:

$$\begin{aligned} \text{Prob}(\text{second ball is red}) &= \\ &\text{Prob}(\text{second ball is red given that the first ball is red}) \times \\ &\text{Prob}(\text{first ball is red}) \\ &+ \text{Prob}(\text{second ball is red given that the first ball is black}) \times \\ &\text{Prob}(\text{first ball is black}) \\ &= (99/199)(100/200) + (100/199)(100/200) = (100/200) \\ &\quad (99/199 + 100/199) = (100/200)(199/199) = \frac{1}{2}. \end{aligned}$$

Now, if this result doesn't surprise you—it's *still* 1/2—well, all I can say is that I used this example in my classes for decades, and it still makes *me* scratch my head!

Or how about this amazing result. Each of a million men puts his hat into a very large box. Every hat has its owner's name on it. The box is given a good shaking, and then each man, one after the other, randomly draws a hat out of the box. What's the probability that at least one of the men gets his own hat back? Most people would answer with “pretty slim,” but in fact the answer is the astonishingly large 0.632! Who would have guessed that?

Probability has much to offer to analysts who fall outside the usual techno-types, too. For example, students of the social implications of new laws can often use the mathematics of chance to explore what might happen in the future. That claim may seem a bit murky, so let me provide a specific example of what I mean. The topic of what to do about undocumented immigrants has been simmering for years in America, and exploded onto center-stage during the intense political debates of the 2012 presidential election year. One proposed law would allow police officers to stop any person, any time, and demand to see citizenship identification papers. There are certainly all sorts of issues one might have with such a law, but let's concentrate on one in particular—how intrusive, that is, how *inconvenient* would such a law be? We can model this question as a traditional probability problem, as follows.

Imagine America as a huge urn and its population as balls: b black balls for legal residents and r red balls for undocumented residents.

A police stop is equivalent to drawing a ball at random from the urn. A black ball (legal) is returned to the urn, while a red ball (undocumented) is permanently discarded (that is, deported). Our inconvenience question is then just this: on average, how many times is each black ball drawn (a legal resident stopped by police) before 50%, or 90%, or 99% of the red balls (undocumented residents) have been removed? If we can answer this question (and we will later in this book), then the specific answers are actually not the central point. What is the central point is that a probability analysis will have given us some numbers to discuss, and not just emotionally charged words for advocates on each side to throw at each other like rocks.

I taught probability theory and its applications to electrical engineering students at both the undergraduate and graduate level (at the University of New Hampshire and the University of Virginia) for nearly thirty years. I sincerely hope, in my retirement, there are in the world still students of mine who, if asked, would say “Yes, Professor Nahin did teach me some stuff that I have found, now and then, to be useful.” But of course, the learning process isn’t just a one-way process. As any honest teacher will admit, it isn’t only the students who learn new things in school, and I was no exception. Two very important lessons I learned during those three decades of lecturing and chalking mathematics on blackboards were:

- (1) “Obvious” results are boring; students are interested in (and pay attention to) calculations that result in nonintuitive and/or surprising (almost “seemingly impossible”) results.
- (2) While theoretical proofs are nice, and indeed to be hoped for, students are inherently skeptical. They love to ask, “But how can you be *sure* you haven’t overlooked something, or haven’t made some subtle error of reasoning?”

There are numerous examples of both cases in modern books on probability theory and its applications. To illustrate case (1), for example, the well-known birthday problem is virtually certain to appear in any undergraduate probability book pulled at random from the stacks of a college library. The problem is easy to state. Assume birthdays are randomly distributed uniformly over all the days in the year. Then, if students are asked how many people would have to gather together to achieve a probability of at least 0.5 that two (or more) of them have

xviii Preface

the same birthday (month and day), most guess a “large” number such as 183 (the first integer greater than half of 365). Literally *everybody* is astonished at how small the actual value is (just 23). If we raise the probability to 0.99, the number of people required increases only to 55, still an astonishingly small value. A twist on the birthday problem, with an equally surprising answer, is to ask how many people you’d have to ask to have a probability of at least $1/2$ to find somebody with *your* birthday. Now the answer *is* large, even greater than 183 (it’s 253).

I have often used the first birthday question in large introductory classes (say, 40 to 50 students) by having the students, one after the other, call out their birthday; to hear the collective gasp of astonishment when, almost always, there would be a match (often very quickly, too) was a lot of fun for all. (Once we had a triple match!) The students were intrigued by the possibility of actually being able to calculate such an amazing result, and thereafter paid more attention to what I said in class—or at least they usually did until the end of that lecture! So the two birthday problems are undeniably great problems to include in any beginning course on probability theory. But that’s the very reason they are not analyzed here: they are simply too well known to be included in this book. To get into this book, a case (1) problem has to be both amazing and not well known. For this reason, too, the wonderful Buffon needle problem for “experimentally” determining the value of π is also a no-show here.

In some respects, this book resembles the 1965 classic by the late Harvard mathematician Frederick Mosteller, *Fifty Challenging Problems in Probability with Solutions*. Nearly half a century has passed since its appearance, and most of the problems in that book have become regulars in textbooks. They are all great problems, but they are no longer “curiously odd and strange” problems. The million-hats-in-a-box puzzle I mentioned earlier, for example, is discussed in Mosteller’s book, but even when he wrote, the problem was already centuries old, dating back to the early 1700s.

I’ve already used the balls-in-urns imagery twice in this preface, so here’s another fun example of what I do think amazing, in the form of another balls-in-urns puzzle for you to mull over as you finish reading this essay. I’ll give you the answer at the end. Suppose you have 10 white balls and 10 black balls that you can distribute any way you like between

two urns. After you've done that, you give a friend a fair coin and he selects an urn at random (tails he picks one, heads he picks the other). From that urn he draws a ball at random. How should you have distributed the balls to maximize the probability your friend draws a white ball? If, for example, you had put all the white balls in one urn and all the black balls in the other urn, then the probability of drawing a white ball would be $1/2(10/10) + 1/2(0/10) = 1/2$. On the other hand, if you had put 5 black balls and 5 white balls in each urn, the probability of drawing a white ball would still be $1/2$ because $1/2(5/10) + 1/2(5/10) = 1/2$. You can do a lot better than this, however, with a different distribution, and I think you'll be amazed at how much better you can do. And once you've found the answer, can you then see what distribution of the balls *minimizes* the probability that your friend draws a white ball?

Here's another probability puzzler with an amazing answer, made doubly amazing because no math is required, just logical reasoning. Suppose there are 100 people at an airport waiting to board a plane. All have a boarding pass, with an assigned seat. It's a full flight; there are exactly 100 seats on the plane. The first person to board is a free spirit and simply takes a seat at random. It might in fact *be* his assigned seat, but if it is, it's just by chance. Thereafter, however, all other persons boarding, one after the other and being more willing to follow the rules, each do go to their assigned seat and take it unless it is already occupied. In that case they too simply take an unoccupied seat at random. What is the probability that the last person to board still gets his assigned seat? The answer is at the end of this preface.

What counts as amazing does not include "surprising" results arrived at by erroneous reasoning. Here are two examples of what I mean by that perhaps curious comment. For my first example, in her July 31, 2011, "Ask Marilyn" column in *Parade Magazine*, the famous (some mathematicians might substitute the word *infamous*) puzzler Marilyn vos Savant asked her readers the following probability question:

Say you plan to roll a [fair] die 20 times. Which of these results is more likely:

- (a) 11111111111111111111
- (b) 66234441536125563152?

xx Preface

The answer given by vos Savant was,

In theory, the results are equally likely. Both specify the number that must appear each time the die is rolled. . . . Each number (1 through 6) has the same chance of landing face-up [that is, $1/6$]. But let's say you tossed a die out of my view and then said that the results were one of the above. Which series is more likely to be the one you threw? Because the roll has already occurred, the answer is (b). It's far more likely that the roll produced a mixed bunch of numbers than a series of 1's.

Is vos Savant's answer to the original question correct? Is vos Savant's answer to her extended version of the original question correct? Think about this for now. I'll give you the correct reasoning in just a bit.

For a second vos Savant example of really bad mathematical reasoning, consider her Christmas 2011 column in *Parade Magazine*, in which she printed the following letter from a reader:

I manage a drug-testing program for an organization with 400 employees. Every three months, a random-number generator selects 100 names for testing. Afterward, these names go back into the selection pool. Obviously, the probability of an employee being chosen in one quarter is 25 percent.

But what's the likelihood of being chosen over the course of a year?

Marilyn's answer was,

The probability remains 25 percent, despite the repeated testing. One might think that as the number of tests grows, the likelihood of being tested increases, but as long as the size of the pool remains the same, so does the probability. Goes against your intuition, doesn't it?

Yes, it certainly does, perhaps because her comments are so astonishingly, indeed breathtakingly, wrong. Again, think about her correspondent's question while you read, and I'll give you the correct solution in just a bit.

Now, what about case (2) that I mentioned earlier? What I did when teaching, and have done in two previous probability books published

by Princeton University Press (*Duelling Idiots and Other Probability Puzzlers* and *Digital Dice*), was endorse the use of computer simulation to check theoretical results. If a computer simulation of a random process agrees (to some acceptable degree) with a theoretical result, then I think one's confidence in both approaches is enhanced. Such an agreement doesn't, of course, prove that either result is correct, but surely one would then have to believe that a remarkable coincidence had instead occurred.

So computer simulations play a big role in this book, but I want to emphasize that it is a book on analysis and not a programming book. I use MATLAB[®], but I've written all the computer codes in such a low-level way that you should have little trouble converting any of them to your favorite language. MATLAB[®] experts may have to bite their tongues, however, at my avoidance of the advantages of the powerful vector/matrix structure of the language. That structure does indeed allow spectacular decreases in computational times, when compared to what can be done with *for*, *if/else*, and *while* loops. I've made extensive use of such loops, however, precisely because I don't want to restrict this book's codes to being MATLAB[®]-only codes.

All Monte Carlo codes, of course, use a random number generator, a feature available in any modern scientific programming language (every time you call the generator, you get back a number from a distribution uniformly spread from 0 to 1), and MATLAB[®] is no exception. For more on how random number generators actually work—something you really don't need to know to use them—see *Duelling Idiots* (pp. 175–197). You may find the brief Technical Note at the end of this book helpful, too.

There is an interesting feature to doing computer simulations that I have noticed, after decades of writing computer codes in different languages to implement them. Problems that are hard to do theoretically may require only an easy simulation; the converse may also be true, that is, a problem easy to analyze theoretically may require a complicated simulation code. In the introduction you'll find examples of the use of a computer simulation to check a theoretical derivation, as well as an amusing problem (until recently unsolved) for which I think a computer simulation would be very difficult to write (I don't even try, but don't let that stop *you*).

In the rest of this book I'll assume that you have had some encounters with probability arguments in the past. But I'll not assume it was

xxii Preface

necessarily the recent past! For example, I'll not bother with defining a binomial coefficient, or with extensive explanations either of Bayes' theorem from conditional probability or of why the expected value of a discrete random variable X that takes its possible values from the non-negative integers is given by

$$E[X] = \sum_{j=0}^{\infty} j \text{Prob}(X = j).$$

On the other hand, when a discussion includes the concepts of the joint probability density and distribution function for the continuous random variables X and Y , $f_{X,Y}(x,y)$ and $F_{X,Y}(x,y)$, respectively, I will remind you that

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

What to assume you already know and what you might need a reminder of is a bit of a guess on my part, and I can only hope I've been right more often than wrong. In any case, where I've guessed wrong, I assume you can look things up on your own.

Now, what about that die-rolling question posed by Marilyn vos Savant? In fact, vos Savant was correct with her first answer but wrong with her second. Indeed, her proper explanation for the first part is precisely why she is wrong in the second part. Both sequences are still equally likely. I finally understood the origin of her confusion when I read her follow-up column in *Parade Magazine* on October 23, 2011. In that column she printed a letter from a reader who correctly claimed that vos Savant was wrong. Vos Savant continued to claim she was right, and then presented a meaningless, almost incoherent defense; at one point she argued that she had actually rolled, once, a die 20 times and, since she did get a jumbled sequence of numbers, that proved her right. Her last line was the clue for me to where she had stumbled: after rolling a die 20 times the result is, she stated, "far more likely to be . . . a jumble of numbers."

And, of course, that is correct, because there are a lot of jumbled sequences that could occur but only one sequence of all 1s. But that's all beside the point because that wasn't the original question. The original question asked about comparing the probability of the sequence of

20 1s with the probability of *one particular* jumbled sequence of 20 digits, namely, 66234441536125563152. Rolling the die “out of her view” is simply irrelevant. And after rereading her comments in the original July 31 column, I realized she had already made the “mixed bunch of numbers” assertion. When you present a problem, ask for a solution, and then correctly answer a *different* question, that doesn’t give you the right to claim everybody else is wrong about the original question.

What about vos Savant’s drug testing question? Although I suspected it was going to be a waste of my time, I couldn’t resist sending the following e-mail to her via *Parade Magazine*.

Dear Marilyn,

*If the probability of being selected for a quarterly drug test is 0.25, then the probability of **not** being selected is 0.75. Thus, the probability of **not** being selected four straight times (four quarters in a year) is $(0.75)^4 = 0.3164$.*

That means the probability of being selected at least once is $1 - 0.3164 = 0.6836$, not the 0.25 you stated.

Best, Paul

Paul J. Nahin

Professor Emeritus of Electrical Engineering/University of New Hampshire

Not surprisingly, I never received a reply, but in her January 22, 2012, column she finally admitted her mistake, writing, “My neurons must have been napping.” There then followed some hand-waving comments that seemed to imply she was thinking of the constant probability of being selected for any given quarterly test, but why should that “go against your intuition”? It is, in fact, quite clear that this trivial interpretation is not what she had in mind when she wrote her original answer.

This all just goes to show that, when it comes to probability questions, even a person with the “highest recorded IQ” can tumble headfirst into a tar pit of nonsense, even with as elementary a question as the die-tossing problem, thus illustrating for us all the value and importance of due diligence. In the introduction you’ll see how even the super-genius of Isaac Newton took a tumble over a probability problem involving dice. As the great Victorian mathematician Augustus De Morgan (1806–1871) once wrote, “Everyone makes errors in probabilities, at times, and big ones.”

One tumble that De Morgan might have had in mind was mentioned by one of his fellow countrymen, Isaac Todhunter (1820–1884), in his 1865 classic, *A History of the Mathematical Theory of Probability*. There, commenting on the faulty reasoning by the French mathematician Jean Le Rond d’Alembert (1717–1783) in analyzing a coin-flipping problem, Todhunter wrote, “Any person who wishes to see all that a great mathematician could produce on the wrong side of a question should consult [D’Alembert’s 1754 essay].” So, even mathematicians make mistakes in their arithmetic (see Problem 24 for a discussion of D’Alembert’s goof). I do suspect, however, that the ghosts of De Morgan and Todhunter would swallow *very* hard at Vos Savant’s prize-winning goofs!

A final comment on the nature of the problems you’ll find in this book. With few exceptions, they are not simply “fun puzzles for a party,” as are the birthday problems. The level of difficulty varies over a pretty wide spectrum. Some are straightforward applications of not much more than high school algebra (but with surprising conclusions), while others—not to be overly dramatic—are mathematically very involved. These latter problems are more properly thought of as “puzzles at the edge of the doable.” That, I think, is their special charm, that they have extreme complexity that stays just within the borderline reach of undergraduate mathematics. Now, lest the book be thought all serious business, a few of the puzzles perhaps are more in the party category than not, in particular the penultimate one on chickens in boxes that I’ve taken from an old Marilyn vos Savant column (a puzzle question with which she challenged readers but, as far as I know, she never answered). So, I think that both partygoers and dedicated analysts will find stimulating problems on the pages that follow.

Solution to the Ball-Distribution Puzzle

Put b black balls and w white balls into one urn, and the remaining $10 - b$ black balls and $10 - w$ white balls into the other urn. The probability of drawing a white ball from a randomly selected (using a fair coin) urn is then given by

$$\left(\frac{w}{w+b}\right)\frac{1}{2} + \left(\frac{10-w}{20-b-w}\right)\frac{1}{2}.$$

There are just enough possibilities to check to make it a pain to do them all by hand, but for a computer code it's duck soup. And that's what I did, letting the code run through all the possibilities of w and b , each independently varying from 0 to 10 (a total of 121 combinations); the code found that the maximum probability occurs when $b = 0$ and $w = 1$. That is, one urn has just a single white ball, while the other urn has all 10 black balls and the remaining 9 white balls. This gives the probability for drawing a white ball as

$$(1)\left(\frac{1}{2}\right) + \left(\frac{9}{19}\right)\left(\frac{1}{2}\right) = \frac{14}{19} = 0.7368, \quad (1)$$

a value significantly greater than 0.5. Notice that minimizing the probability of drawing a white ball is equivalent to maximizing the probability of drawing a black ball. By symmetry with the original question we know that $b = 1$ and $w = 0$ is the distribution that maximizes the probability of drawing a black ball (which is, of course, again $14/19$). But that means the minimum probability for drawing a white ball must be $1 - 14/19 = 5/19 = 0.2632$.

Solution to the Airplane-Seating Puzzle

The answer is $1/2$, and not just for 100 people but for any number of people. The key to understanding this problem is realizing that the last person to board will not end up with just any seat out of the 100 (or whatever the number may be) seats on the plane. Instead, he will get either his assigned seat or the seat assigned to the first person who boards. This claim initially strikes most people as surprising, but here's how it works. If the first person to board does take (by chance) his assigned seat then the last person to board definitely gets his assigned seat because all the other people (following the rules) go to their assigned seats. If, on the other hand, the first person to board takes, let's say, seat 10, then the next persons to board with seats 2 through 9 take their seats, too. The next person, the 10th person, then has three possibilities: (1) he takes the last person's seat, (2) he takes the first person's seat, or (3) he takes some other still empty seat. If he takes the first person's seat, then all the other people (including the last person) get their assigned seat. If he takes the last person's seat, then all the other

xxvi Preface

people get their assigned seat except for the last person who has to take the lone remaining seat, the seat assigned to the first person. Finally, if he takes some other seat, say seat 27, then we are back in the same situation with persons 11 to 26 (who simply go to their assigned seats) and the fellow with seat 27 playing the same role as did the person with seat 10. The end result is that the last person to board ends up either with his assigned seat or with the assigned seat of the first person. Every time a person who is boarding can't take his assigned seat because it's occupied, he chooses randomly among all the remaining seats, which includes the seats of the first and the last persons to board. So, at all times there is nothing that distinguishes these two particular seats, and thus they have the same probability of being the seat of the last person to board. And since they are the *only* possible seats for the last person to board, each has probability $1/2$.