Chapter 10

In All Probability

Dealing with probabilities is intuitive for some folks, less so for others. But there are just a few basic ideas, which, if mastered, convey a lot of power.

A thing that may or may not happen is called an *event*, and its probability measures the likelihood that it will happen; if the "trial" that could make the event happen is repeatable (like rolling a die), then the event's probability is the fraction of time that it occurs in many trials.

If the trial has n possible outcomes that are equally likely (perhaps because of symmetry) then the probability of the event A is the number of outcomes that result in A occurring, divided by n. For example, if we roll a die and A is the event that an even number appears on top, then the probability $\mathbb{P}(A)$ that A occurs is $3/6 = 1/2$, since three of the outcomes are even numbers. The fact that the outcomes are equally likely relies on the assumption that the die behaves like a true, homogeneous, *symmetrical*, cube.

If we roll a pair of dice, there are 36 equally likely outcomes; note that (2,3) is different from (3,2) because even if *you* can't distinguish between the two dice, Tyche (the goddess of chance) can. Thus, for example, there are five ways to roll a sum of 6—(1,5), (2,4), (3,3), (4,2), and (5,1)—so $\mathbb{P}(\text{two dice roll a sum of } 6) = 5/36$.

If two events A and B are *independent*, that is, the occurrence of one does not affect the probability of the other, then the probability $\mathbb{P}(A \wedge B)$ that they both occur is equal to the product $\mathbb{P}(A)\mathbb{P}(B)$ of the probabilities of the two events. Without the assumption of independence, we can only say $\mathbb{P}(A \wedge B) = \mathbb{P}(A)\mathbb{P}(B|A)$ where $\mathbb{P}(B|A)$, read "the probability of B given A," has intuitive meaning but is often defined simply as the quantity that makes the equation true.

The probability $\mathbb{P}(A \vee B)$ that at least one of A and B occurs is always equal to $\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \wedge B)$, as you can see from a Venn diagram. This reduces to $\mathbb{P}(A \vee B) = \mathbb{P}(A) + \mathbb{P}(B)$ only in the case where A and B are mutually exclusive—that is, they can't both occur.

Let's try a few amusing applications of these ideas.

Winning at Wimbledon

As a result of temporary magical powers (which might need to be able to change your gender), you have made it to the women's singles tennis finals at Wimbledon and are playing Serena Williams for all the marbles. However, your powers cannot last the whole match. What score do you want it to be when they disappear, to maximize your chances of notching an upset win?

Solution:

Naturally, you want to have won the first set of the best-of-three-sets match, and be far ahead in the second. Your first thought might be that you'd like to be up five games to love and serving at 40-love (or, if your serve resembles mine, up 5-0 in games and *receiving* at love-40). Either gives you three bites at the apple, that is, if your probability of winning a point is some small quantity ε then your probability of winning is about 3ε.

(Technically, if the results of the next three points are independent, and you discount the possibility of winning the match despite losing them all, your success probability is $1-(1-\varepsilon)^3 = 3\varepsilon - 3\varepsilon^2 + \varepsilon^3$.)

But you can do better. Let the game score be 6-6, and suppose you are ahead 6-0 in the seven-point tiebreaker. Then you have six chances to win, almost doubling your probability when ε is small.

As I write this, you can do even better in the Australian Open, where the final-set tiebreaker is played to 10 points. The lesson, either way, is this: If you need a miracle, try to arrange for as many shots at that miracle as possible.

Birthday Match

You are on a cruise where you don't know anyone else. The ship announces a contest, the upshot of which is that if you can find someone who has the same birthday as yours, you (both) win a Beef Wellington dinner.

How many people do you have to compare birthdays with in order to have a better than 50% chance of success?

Solution:

Let's assume that (a) neither you (important!) nor anyone else on board (less important) was born on Feb. 29, (b) other dates are equally likely to be a given shipmate's birthday, and (c) there are no sets of twins (or triplets etc.) on board that will cause you to waste queries. Then the probability that a given shipmate *fails* to match your birthday is 364/365, thus if you ask n people, the probability that you have no luck is $(364/365)^n$. You want this number to dip *below* 50%, which it does when n reaches 253; in fact $(364/365)^{253} = 0.49952284596...$.

Is 253 higher than you guessed? You'd only need 183 queries to get your success probability over 1/2, if the people you asked were guaranteed to have *different* birthdays. The trouble is, you will probably hear a lot of duplicated answers that don't help you. (Well, it could help you a bit if you demand a bite of their beef Wellington in return for telling each about the other).

If you thought the answer to the puzzle was 23, you confused the question with the more famous problem of how many people you need in a room to have a better than 50% probability that *some* pair of people share a birthday. In fact, you can get the answer "23" from our answer; the number of pairs of people in that room is $\binom{23}{2} = 253$. (The pairs are not precisely independent, but they're close enough to make this work.)

By the way, how do you determine that 253 is the least n for which $(364/365)^n < 1/2$ without a lot of trial and error? My favorite way is to use the approximation $(1 - 1/m)^m \sim e^{-m}$ for large m. If you take this as an equality and solve $(1 - 1/365)^n = 1/2$ you get $n = \log_e(2) \cdot 365 \sim 252.9987$.

When the event A is decided before the event B, the conditional probability of B given A is often easy to deduce but $\mathbb{P}(A|B)$ is less intuitive. The laws of probability don't care about the order of events, though.

Other Side of the Coin

A two-headed coin, a two-tailed coin, and an ordinary coin are placed in a bag. One of the coins is drawn at random and flipped; it comes up "heads." What is the probability that there is a head on the other side of this coin?

Solution:

The temptation is to say that the drawn coin is either the two-headed coin or the fair coin, so the probability that its other side is heads is 1/2. But is it? Imagine that the coin is flipped many times, and comes up heads every time. It could still be either coin, but you can't help thinking it's more likely to be the two-headed coin. That kind of reasoning is basically how we learn from statistical observations.

In fact, the inference that the drawn coin is probably two-headed exists already after one flip. We want to compute $\mathbb{P}(A|B)$ where A is the event that the two-headed coin was drawn, and B the event that heads was flipped. For any two events A and B, we calculate

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \wedge B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}((A \wedge B) \vee (\overline{A} \wedge B))}$$

$$= \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A \wedge B) + \mathbb{P}(\overline{A} \wedge B)}$$

$$= \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(\overline{A})\mathbb{P}(B|\overline{A})}$$

where $\overline{A}$ is the event that A does *not* occur. This equation is known as Bayes' Rule, and if we plug in the values from our experiment, we get

$$\mathbb{P}(A|B) = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot \frac{1}{4}} = \frac{2}{3} \,.$$

If all the above fractions leave you unconvinced, here's another way to get the answer. Steal a crayon or marker from your kid and label each of the six coin-faces as follows: "1" and "2" on the two sides of the two-headed coin, "3" on the heads side of the fair coin, "4" on its tails side, and finally "5" and "6" on the two sides of the two-tailed coin.

You will agree, I hope, that drawing a coin and flipping it has the same effect as rolling a fair die: You are equally likely to come up with any number from 1 to 6. Given that you flipped a head, then, you are equally likely to have seen 1, 2, or 3. Two of those three faces have a head on the other side.

"Reverse" conditional probabilities are crucial in statistics, where you need to compute $\mathbb{P}(A|B)$ where A is the hypothesis you are testing, and B the result of your experiment. Here are some more puzzles involving that kind of calculation.

Boy Born on Tuesday

Mrs. Chance has two children of different ages. At least one of them is a boy born on a Tuesday. What is the probability that both of them are boys?

Solution:

We consider first the classic version, where the information is that at least one of Mrs. Chance's children is a boy. Here, the solution seems simple: *a priori*, the children (in age order) are equally likely to be boy-boy, boy-girl, girl-boy or girl-girl. The last of these is ruled out, so the probability both children are boys is $1/3$.

The difficulty is that in *almost* every way that you are likely to find out that at least one of Mrs. Chance's kids is a boy, there is an additional inference which raises the probability of two boys from $1/3$ to $1/2$.

For example, if you find that the older child is a boy, then the girl-boy combination above is ruled out and it's clear that the probability that the younger child is a boy is 1/2. But that reasoning also applies if your information is that the taller child is a boy, or the one with darker hair is a boy, or the child you saw Mrs. Chance with yesterday is a boy, or even if you happen to catch Mrs. Chance in the store buying boys' clothes.

Why the latter? Because if Mrs. Chance's other child were a girl, you'd be less likely to find her buying boys' clothes. It's quite similar to the previous puzzle; the fact that if the coin were fair it might not have come up "heads" affords an inference that the coin is the two-headed coin, raising the probability of two heads from 1/2 to 2/3. Here the effect is to raise the probability that the children are both boys from 1/3 to 1/2.

Why "almost"? Is there *any* way you could find out that at least one of Mrs. Chance's kids is a boy, that would lead to the conclusion that the probability they are both boys is 1/3? As it happens, I have a friend whose wife became pregnant with fraternal twins, and an amniotic test revealed the presence of Y chromosomes in the placenta. The conclusion was that at least one of the twins was a boy, and naturally my friend and his wife hoped that the other was a girl—but assumed that the probability of this event was 1/2. In fact, it really was 2/3; I inquired and was told that the test always finds Y chromosomes as long as one or more fetuses is male. (In fact, the outcome was mixed twins.)

Now back to the boy born on a Tuesday. This is a very delicate question and you need to assume that the information has come to you in with no additional inferences—for example, you actually chose Mrs. Chance randomly among many mothers of two (non-twin) children and asked her if at least one was a boy born on a Tuesday, and she said yes.

Compared to the classic version, here there is a substantial inference that both kids are boys because having two boys, instead of only one, *nearly* doubles the probability that some boy was born on a Tuesday. But not quite, because if they were both born on Tuesday they count only once. Thus you'd expect the answer to be nearly 1/2, and that is indeed the case. The calculation is straightforward. If we classify all two-kid families according to gender and birth-day-of-the-week we get $2\times7\times2\times7 = 196$ possibilities, of which $1\times2\times7+2\times7\times1-1\times1 = 27$ contain a boy born on a Tuesday (the -1 comes from over-counting the case of two boys born on Tuesdays). Of these, $1\times1\times7+1\times7\times1-1\times1 = 13$ involve two boys, so the probability we seek is 13/27.

Whose Bullet?

Two marksmen, one of whom ("A") hits a certain small target 75% of the time and the other ("B") only 25%, aim simultaneously at that target. Just one bullet hits. What's the probability that it came from A?

Solution:

You might think that since A is three times as good a shot as B, he is three times as likely to have been the one whose bullet hit; in other words, the probability that he deserves the credit is 75%.

But there are two things happening here: Hitting and missing. Since the probability that A hits and B misses is $3/4 \times 3/4 = 9/16$ while the probability that B hits but A misses is only $1/4 \times 1/4 = 1/16$, the probability that the successful bullet was A's is actually a full 90%.

Conditional probabilities can sometimes defy your intuition. Another example:

Second Ace

What is the probability that a poker hand (five cards dealt at random from a 52-card deck) contains at least two aces, given that it has at least one? What is the probability that it contains two aces, given that it has the Ace of Spades? If your answers are different, then: What's so special about the Ace of Spades?

Solution:

There are $\binom{52}{5}$ poker hands of which $\binom{52}{5} - \binom{48}{5}$ have at least one ace, and $\binom{52}{5} - \binom{48}{5} - \binom{4}{1} \cdot \binom{48}{4}$ have at least two aces; dividing the last of these expressions by the second gives our first answer, 0.12218492854.

There are $\binom{51}{4}$ hands containing the Ace of Spades and $\binom{51}{4} - \binom{48}{4}$ containing another ace as well, giving our second answer, 0.22136854741, quite a lot larger than the first answer.

Of course, there's nothing special about the Ace of Spades; specifying the ace held changes the conditional probabilities.

Here's a simpler example that you don't need a calculator for. Suppose a third of the families in town have two children, a third have one, and the remaining third none. Then the probability that a family has a second child, given that it has at least one, is $1/2$. But the probability that a family has a second child given that it has a *girl* is $3/5$, because only half the families with one child have a girl, but $3/4$ of those with two children have a girl.

Stopping After the Boy

In a certain country, a law is passed forbidding families to have another child after having a boy. Thus, families might consist of one boy, one girl and one boy, five girls and one boy... How will this law affect the ratio of boys to girls?

Solution:

It might seem like limiting families to one boy should cut down the boy-girl ratio, but assuming that gender is independent among siblings, the child whose birth is prevented by having an older brother is as likely to be a girl as a boy. Thus the sex ratio should not be affected by this law.

One could make the case that, on account of the possibility of identical twins, gender among siblings is not precisely independent. However, in the statement of the puzzle, the law's effect when twin boys are born has not been made clear (and is perhaps a bit horrifying to think about). Thus we solvers are pretty much forced to discount that possibility.

Up until now we've been talking about *discrete* probability. Things get a bit more subtle when a random value is chosen from an interval of real numbers. Awkwardly, we have to assign probability 0 to the event that any particular number is chosen, even though some event of that kind is guaranteed to happen. Typically, we deal instead with the probability that the chosen point lies in a particular subinterval. For example, if we choose a point uniformly at random from the unit interval, the probability that it falls between 0.3 and 0.4 is 1/10 (regardless of whether we count 0.3 and 0.4 as part of the interval).

Points on a Circle

What is the probability that three uniformly random points on a circle will be contained in some semicircle?

Solution:

It's not hard to see that no matter where we put the first two points, assuming they don't coincide, it is more likely that the third point will be in some semicircle with the first two, than not. Thus, in fact, the answer to the puzzle should be more than 1/2. But how do we compute it, without the bother of trying to condition on the distance between the first two points?

The answer is (as is often the case in probability problems) to choose the points in a different way—but, of course, one that still results in uniformly random points. Here, let us choose three random *diameters* of our circle. Each hits the circle at two (antipodal) points, and we use three coinflips do decide which three points to use.

This may seem like an unnecessarily complicated way to choose our points—why in six steps, when we can do it in three steps? To see the answer, draw *any* three diameters and the six points where they hit the circle. Observe that of those six, if three *consecutive* points are chosen, they will certainly lie in a semicircle; otherwise, they won't!

Well, there are $2^3 = 8$ ways to pick the points, using our coinflips, and 6 of those result in consecutive points being chosen. So the probability of their being contained in some semicircle is $6/8 = 3/4$. ♡

(If you found that puzzle too easy, try to determine the probability that four uniformly random points on a sphere are all contained in some hemisphere!)

Comparing Numbers, Version I

Paula (the perpetrator) takes two slips of paper and writes an integer on each. There are no restrictions on the two numbers except that they must be different. She then conceals one slip in each hand.

Victor (the victim) chooses one of Paula's hands, which Paula then opens, allowing Victor to see the number on that slip. Victor must now guess whether that number is the larger or the smaller of Paula's two numbers; if he guesses right, he wins \$1, otherwise he loses \$1.

Clearly, Victor can at least break even in this game, for example, by flipping a coin to decide whether to guess "larger" or "smaller"—or by always guessing "larger," but choosing the hand at random. The question is: Not knowing anything about Paula's psychology, is there any way he can do better than break even?

Solution:

Amazingly, there is a strategy which guarantees Victor a better than 50% chance to win.

Before playing, Victor selects a probability distribution on the integers that assigns positive probability to each integer. (For example, he plans to flip a coin until a "head" appears. If he sees an even number $2k$ of tails, he will select the integer k; if he sees $2k-1$ tails, he will select the integer $-k$.)

If Victor is smart, he will conceal this distribution from Paula, but as you will see, Victor gets his guarantee even if Paula finds out. It is perhaps worth noting that although Victor might like the idea of assigning the *same* probability to every integer, this is not possible—there is no such probability distribution on the integers!

After Paula picks her numbers, Victor selects an integer from his probability distribution and adds $\frac{1}{2}$ to it; that becomes his "threshold" t. For example, using the distribution above, if he flips five tails before his first head, his random integer will be -3 and his threshold t will be $-2\frac{1}{2}$.

When Paula offers her two hands, Victor flips a fair coin to decide which hand to choose, then looks at the number in that hand. If it exceeds t, he guesses that it is the larger of Paula's numbers; if it is smaller than t, he guesses that it is the smaller of Paula's numbers.

So, why does this work? Well, suppose that t turns out to be larger than either of Paula's numbers; then Victor will guess "smaller" regardless of which number he gets, and thus will be right with probability exactly $\frac{1}{2}$. If t undercuts both of Paula's numbers, Victor will inevitably guess "larger" and will again be right with probability $\frac{1}{2}$.

But, *with positive probability*, Victor's threshold t will fall *between* Paula's two numbers; and then Victor wins regardless of which hand he picks. This possibility, then, gives Victor the edge which enables him to beat 50%. ♡

Neither this nor any other strategy enables Victor to guarantee, for some fixed $\varepsilon > 0$, a probability of winning greater than $50\% + \varepsilon$. A smart Paula can choose randomly two consecutive multidigit integers, and thereby reduce Victor's edge to an insignificant smidgen.

Comparing Numbers, Version II

Now let's make things more favorable to Victor: Instead of being chosen by Paula, the numbers are chosen independently at random from the uniform distribution on [0,1] (two outputs from a standard random number generator will do fine).

To compensate Paula, we allow her to examine the two random numbers and *to decide which one Victor will see*. Again, Victor must guess whether the number he sees is the larger or smaller of the two, with \$1 at stake. Can he do better than break even? What are his and Paula's best (i.e., "equilibrium") strategies?

Solution:

It looks like the ability to choose which number Victor sees is paltry compensation to Paula for not getting to pick the numbers, but in fact *this* version of the game is strictly fair: Paula can prevent Victor from getting any advantage at all.

Her strategy is simple: Look at the two random real numbers, then feed Victor the one which is closer to $\frac{1}{2}$.

To see that this reduces Victor to a pure guess, suppose that the number x revealed to him is between 0 and $\frac{1}{2}$. Then the unseen number is uniformly distributed in the set $[0,x] \cup [1-x,1]$ and is, therefore, equally likely to be smaller or greater than x. If $x > \frac{1}{2}$, then the set is $[0,1-x] \cup [x,1]$ and the argument is the same.

Of course, Victor can guarantee probability $\frac{1}{2}$ against any strategy by ignoring his number and flipping a coin, so the game is completely fair. ♡

This amusing game was brought to my attention at a restaurant in Atlanta. Lots of smart people were present and were stymied, so if you failed to spot this nice strategy of Paula's, you're in good company.

Biased Betting

Alice and Bob each have \$100 and a biased coin that comes up heads with probability 51%. At a signal, each begins flipping his or her coin once a minute and bets \$1 (at even odds) on each outcome, against a bank with unlimited funds. Alice bets on heads, Bob on tails. Suppose both eventually go broke. Who is more likely to have gone broke first?

Suppose now that Alice and Bob are flipping the *same* coin, so that when the first one goes broke the second one's stack will be at \$200. Same question: Given that they both go broke, who is more likely to have gone broke first?

Solution:

In the first problem, Alice and Bob are equally likely to have gone broke first. Bob always goes broke (as noted later, his average time to go broke is finite). Alice may never go broke, but if we condition on her having gone broke, it turns out that Alice and Bob have exactly the same probability of having gone broke at any particular time t.

To see this, pick any win-loss sequence s that ends by going broke; suppose s has n wins, thus $n+100$ losses. It will have probability $p^n q^{n+100}$ for Alice, but $p^{n+100} q^n$ for Bob, where here $p = 51\%$ and $q = 49\%$. The ratio of these probabilities is the constant $(q/p)^{100}$, thus $\mathbb{P}(\text{Alice goes broke}) = (q/p)^{100}$ (which happens to be about 0.0183058708; call it 2%). Once we divide Alice's probability of encountering s by this number, her probability is the same as Bob's.

For the second problem, where Alice and Bob were flipping the same coin, your intuition might tell you that Alice probably went broke first. The reasoning is that if Bob went broke first, Alice would have to have gone broke after building up to \$200, an unlikely occurrence. But is this true? Can't we argue as above, comparing win-lose sequences that result in both going broke?

No. The proof above doesn't work because here, Alice's going broke also affects Bob's longevity. In fact, the same ratio $(q/p)^{100}$ applies when comparing a both-broke win-loss sequence and the conclusion is that the probability that Bob goes broke first is $(q/p)^{100}$ times the probability that Alice goes broke first. Intuition was correct; Alice is more than 50 times more likely to have gone broke first!

It's worth noting that, putting Bob aside, we can deduce from the first calculation how long it takes, on average, for Alice to go broke *given that she goes broke.* The reason is that under this condition, we have shown Alice's expected time to ruin is the same as Bob's. But Bob *always* goes broke and since he loses 2 cents per toss on average, his expected time to ruin is $100/0.02 = 5000$ tosses. At 60 tosses per hour that comes to $83\frac{1}{3}$ hours, or about $3\frac{1}{2}$ days.

Home-Field Advantage

Every year, the Elkton Earlies and the Linthicum Lates face off in a series of baseball games, the winner being the first to win four games. The teams are evenly matched but each has a small edge (say, a 51% chance of winning) when playing at home.

Every year, the first three games are played in Elkton, the rest in Linthicum.

Which team has the advantage?

Solution:

The series will last 4, 5, 6, or 7 games; it seems that the average should be somewhere between 5 and 6 games, in which case more games are played on average in Elkton than in Linthicum; therefore, Elkton has the advantage.

First issue: Are more games really played in Elkton, on average? To verify that we'll need to compute the probabilities that the series will last 4, 5, 6, or 7 games. If all the games were 50-50, the probabilities that a series will last 4, 5, 6, or 7 games would be, respectively, 1/8, 1/4, 5/16 and 5/16. Thus the average number of games played would be

$$4 \cdot \frac{1}{8} + 5 \cdot \frac{1}{4} + 6 \cdot \frac{5}{16} + 7 \cdot \frac{5}{16} = \frac{93}{16} = 5\frac{13}{16} = 5.8125$$

and since three games are always played in Elkton, on average only 2 13/16 would be played in Linthicum.

As one might expect, re-calculating the above odds taking into account the 1% home-field advantage changes the numbers very little. The average number of games played shifts upward a teensy bit to 5.81267507002.

So, indeed, more games are played in Elkton, on average. Over a very long period of time, this will be reflected in the game statistics; you can expect Elkton to have won about 50.03222697428% of the games over a million-year period.

However, shockingly, you can expect Linthicum to have won most of the series! Think of it this way: Imagine that all seven games are played (with equal ferocity) regardless of the outcomes. This doesn't change the probability of winning the series (if it did, we would never be able to quit after fewer than seven games). Those seven consist of four games in Linthicum and only three in Elkton, giving the advantage to Linthicum: their probability of winning the series works out to 50.31257503002%.

If it seems odd to you that the home-field advantage in games 5, 6, and 7, which might not get played, counts as much as home-field advantage in earlier games, that's understandable. But the point is that the late home-field advantage *is there when Linthicum needs it*, and something that's there when you need it is just as good as its being there all the time.

Have you ever wondered about the effect of different rules, from sport to sport, concerning which player or team serves next? Maybe the next puzzle will answer your questions.

Service Options

You are challenged to a short tennis match, with the winner to be the first player to win four games. You get to serve first. But there are options for determining the sequence in which the two of you serve:

1. Standard: Serve alternates (you, her, you, her, you, her, you).
2. Volleyball Style: The winner of the previous game serves the next one.
3. Reverse Volleyball Style: The winner of the previous game receives in the next one.

Which option should you choose? You may assume it is to your advantage to serve. You may also assume that the outcome of any game is independent of when the game is played and of the outcome of any previous game.

Solution:

Again using the idea (from the previous puzzle's solution) that it doesn't hurt to assume extra games are played, assume that you play lots of games—maybe more than are needed to determine the match winner. Let A be the event that of the first four served by you and the first three served by your opponent, at least four are won by you. Then, it is easily checked that no matter which service option you choose, you will win if A occurs and lose otherwise. Thus, your choice makes no difference. Notice that the independence assumptions mean that the probability of the event A, since it always involves four particular service games and three particular returning games, does not depend on when the games are played, or in what order.

If the game outcomes were not independent, the service option could make a difference. For example, if your opponent is easily discouraged when losing, you might benefit by using the volleyball scheme, in which you keep serving if you win.

The lesson learned above from *Home-Field Advantage* will again come in handy in the next puzzle.

Who Won the Series?

Two evenly-matched teams meet to play a best-of-seven World Series of baseball games. Each team has the same small advantage when playing at home. As is customary for this event, one team (say, Team A) plays games 1 and 2 at home, and, if necessary, plays games 6 and 7 at home. Team B plays games 3, 4 and, if needed, 5 at home.

You go to a conference in Europe and return to find that the series is over, and six games were played. Which team is more likely to have won the series?

Solution:

You could work this about by assigning a variable $p > 1/2$ to the probability of winning a game at home, then doing a bunch of calculations. But no calculations are actually necessary, if you remember that home-field advantage in *potential* games is as good as it is in games that are always played.

It follows that if you know that *at most* six games were played, you can conclude that teams A and B are equally likely to have won. On the other hand, if you know that at most 5 had been played, then (as host of 3 games) Team B is more likely to have won.

From these two facts we deduce that if *exactly* six games were played, Team A is the more likely winner.

It sometimes happens that there's alternative way to solve a puzzle if you trust the poser, and this is the case here. Since the degree of advantage experienced by the home team is not specified, you might choose to believe that the answer should be the same even if that degree is huge—in other words, if it requires a virtual miracle to win an away game.

In that case, with high probability there was just one upset (game won by the visitor) in the six games that were played, but that upset cannot have occurred in Game 1 or Game 2 else Team B would have won in only five games. Thus it happened in Game 3, 4, 5 or 6 and in the first three of those cases it is Team A who benefitted.

Next is the first of several probability-comparison puzzles.

Dishwashing Game

You and your spouse flip a coin to see who washes the dishes each evening. "Heads" she washes, "tails" you wash.

Tonight she tells you she is imposing a different scheme. You flip the coin 13 times, then she flips it 12 times. If you get more heads than she does, she washes; if you get the same number of heads or fewer, you wash.

Should you be happy?

Solution:

If "should you be happy?" means anything at all, it should mean "are you better off than you were before?".

Here, the easy route is to imagine that first you and your spouse flip just 12 times each. If you get different numbers of heads, the one with fewer will be washing dishes regardless of the outcome of the next flip; so those scenarios cancel. The rest of the time, when you tie, the final flip will determine the washer. So it's still a 50-50 proposition, and you should be indifferent to the change in procedure (unless you dislike flipping coins).

Now suppose that some serious experimentation with the coin in question shows that it is actually slightly biased, and comes up heads 51% of the time. Should you still be indifferent to the new decision procedure? It might seem that you would welcome it now: Heads are good for you, so, the more flipping, the better.

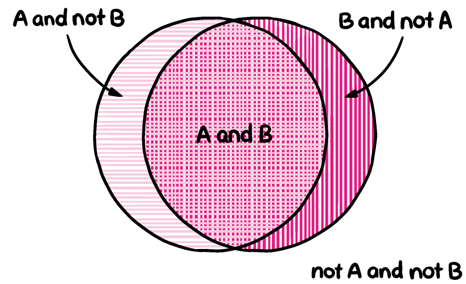
But re-examination of the above argument shows the opposite. When you and your spouse flip 12 times each, it's still equally likely that you flip more heads than she does, or the reverse. Only if the two of you flip the same number of heads do you get your 51% advantage from the final flip. Thus, overall, your probability of ducking the dishwashing falls to somewhere between 50% and 51%.

The analysis suggests a third procedure: You and your spouse alternate flipping the coin until you reach a point where you have both flipped the same number of times but one of you (the winner) has flipped more heads. The advantage of this scheme? It's fair even if the coin is biased!

Many puzzles that involve comparing two probabilities can be approached by a technique called *coupling*, in which two events associated with different conditions are somehow built into a single experiment for which both events make sense.

Then, if you want to determine which of the two events A and B is more probable, you can ignore outcomes where they both occur or neither occurs, and just compare the probability of A happening without B with the probability of B happening without A. This amounts to comparing the areas of the two crescents in the Venn diagram below.

OK, that's a lot of abstract theory. Let's see some examples.



Random Judge

After a wild night of shore leave, you are about to be tried by your U.S. Navy superiors for unseemly behavior. You have a choice between accepting a "summary" court-martial with just one judge, or a "special" court-martial with three judges who decide by majority vote.

Each possible judge has the same (independent) probability—65.43%—of deciding in your favor, except that one officer who would be judging your special court-martial (but not the summary) is notorious for flipping a coin to make his decisions.

Which type of court-martial is more likely to keep you out of the brig?

Solution:

It's not that hard to "do the math" on this one and simply compute your chance of getting off in the special court-martial, then compare it to your 65.43% chance of surviving the summary court-martial. (You might save yourself some arithmetic if you replace 65.43% by a variable, say p, and put the number back in later if necessary.)

But let's try coupling. You may as well suppose officer A would be on both panels and would vote the same way on each; officer C is the random one. If you go with the special court-martial you will regret it precisely when A votes innocent and B and C vote guilty, but thank your stars if A votes guilty while B and C vote innocent.

Without any calculations you know that these events have the same probability, just by exchanging the roles of A and B and reversing C's coinflip. So your two choices are equally good.

Suppose there's also an option of a "general" court-martial with five judges, two of whom are coin-flippers. You can apply the coupling method again, but this time it tells you to choose the five judges if p (as here) is greater than 1/2.

In fact, the Law of Large Numbers tells you that even if 90% of the judges were coin-flippers, the remaining 10% voting in your favor with some fixed probability $p > 1/2$, you could boost your chances to 99% with enough judges.

Wins in a Row

You want to join a certain chess club, but admission requires that you play three games against Ioana, the current club champion, and win two in a row.

Since it is an advantage to play the white pieces (which move first), you alternate playing white and black.

A coin is flipped, and the result is that you will be white in the first and third games, black in the second.

Should you be happy?

Solution:

You can answer this question by assigning a variable w to the probability of winning a game as White, and another variable $b < w$ to the probability of winning as Black; then doing some algebra.

Using coupling, you can get the answer without algebra—even if the problem is modified so that you have to win two in a row out of *seventeen* games, or m in a row out of n. (If n is even, it doesn't matter who plays white first; if n is odd, you want to be black first when m is even, white first when m is odd.)

The coupling argument in the original 2-out-of-3 puzzle goes like this. Imagine that you are going to play *four* games against Ioana, playing white, then black, then white, then black. You still need to win two in a row, but you must decide *in advance* whether to discount the first game, or the last.

Obviously turning the first game into a "practice game" is equivalent to playing BWB in the original problem, and failing to count the last game is equivalent to playing WBW, so the new problem is equivalent to the old one.

But now the events are in the same experiment. For it to make a difference which game you discounted, the results must be either WWLX or XLWW. In words: If you win the first two games, and lose (or draw) the third, you will wish that you had discounted the last game; if you lose the second but win the last two, you will wish that you had discounted the first game.

But it is easy to see that XLWW is more likely than WWLX. The two wins in each case are one with white and one with black, so those cases balance; but the loss in XLWW is with black, more likely than the loss in WWLX with white. So you want to discount the first game, that is, start as black in the original problem.

A slightly more challenging version of this argument is needed if you change the number of games played, and/or the number of wins needed in a row.

Split Games

You are a rabid baseball fan and, miraculously, your team has won the pennant—thus, it gets to play in the World Series. Unfortunately, the opposition is a superior team whose probability of winning any given game against your team is 60%.

Sure enough, your team loses the first game in the best-of-seven series and you are so unhappy that you drink yourself into a stupor. When you regain consciousness, you discover that two more games have been played.

You run out into the street and grab the first passer-by. “What happened in games 2 and 3 of the World Series?”

“They were split,” he says. “One game each.”

Should you be happy?

Solution:

In my experience, presented with this puzzle, about half of respondents answer “Yes—if those games *hadn't* been split, your team would probably have lost them both.”

The other half argue: “No—if your team keeps splitting games, they will lose the series. They have to do better.”

Which argument is correct—and how do you verify the answer without a messy computation?

The task at hand is to determine, after hearing that games 2 and 3 were split, whether you are better or worse off than before. In other words, is your team's probability of winning the series better now, when it needs three of the next four games, than it was before, when it needed four out of six?

Computing and comparing tails of binomial distributions is messy but not difficult. Before the news, your team needed to win 4, 5, or 6 of the next six games. (Wait, what if fewer than seven games are played? As in both *Home-Field Advantage* and *Who Won the Series?*, it costs nothing to imagine that all seven games are played no matter what.)

The probability that your team wins exactly four of six games is “6 choose 4” (the number of ways that can happen) times 0.4^4 (the probability that your team wins a particular four games) times 0.6^2 (the probability that the other guys win the other two). Altogether, the probability that your team wins at least four of six is

$$\binom{6}{4} \cdot 0.4^4 \cdot 0.6^2 + \binom{6}{5} \cdot 0.4^5 \cdot 0.6 + \binom{6}{6} \cdot 0.4^6$$

$$= 15 \cdot 2^4 \cdot 3^2/5^6 + 6 \cdot 2^5 \cdot 3/5^6 + 2^6/5^6 = 112/625 \ .$$

After the second and third games are split, your team needs at least three of the remaining four. The probability of winning is now:

$$\binom{4}{3} \cdot 0.4^3 \cdot 0.6 + \binom{4}{4} \cdot 0.4^4$$

$$= 4 \cdot 2^3 \cdot 3/5^4 + 2^4/5^4 = 112/625 \ (!)$$

So you should be exactly indifferent to the news! The two arguments (one backward-looking, suggesting that you should be happy, the other forward-looking, suggesting that you should be unhappy) seem to have cancelled each other precisely. Can this be a coincidence? Is there any way to get this result "in your head"?

Of course there is—by coupling. To set this up a little imagination helps. Suppose that, after game 3, it is discovered that an umpire who participated in games 2 and 3 had lied on his application. There is a movement to void those two games, and a counter-movement to keep them. The commissioner of baseball, wise man that he is, appoints a committee to decide what to do with the results of games 2 and 3; and in the meantime, he tells the teams to keep playing.

Of course, the commissioner hopes—and so do we puzzle-solvers—that *by the time the committee makes its decision, the question will be moot.*

Suppose five(!) more games are played before the committee is ready to report. If your team wins four or five of them, it has won the series regardless of the disposition of the second and third game results. On the other hand, if the opposition has won three or more, they have won the series regardless. The only case where the committee can make a difference is where your team won exactly three of those five new games.

In that case, if the results of games 2 and 3 are voided, one more game needs to be played; your team wins the series if they win that game, which happens with probability 2/5.

If, on the other hand, the committee decides to count games 2 and 3, the series ended before the last game, and whoever *lost* that game is the World Series winner. Sounds good for your team, no? Oops, remember that your team won three of those five new games, so the probability that the last game is among the two they lost is again 2/5. ♡

Angry Baseball

As in *Split Games*, your team is the underdog and wins any given game in the best-of-seven World Series with probability 40%. But, hold on: This time, whenever your team is behind in the series, the players get angry and play better, raising your team's probability of winning that game to 60%.

Before it all begins, what is your team's probability of winning the World Series?

Solution:

It would be a pain in the neck to work out all the possible outcomes and their probabilities, but coupling again comes to our rescue. Let k be the number of the first game played *after the last time the teams were tied.* Since the teams were tied 0-0 at the start, k could be 1, but it could also be 3, 5, or 7. (It turns out, you don't care!)

Suppose team X wins game k. Let us represent the winners of the rest of the games by a sequence of X's and Y's. Since there are no more ties, X must have won the series, each win by X having had probability 40% and each win by Y, 60%.

If it was Team Y that won game k, the same reasoning applies and the same probability attaches to the subsequent win-sequence except with X's and Y's exchanged. In effect, we couple win-sequences after X wins game k with the complementary win-sequences after Y wins game k.

We conclude that the probability of winning the series for Team X is exactly the probability that Team X wins game k. That probability is 40% when Team X is your team. Since that does not depend on the value of k, your team's probability of winning the series is 40% from the start.

Note that unlike *Split Games*, this puzzle works with any game-winning probability p (changing to $1-p$ when the team is behind). The answer is then p.

Let's move on to (American) football. Are you ready to use probability theory to be an effective coach?

Two-point Conversion

As coach of the Hoboken Hominids (American) football team, you saw your team fall 14 points behind the Gloucester Great Apes until, with just a minute to go in the game, the Hominids scored a touchdown. You have a choice of kicking an extra point (95% success rate) or going for two (45% success rate). Which should you do?

Solution:

You must assume the Hominids will be able to score a second touchdown; and it's reasonable to postulate that if the game goes to overtime, either team is equally likely to win it.

Whatever you do, if it doesn't work you'll need to go for two points after the second TD, winning with probability $0.45 \cdot \frac{1}{2} = 0.225$.

If you do succeed with the first conversion, you'll want to go for just one the next time, winning with probability $\frac{1}{2} \cdot 0.95 = 0.475$ if you had kicked last time but a full 0.95 if you had gotten two extra points last time.

If you succeeded with your two-point conversion but missed the subsequent point after, you still have to win the playoff; that adds another $0.05 \cdot 0.5 = 0.025$ to give a total probability of 0.975 for winning, after your two-point conversion.

So, altogether (assuming that you did get that second touchdown) your probability of winning with the "kick after first TD" option is

$$0.95 \cdot 0.475 + 0.05 \cdot 0.225 = 0.4625,$$

while your probability of winning if you went for two after the first TD is

$$0.45 \cdot 0.975 + 0.55 \cdot 0.225 = 0.5625,$$

so going for two is much better.

It might be easier to see why this is so if you instead imagine that the kick is a sure thing while the two-point conversion has probability of success 1/2. Then if you kick, you are headed for overtime for sure (given, of course, the second touchdown) where you win with probability 1/2. If you go for two and succeed you win for sure, so already you get probability 1/2 of winning the game; but if you miss it ain't over. Altogether the two-point conversion strategy wins with 5/8 of the time.

So why do coaches frequently try to kick the extra point in these situations? Do they think missing the conversion will demoralize their team and make the second touchdown unlikely?

My own theory is that coaches tend to make conservative decisions, making it harder for their managers to point at a coach's decision that cost the game. In this situation, for example, if the team misses both two-point conversions, blame is likely to fall on the coach; but if they kick both extra points and lose in overtime, most of the blame will fall on whichever player screwed up in overtime.

Random Chord

What is the probability that a random chord of a circle is longer than a side of an equilateral triangle inscribed in the circle?

Solution:

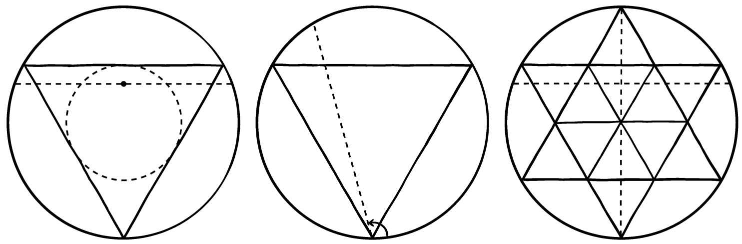
Call a chord "long" if it is longer than the side of our inscribed triangle.

Suppose we choose our random chord by fixing some direction—horizontal is as good as any. Then we can draw a vertical diameter in our circle, pick a point on it uniformly at random, and draw the horizontal chord through that point.

In the left-hand third of the figure below, the inscribed triangle has been drawn with its base up, so that the diameter we just drew is perpendicular to the base. We claim the triangle's base is exactly halfway between the circle's center and the bottom point of the diameter; one way to see this is to draw another inscribed triangle, this one upside-down, so that the two triangles make a "star of David." The hexagon in the middle of the star can be broken into six equilateral triangles and once we do that, it is evident that the part of our diameter that yields a long chord is the same length as the rest. Thus the probability that the random chord is long is 1/2.

But . . . we could instead fix one end of the chord, say at the triangle's apex, and choose its angle uniformly at random, as in the middle third of the figure. Since the angle made by the triangle is $60°$ and our random angle is between $0°$ and $180°$, we deduce that the probability the chord is long is 1/3.

Wait, here's yet another way to solve the problem. Except for diameters (which have probability zero), every chord has a different midpoint and every point inside the circle, except for the center, is the midpoint of a unique chord. So let's pick a point inside the circle uniformly at random, and use the chord



with that midpoint. It's easy to see that the chord will be long exactly if our point is closer to the center of the circle than the midpoint of a side of the triangle (see right-hand third of figure). Thus, we get a long chord if and only if the chosen point lies inside the circle that's inscribed *in* our triangle. But this circle has half the radius, thus one-quarter the area, of the original circle; therefore the probability that our chord is long is 1/4.

What gives? What't the right answer, 1/2, 1/3, 1/4 or something else?

The answer is that the concept of "uniform randomness" for an object chosen from an infinite set is not automatically defined. True, for point chosen from a geometric shape (e.g., an interval or a polygon) we generally take uniformity to mean that the probability that the point lands in a particular subset is proportional to the measure (e.g., length or area) of that subset.

For chords, we don't have a standard notion of uniform randomness and as we've just seen, different attempts at defining uniformity can lead to different answers.

Random Bias

Suppose you choose a real number p between 0 and 1 uniformly at random, then bend a coin so that its probability of coming up heads when you flip it is precisely p. Finally you flip your bent coin 100 times. What is the probability that after all this, you end up flipping exactly 50 heads?

Solution:

This puzzle again calls for a form of coupling; the idea in this case is couple the coinflips even though they involve different coins, depending on the choice of p. To do this we choose, in advance, a uniformly random threshold $p_i \in [0,1]$ for the ith coinflip, independently for each i. Then, after p is chosen, the ith coinflip is deemed to be heads just when $p_i < p$.

Now we need only observe that getting 50 heads is tantamount to p being the middle number, that is, the 51st largest, of the 101 values $p_1, p_2, \ldots, p_{100}, p$. Since these numbers are all uniform and independent, their order is uniformly random and the probability that p ends up in the middle position is exactly 1/101.

Coin Testing

The Unfair Advantage Magic Company has supplied you, a magician, with a special penny and a special nickel. One of these is supposed to flip "Heads" with probability 1/3, the other 1/4, but UAMCO has not bothered to tell you which is which.

Having limited patience, you plan to try to identify the biases by flipping the nickel and penny one by one until one of them comes up Heads, at which point that one will be declared to be the 1/3-Heads coin.

In what order should you flip the coins to maximize the probability that you get the correct answer, and at the same time to be fair, that is, to give the penny and the nickel the same chance to be designated the 1/3-Heads coin?

Solution:

From the point of view of accuracy, at any point when you've flipped each coin the same number of times, it doesn't matter which one you flip next. At any point when you've flipped the penny more times than the nickel, you need to flip the nickel next, and vice versa.

Thus, for the purpose of accuracy, the order doesn't matter as long as for every i, flips $2i-1$ and $2i$ are of different coins.

However, if you want a *fair* test—one equally likely to damn either coin—you need to be more careful. Imagine that you make an extra flip when the first Head turns up at an odd flip, and declare the result to be inconclusive if the succeeding flip (necessarily, of the other coin) is again a Head. This would then certainly be fair to the two coins, but it may mean more flipping. And if you're reading this book, you're likely to be the kind of person that would rather spend an hour figuring something out then an extra minute actually doing it.

It follows from considering this tie scenario that the bias toward the penny (say) is the probability that the "tie" happens with the penny having flipped Heads first, and the nickel immediately after; and vice-versa.

To arrange these biases so that they cancel, we need to order the flips so that if a tie "would have" occurred, it is equally like to have been with the penny or the nickel coming up Heads first.

Suppose we flip the penny first, that is, we begin "PN". The probability of getting Heads already on the first two flips is $\frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$.

What's the probability that the modified game ends in a tie? You will be flipping in pairs—odd flip, then even flip—until you either get HT, TH, or HH, and the tie occurs in the last of these cases, which has probability $(\frac{1}{3} \cdot \frac{1}{4})/(\frac{2}{3} \cdot \frac{1}{4} + \frac{3}{4} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{4}) = (\frac{1}{12})/(\frac{6}{12}) = \frac{1}{6}$.

Therefore, in order to cancel the penny's first-flip advantage, *all* the rest of the odd flips must be performed by the nickel!

In conclusion, the flip order should either be PNNPNPNPNPNP... or NPP-NPNPNPNPN... .

Coin Game

You and a friend each pick a different heads-tails sequence of length 4 and a fair coin is flipped until one sequence or the other appears; the owner of that sequence wins the game.

For example, if you pick HHHH and she picks TTTT, you win if a run of four heads occurs before a run of four tails.

Do you want to pick first or second? If you pick first, what should you pick? If your friend picks first, how should you respond?

Solution:

At first it seems as if this must be a fair game; after all, any particular sequence of four flips has the same probability of appearing (namely, $\frac{1}{16}$) as any other. Indeed, if the coin is flipped four times and then the game begins anew, nothing of interest arises. But because the winning sequence could start with any flip, including the second, third or fourth, significant imbalances appear.

For example, if you pick HHHH (a miserable choice) as your sequence and your friend counters with THHH, you will lose unless the game begins with four heads in a row (only a $\frac{1}{16}$ probability). Why? Because otherwise the first occurrence of HHHH will be preceded by a tail, thus your friend will get there first with her THHH.

How can you determine how to play optimally? For that you need a formula for the probability that string B beats string A. There is, in fact, a very nice one, discovered by the late, indominable John Horton Conway. Here's how it works.

We use the expression "A·B" to denote the degree to which A and B can overlap when A begins first (or they begin simultaneously). We measure this with a four-digit binary number $x_4x_3x_2x_1$ where x_i is 1 if the first i letters of B match the last i of A, otherwise 0. For example, HHTH·HTHT = 0101(binary) = 5 since H[HTH]T gives an overlap of 3 while HHT[H]THT gives an overlap of 1. The left-most digit, x_4, can only be 1 if A=B; thus A·B is equal to or greater than 8 if and only if A=B.

Now if your friend has chosen A as her sequence, how should you choose B? You want B to have low waiting time, thus B·B should be as low as possible. You want B·A to be high, so that you maximize your probability of sneaking in ahead of your friend; and you want A·B to be low to minimize the probability that she sneaks in ahead of you. The formula says that the odds that B beats A are (A·A − A·B):(B·B − B·A).

For example, suppose your friend picks A = THHT as her sequence (then A·A = 9). Your best choice is (always) to tack an H or T to the front of her choice and drop the last letter; this ensures that B·A is at least 4. If you pick B = HTHH you get B·B = 9, B·A = 4, and A·B = 2, so Conway's formula says the odds in your favor are (9-2):(9-4) = 7:5. If instead you pick B = TTHH you get B·B = 8, B·A = 4 and A·B = 1, and the formula says the odds in your favor are (9-1):(8-4) = 2:1. So in this case B = TTHH is the better choice.

Was A = THHT a poor choice by your friend? Actually, it was not bad. Slightly better is THTT or HTHH, where you can't get more than a 9:5 advantage.

Conway's formula continues to work beautifully for strings of length k, for any integer $k \geq 2$, and also to non-binary randomizers such as dice. In the case of dice, the overlap vectors still consist of zeroes and ones, but are now interpreted base 6.

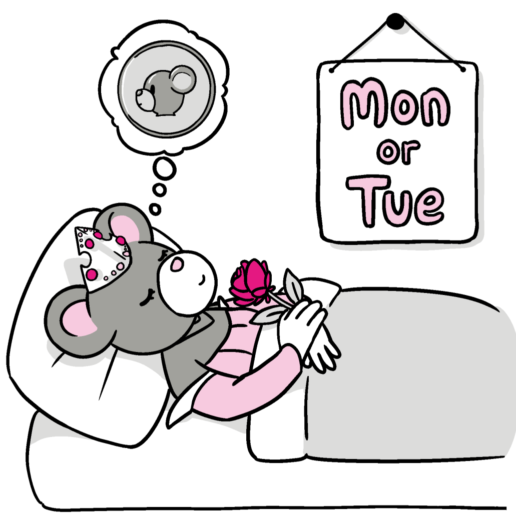
Sleeping Beauty

Sleeping Beauty agrees to the following experiment. On Sunday she is put to sleep and a fair coin is flipped. If it comes up Heads, she is awakened on Monday morning; if Tails, she is awakened on Monday morning and again on Tuesday morning. In all cases, she is not told the day of the week, is put back to sleep shortly after, and will have no memory of any Monday or Tuesday awakenings.

When Sleeping Beauty is awakened on Monday or Tuesday, what—to her—is the probability that the coin came up Heads?

Solution:

This puzzle, introduced by philosopher Adam Elga in 2001, became the subject of major controversy in the philosophy community over the next 20 years. Many were persuaded by the following argument: On Sunday SB (Sleeping Beauty) knows that the probability of Heads is 1/2. She knows she will be awakened, so when she *is* awakened (on Monday or Tuesday), she has no new information and thus no reason to change her mind. So the answer is 1/2.



The flaw in this reasoning is that there is indeed new information: That she is awake *now*. Many arguments show that the correct answer is 1/3; one very nice example, offered by philosophers Cian Dorr and Frank Arntzenius, runs as follows.

Suppose that SB is awakened both days, regardless of the coin, but if the coin comes up Heads, then on Tuesday, 15 minutes after she is awakened, she will be told "It's Tuesday and the coin came up Heads." She knows all this in advance.

When she is awakened, it's equally likely to be Monday or Tuesday (by symmetry) and equally likely to be Heads or Tails (again by symmetry). Thus, the four events "Monday, Tails," "Monday, Heads," "Tuesday, Tails" and "Tuesday, Heads" are all equally likely.

When, 15 minutes later, SB is *not* told that it's "Tuesday, Heads" she rules out that last option and the rest remain equally likely. ♡

You might, on the other hand, be more persuaded by a frequency argument. If the experiment is run 100 times, you'd expect about 150 Monday/Tuesday awakenings, about 1/3 of which would be after Heads was flipped.

The last puzzle will lead us to our theorem for this chapter.

Leading All the Way

Running for local office against Bob, Alice wins with 105 votes to Bob's 95. What is the probability that as the votes were counted (in random order), Alice got the first vote and then led the whole way?

Solution:

Of course, what we'd really like to do is solve this problem in general (when Alice has won with a votes to Bob's b, where a and b are arbitrary positive integers with $a > b$).

The answer is given by what is often known as Bertrand's Ballot Theorem (although it was apparently first proved by W. A. Whitworth in 1878, not J. L. F. Bertrand in 1887).

Theorem. *Let $a > b > 0$ and let $x = (x_1, x_2, \ldots, x_{a+b})$ be a uniformly random string of 1's and -1's, containing a 1's and b -1's. For each k between 1 and $a+b$, let s_k be the sum of the first k elements of x, that is,*

$$s_k = \sum_{i=1}^{k} x_i.$$

Then the probability that $s_1, s_2, \ldots, s_{a+b}$ are all positive is $(a-b)/(a+b)$.

Let's first check that this statement tells us what we want to know. The string x codes the ballot-counting process; each 1 represents a vote for Alice, each -1 a vote for Bob. The sum s_k tells us by how much Alice leads after the

kth vote has been counted; for example, if $s_9 = -3$ we conclude that Alice was actually *behind* by three votes after the ninth vote was counted.

If we apply the theorem to the puzzle above, we get the answer $(105 - 95)/(105 + 95) = 1/20 = 5\%$. So how can we prove the theorem?

There's a famous and clever proof that uses reflection of random walk, but doesn't really explain why the probability turns out to be the vote difference divided by the total vote. Instead, consider the following argument.

One way to choose the random order in which the ballots are counted is first to place the ballots randomly in a circle, then to choose a random ballot in the circle and start counting (say, clockwise) from there. Since there are $a+b$ ballots, there are $a+b$ places to start, and the claim is that no matter what the circular order is, exactly $a-b$ of those starting points are "good" in that they result in Alice leading all the way.

To see this we observe that any occurrence of $1,-1$ (i.e., a vote for Alice followed immediately by a vote for Bob in clockwise order) can be deleted without changing the number of good starting positions. Why? First, you can't start with *that* 1 or -1, since if you start with the 1 the candidates are tied after the second vote, and if you start with the -1, Alice actually falls behind. Secondly, the $1,-1$ pair has no influence upon the the eligibility of any other starting point, since all it does is raise the sum (Alice's lead) by one and then lower it again.

So we can keep deleting $1,-1$'s until all that's left is $a-b$ 1's, each of which is obviously a good starting point. We conclude that all along there were $a-b$ good starting points, and we are done! ♡