Chapter 14

Great Expectation

The *expectation* or *expected value* of a random number is the average value that is anticipated (via the Law of Large Numbers) when the random number is "re-drawn" independently many times. If the random number can take only finitely or countably many values (e.g., if it takes only integer values), we can define its expectation as the average of the values, weighted by their probabilities.

Numbers whose values are determined by chance are called by probabilists *random variables*, and denoted typically by boldface capital letters from the end of the Latin alphabet, such as $\mathbf{X}$. If $\mathbf{X}$'s possible values are $x_1, x_2, \dots$ with corresponding probabilities $p_1, p_2, \dots$, then the expectation of $\mathbf{X}$ is

$$\mathbb{E}\mathbf{X} := \sum_i p_i x_i.$$

The terminology is a bit misleading; for example, the expectation when you roll a fair die is $3\frac{1}{2}$, but you don't "expect" the result $3\frac{1}{2}$. It would be more accurate to call $3\frac{1}{2}$ the "expected average" rather than the "expectation" or "expected value." But convenience triumphs.

Expectation is often easier to work with than probability. Suppose $\mathbf{X}$ and $\mathbf{Y}$ are two random variables. If we want to know the probability that $\mathbf{X} + \mathbf{Y} > 5$ (say) it is not enough to know everything there is to know about each of $\mathbf{X}$ and $\mathbf{Y}$ *individually*; you will also need to know how $\mathbf{X}$ and $\mathbf{Y}$ interact with each other.

But if you only want the *expectation* of $\mathbf{X} + \mathbf{Y}$, it is guaranteed to satisfy

$$\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}\mathbf{X} + \mathbb{E}\mathbf{Y}$$

so all you need to know are the expectation of $\mathbf{X}$ and the expectation of $\mathbf{Y}$. For example, suppose $\mathbf{X}$ and $\mathbf{Y}$ each take values 1, 2, 3, 4, 5, or 6 with equal probability, so that each has expectation $3\frac{1}{2}$. What does the random variable $\mathbf{X} + \mathbf{Y}$ look like? Well, if $\mathbf{X}$ and $\mathbf{Y}$ are independent—for example, if $\mathbf{X}$ is the first roll of a die and $\mathbf{Y}$ the second—then $\mathbf{X} + \mathbf{Y}$ takes on values $2,\ 3, \dots, 12$ with probabilities $1/36,\ 2/36, \dots, 5/36,\ 6/36,\ 5/36, \dots, 1/36$. But if $\mathbf{X}$ and $\mathbf{Y}$

are both the outcome of the first roll, that is, if $\mathbf{X} = \mathbf{Y}$, then $\mathbf{X}+\mathbf{Y} = 2\mathbf{X}$ takes on only even values from 2 to 12, each with probability $\frac{1}{6}$. And if $\mathbf{Y}$ happens to be defined as $7 - \mathbf{X}$, we find that $\mathbf{X} + \mathbf{Y}$ is the constant value 7. All very different, but notice that in every case $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = 3\frac{1}{2} + 3\frac{1}{2} = 7$.

This miracle is a special case of what mathematicians call "linearity of expectation," whose general formulation is

$$\mathbb{E}\sum_i c_i \mathbf{X}_i = \sum_i c_i \mathbb{E}\mathbf{X}_i,$$

that is, the expectation of a weighted sum of random variables is the weighted sum of their expectations.

Let's start with a calculation of expectation in the service of making a strategic decision.

Bidding in the Dark

You have the opportunity to make one bid on a widget whose value to its owner is, as far as you know, uniformly random between \$0 and \$100. What you do know is that you are so much better at operating the widget than he is, that its value to *you* is 80% greater than its value to him.

If you offer more than the widget is worth to the owner, he will sell it. But you only get one shot. How much should you bid?

Solution:

You should not bid. If you do bid $\$x$, then the expected value to the widget's owner, *given that he sells*, is $\$x/2$; thus, its expected value to you, if you get it, is $1.8 \cdot \$x/2 = \$0.9x$. Thus you lose money on average if you win, and of course, you gain or lose nothing when you do not, so it is foolish to bet. ♡

One common setting for expectation is waiting time; in particular, the average number of times you expect to have to try something until you succeed. For example, how many times, on average, do you need to roll a die until you get a 6? We encountered this sort of problem in *Odd Run of Heads* in Chapter 5. Let x be the expected number of rolls. "If at first you don't succeed" (probability $\frac{5}{6}$) you still have an average of x rolls to "try, try again," so $x = 1 + \frac{5}{6} \cdot x$ giving $x = 6$. The same argument shows that in general, if you conduct a series of "trials" each of which has probability of success p, then the average number of trials you need to get your first success is $1/p$.

Let's apply this observation to a slightly trickier dice puzzle.

Rolling All the Numbers

On average, how many times do you need to roll a die before all six different numbers have turned up?

Solution:

Often the most important step in applying linearity of expectation is simply recognizing that what you are being asked for is the expectation of a sum of random variables. The key for this puzzle is to think of rolling all the numbers as a 6-stage process, where at Stage i you are going for your ith new number. In other words, let $\mathbf{X}_i$ be the number of rolls needed to get your ith new number, given that you have already seen $i-1$ different numbers.

Then the total number of rolls needed is $\mathbf{X} := \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_6$ and what we are asked for is $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{X}_1 + \mathbb{E}\mathbf{X}_2 + \cdots + \mathbb{E}\mathbf{X}_6$. We already know that $\mathbb{E}\mathbf{X}_1 = 1$ since it always takes exactly one roll to get the first new number. The second will come with probability $5/6$ from any given roll, so takes $6/5$ rolls on average. Continuing in this manner, we have

$$\mathbb{E}\mathbf{X} = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 14.7$$

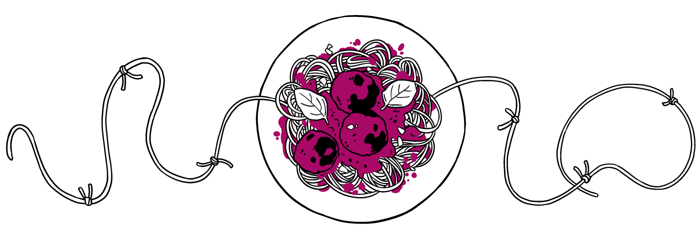
and we are done.

Warning: Some experimentation will remind you that $\mathbf{X}$ is by no means guaranteed to be near 14.7; the last new number, in particular, could take a frustratingly long time to roll. Later we'll be reminded that the Law of Large Numbers sometimes takes a while to kick in.

Next, see if you can set up a sum of random variables for this puzzle.

Spaghetti Loops

The 100 ends of 50 strands of cooked spaghetti are paired at random and tied together. How many pasta loops should you expect to result from this process, on average?



Solution:

The key here is to recognize that you start with 100 spaghetti-ends and each tying operation reduces the number of spaghetti-ends by two. The operation results in a new loop only if the first end you pick up is matched to the other end of the same strand. At step i there are $100-2(i-1)$ ends, thus $100-2(i-1)-1 = 101 - 2i$ *other* ends you could tie any given end to. In other words, there are

$101-2i$ choices of which only one makes a loop; the others just make two strands into one longer one.

It follows that the probability of forming a loop at step i is $1/(101-2i)$. If $\mathbf{X}_i$ is the number of loops (1 or 0) formed at step i, then the expected value of $\mathbf{X}_i$ is $1/(101-2i)$. Since $\mathbf{X} := \mathbf{X}_1 + \cdots + \mathbf{X}_{50}$ is the final number of loops, we apply linearity of expectation to get

$$\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{X}_1 + \cdots + \mathbb{E}\mathbf{X}_{50} = \frac{1}{99} + \frac{1}{97} + \frac{1}{95} + \cdots + \frac{1}{3} + \frac{1}{1} = 2.93777485\ldots\ ,$$

less than three loops! ♡

For the last two puzzles, the summands were independent random variables—we didn't actually need to know that independence is not required for linearity of expectation. Independence of two events A and B means that the probability they both occur is the product of their individual probabilities, so, for example, if we wanted to know the probability of getting all six numbers in only six rolls of a die, we could have obtained it as

$$1 \cdot \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6} = \frac{6!}{6^6} \sim 1.5\%\ .$$

For all the spaghetti to end up in one big loop, you must never tie a strand to its other end except at the last step; in other words, $\mathbf{X}_1, \ldots, \mathbf{X}_{49}$ must all be zero. This happens with probability

$$\frac{98}{99} \cdot \frac{96}{97} \cdot \frac{94}{95} \cdot\ \cdots\ \cdot \frac{2}{3}$$
$$= \frac{49! \cdot 2^{49}}{99!/(49! \cdot 2^{49})} = \frac{49!^2 \cdot 2^{98}}{99!} = 0.12564512901\ldots\ ,$$

so about one time in eight.

Ping-Pong Progression

Alice and Bob play table tennis, with Bob's probability of winning any given point being 30%. They play until someone reaches a score of 21. What, approximately, is the expected number of points played?

Solution:

With high probability, Alice will win, and thus the expected number of points won by Bob is $(3/7)21 = 9$ and therefore the expected total number of points is close to 30. Another way to make that calculation: Since Alice wins any given point with probability 70%, it takes on average $1/0.7 = 10/7$ rallies to earn Alice a point, thus $21 \times 10/7 = 30$ rallies on average for her to get to 21.

The actual answer is a bit less than 30, because playing until Alice scores 21 points is sufficient, but not always necessary, to finish the game. But Bob wins so rarely that the effect is negligible.

Do you recall a puzzle called "Finding a Jack" from a Chapter 8? You can use its solution technique—twice!—for the next puzzle.

Drawing Socks

You have 60 red and 40 blue socks in a drawer, and you keep drawing a sock uniformly at random until you have drawn all the socks of one color. What is the expected number of socks left in the drawer?

Solution:

It's useful to imagine that the sock-drawing is done by arranging all the socks in a uniformly random order, then removing them, starting from the beginning of the order, until all of one color have been removed. If the color of the last sock is blue (probability 0.4), then it must have been the red socks that were all removed, and the number of socks "left in the drawer" is the number of blue socks encountered starting from the *end* of the arrangement before any red sock is found.

If indeed the last sock is blue, there are 39 other blue socks distributed in 61 *slots* (before the first red sock, between the first and second red socks, and so forth, finally after the 60th red sock). Thus, we expect 39/61 blue socks in that last slot; adding the final blue sock gives on average of 100/61 blue socks left in the drawer. Similarly, if the last sock is red (probability 0.6), we have 59 other red socks in 41 slots so on average 59/41 in the last slot; that leaves 1 + 59/41 = 100/41 red socks behind. Putting these together, we have that the expected number of socks left is $0.4 \times 100/61 + 0.6 \times 100/41 \sim 2.12$.

If there are m socks of one color and n of the other, the above argument gives you an average of $m/(n{+}1) + n/(m{+}1)$ socks left in the drawer.

Linearity of expectation applies even when there are continuum-many random variables. Modern mathematicians would replace the sum by an integral, but sometimes an Archimedean approach will do the trick.

Random Intersection

Two unit-radius balls are randomly positioned subject to intersecting. What is the expected volume of their intersection? For that matter, what is the expected *surface area* of their intersection?

Solution:

We can assume that the center of the first ball is at the origin of a three-dimensional coordinate system; the first ball thus consists of all points in space within distance 1 of the origin. The center C of the second ball must be somewhere in the ball of radius *two* about the origin, in order for it to intersect the first ball. The conditions of the problem imply that C is uniformly random subject to this constraint.

(Note that this is not at all the same as taking the center of the second ball to be a random point between -2 and 2 on the x-axis. The latter assumption would give a much-too-big expected intersection size of $\pi/2$.)

Now consider a point P inside the first ball: What's the probability that P will be in the intersection of the two balls? That will happen just when C falls in the unit ball whose center is P. This ball is inside the ball of radius 2 about the origin that C is chosen from, with $1/2^3 = 1/8$ of the volume. Thus the probability that P lies in our intersection is $1/8$, and it follows from linearity of expectation that the expected volume of the intersection is $1/8$ times the volume of the first ball, that is, $1/8 \times 4\pi/3 = \pi/6$.

If it seems fishy to you to be adding up points P like this, good for you! Think of P not as a point but a small cell, or "voxel," inside the ball. If the voxel's volume is $(4\pi/3)/n$ with the ball divided into n voxels, n being some large integer, then an ordinary finite application of linearity of expectation gives approximately the same answer, $\pi/6$. It's not exact owing to the fact that a few voxels will lie partly inside and partly outside the intersection. As the voxels get smaller the approximation gets better.

For the surface area of the intersection, the point P—now representing a two-dimensional *pixel*—is taken on the surface of the unit ball centered about the origin. This will lie on the surface of the intersection if it's inside the second ball, that is, again, if C lies in the unit ball centered at P. We've already seen that this happens with probability $1/8$, but note that the intersection of the two spheres has *two* surfaces of the same area, one that was part of the surface of the first ball—that's the part that could contain P—and one that was part of the surface of the second ball. Thus the expected surface area of the intersection is $2 \times \frac{1}{8} \times 4\pi = \pi$.

You may have noticed, by the way, that all this can be done in any dimension; with disks on the plane, for example, you get expected area $\pi/4$ and perimeter π. You can even do it with (some) other shapes.

One place where linearity of expectation arises frequently is in gambling. A game is called "fair" if the expected return for each player is 0, and we can deduce that any combination of fair games is fair. This applies even if your choice of game, or strategy within a game, is made "as you play" (as long as you can't foresee the future).

Of course, in a gambling casino, the games are not fair; your expectation is negative, the casino's positive. For example, in (American) roulette there are 38 numbers (0, 00, and 1 through 36) on the wheel, but if you bet \$1 on a single number and win, you earn only \$36, not \$38, in return for your dollar. The result is that your expectation is

$$\frac{1}{38} \cdot \$35 + \frac{37}{38} \cdot (-\$1) = -\$\frac{1}{19},$$

so that you lose, on average, about a nickel per dollar bet. This same negative expectation applies also to bets made on groups of numbers or colors.

However:

Roulette for the Unwary

Elwyn is in Las Vegas celebrating his 21st birthday, and his girlfriend has gifted him with twenty-one \$5 bills to gamble with. He saunters over to the roulette table, noting that there are 38 numbers (0, 00, and 1 through 36) on the wheel. If he bets \$1 on a single number, he will win with probability 1/38 and collect \$36 (in return for his \$1 stake, which still goes to the bank). Otherwise, of course, he just loses the dollar.

Elwyn decides to use his \$105 to make 105 one-dollar bets on the number 21. What, approximately, is the probability that Elwyn will come out ahead? Is it better than, say, 10%?

Solution:

We know from linearity of expectation that on average, Elwyn should expect to lose $\$\frac{105}{19} \sim \$5.53$. But this says very little about the probability that Elwyn comes out ahead. A random variable with expectation -5.53 could be anywhere from always negative, to almost never negative!

In this case, we need to work out the probability that Elwyn wins three or more of his 105 bets, because if he wins three times he pulls in $3 \cdot \$36 = \$108$ and comes out ahead.

That takes a little work. The probability that Elwyn loses every time is $(37/38)^{105} \sim 0.06079997242$. To compute the probability that he wins exactly once, we multiply the number of bets by the probability of winning a given bet and losing the rest; this gives us

$$105 \cdot \frac{1}{38} \cdot \left(\frac{37}{38}\right)^{104} \sim 0.17254046227.$$

The probability that Elwyn wins exactly twice is the number of *pairs* of bets times the probability of winning two specified bets and losing the rest, which works out to

$$\frac{105 \cdot 104}{2} \cdot \left(\frac{1}{38}\right)^{2} \cdot \left(\frac{37}{38}\right)^{103} \sim 0.24248929833.$$

Adding these three numbers and subtracting from 1 gives the probability that Elwyn comes out ahead: 0.52417026698, better than 50%!

No, this does not mean that Elwyn has figured out how to beat 'Vegas. He comes out *slightly* ahead rather often but *far* behind quite often as well, since winning only twice leaves him with just \$72 to show for his \$105 stake. Elwyn's negative expectation is the more relevant statistic here.

If you want to go to Vegas and be able to tell your friends that you made a profit at roulette, here's a suggestion: Pick a number of dollars that's one less than a power of 2, and that you can afford to lose. Suppose it's \$63. Then bet \$1 on "red." (Half the positive numbers are red, so you get your dollar back and win an additional \$1 with probability 18/38, but lose your dollar otherwise.) If you do win, go home. If you lose, bet \$2 on red; if you win, go home having made a \$1 profit. If you lose, bet \$4 next time, etc., until you win or have lost your sixth straight bet (\$32). At this point, you must give up or risk ruin; fortunately the probability that you will be this unlucky is only $(20/38)^6$, about 2%. Your *expectation* for the whole experiment is of course still negative.

An Attractive Game

You have an opportunity to bet \$1 on a number between 1 and 6. Three dice are then rolled. If your number fails to appear, you lose your \$1. If it appears once, you win \$1; if twice, \$2; if three times, \$3.

Is this bet in your favor, fair, or against the odds? Is there a way to determine this without pencil and paper (or a computer)?

Solution:

This game seems pretty decent, maybe even favorable. You have three chances to get your number; if you were guaranteed to roll three different numbers, the probability that one would be yours would be 50%, making the game perfectly fair. If you *don't* roll three different numbers, you could end up with a bonus. So what's not to like?

In fact this is a well known game, often called "Chuck-a-Luck" or "Birdcage" (the latter because the dice are typically kept inside a cage and rolled by shaking the cage). Like any gambling game it is designed to look attractive but to make money for the house, and indeed that is the case.

The easiest way to see this is to imagine that there are six players each of whom bets on a different number. The expectation for each player is of course the same, since the rules don't discriminate. If that expectation is x then the house's expectation must be $-6x$ by linearity of expectation—this is a "zero-sum game," since no money leaves or enters from outside.

But now if the three rolls are different, three players win and three lose so the house breaks even. If two numbers are the same, the house pays \$2 to one player and \$1 to another while picking up \$1 from each of the remaining four players, so makes a \$1 profit. If all three dice come up the same, the house pays out \$3 and collects \$5, making a \$2 profit. So the house's expectation is positive and therefore the players' is negative.

(It's not hard to work out that the probability the three rolls are different is $(6\cdot5\cdot4)/6^3 = 5/9$, the probability that they are all the same is $1/6^2 = 1/36$, and thus the probability that just two of them match is the remaining $15/36 = 5/12$. Thus the house's expectation is $(5/12)\cdot\$1+(1/36)\cdot\$2 = \$17/36$ and each player's expectation is thus $(-1/6)\cdot\$17/36 \sim -\$0.0787037037$, so a player loses on the average about 8 cents per dollar bet—worse even than American roulette, about which more later.)

Here's a game which is definitely favorable; the question is, to what extent can you turn positive expectation into a sure thing?

Next Card Bet

Cards are turned face up one at a time from the top of a well-shuffled deck. You begin with a bankroll of \$1, and can bet any fraction of your current worth, prior to each revelation, on the color of the next card. You get even odds regardless of the current composition of the deck. Thus, for example, you can decline to bet until the last card, whose color you will of course know, then bet everything and be assured of going home with \$2.

Is there any way you can *guarantee* to finish with more than \$2? If so, what's the maximum amount you can assure yourself of winning?

Solution:

With a little thought you can improve on \$2 by waiting to bet until three cards remain. If they're all one color, great—you can bet everything for the remaining three rounds and go home with \$16. If one is red and the others black, bet 1/3 of your money on black; if you're right you now have $\$1\frac{1}{3}$ which you can double at the last card for a final tally of $\$2\frac{2}{3}$. If you're wrong you're down to $\$\frac{2}{3}$ but the remaining cards are both black, so you can double this twice to again end with $\$2\frac{2}{3}$. Since a similar strategy works when the remaining cards are two red and one black, you have found a way to guarantee $\$2\frac{2}{3}$.

To extend that backward and guarantee even more starts to look messy, so let's look at the problem another way. Whatever betting strategy you use, you'll certainly want to bet everything at every round as soon as the remaining deck becomes monochromatic; let's call strategies with this property "sensible."

It is useful first to consider which of your strategies are optimal in the sense of expectation, that is, which maximize your expected return. It is easy to see that all such strategies are sensible.

Surprisingly, the converse is also true: No matter how crazy your betting is, as long as you come to your senses when the deck becomes monochromatic, your expectation is the same! To see this, consider first the following *pure* strategy: Imagine some fixed specific distribution of red and black in the deck, and bet everything you have on that distribution at every turn.

Of course, you will nearly always go broke with this strategy, but if you win you can buy the earth—your take-home is then $2^{52} \times \$1$, around 4.5 quadrillion dollars. Since there are $\binom{52}{26}$ ways the colors can be distributed in the deck, your *expected* return is $\$2^{52}/\binom{52}{26} = \$9.0813$.

Of course, this strategy is not realistic, but it is "sensible" by our definition, and, most importantly, *every sensible strategy is a combination of pure strategies of this type.* To see this, imagine that you have $\binom{52}{26}$ assistants working for you, each playing a different one of the pure strategies.

We claim that every sensible strategy amounts to distributing your original \$1 stake among these assistants, in some way. If at some point your collective assistants bet $\$x$ on red and $\$y$ on black, that amounts to you yourself betting $\$x - \$y$ on red (when $x > y$) and $\$y - \$x$ on black (when $y > x$).

Each sensible strategy can be implemented by a distribution of money to the assistants, as follows. Say you want to bet \$0.08 that the first card is red; this means that the assistants who are guessing "red" first get a total of \$0.54 while the others get only \$0.46. If, on winning, you plan next to bet \$0.04 on black, you allot \$0.04 more of the \$0.54 total to the "red-black" assistants than to the "red-red" assistants. Proceeding in this manner, eventually each individual assistant has his assigned stake.

Now, any mix of strategies with the same expectation shares that expectation, hence every sensible strategy has the same expected return of \$9.08 (yielding an expected profit of \$8.08). In particular, all sensible strategies are optimal.

But one of these strategies *guarantees* \$9.08; namely, the one in which the \$1 stake is divided equally among the assistants. Since we can never guarantee more than the expected value, this is the best possible guarantee.

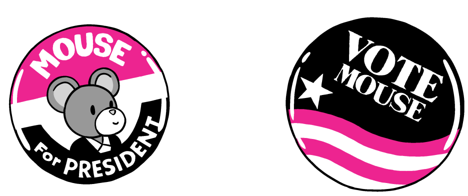
This strategy is actually quite easy to implement (assuming as we do that US currency is infinitely divisible). If there are b black cards and r red cards remaining in the deck, where $b \geq r$, you bet a fraction $(b-r)/(b+r)$ of your current worth on black; if $r > b$, you bet $(r-b)/(b+r)$ of your worth on red. You will be entirely indifferent to the outcome of each bet, and can relax and collect your \$8.08 profit at the end.

Sometimes the notion of fair game comes in handy even when there is no mention of a game.

Serious Candidates

Assume that, as is often the case, no one has any idea who the next nominee for President of the United States will be, of the party not currently in power. In particular, at the moment no person has probability as high as 20% of being chosen.

As the politics and primaries proceed, probabilities change continuously and some candidates will exceed the 20% threshold while others will never do so. Eventually one candidate's probability will rise to 100% while everyone else's drops to 0. Let us say that after the convention, a candidate is entitled to say



he or she was a "serious" candidate if at some point his or her probability of being nominated exceeded 20%.

Can anything be said about the expected number of serious candidates?

Solution:

It seems that you'd need to know a lot about the political process, and the current circumstances, to answer this question. For example, doesn't it seem plausible that under some conditions, the first candidate to exceed 20% is likely to continue rising and eventually secure the nomination? Wait, that can't be right; if he/she really had probability only 20% at the critical point, then most likely someone else would end up with the nomination.

In fact, the expected number of "serious" candidates has to be 5. To see this, imagine that you enter the prediction market with the following strategy: Anytime a candidate hits 20% for his or her first time, you bet \$1 on that candidate. If the market's probabilities are "correct" and the market is fair, you should win \$5 if that candidate is eventually nominated.

In this game, you are bound to win \$5; no matter which candidate gets nominated, you will have bet on him or her at some point. So the only variable is how many bets you made, and since the game is presumed to be fair, your expected expenditure must be \$5.

Why is the word "correct" in quotes above? The problem is that when everything is done and (say) Smith has won the nomination, one might argue that a good political prognosticator would, or should, have known all along that that would be the outcome. Thus when Smith hit 20% he or she should probably have been rated at 100%, or at least a lot higher than 20%. Without getting into the philosophy of probability, let's just say that a candidate's rating represents some sort of consensus which you, the bettor, have no reason to doubt one way or the other.

Rolling a Six

How may rolls of a die does it take, on average, to get a 6—given that you didn't roll any odd numbers *en route*?

Solution:

Obviously, three. Right? It's the same as if the die had only three numbers, 2, 4 and 6, each equally likely to appear. Then the probability of rolling a 6 is $1/3$, and as we have seen in the puzzle *Rolling All the Numbers*, in an experiment that has probability of success p, it takes, on average, $1/p$ trials to get a success.

Except that it's not the same at all. Imagine that you arrange to compute your answer to this puzzle via multiple experiments, making use of the law of large numbers. You start rolling your die, counting trials as you go. If you hit a "6" before you get an odd number, you record the number of trials and repeat. (For example, if you roll "4,4,6" you record a 3.) What happens if you do roll an odd number? Then you can't simply throw out that roll and continue counting trials (if you did that, you would eventually converge to 3 as your answer). Instead, you must throw out that whole experiment, recording nothing, and start anew, beginning with roll #1 of your new experiment. That will produce a smaller average than 3, but how much smaller?

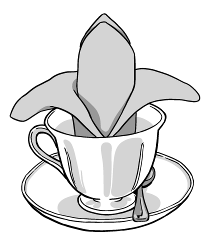
An intuitive way to look at it: When you've rolled a 6 without rolling an odd number, there's an inference that you succeeded quickly—since if you took a long time to get your 6, you'd probably have hit an odd number.

To figure out the correct answer, it helps to think about what you're doing in your multi-experiment. Basically, each sub-experiment consists of rolling the die until you get a 6 *or* and odd number. If you counted rolls in all of these, you'd discover their average length to be $3/2$, since the probability of rolling a 6, 1, 3, or 5 is $2/3$.

In your *gedankenexperiment* you're only counting the sub-experiments that end in a 6, but does that matter? The finishing roll (be it 1, 3, 5, or 6) is independent of all that has gone before, and independence is a reciprocal notion; thus, the number of rolls "doesn't care" what the sub-experiment ended with. It follows that the average length of the subexperiments that ended with a 6 is that same $3/2$.

Napkins in a Random Setting

At a conference banquet for a meeting of the Association for Women in Mathematics, the participants find themselves assigned to a big circular table. On the table, between each pair of settings, is a coffee cup containing a cloth napkin. As each mathematician sits down, she takes a napkin from her left or right; if both napkins are present, she chooses randomly. If the seats are occupied in



random order and the number of mathematicians is large, what fraction of them (asymptotically) will end up without a napkin?

Solution:

We want to compute the probability that the diner in position 0 (modulo n) is deprived of a napkin. The limit of this quantity, as $n \to \infty$, is the desired limiting fraction of napkinless diners.

We may assume that everyone decides in advance whether to go for her right or left napkin, in case both are available; later, of course, some will have to change their minds or go without.

Say that diners $1, 2, \ldots, i-1$ choose "right" (away from 0) while i chooses "left"; and diners $-1, -2, \ldots, -j+1$ choose "left" (again, away from 0) while diner $-j$ chooses "right." Note that i and j might both be equal to 1, meaning that the diners on both sides of diner 0 are planning to go for the napkin between them and diner 0.

If $k = i+j+1$, then as long as $k \leq n$ (which is the case with high probability), the probability of this configuration is 2^{1-k}.

Observe that diner 0 loses out only when she is last to pick among $-j, \ldots, i$ and *none* of the diners $-j+1, \ldots, -2, -1; 1, 2, \ldots, i-1$ get the napkins they wanted. If $t(x)$ is the time at which diner x makes her grab, then this happens exactly when $t(0)$ is the unique local maximum of t in the range $[-j, -j+1, \ldots, 0, \ldots, i-1, i]$.

If t is plotted on this interval, it looks like a mountain with $(0, t(0))$ on top; more precisely, $t(-j) < t(-j+1) < \cdots < t(-1) < t(0) > t(1) > t(2) > \cdots > t(i)$.

Instead of evaluating the probability of this event for fixed i and j, it is convenient to lump all pairs (i, j) together which satisfy $i+j+1 = k$ for fixed k. Altogether there are $k!$ ways the values $t(-j), \ldots, t(i)$ can be ordered. If T is the set of all k grabbing times and $t_{\max}$ is the last of these, then each mountain-ordering is uniquely identified by the nontrivial subset of $T \setminus \{t_{\max}\}$ which constitutes the values $\{t(1), \ldots, t(i)\}$. Thus, the number of orderings that make a valid mountain is $2^{k-1} - 2$.

Finally, the total probability that diner 0 is deprived of her napkin is

$$\sum_{k=3}^{\infty} \frac{2^{1-k} \cdot (2^{k-1} - 2)}{k!} = \left(2 - \sqrt{e}\right)^2 \approx 0.12339675. \heartsuit$$

Roulette for Parking Money

You're in Las Vegas with only \$2 and in desperate need of \$5 to feed a parking meter. You run in through the nearest door and find yourself at a roulette table. You can bet any whole dollar amount on any allowable set of numbers. What roulette-betting strategy will maximize your probability of walking out with \$5?

Solution:

There are two major considerations in trying to maximize your probability of boosting your fortune to \$5. One is to avoid waste: That is, try to hit \$5 exactly. Overshooting gains nothing and since your expectation is limited by your initial \$2, minus what the "house" is expecting to take from you, you don't want to spend it on overshooting your goal.

The second is speed. As we saw above in *Roulette for the Unwary*, it costs you about a nickel in expectation every time you bet \$1 playing roulette, no matter what you bet on.

Of course, you must do *some* betting. Ideal would be if there were a roulette bet that paid off 5:2, so you could just put your \$2 down once and win what you need or lose everything in one shot.

Unfortunately there is no such bet in roulette, so it looks like you are doomed to either risk overshooting, or risk having to make many bets. Example: You could bet \$1 on the numbers 1 through 8. If you win you're home, otherwise you could bet your remaining \$1 on red and if it wins, start over. That gives you probability p of success where

$$p = \frac{8}{38} + \frac{30}{38} \cdot \frac{18}{38} \cdot p \ .$$

That gives $p = \frac{8}{38}/(1 - \frac{30}{38} \cdot \frac{18}{38}) = 0.33628318584$, about $1/3$.

But you can do better by making correlation work for you. How? By making simultaneous bets of \$1 each of the right kind. Specifically: Bet one of your dollars on the numbers 1 through 12, and *simultaneously*, the other on 1 through 18. Then, if you do get a number between 1 and 12, you hit your \$5 on the nose. If you hit a high number, you lose it all (but at least you've lost it all quickly). If you happen to hit 13, 14, 15, 16, 17, or 18, you again have \$2 and can repeat your double bet.

The probability p of success for this scheme is given by

$$p = \frac{12}{38} + \frac{6}{38} \cdot p,$$

which yields $p = \frac{12}{38}/(1 - \frac{6}{38}) = 3/8 = 0.375$, quite a bit better. In fact this scheme can be shown to be optimal.

Next is a puzzle that you might find familiar. Posed and solved by Georges-Louis Leclerc, Comte de Buffon in 1733, it is said to be the very first solved problem in geometric probability.

But to solve it we will use almost no geometry, relying instead on linearity of expectation.

Buffon's Needle

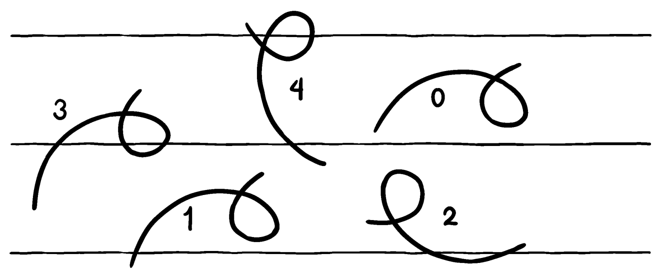
A needle one inch in length is tossed onto a large mat marked with parallel lines one inch apart. What is the probability that the needle lands across a line?

Solution:

Let C be a "noodle": Any smooth plane curve—self-intersections permitted—of length, say, ℓ. Toss the noodle randomly onto the mat, and let $\mathbf{X}_C$ be the number of crossings you get, that is, points where C crosses a line. We claim that the average number of crossings, $\mathbb{E}\mathbf{X}_C$, is proportional to ℓ, irrespective of the choice of C! The figure below shows the results of five tosses of a particular noodle, each accompanied by its number of crossings.

To verify the claim, imagine that C is composed of many short line segments, each of length ε. Throwing just one line segment L of length $\varepsilon < 1$ onto the mat produces a crossing with some probability p, but never more than one crossing, thus $\mathbb{E}\mathbf{X}_L = p$. At the moment, we don't know what p is.

But, by linearity of expectation, we know that if C is composed of n such segments, then $\mathbb{E}\mathbf{X}_C = np$. It follows, by letting ε shrink, that $\mathbb{E}\mathbf{X}_C = \alpha\ell$ for some constant α, since a smooth C is approximable by a string of short line



segments. If we can determine what α is, we can infer that $\mathbb{E}\mathbf{X}_N = \alpha$, since Buffon's needle N has length 1.

To find α we only have to pick C cleverly, and this is how: let C be a circle of diameter 1! Then no matter how C lands on the mat, it produces exactly 2 crossings. Thus $\mathbb{E}\mathbf{X}_C = 2 = \alpha \cdot \pi$, and we solve to get $\alpha = 2/\pi$. We deduce that $\mathbb{E}\mathbf{X}_N = 2/\pi$, and since (with probability 1) the needle either produces one crossing or none, the probability that Buffon's needle crosses a line is that same $2/\pi$. ♡

This wonderful proof was published by T. F. Ramaley in 1969, in a paper entitled "Buffon's Noodle Problem."

Here are a couple of more modern geometry problems, that seem not to have any probability in them.

Covering the Stains

Just as a big event is about to begin, the queen's caterer notices, to his horror, that there are 10 tiny gravy stains on the tablecloth. All he has the time to do is to cover the stains with non-overlapping plates. He has plenty of plates, each a unit disk. Can he do it, no matter how the stains are distributed?

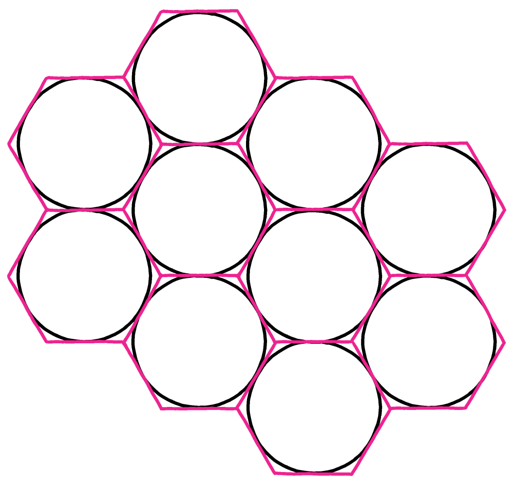
Solution:

The naked mathematical question being asked here is: Can any set of 10 points in the plane be covered by disjoint unit disks? Sure, if the points are close together, you can cover them with one disk; if they're spread widely apart, you can cover each with its own disk. It's the *medium* distances that could cause problems.

It's a good idea, as advertised in Chapter 7, to think about replacing 10 by other numbers. It's pretty obvious that any one or two points can be covered, and easy to show that any three can be covered.

On the other hand, the answer is certainly "no" if 10 is replaced by 10,000. The reason is that you can't "tile" the plane with unit disks; in other words, you can't arrange disjoint unit disks so as to cover the whole plane (as you could, say, with unit squares). If you put down 10,000 points in a 100×100 grid pattern covering a square somewhat bigger than a unit disk, it won't be possible to cover every grid point with disjoint unit disks.

So it's a question of quantity—is 10 more like 3, or more like 10,000? Here's a thought—how much of the plane *can* be covered by disjoint unit disks? You would probably guess that the best way to pack unit disks is in a hexagonal array, as shown in the figure below. In fact, it is provable that this is the best packing.



How good is it—in other words, if you choose a large region of the plane (say, a $1{,}000{,}000 \times 1{,}000{,}000$) square) and pack it hexagonally with disks, what fraction of the region will be covered? In the limit, that fraction must be the fraction of the area of one of those hexagonal cells that is occupied by an inscribed unit disk. Of course, the disk, having radius 1, has area π. The hexagon is made up of six equilateral triangles of altitude 1 and area $1/\sqrt{3}$, thus total area $6/\sqrt{3} = \sqrt{12}$. It follows that the fraction of area covered by the disk is $\pi/\sqrt{12} \sim 0.9069$.

Hmm. The disks cover more than 90% of the plane, that is, more than 9 points out of 10. Can we make use of that? Trouble is, the gravy stains are not random points, they are (in the worst case) arranged as if by an adversary.

But we can shift the hexagonal disk-packing randomly. To do this, just fix some hexagonal cell and pick a point in it uniformly at random; then shift (without rotating) the whole picture by moving the center of the cell to the selected point. Now we can say that for any *fixed* point of the plane, the probability that it is covered by some disk of the randomly shifted configuration is 0.9096.

It follows by linearity of expectation that the *expected* number of our 10 gravy stains covered by disks of the randomly shifted hexagonal packing is 9.096 stains, that is, on average more than 9 of the 10 points are covered. But then some configurations must cover all 10 points, and the problem is solved. ♡

This technique, in which we prove that something can be done by showing that if we do it randomly it succeeds with probability > 0, is often known by the fancy name "*the probabilistic method*" and associated with Paul Erdős and Alfréd Rényi. We'll see another example in the theorem at the end of this chapter.

The next puzzle provides another example where introduction of randomness is useful.

Colors and Distances

In the town of Hoegaarden, Belgium, exactly half the houses are occupied by Flemish families, the rest by French-speaking Walloons. Can it be that the town is so well mixed that the sum of the distances between pairs of houses of like ethnicity exceeds the sum of the distances between pairs of houses of different ethnicity?

Solution:

No. Represent each house in Hoegaarden as a distinct point in the plane, and fix a large disk that contains all the points. Now consider a random line that intersects the disk. (To get such a line, choose a radius of the disk uniformly at random, then take the line perpendicular to the radius that crosses it at a uniformly random point along the radius.)

Your random line will cut the house-points into two sets (one of which could be empty); we claim it separates at least as many house-pairs of different ethnicity as pairs of the same ethnicity. Why? Suppose there are n houses of each ethnicity, and that f Flemish houses and w Walloon houses lie on one side of the line. Then the number of heterogeneous crossings is $(n-w)f + (n-f)w$ while the number of homogeneous crossings is $(n-w)w + (n-f)f$; the former minus the latter is $f^2 + w^2 - 2wf = (f-w)^2 \geq 0$.

Now we make use of the critical fact that for any two particular points, the probability that they are separated by our random line is proportional to the distance between the points. Adding expectations, we deduce that the total distance between heterogeneous (resp., homogeneous) pairs of points is proportional to the expected number of separated heterogeneous (resp., homogeneous) pairs. Since every line cuts at least as many heterogeneous as homogeneous pairs, we conclude that the former distances add up to at least as much as the latter.

Note that for distinct points the inequality is strict, because as long as there is one point which is not simultaneously of both ethnicities, there is positive probability that a line will result in $f \neq w$ causing an imbalance in favor of heterogeneous pairs.

The next puzzle also has a geometric component, but here it is less surprising that we use linearity of expectation.

Painting the Fence

Each of n industrious people chooses a random point on a circular fence, and begins painting toward the farther of her two neighbors until she encounters a painted section. On average, how much of the fence gets painted? How about if instead, each person paints toward her *nearer* neighbor?

Solution:

For $n = 2$, the answers are easily determined to be $3/4$ and $1/4$, respectively. For larger n, we consider any interval I between neighboring painters and compare it to its two adjacent intervals. In Case (A), where each painter paints toward her farther neighbor, I will go unpainted exactly when it is the shortest of the three; in Case (B), when it is the longest. Each of these events happens with probability $1/3$.

It seems like we're going to need to know the *expected* length of the shortest (resp., longest) of three intervals chosen by cutting a given interval J at two random points. You can get this noting that choosing two random points from the unit interval is equivalent to choosing a uniformly random point in an equilateral triangle; the barycentric coordinates of the point give your interval lengths. Then, if you identify the regions of the triangle which result in the middle third being shortest (resp. longest) of the three intervals and compute conditional expectations, you get $1/9$ and $11/18$, respectively.

That calculation was for the unit interval, so if the length of I together with its two neighbor intervals is x, then the expected length of I given that it is the shortest is $x/9$ and given that it is the longest, $11x/18$.

On average, the total length of I and its two neighbors is $3/n$ since adding up this figure for all I counts every interval three times. But in both versions, I is unpainted only a third of the time (in expectation). It follows that in Case (A) the expected amount of unpainted fence is $\frac{n}{3} \cdot \frac{3/n}{9} = 1/9$, and in Case (B), $\frac{n}{3} \cdot \frac{11 \cdot 3/n}{18} = 11/18$.

For the next puzzle, it is useful to have an alternative formula for expectation. Suppose $\mathbf{X}$ is a "counting" random variable, that is, a random variable that takes values in the non-negative integers. Then

$$\mathbb{E}\mathbf{X} = \sum_{i=0}^{\infty} \mathbb{P}(\mathbf{X} > i).$$

To see that this formula is equivalent to the definition at the top of the chapter, we just need to verify that for each j, the probability that $\mathbf{X} = j$ is counted j times in the sum. But that's easy: It's counted once for each i between 0 and $j-1$, so j times as required.

Filling the Cup

You go to the grocery store needing one cup of rice. When you push the button on the machine, it dispenses a uniformly random amount of rice between nothing and one cup. On average, how many times do you have to push the button to get your cup?

Solution:

We know that at least two pushes are necessary, and in fact two will suffice just half the time; and sometimes more than three will be required. So the answer must be more than $2\frac{1}{2}$. Would you believe 2.718281828459045?

Let $\mathbf{X}_1$, $\mathbf{X}_2$ etc., be the amount dispensed, and let $\mathbf{Y}_i$ be the fractional part of $\mathbf{X}_1 + \cdots + \mathbf{X}_i$. Then the $\mathbf{Y}_i$'s, like the $\mathbf{X}_i$'s, are independent and uniformly random in the unit interval.

The first decrease in the values of the $\mathbf{Y}_i$'s signals the point when the amount of rice exceeds 1 cup. Thus, the question at hand is equivalent to: What is the expected value of $\mathbf{Z}$, the length of the longest increasing initial sequence of the $\mathbf{Y}_i$'s?

The probability that the first j values are increasing is $1/j!$, since the increasing order is just one of $j!$ permutations. Here's where we use the special formula above for the expected value of a counting random variable:

$$\mathbb{E}\mathbf{Z} = \sum_{j=0}^{\infty} \mathbb{P}(\mathbf{Z} > j) = \sum_{j=0}^{\infty} 1/j!$$

which is just exactly e, Euler's number (also known as Napier's constant, but, naturally, discovered by yet someone else—Jacob Bernoulli).

It's time for our theorem, which, as promised, will be proved using the probabilistic method.

Arguably the most famous theorem in combinatorics, the following remarkable fact was proved in 1930 by Frank Plumpton Ramsey, scion of a famous family of British intellectuals (and brother of Arthur Michael Ramsey, Archbishop of Canterbury). In its simplest finite form, Ramsey's Theorem states that for any positive integer k there is a number n such that if you color every unordered pair of numbers from the set $\{1, \ldots, n\}$ either red or green, then there must be a set S of k numbers which is "homogeneous" in the sense that every pair from S is the same color.

The least n for which this is true is called the "kth Ramsey number" and denoted $R(k, k)$. For example, $R(4, 4)$ turns out to be 18; that means that for any two-coloring of the pairs of numbers between 1 and 18, there's a homogeneous set of size 4; but this does not hold for all colorings of pairs from the set $\{1, \ldots, 17\}$. But $R(5, 5)$ is not known! Paul Erdős, who wrote more papers than any other mathematician in history and was fascinated by Ramsey's theorem, enjoyed giving the following advice. If a powerful alien force demands, on penalty of Earth's annihilation, that we tell them the value of $R(5, 5)$, we should put all the computing power on the planet to work and perhaps succeed in computing that number.

But if they ask for $R(6, 6)$, we should attack them before they attack us.

It is not hard to prove—and, indeed, we will prove it in Chapter 15—that $R(k,k)$ is at most $\binom{2k-2}{k-1}$, which in turn is less than 2^{2k}. What we're going to do here is use linearity of expectation to get an exponential *lower* bound for $R(k,k)$.

Theorem. *For all* $k > 2$, $R(k,k) > 2^{k/2}$.

The idea of the proof, published by Erdős in 1947, is to show that a *random* coloring of the pairs of numbers from 1 to $n = 2^{k/2}$ will, with positive probability, have no monchromatic set of size k. Thus, such colorings exist (although the proof does not exhibit one).

By a random coloring, we mean this: For each pair $\{i,j\}$ of numbers between 1 and n, we flip a fair coin to decide whether to color it red or green. If we pick a fixed set S of size k, it contains $\binom{k}{2} = k(k-1)/2$ pairs and thus will be homogeneous with probability $2/2^{k(k-1)/2} = 2^{(k+1)/2-k^2/2}$.

The number of candidates for the set S, i.e., the number of subsets of $\{1,\ldots,n\}$ of size k, is $\binom{n}{k} < (2^{k/2})^k/k! = 2^{k^2/2}/k!$, thus by linearity of expectation, the *expected* number of homogeneous sets is less than

$$2^{(k+1)/2-k^2/2} \cdot 2^{k^2/2}/k! = \frac{2^{(k+1)/2}}{k!}$$

. which is equal to $\sqrt{2}$ when $k = 2$ and declines as k increases.

So the expected number of homogeous sets is less than 1, and we claim that means the probability that there are no homogeneous sets is greater than zero. One way to see this is to use the formula for expectation of a counting random variable from *Filling the Cup* above. The expected value of $\mathbf{X}$, if $\mathbf{X}$ is the number of homogeneous sets, equals $\Pr(\mathbf{X} > 0)$ plus more non-negative terms; so if $\mathbb{E}\mathbf{X}$ is less than 1, so is $\Pr(\mathbf{X} > 0)$. ♡

The theorem tells us that $R(5,5)$ is at least $\lceil 2^{5/2} \rceil = 6$ and our later induction proof implies an upper bound of $\binom{8}{4} = 70$. In fact $R(5,5)$ is, at the time of this writing, known to be between 43 and 48 inclusive.