Introduction to Probability

Probability is the last topic in this course and perhaps the most important. Many algorithms rely on randomization. Investigating their correctness and performance requires probability theory. Moreover, many aspects of computer systems, such as memory management, branch prediction, packet routing, and load balancing are designed around probabilistic assumptions and analyses. Probability also comes up in information theory, cryptography, artificial intelligence, and game theory. Beyond these engineering applications, an understanding of probability gives insight into many everyday issues, such as polling, DNA testing, risk assessment, investing, and gambling.

So probability is good stuff.

1 Monty Hall

In the September 9, 1990 issue of *Parade* magazine, the columnist Marilyn vos Savant responded to this letter:

> *Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?*
>
> Craig. F. Whitaker
> Columbia, MD

The letter roughly describes a situation faced by contestants on the 1970's game show *Let's Make a Deal*, hosted by Monty Hall and Carol Merrill. Marilyn replied that the contestant should indeed switch. But she soon received a torrent of letters— many from mathematicians— telling her that she was wrong. The problem generated thousands of hours of heated debate.

Yet this is is an elementary problem with an elementary solution. Why was there so much dispute? Apparently, most people *believe* they have an intuitive grasp of probability. (This is in stark contrast to other branches of mathematics; few people believe they have an intuitive ability to compute integrals or factor large integers!) Unfortunately, approximately 100% of those people are *wrong*. In fact, everyone who has studied probability at

length can name a half-dozen problems in which their intuition led them astray— often embarassingly so.

The way to avoid errors is to distrust informal arguments and rely instead on a rigorous, systematic approach. In short: intuition *bad*, formalism *good*. If you insist on relying on intuition, then there are lots of compelling financial deals we'd love to offer you!

1.1 The Four-Step Method

Every probability problem involves some sort of randomized experiment, process, or game. And each such problem involves two distinct challenges:

1. How do we model the situation mathematically?
2. How do we solve the resulting mathematical problem?

In this section, we introduce a four-step approach to questions of the form, "What is the probability that —— ?" In this approach, we build a probabilistic model step-by-step, formalizing the original question in terms of that model. Remarkably, the structured thinking that this approach imposes reduces many famously-confusing problems to near triviality. For example, as you'll see, the four-step method cuts through the confusion surrounding the Monty Hall problem like a Ginsu knife. However, more complex probability questions may spin off challenging counting, summing, and approximation problems—which, fortunately, you've already spent weeks learning how to solve!

1.2 Clarifying the Problem

Craig's original letter to Marilyn vos Savant is a bit vague, so we must make some assumptions in order to have any hope of modeling the game formally:

1. The car is equally likely to be hidden behind each of the three doors.
2. The player is equally likely to pick each of the three doors, regardless of the car's location.
3. After the player picks a door, the host *must* open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
4. If the host has a choice of which door to open, then he is equally likely to select each of them.

In making these assumptions, we're reading a lot into Craig Whitaker's letter. Other interpretations are at least as defensible, and some actually lead to different answers. But let's accept these assumptions for now and address the question, "What is the probability that a player who switches wins the car?"

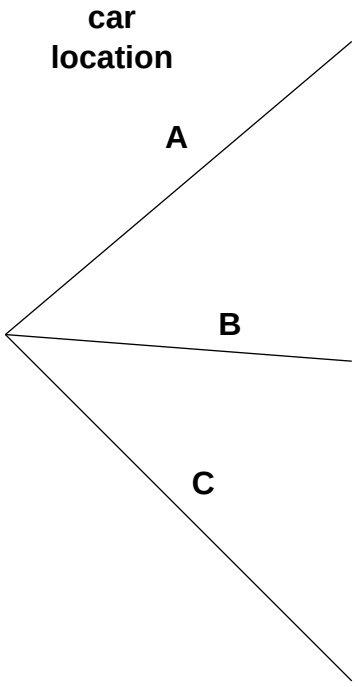
1.3 Step 1: Find the Sample Space

Our first objective is to identify all the possible outcomes of the experiment. A typical experiment involves several randomly-determined quantities. For example, the Monty Hall game involves three such quantities:

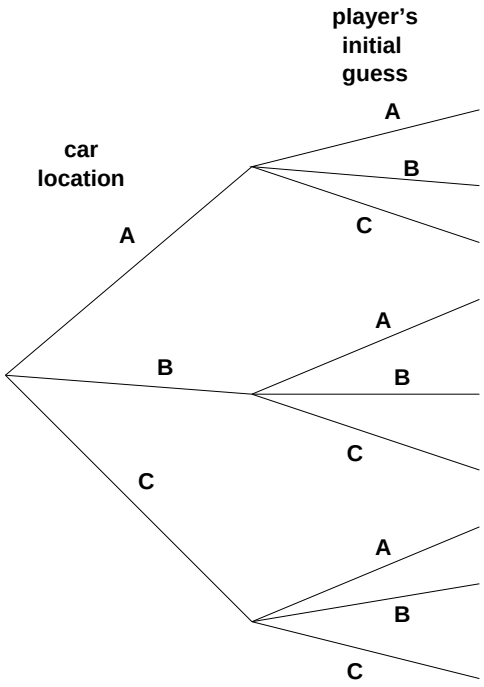
1. The door concealing the car.
2. The door initially chosen by the player.
3. The door that the host opens to reveal a goat.

Every possible combination of these randomly-determined quantities is called an ***outcome***. The set of all possible outcomes is called the ***sample space*** for the experiment.

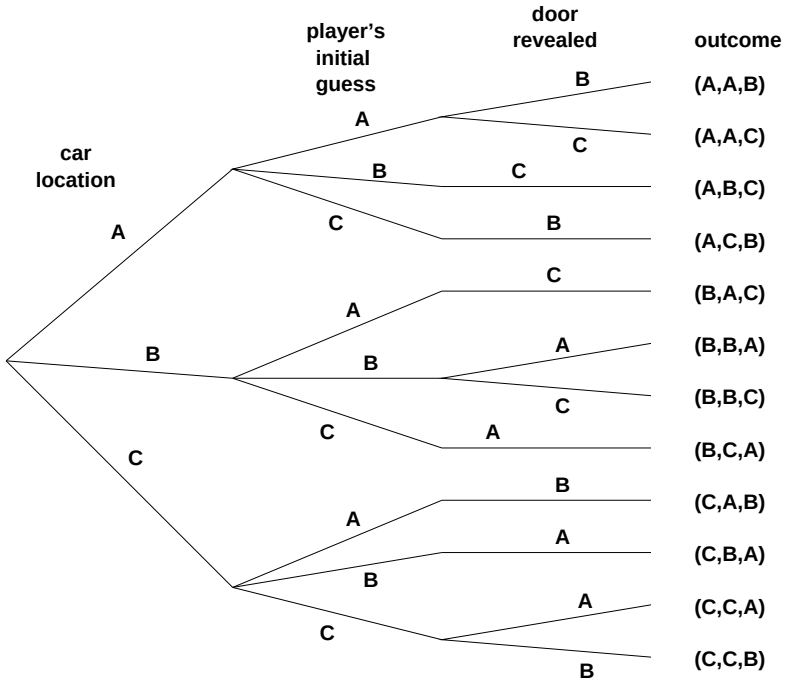
A ***tree diagram*** is a graphical tool that can help us work through the four-step approach when the number of outcomes is not too large or the problem is nicely structured. In particular, we can use a tree diagram to help understand the sample space of an experiment. The first randomly-determined quantity in our experiment is the door concealing the prize. We represent this as a tree with three branches:



In this diagram, the doors are called A, B, and C instead of 1, 2, and 3 because we'll be adding a lot of other numbers to the picture later. Now, for each possible location of the prize, the player could initially chose any of the three doors. We represent this by adding a second layer to the tree:



Finally, the host opens a door to reveal a goat. The host has either one choice or two, depending on the position of the car and the door initially selected by the player. For example, if the prize is behind door A and the player picks door B, then the host must open door C. However, if the prize is behind door A and the player picks door A, then the host could open either door B or door C. All of these possibilities are worked out in a third layer of the tree:



Now let's relate this picture to the terms we introduced earlier: the leaves of the tree represent *outcomes* of the experiment, and the set of all leaves represents the *sample space*. Thus, for this experiment, the sample space consists of 12 outcomes. For reference, we've labeled each outcome with a triple of doors indicating:

$$(\text{door concealing prize},\ \text{door initially chosen},\ \text{door opened to reveal a goat})$$

In these terms, the sample space is the set:

$$S = \left\{ \begin{array}{llllll} (A,A,B), & (A,A,C), & (A,B,C), & (A,C,B), & (B,A,C), & (B,B,A), \\ (B,B,C), & (B,C,A), & (C,A,B), & (C,B,A), & (C,C,A), & (C,C,B) \end{array} \right\}$$

The tree diagram has a broader interpretation as well: we can regard the whole experiment as "walk" from the root down to a leaf, where the branch taken at each stage is randomly determined. Keep this interpretation in mind; we'll use it again later.

1.4 Step 2: Define Events of Interest

Our objective is to answer questions of the form "What is the probability that —— ?", where the horizontal line stands for some phrase such as "the player wins by switching", "the player initially picked the door concealing the prize", or "the prize is behind door C". Almost any such phrase can be modeled mathematically as an ***event***, which is defined to be a subset of the sample space.

For example, the event that the prize is behind door C is the set of outcomes:

$$\{(C, A, B), (C, B, A), (C, C, A), (C, C, B)\}$$

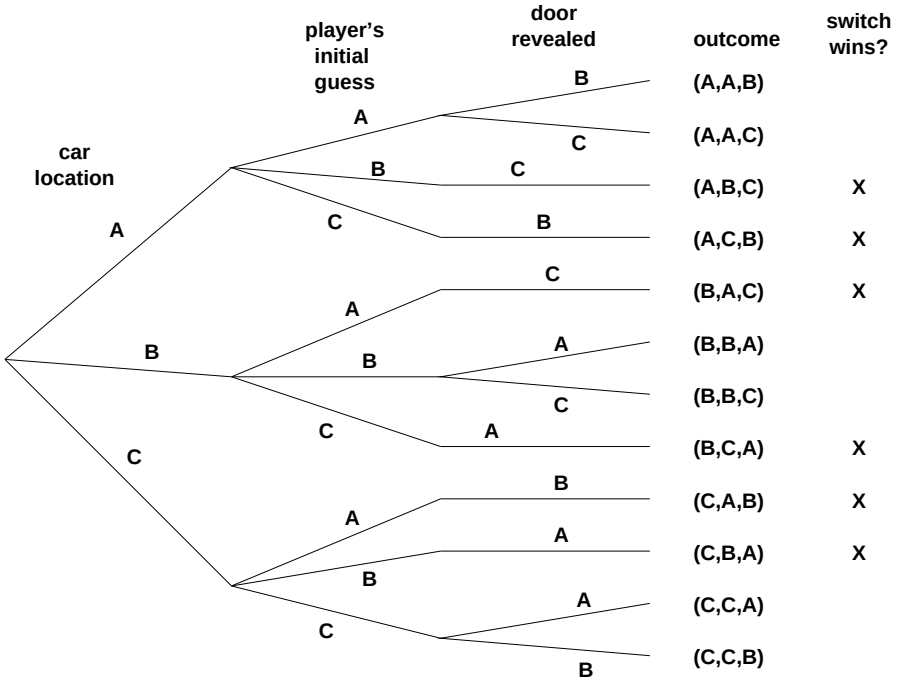
The event that the player initially picked the door concealing the prize is the set of outcomes:

$$\{(A, A, B), (A, A, C), (B, B, A), (B, B, C), (C, C, A), (C, C, B)\}$$

And what we're really after, the event that the player wins by switching, is the set of outcomes:

$$\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}$$

Let's annonate our tree diagram to indicate the outcomes in this event.



Notice that exactly half of the outcomes are marked, meaning that the player wins by switching in half of all outcomes. You might be tempted to conclude that a player who switches wins with probability $\frac{1}{2}$. *This is wrong.* The reason is that these outcomes are not all equally likely, as we'll see shortly.

1.5 Step 3: Determine Outcome Probabilities

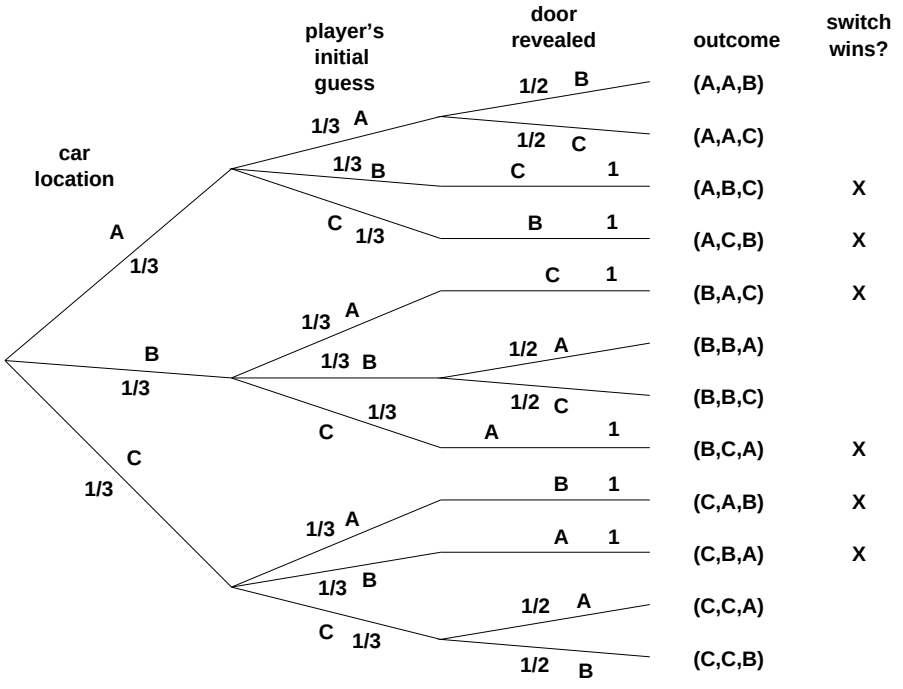
So far we've enumerated all the possible outcomes of the experiment. Now we must start assessing the likelihood of those outcomes. In particular, the goal of this step is to assign

each outcome a probability, which is a real number between 0 and 1. The sum of all outcome probabilities must be 1, reflecting the fact that exactly one outcome must occur.

Ultimately, outcome probabilities are determined by the phenomenon we're modeling and thus are not quantities that we can derive mathematically. However, mathematics can help us compute the probability of every outcome *based on fewer and more elementary modeling decisions.* In particular, we'll break the task of determining outcome probabilities into two stages.

1.5.1 Step 3a: Assign Edge Probabilities

First, we record a probability on each *edge* of the tree diagram. These edge-probabilities are determined by the assumptions we made at the outset: that the prize is equally likely to be behind each door, that the player is equally likely to pick each door, and that the host is equally likely to reveal each goat, if he has a choice. Notice that when the host has no choice regarding which door to open, the single branch is assigned probability 1.



1.5.2 Step 3b: Compute Outcome Probabilities

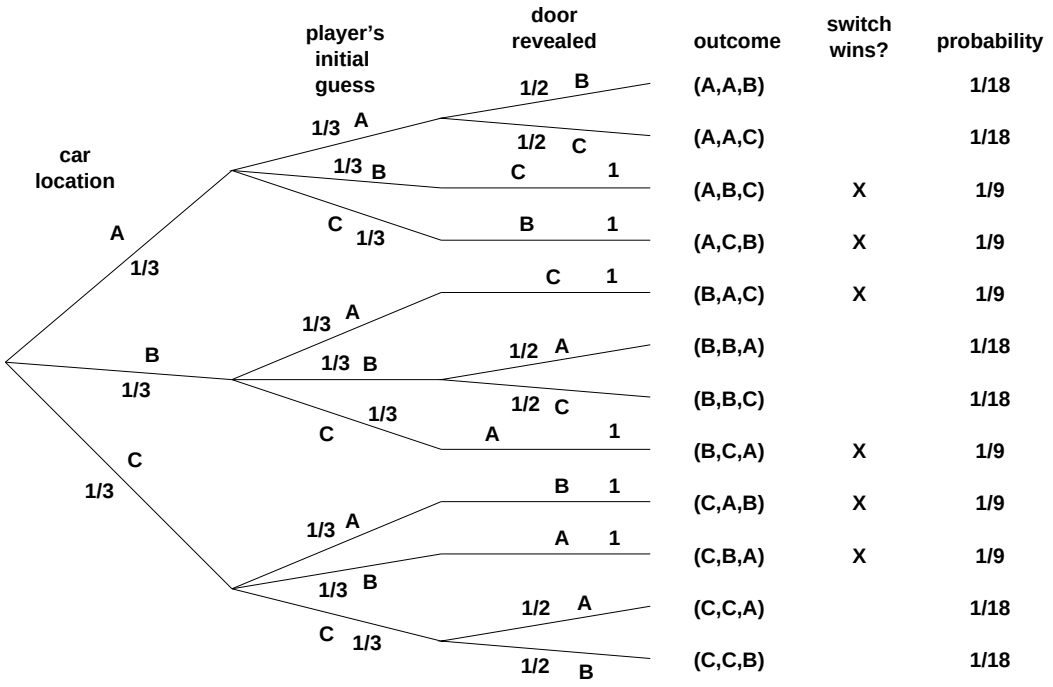
Our next job is to convert edge probabilities into outcome probabilities. This is a purely mechanical process: *the probability of an outcome is equal to the product of the edge-probabilities*

on the path from the root to that outcome. For example, the probability of the topmost outcome, (A, A, B) is $\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}$.

We'll justify this process formally next time. In the meanwhile, here is a nice informal justification to tide you over. Remember that the whole experiment can be regarded as a walk from the root of the tree diagram down to a leaf, where the branch taken at each step is randomly determined. In particular, the probabilities on the edges indicate how likely the walk is to proceed along each path. For example, a walk starting at the root in our example is equally likely to go down each of the three top-level branches.

Now, how likely is such a walk to arrive at the topmost outcome, (A, A, B)? Well, there is a 1-in-3 chance that a walk would follow the A-branch at the top level, a 1-in-3 chance it would continue along the A-branch at the second level, and 1-in-2 chance it would follow the B-branch at the third level. Thus, it seems that about 1 walk in 18 should arrive at the (A, A, B) leaf, which is precisely the probability we assign it.

Anyway, let's record all the outcome probabilities in our tree diagram.



Specifying the probability of each outcome amounts to defining a function that maps each outcome to a probability. This function is usually called **Pr**. In these terms, we've

just determined that:

$$\Pr(A, A, B) = \frac{1}{18}$$
$$\Pr(A, A, C) = \frac{1}{18}$$
$$\Pr(A, B, C) = \frac{1}{9}$$
$$\text{etc.}$$

Earlier, we noted that the sum of all outcome probabilties must be 1 since exactly one outcome must occur. We can now express this symbolically:

$$\sum_{x \in S} \Pr(x) = 1$$

In this equation, S denotes the sample space.

Though $\Pr$ is an ordinary function, just like your old friends f and g from calculus, we will subject it to all sorts of horrible notational abuses that f and g were mercifully spared. Just for starters, all of the following are common notations for the probability of an outcome x:

$$\Pr(x) \qquad \Pr(x) \qquad \Pr[x] \qquad \Pr x \qquad p(x)$$

A sample space S and a probability function $\Pr : S \to [0, 1]$ together form a ***probability space***. Thus, a probability space describes all possible outcomes of an experiment *and* the probability of each outcome. A probability space is a complete mathematical model of an experiment.

If for any two outcomes x and y, we have $\Pr(x) = \Pr(y)$, then we say the sample space is a *uniform sample space*. In this case, since $\sum_{x \in S} \Pr(x) = 1$, it must be the case that $\Pr(x) = \frac{1}{|S|}$ for all outcomes x. It turns out that the Monty Hall sample space is *not* uniform.

1.6 Step 4: Compute Event Probabilities

We now have a probability for each *outcome*, but we want to determine the probability of an *event*. We can bridge this gap with a definition: *the probability of an event is the sum of the probabilities of the outcomes it contains*. As a notational matter, the probability of an event $E \subseteq S$ is written $\Pr(E)$. Thus, our definition of the probability of an event can be written:

$$\Pr(E) = \sum_{x \in E} \Pr(x)$$

For example, the probability of the event that the player wins by switching is:

$$\begin{aligned}\Pr(\text{switching wins}) &= \Pr(A,B,C) + \Pr(A,C,B) + \Pr(B,A,C) + \\ &\quad \Pr(B,C,A) + \Pr(C,A,B) + \Pr(C,B,A) \\ &= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \\ &= \frac{2}{3}\end{aligned}$$

It seems Marilyn's answer is correct; a player who switches doors wins the car with probability $2/3$! In contrast, a player who stays with his or her original door wins with probability $1/3$, since staying wins if and only if switching loses.

If in the Monty Hall problem, the sample space were uniform, then, since a player wins by switching in exactly half of the outcomes, the player who switches would win with probability $\frac{1}{2}$. However, as we've seen, *this is wrong* since the Monty Hall sample space is not uniform.

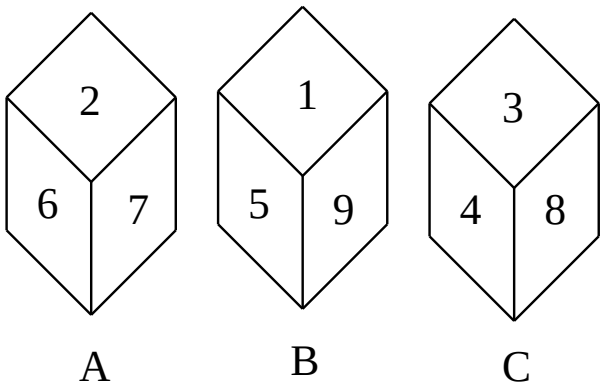
We're done with the problem! We didn't need any appeals to intuition or ingenious analogies. In fact, no mathematics more difficult than adding and multiplying fractions was required. The only hard part was resisting the temptation to leap to an "intuitively obvious" answer.

1.7 An Alternative Interpretation of the Monty Hall Problem

Was Marilyn really right? A more accurate conclusion is that her answer is correct *provided we accept her interpretation of the question*. There is an equally plausible interpretation in which Marilyn's answer is wrong. Notice that Craig Whitaker's original letter does not say that the host is *required* to reveal a goat and offer the player the option to switch, merely that he *did* these things. In fact, on the *Let's Make a Deal* show, Monty Hall sometimes simply opened the door that the contestant picked initially. Therefore, if he wanted to, Monty could give the option of switching only to contestants who picked the correct door initially. If this case, switching never works!

2 Strange Dice

Let's play *Strange Dice*! The rules are simple. There are three dice, A, B, and C. Not surprisingly, the dice are numbered *strangely*, as shown below:



The number on each concealed face is the same as the number on the opposite, exposed face. The rules are simple. You pick one of the three dice, and then I pick one of the two remainders. We both roll and the player with the higher number wins.

Which of the dice should you choose to maximize your chances of winning? Die B is appealling, because it has a 9, the highest number overall. Then again, die A has two relatively large numbers, 6 and 7. But die C has an 8 and no very small numbers at all. Intuition gives no clear answer!

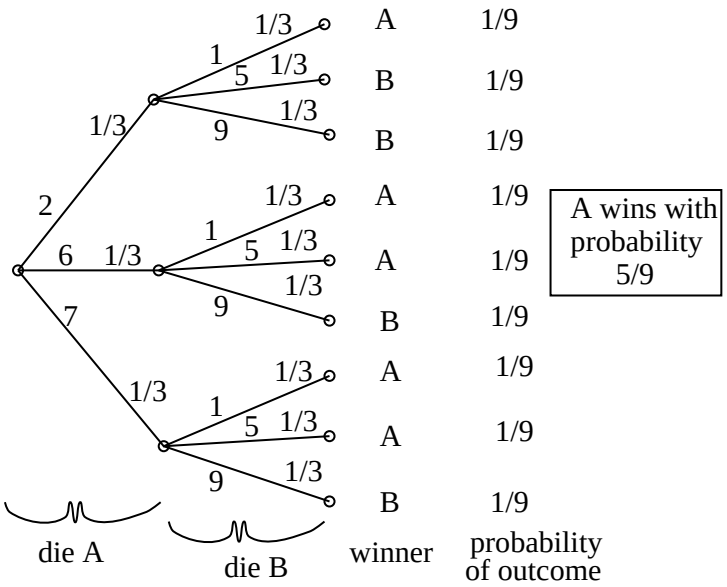
2.1 Analysis of Strange Dice

We can analyze Strange Dice using our standard, four-step method for solving probability problems. To fully understand the game, we need to consider three different experiments, corresponding to the three pairs of dice that could be pitted against one another.

2.1.1 Die *A* versus Die *B*

First, let's determine what happens when die A is played against die B.

Step 1: Find the sample space. The sample space for this experiment is worked out in the tree diagram show below. (Actually, the whole probability space is worked out in this one picture. But pretend that each component sort of fades in— nyyyrrroom!— as you read about the corresponding step below.)



For this experiment, the sample space is a set of nine outcomes:

$$S = \{\ (2,1),\ (2,5),\ (2,9),\ (6,1),\ (6,5),\ (6,9),\ (7,1),\ (7,5),\ (7,9)\ \}$$

Step 2: Define events of interest. We are interested in the event that the number on die A is greater than the number on die B. This event is a set of five outcomes:

$$\{\ (2,1),\ (6,1),\ (6,5),\ (7,1),\ (7,5)\ \}$$

These outcomes are marked A in the tree diagram above.

Step 3: Determine outcome probabilities. To find outcome probabilities, we first assign probabilities to edges in the tree diagram. Each number on each die comes up with probability $1/3$, regardless of the value of the other die. Therefore, we assign all edges probability $1/3$. The probability of an outcome is the product of probabilities on the corresponding root-to-leaf path, which means that every outcome has probability $1/9$. These probabilities are recorded on the right side of the tree diagram.

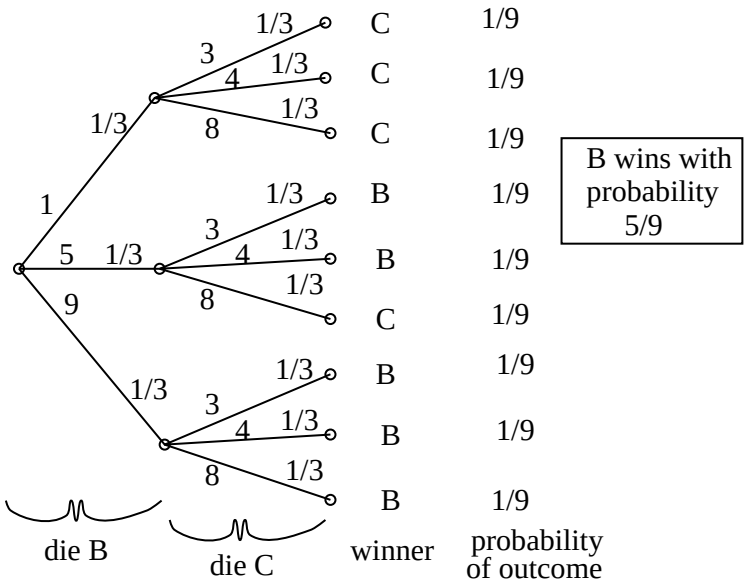
Step 4: Compute event probabilities. The probability of an event is the sum of the probabilities of the outcomes in that event. Therefore, the probability that die A comes up greater than die B is:

$$\begin{aligned}\Pr(A > B) &= \Pr(2,1) + \Pr(6,1) + \Pr(6,5) + \Pr(7,1) + \Pr(7,5) \\ &= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \\ &= \frac{5}{9}\end{aligned}$$

Therefore, die A beats die B more than half of the time. You had better not choose die B or else I'll pick die A and have a better-than-even chance of winning the game!

2.1.2 Die *B* versus Die *C*

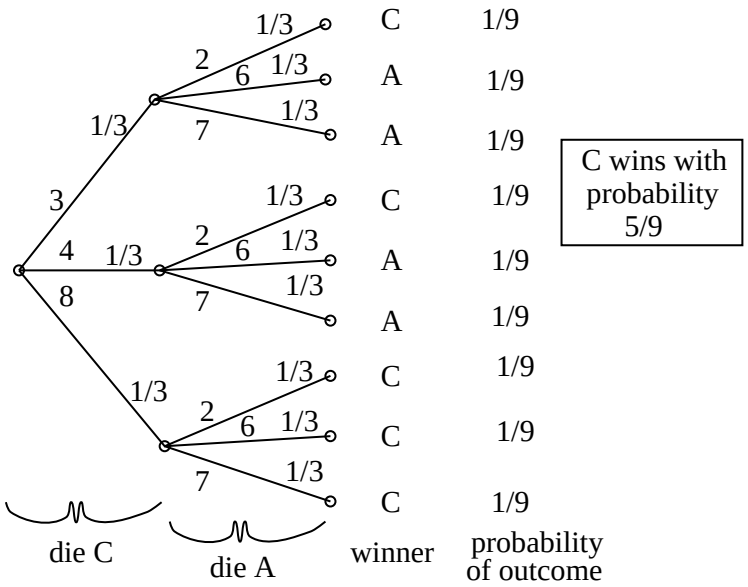
Now suppose that die B is played against die C. The tree diagram for this experiment is shown below.



The analysis is the same as before and leads to the conclusion that die B beats die C with probability $5/9$ as well. Therefore, you had beter not choose die C; if you do, I'll pick die B and most likely win!

2.1.3 Die *C* versus Die *A*

We've seen that A beats B and B beats C. Apparently, die A is the best and die C is the worst. The result of a confrontation between A and C seems a forgone conclusion. A tree diagram for this final experiment is worked out below.



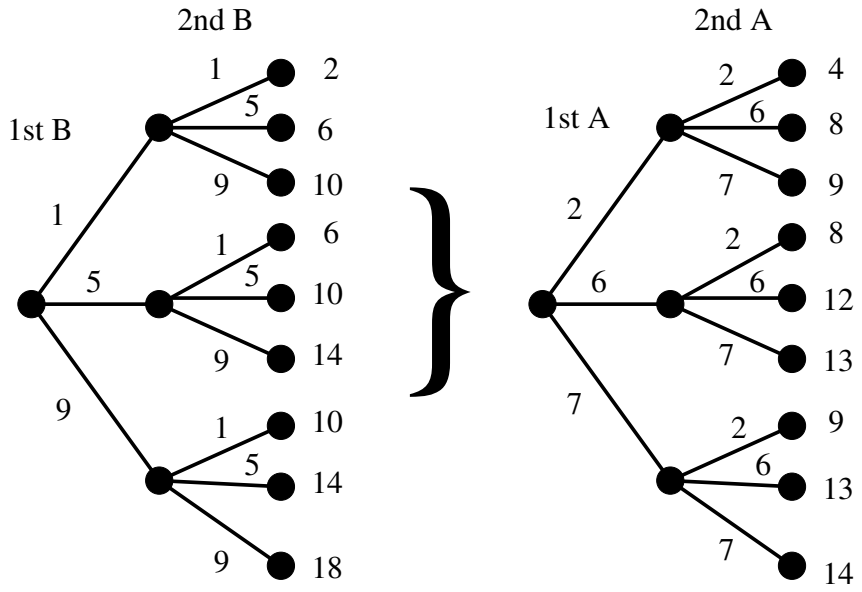
Surprisingly, die C beats die A with probability $5/9$!

In summary, die A beats B, B beats C, and C beats A! Evidently, there is a relation between the dice that is *not transitive*! This means that no matter what die the first player chooses, the second player can choose a die that beats it with probability $5/9$. The player who picks first is always at a disadvantage!

Challenge: The dice can be renumbered so that A beats B and B beats C, each with probability $2/3$, and C still beats A with probability $5/9$. Can you find such a numbering?

All right, we will play one more game. This time we'll each roll our die twice and add the result. The highest result wins. I will pick die B and you will pick die A, since intuitively, A beats B with 1 roll, so you can beat me by choosing die A. Let's argue about this formally, and see if you are correct.

We first write down the tree for the sample space.



The sample space is a little more complicated and hard to write out the whole tree this time. In the figure, it should be understood that the tree corresponding to A is connected to each leaf of the tree corresponding to B.

First of all, how many leaves are there in the whole tree? There are 81. What is the probability of each leaf? Well this is a uniform sample space, so it is $(1/3)^4 = 1/81$. Let's work out the chances of winning. The sum of the two rolls of the A die is equally likely to be any element of the following multiset:

$$S_A = \{4, 8, 8, 9, 9, 12, 13, 13, 14\}.$$

The sum of the two rolls of the B die is equally likely to be any element of the following multiset:

$$S_B = \{2, 6, 6, 10, 10, 10, 14, 14, 18\}.$$

We can treat each outcome as a pair $(x, y) \in S_A \times S_B$, where B wins iff $y > x$. If $y = 2$, there is no x for which $y > x$. If $y = 6$, there is 1 value of x, namely $x = 4$, for which $y > x$. Continuing the count in this way, the number of pairs for which $y > x$ is

$$0 + 1 + 1 + 5 + 5 + 5 + 8 + 8 + 9 = 42,$$

while a similar count shows that there are only 37 pairs for which $x > y$, and there are two pairs $((14, 14), (14, 14))$ which result in ties.

Thus, rolling die B twice is more likely to win than rolling die A twice! How can this be? We say that A is more likely than B to win 1 roll, but B is more likely to win 2 rolls ?!? Well, why not? The only reason we'd think otherwise is in fact our faulty intuition. In fact, the strength reverses no matter which two die we picked. So for 1 roll, we had

$$A > B > C > A,$$

and for 2 rolls you will show in the homework that

$$A < B < C < A.$$

Challenge: What happens for 3 or 4 rolls? Extra credit if someone wants to write code to figure it out. Maybe then we'll prove a general theorm for any number k of rolls.

Appendix: The Four-Step Method

This is a good approach to questions of the form, "What is the probability that ——–?" Intuition *will* mislead you, but this formal approach gives the right answer every time.

1. Find the sample space. (Use a tree diagram.)
2. Define events of interest. (Mark leaves corresponding to these events.)
3. Determine outcome probabilities:
   (a) Assign edge probabilities.
   (b) Compute outcome probabilities. (Multiply along root-to-leaf paths.)
4. Compute event probabilities. (Sum the probabilities of all outcomes in the event.)