Lecture Notes

Probability will be the topic for the rest of the term.
Probability theory is one of the most important subjects in all of mathematics and computer science.
Most upper level computer science courses require probability in some form.
It is used in analysis of algorithms and data structures; it is used in information theory, cryptography, and game theory.

Probability is also frequently an issue in daily life.
For example, last month there was national controversy about the probability that the nuclear-powered spacecraft Cassini would accidentally coat the earth with super-toxic plutonium dioxide.
Not an entirely academic question.
In a press release, NASA asserted:

> Jet Propulsion Laboratory has conducted an in-depth analysis, which incorporated human error[1] and historical JPL spacecraft data, to determine the probability of an inadvertent reentry.
> This analysis determined that the probability of an inadvertent Earth reentry is less than one in one million.

Yet some people felt that the probability of disaster was sufficiently high that they staged vigorous protests and filed suit in federal court to prevent the launch.
How could there be such intense dispute about a probability question?

Part of the problem is that probability is a tricky subject.
In principle, probability boils down to a few simple rules.
The difficulty is that these rules often produce highly counterintuitive results; "common sense" often leads to the wrong answer! We will see two examples of this today.

1 Basic Concepts

One intuition about probability is that we want to predict how likely it is for a given experiment to have a certain kind of outcome.
Asking this question invariably involves four distinct steps:

[1] This is just a grammar bug.
Yeah.
Well, hopefully.

Find the sample space. Determine all the possible outcomes of the experiment.

Define the event of interest. Determine which of those possible outcomes is "interesting.
"

Determine the individual outcome probabilities. Decide how likely each individual outcome is to occur.

Determine the probability of the event. Combine the probabilities of "interesting" outcomes to find the overall probability of the event we care about.

In order to understand these four steps, we will begin with a toy problem.
We consider rolling three dice, and try to determine the probability that we roll exactly two sixes.

Step 1: Find the Sample Space

Every probability problem involves some experiment or game.
The key to most probability problems is to look carefully at the *sample space* of the the experiment.
This term is defined below.

Definition The *sample space* of an experiment is the set of all possible outcomes.

Definition An *outcome* of an experiment consists of the total information about the experiment after it has been performed.
This includes the values of all random choices that were made.

(An outcome is also called a "sample point" or an "atomic event".
JPL may have preferred the latter term in its Cassini analysis.
)

In our die rolling experiment, a particular outcome can be expressed as a triple of numbers from 1 to 6.
For example, the triple $(3, 5, 6)$ indicates that the first die rolled 3, the second rolled 5, and the third rolled 6.

Step 2: Define Events of Interest

We usually declare some subset of the possible outcomes in the sample space to be "good.
" Any subset of the sample space is called an *event*.

Definition An *event* is a subset of the sample space.

For example, the event that all dice are the same consists of six possible outcomes

$$\{(1,1,1),(2,2,2),(3,3,3),(4,4,4),(5,5,5),(6,6,6)\} .$$

The event that we roll exactly two sixes consists of $3*5=15$ possible outcomes (we need to choose which die is not a six, and then we need to choose a value for that die):

$$\begin{aligned} T = \{&(1,6,6),(2,6,6),(3,6,6),(4,6,6),(5,6,6),\\ &(6,1,6),(6,2,6),(6,3,6),(6,4,6),(6,5,6),\\ &(6,6,1),(6,6,2),(6,6,3),(6,6,4),(6,6,5)\} \end{aligned}$$

Our goal is to determine the probability that our experiment yields one of the outcomes in this set T.

Step 3: Compute Outcome Probabilities

The outcomes of an experiment and their probabilities are described by a *probability space*.

Definition The *probability space* of an experiment consists of the sample space S and a probability function $\Pr : S \mapsto \mathcal{R}$ such that:

1. For every outcome w in the sample space S, we have $0 \leq \Pr(w) \leq 1$.

2. The sum of the probabilities of all outcomes is 1; that is, $\sum_{w \in S} \Pr(w) = 1$.

We interpret $\Pr(w)$ to be the probability that w is the outcome of the experiment.

1.0.1 Assumptions

Assigning probabilities to the atomic outcomes is a *axiomatic* action.
In our experiment, it seems quite plausible to say that all the possible outcomes are equally likely.

Definition 1.1 A *uniform* sample space S is one in which all the atomic events are equally likely, meaning $\Pr(w) = 1/|S|$ for every $w \in S$.

Since there are $6^3 = 216$ possible outcomes, we axiomatically declare that each occurs with probability $1/216$.

Step 4: Compute Event Probabilities

We now have a probability for each outcome.
All that remains is to compute the probability of T, the event that we get exactly two sixes.
The following definition explains how to compute the probability of an event given the probability of each outcome.

Definition Let S be the sample space of an experiment, and let Pr be the probability function. Then for any event $T \subseteq S$:

$$\Pr(T) = \sum_{w \in T} \Pr(w)$$

In words, the probability of an event is defined to be the sum of the probabilities of all the outcomes in that event.

In our example, since there are 15 outcomes in T, each with probability $1/216$, we can deduce that $\Pr(T) = 15/216$.

1.1 Coin Tossing

Here's another simple example.
Imagine that perform 100 tosses of a fair coin.
What is the probability that we get an even number of heads? We apply the four-step analysis:

Define the sample space. We can represent a sequence of 100 tosses as a length-100 sequence of letters H and T; for example the an outcome of alternating heads and tails can be written as $HTHTHTHT\cdots$.

Define the interesting event. We are interested in sequences with an even number of H symbols in them.

Assign probabilities to the atomic events. We will assume the coin is fair and is tossed fairly, so we declare that all the outcomes are equally likely.
Since our sample space has 2^100 elements, each has probability 2^{-100}.
This assumption is a strong one: a certain famous mathematician can get heads 3 times out of 4 on a fair coin (this took lots of practice using a strobe light).

Determine the interesting event probability. Since the sample space is uniform, this comes down to determining the number of sequences with an even number of heads.
We have seen this problem before in another guise.
For each sequence, consider a corresponding set: the set of coin-toss-numbers (between 1 and 100) that yielded heads.

There is an even number of heads precisely when the corresponding set is even.
So we just need to count the number of even subsets of $\{1, \ldots, 100\}$.
We saw before that this is $2^9 9$.

So, the overall probability of our event is just $2^{99} \cdot 2^{-100}$.

As we can see from the above examples, probability on uniform sample spaces is pretty much the same as counting.
Instead of asking how many distinct full houses there are in poker, we can ask about the probability that a "random" poker hand is a full house.
Probability provides a kind of normalization that often makes for answers that are more pleasant to write down.
For example, our analysis above really says that for any number of coin tosses, the probability that I get an even number of heads is 1/2—I have normalized away any dependence on the number of coin tosses.
I can obviously switch back to the counting version if I just multiply the probability by the number of outcomes.

2 The Monty Hall Problem

In the 1970's, there was a game show called Let's Make a Deal, hosted by Monty Hall and his assistant Carol Merryl.
At one stage of the game, a contestant is shown three doors.
The contestant knows there is a prize behind one door and that there are goats behind the other two.
The contestant picks a door.
To build suspense, Carol always opens a *different* door, revealing a goat.
The contestant can then stick with his original door or switch to the other unopened door.
He wins the prize only if he now picks the correct door.
Should the contestant "stick" with his original door, "switch" to the other door, or does it not matter?

This was the subject of an "Ask Marilyn" column in *Parade* Magazine a few years ago.
Marylin wrote that your chances of winning were 2/3 if you switched — because if you switch, then you win if the car was originally behind either of the two doors you didn't pick.
Now, Marylin is listed in the *Guiness Book of World Records* as having the world's highest IQ, but for this answer she got a tidal wave of angry mail, much of it from people with Ph.
D.
's in mathematics, telling her she was wrong, and that the answer was 1/2 — because after Monty opens a door, the car is now equally likely to be behind one of the two remaining doors.

It turned out that Marilyn was right.
But both of the arguments above are based on intuition, and the answer is actually pretty subtle.
So rather than try to come up with our own explanation in words, let's use the definitions of probability.

We will analyze the probability that the contestant wins with the "switch" strategy; that is, the contestant chooses a random door initially and then always switches after Carol reveals a goat behind one door.
The analysis can be broken down into the standard four steps.

Step 1: Find the Sample Space

In the Monty Hall problem, an outcome is a triple of door numbers:

1. The number of the door concealing the prize.
2. The number of the door initially chosen by the contestant.
3. The number of the door Carol opens to reveal a goat.

For example, the outcome $(2,1,3)$ represents the case where the prize is behind door 2, the contestant initially chooses door 1, and Carol reveals the goat behind door 3.
In this case, a contestant using the "switch" strategy wins the prize.

Not every triple of numbers is an outcome; for example, $(1,2,1)$ is not an outcome, because Carol never opens the door with the prize.
Similarly, $(1,2,2)$ is not an outcome, because Carol does not open the door initially selected by the contestant, either.

As with counting, a tree diagram is a standard tool for studying the sample space of an experiment.
The tree diagram for the Monty Hall problem is shown in Figure 1.
Each vertex in the tree corresponds to a state of the experiment.
In particular, the root represents the initial state, before the prize is even placed.
Internal nodes represent intermediate states of the experiment, such as after the prize is placed, but before the contestant picks a door.
Each leaf represents a final state, an outcome of the experiment.
One can think of the experiment as a walk from the root (initial state) to a leaf (outcome).
In the figure, each leaf of the tree is labeled with an outcome (a triple of numbers) and a "W" or "L" to indicate whether the contestant wins or loses.

Step 2: Define Events of Interest

For the Monty Hall problem, let S denote the sample space, the set of all 12 outcomes shown in Figure 1.
The event $S_w \subseteq S$ that the contestant wins with the "switch" strategy consists of six outcomes:

$$S_w = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$$

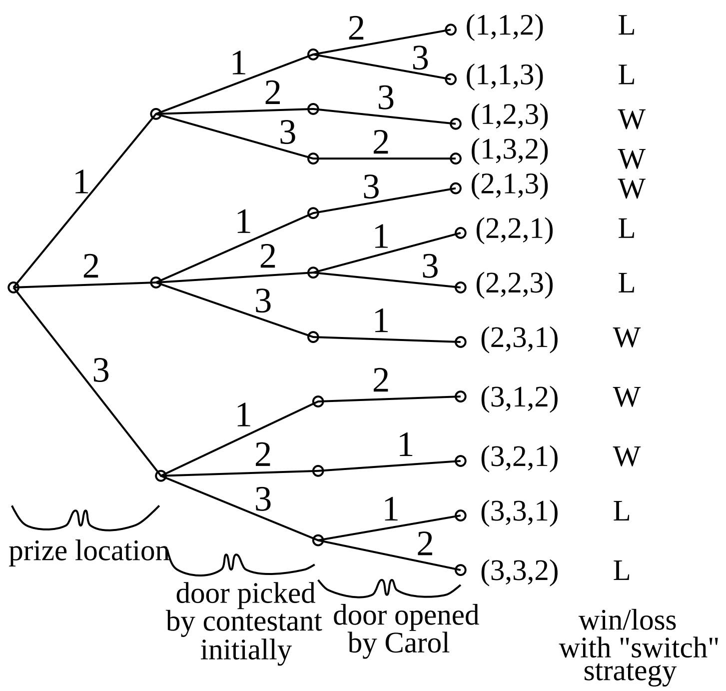

Figure 1: *This is a tree diagram for the Monty Hall problem. Each of the 12 leaves of the tree represents an outcome. A "W" next to an outcome indicates that the contestant wins, and an "L" indicates that he loses.*

The event $S_l \subseteq S$ that the contestant loses is the complementary set:

$$S_l = \{(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}$$

Our goal is to determine the probability of the event S_w; that is, the probability that the contestant wins with the "switch" strategy.

Well, the contestant wins in 6 outcomes and loses in 6 outcomes.
Does this not imply that the contestant has a $\frac{6}{12} = \frac{1}{2}$ chance of winning? No! Under our natural assumptions, this sample space is not uniform! Some outcomes may be more likely than others.
We must compute the probability of each outcome.

Step 3: Compute Outcome Probabilities

2.0.1 Assumptions

To assign a meaningful probability to each outcome in the Monty Hall problem, we must make some assumptions.
The following three are sufficient:

1. The prize is placed behind each door with probability $\frac{1}{3}$.

2. No matter where the prize is placed, the contestant picks each door with probability $\frac{1}{3}$.

3. No matter where the prize is placed, if Carol has a choice of which door to open, then she opens each possible door with equal probability.

The first two assumptions capture the idea that the contestant initially has no idea where the prize is placed.
The third assumption eliminates the possibility that Carol somehow secretly communicates the location of the prize by which door she opens.
Assumptions of this sort almost always arise in probability problems; making them explicit is a good idea.
We will see later that not all of these assumptions are absolutely necessary.

2.0.2 Computing Probabilities

With these assumptions, we can assign probabilities to outcomes in the Monty Hall problem.
The calculations are shown in Figure 2 and described below.
There are two steps.

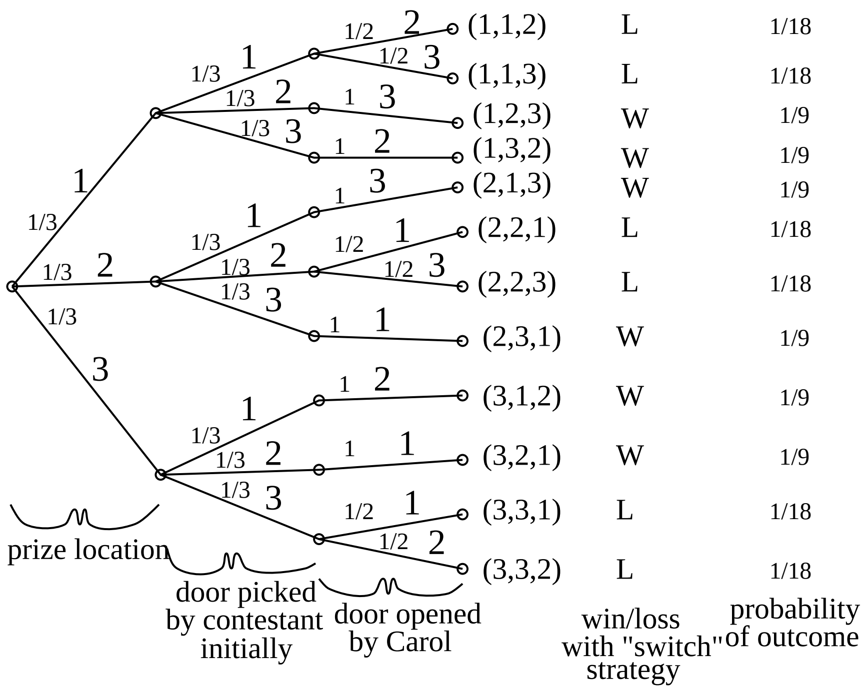

Figure 2: *This is the tree diagram for the Monty Hall problem, annotated with probabilities for each outcome.*

The first step is to record a probability on each edge in the tree diagram.
Recall that each node represents a state of the experiment, and the whole experiment can be regarded as a walk from the root (initial state) to a leaf (outcome).

The probability recorded on an edge is the probability of moving from the state corresponding to the parent node to the state corresponding to the child node.
These edge probabilities follow from our three assumptions about the Monty Hall problem.

Specifically, the first assumption says that there is a $\frac{1}{3}$ chance that the prize is placed behind each of the three doors.
This gives the $\frac{1}{3}$ probabilities on the three edges from the root.
The second assumption says that no matter how the prize is placed, the contestant opens each door with probability $\frac{1}{3}$.
This gives the $\frac{1}{3}$ probabilities on edges leaving the second layer of nodes.
Finally, the third assumption is that if Carol has a choice of what door to open, then she opens each with equal probability.
In cases where Carol has no choice, edges from the third layer of nodes are labeled with probability 1.
In cases where Carol has two choices, edges are labeled with probability $\frac{1}{2}$.

The second step is to use the edge weights to compute a probability for each outcome.
We use the following rule: *the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path.*

The formal justification for this rule will be given in the next lecture, but here is the intuition.
Suppose we want the probability of outcome $(2, 1, 3)$.
In $\frac{1}{3}$ of the experiments, the prize is behind the second door.
Then, in $\frac{1}{3}$ of these experiments, or in $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ of all experiments, the prize is behind the second door, and the contestant opens the first door.
Among all experiments of this type, or among $\frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}$ of all experiments, the prize is behind the second door, the contestant picks the first door, and Carol opens the third door.
Therefore, the probability out of the outcome is the product of the edge probabilities, which is $\frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}$.

For example, the probability of outcome $(2, 2, 3)$ is the product of the edge probabilities on the path from the root to the leaf labeled $(2, 2, 3)$.
Therefore, the probability of the outcome is $\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}$.
Similarly, the probability of outcome $(3, 1, 2)$ is $\frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}$.
The other outcome probabilities are worked out in Figure 2.

Step 4: Compute Event Probabilities

We now have a probability for each outcome.
All that remains is to compute the probability of S_w, the event that the contestant wins with the "switch" strategy.
The following definition explains how to compute the probability of an event given the probability of each outcome.

Definition Let S be the sample space of an experiment, and let Pr be the probability function. Then for any event $S' \subseteq S$:

$$\Pr(S') = \sum_{w \in S'} \Pr(w)$$

In words, the probability of an event is defined to be the sum of the probabilities of all the outcomes in that event.
For example, in the Monty Hall problem, the probability of the contestant winning with the "switch" strategy is the sum of the probabilities of the six outcomes in event S_w.

$$\begin{aligned}
\Pr(S_w) &= \sum_{w \in S_w} \Pr(w) \\
&= \Pr((1,2,3)) + \Pr((1,3,2)) + \Pr((2,1,3)) + \Pr((2,3,1)) + \Pr((3,1,2)) + \Pr((3,2,1)) \\
&= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \\
&= \frac{2}{3}
\end{aligned}$$

The first equation follows from the definition of the probability of an event.
The sum is expanded in the second line; since S_w contains six outcomes, we must sum six probabilities.
The outcome probabilities that we computed are substituted in the third line, and summed in the final line.
The event S_w has probability $\frac{2}{3}$; that is, a contestant employing the "switch" strategy wins two-thirds of the time.

In the same way, we can compute the probability that a contestant loses with the "switch" strategy. This is the probability of event S_l:

$$\begin{aligned}
\Pr(S_l) &= \sum_{w \in S_l} \Pr(w) \\
&= \Pr((1,1,2)) + \Pr((1,1,3)) + \Pr((2,2,1)) + \Pr((2,2,3)) + \Pr((3,3,1)) + \Pr((3,3,2)) \\
&= \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} \\
&= \frac{1}{3}
\end{aligned}$$

The probability of the contestant losing with the switch strategy is $\frac{1}{3}$.
This makes sense; the probability of winning and the probability of losing ought to sum to 1!

We can determine the probability of winning with the "stick" strategy without further calculations. In every case where the "switch" strategy wins, the "stick" strategy loses, and vice versa. Therefore, the probability of winning with the stick strategy is $1 - \frac{2}{3} = \frac{1}{3}$.

Solving the Monty Hall problem formally requires only simple addition and multiplication.
But trying to solve the problem with "common sense" leaves us running in circles!

3 A Dice Problem

There is a game involving three dice and two players.
The dice are not normal; rather, they are numbered as shown in Figure 3.
Each hidden face has the same number as the opposite, exposed face.
As a result, each die has only three distinct numbers, and each number comes up $\frac{1}{3}$ of the time.

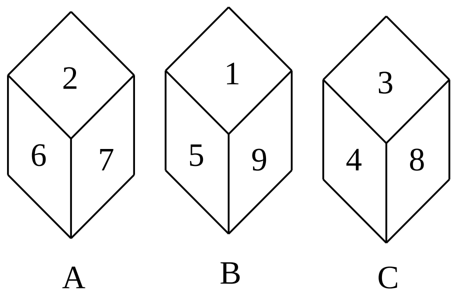

Figure 3: *This figure shows the strange numbering of the three dice in the Dice Problem. The number on each concealed face is the same as the number on the exposed, opposite face.*

In the game, the first player can choose any one of the three dice.
Then the second player chooses one of the two remaining dice.
They both roll and the player with the higher number wins.
Which of the three dice should player one choose? That is, which of the three dice is best?

For example, die B is attractive, because it has a 9, the highest number overall; on the other hand, it also has a 1, the lowest number.
Intuition gives no clear answer! We can solve the problem with the same four-step method used in the Monty Hall problem.

Claim 3.1 Die A beats die B more than half of the time.

Proof. The claim concerns the experiment of throwing dice A and B.

Step 1: Find the Sample Space.
The sample space for this experiment is indicated by the tree diagram in Figure 4.

Step 2: Define Events of Interest.
We are interested in the event that die A comes up greater than die B.
The outcomes in this event are marked "A" in the figure.

Step 3: Compute Outcome Probabilities.
To find outcome probabilities, we first assign probabilities to edges in the tree diagram.
Each number comes up with probability $\frac{1}{3}$, regardless of the value of the other die.
Therefore, we assign all edges probability $\frac{1}{3}$.

The probability of an outcome is the product of probabilities on the corresponding root-to-leaf path; this means that every outcome has probability $\frac{1}{9}$.

Step 4: Compute Event Probabilities.
The probability of an event is the sum of the probabilities of the outcomes in the event. Therefore, the probability that die A comes up greater than die "B" is $\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{5}{9}$.

As claimed, the probability that die A beats die B is greater than half.

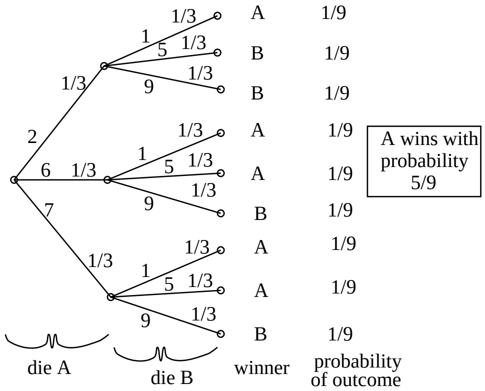

Figure 4: *This is the tree diagram arising when die A is played against die B. Die A beats die B with probability $\frac{5}{9}$.*

∎

Claim 3.2 Die B beats die C more than half of the time.

Proof. See the tree diagram in Figure 5; the proof is the same as for the preceding claim.

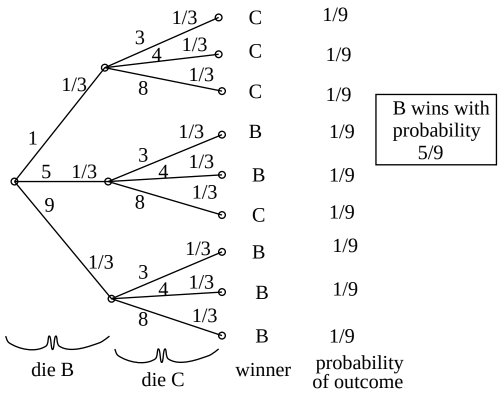

Figure 5: *Die B beats die C with probability $\frac{5}{9}$.*

∎

We have shown that A beats B and that B beats C.
From these results, we might conclude that A is the best die, B is second best, and C is worst.
But this is totally wrong!

Claim 3.3 Die C beats die A more than half of the time!

Proof. See the tree diagram in Figure 6.

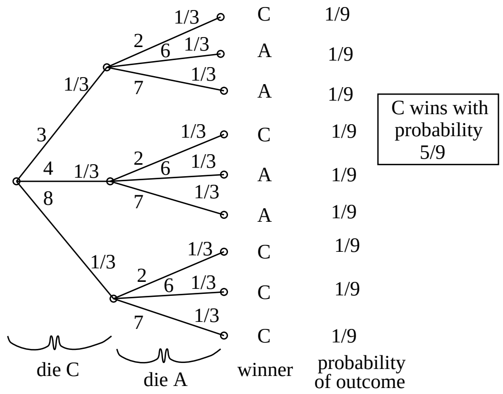

Figure 6: *Die C beats die A with probability* $\frac{5}{9}$. *Amazing!*

■

Die A beats B, B beats C, and C beats A! Apparently, there is no "transitive law" here! This means that no matter what die the first player chooses, the second player can choose a die that beats it with probability $\frac{5}{9}$.
The player who picks first is always at a disadvantage!

The same effect can arise with three dice numbered the ordinary way, but "loaded" so that some numbers turn up more often.
For example, suppose:

$$
\begin{array}{lll}
A & \text{rolls} & 3 \text{ with probability } 1 \\
\\
B & \text{rolls} & 2 \text{ with probability } p = (\sqrt{5}-1)/2 = 0.618\ldots \\
 & \text{rolls} & 5 \text{ with probability } 1-p \\
\\
C & \text{rolls} & 1 \text{ with probability } 1-p \\
 & \text{rolls} & 4 \text{ with probability } p
\end{array}
$$

In this case, A beats B, B beats C, and C beats A, all with probability $p = 0.618\ldots$.
You will prove this fact in homework.
In tutorial, we will see that no matter how the dice are loaded, this probability can be no greater

than $p = 0.666\ldots$.

EXTRA CREDIT: Can you find a way to load the dice to make p greater than 0.618 or prove that this is impossible?