IV Probability

INSERT A goes here

Probability is one of the most important disciplines in ~~computer science, and indeed,~~ ~~in~~ all of the sciences. It is also one of the least well understood. ~~subjects in computer science~~

Probability is especially important in ~~is especially important in~~ ~~computer science & arises~~ in virtually computer science — it arises ~~every branch of computer science~~ ~~for example~~ every branch of the field. In algorithm design and game theory, for example, ~~a~~ randomized ~~randomized~~ algorithms and strategies (those that use a ~~those that use a~~ random number generator ~~to~~ ~~control the~~ as a key input for decision making) frequently ~~typically~~ outperform deterministic algorithms and strategies. In information theory ~~communications,~~ ~~an~~ and signal processing, ~~an an~~ an understanding of randomness ~~probability~~ is critical for filtering out noise and compressing data. In cryptography and digital rights management, probability is crucial for ~~making~~ achieving security. The list of examples is long.

Given the impact ~~that~~ probability has on computer science, it
~~Given its importance, it seems especially~~

A-2

it seems strange ~~Strange~~ that probability should be so misunderstood by so many. Perhaps the trouble is that basic human intuition is wrong as often as it is right when it comes to problems involving random events. As a consequence, many students ~~(and even some researchers and faculty)~~ develop a fear of probability. Indeed, we have witnessed many graduate oral exams where a student will ~~can~~ solve the most horrendous calculation, only to then be tripped up by the simplest probability question. Indeed, ~~even~~ even some faculty will start squirming if you ask them a question that starts off "What is the probability that..."

Our goal in ~~the final part~~ the remaining chapters is to equip you with the tools that will enable ~~you will~~ ~~need~~ you to easily and confidently solve problems involving probability. # We begin in Chapter 14 with the basic definitions and an elementary 4-step process that ~~solves~~ can be used to ~~solve~~ ~~answer~~ ~~a surprisingly large number of questions~~ determine the probability that a specified event ~~such as the~~ occurs. We illustate the method on two famous problems where your intuition will probably fail you.

In Chapter 15, we describe conditional probability and the notion of independence. Both notions are important, and sometimes misused, in practice. We will consider the probability of having a disease given that you tested positive, and the probability that a suspect is guilty given that his blood type ~~is~~ matches the blood found at the scene of the crime.

We study random variables and distributions in chapter 16. Random variables provide a more quantitive way to measure random events. For example, instead of determining the probability that it will rain, we may want to determine <u>how much</u> or <u>how long</u> it is likely to rain. This is closely related to the notion of the expected value of a random variable, which we ~~will~~ consider in chapter 17.

In chapter 18, we examine the probability that a random variable deviates significantly from its expected value. This is especially important in practice, where

A-S

things are generally fine if they are going according to expectation, ~~but then you~~ and you would like to be assured that the probability of deviating from ~~a~~ the expectation is very low.

We conclude in Chapter 19 ~~by with a~~ by combining ~~all~~ the tools we have acquired to solve problems involving ~~more complex~~ more complex random processes. We will see why you ~~should~~ ~~A~~ will probably never get ~~rest~~ very far ahead at the casino, and how two Stanford ~~stu~~ graduate students became ~~a~~ gazillionaires by combining graph theory and probability theory to design a better ~~se web~~ search engine for the web.

CHAPTER 14 Events and Probability Spaces ~~Probability Spaces and Events~~

14.1 Let's Make a Deal

14 Introduction to Probability

Probability plays a key role in the sciences —"hard" and social —including computer science. Many algorithms rely on randomization. Investigating their correctness and performance requires probability theory. Moreover, computer systems designs, such as memory management, branch prediction, packet routing, and load balancing are based on probabilistic assumptions and analyses. Probability is central as well in related subjects such as information theory, cryptography, artificial intelligence, and game theory. But we'll start with a more down-to-earth application: getting a prize in a game show.

14.1 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 describes a situation like one 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. She explained that if the car was behind either of the two unpicked doors —which is twice as likely as the the car being behind the picked door —the contestant wins by switching. But she soon received a torrent of letters, many from mathematicians, telling her that she was wrong. The problem became known as the Monty Hall Problem and it generated thousands of hours of heated debate.

This incident highlights a fact about probability: the subject uncovers lots of examples where ordinary intuition leads to completely wrong conclusions. So until you've studied probabilities enough to have refined your intuition, a way to avoid errors is to fall back on a rigorous, systematic approach such as the Four Step Method, that we will ~~shortly~~ describe shortly. First, let's make sure we really understand the set up for this problem. This is always a good thing to do when ~~it comes~~ you are dealing with probability.

— INSERT B goes here —

(it is text on pp 893-894)

14.2

~~14.1.1~~ The Four Step Method ← full section

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 provides simple solutions to many famously-confusing problems. For example, as you'll see, the four step method cuts through the confusion surrounding the Monty Hall problem like a Ginsu knife. ~~However, m~~More complex probability questions may spin off challenging counting, summing, and approximation problems— which, fortunately, you've already spent weeks learning how to solve.

~~14.1.2 Clarifying the Problem~~

This is insert B and goes to p891

14.1.1 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. For example, we will assume that:

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

this is the end of Insert B

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?"

14.2.1 ~~14.1.3~~ 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; as shown in Figure A1.

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.

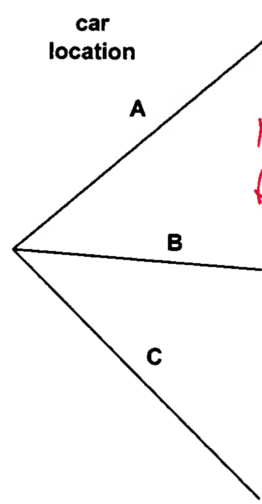


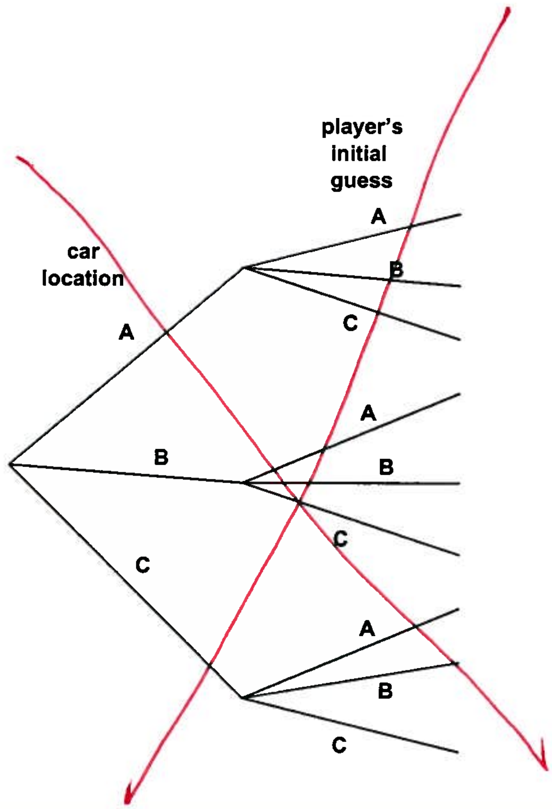
Figure A1: The first level in a tree diagram for the Monty Hall Problem. The branches correspond to the doors ~~the~~ behind which the car is located.

~~Now,~~ For each possible location of the prize, the player could initially choose any of the three doors. We represent this in a second layer added to the tree. Then a third layer represents the possibilities of the final step when the host opens a door to reveal a goat, as shown in Figure A2.

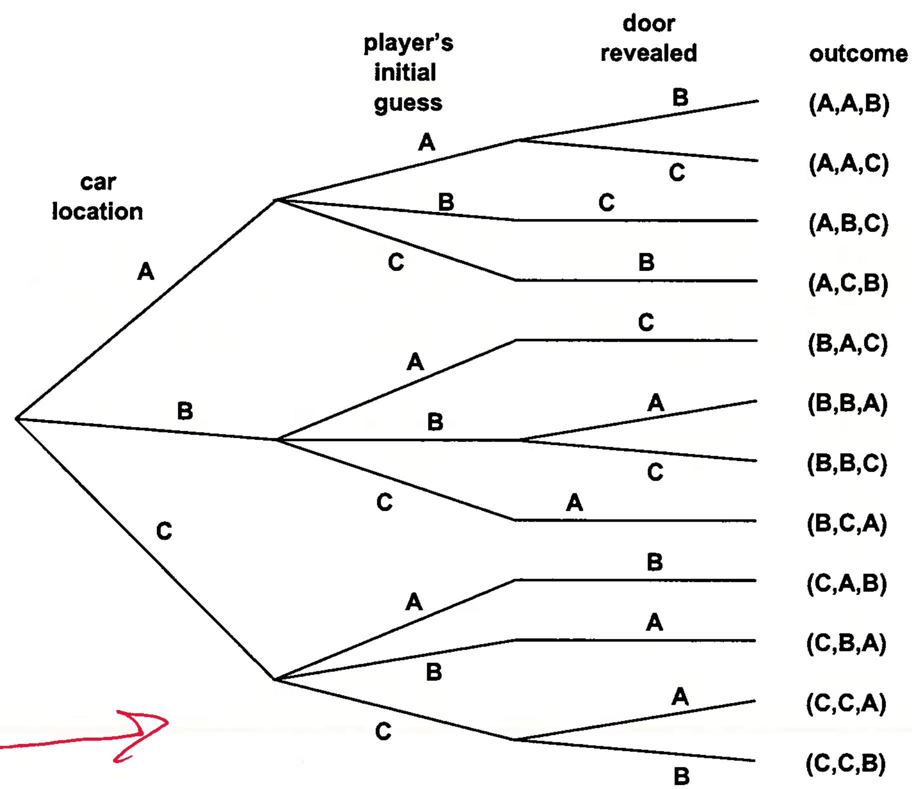
~~EDITING NOTE:~~

David - this figure is the same as the one on p898 but without the "outcome" column of data.

Figure A2: The full tree diagram for the Monty Hall Problem. ~~Level~~ The second level indicates the door initially chosen by the player. The third level indicates the door revealed by Monty Hall.



Notice that the third layer reflects the fact that 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.

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,

Figure 14.3 : ~~The outcomes~~ The tree diagram for the Monty Hall Problem with the ~~leaves~~ outcomes labeled for each path from ~~tree~~ root to leaf. ~~But~~ For example, outcome (A,A,B) corresponds to the car being behind door A, the player initially choosing door A, and Monty Hall revealing the goat behind door B.

in Figure A3

we've labeled each outcome with a triple of doors indicating:

(door concealing prize, door initially chosen, door opened to reveal a goat) .

In these terms, the sample space is the set:

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

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

14.2.2

~~14.1.4~~ Step 2: Define Events of Interest

Our objective is to answer questions of the form "What is the probability that ...?",

where, for example,

~~where~~ the missing phrase might be "the player wins by switching", "the player initially picked the door concealing the prize", or "the prize is behind door C", ~~for example.~~

For example,

Each of these phrases characterizes a set of outcomes~~:~~ the outcomes specified by "the prize is behind door C" is:

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

and it is a subset of the sample space.

A set of outcomes is called an *event*. So the event that the player initially picked the door concealing the prize is the set~~:~~ : ← OK as was

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

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

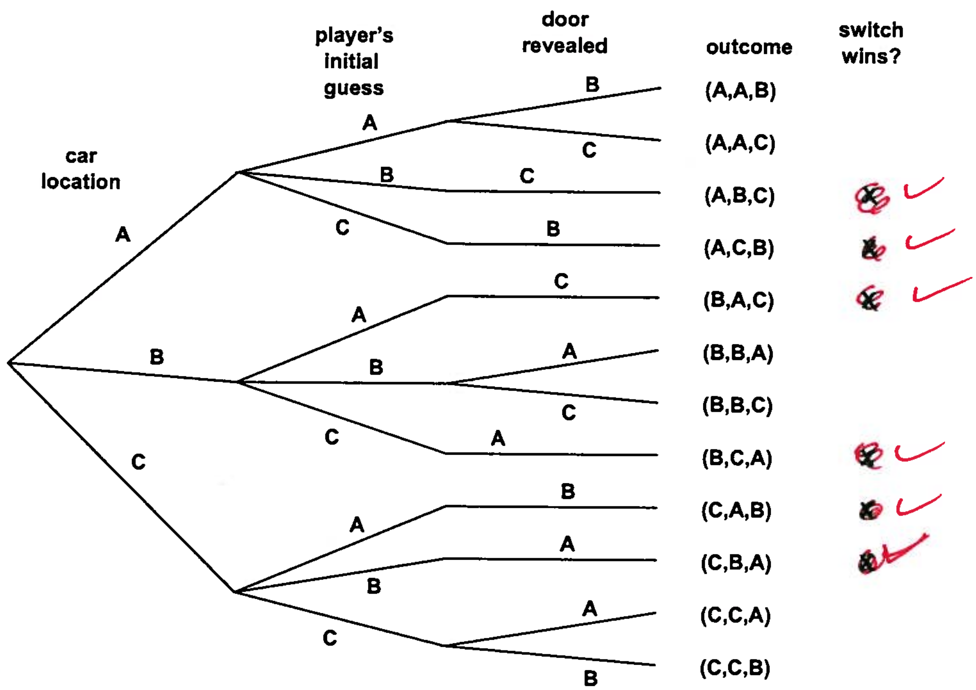
$$\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}$$

~~We have annotated~~

~~Let's annotate our tree diagram to indicate the outcomes in this event, in Figure A4.~~

~~with checkmarks~~

These outcomes are denoted with a check mark in Figure A4.



Notice that exactly half of the outcomes are ~~marked~~ checked, 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 1/2. *This is wrong.* The reason is that these outcomes are not all equally likely, as we'll see shortly.

Figure A4: The tree diagram for the Monty Hall Problem where the ~~the~~ outcomes in the ~~e~~ event where the player wins by switching are denoted with a checkmark.

14.2.3

~~14.1.5~~ 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, indicating the fraction of the time this outcome is expected to occur. The sum of all outcome probabilities must be one, reflecting the fact that there always is an outcome.

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.

Step 3a: Assign Edge Probabilities

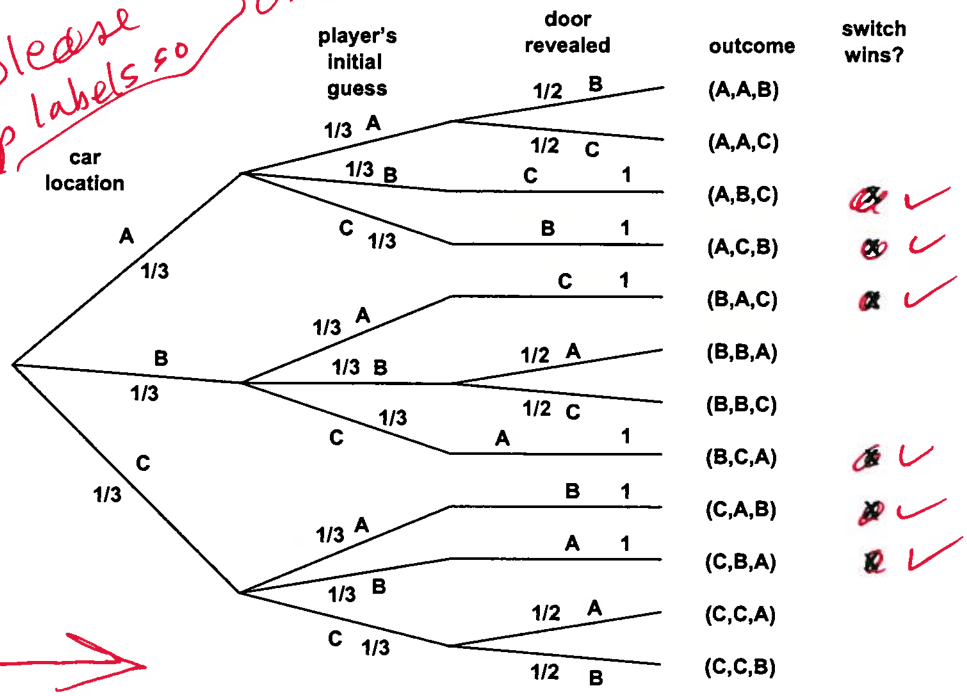
First, we record a probability on each *edge* of the tree diagram. These edge-probabilities are determined by the assumptions we made at the outset: that the prize is equally likely to be behind each door, that the player is equally likely to pick each door, and that the host is equally likely to reveal 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.

For example, see Figure A5.

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 out-

David: please clean up their labels so orientation on the edge's more consistent



come, (A, A, B) is

$$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}.$$

There's an easy, intuitive justification for this rule. As the steps in an experiment progress randomly along a path from the root of the tree to a leaf, the probabilities on the edges indicate how likely the ~~walk~~ path is to proceed along each branch. For example, a

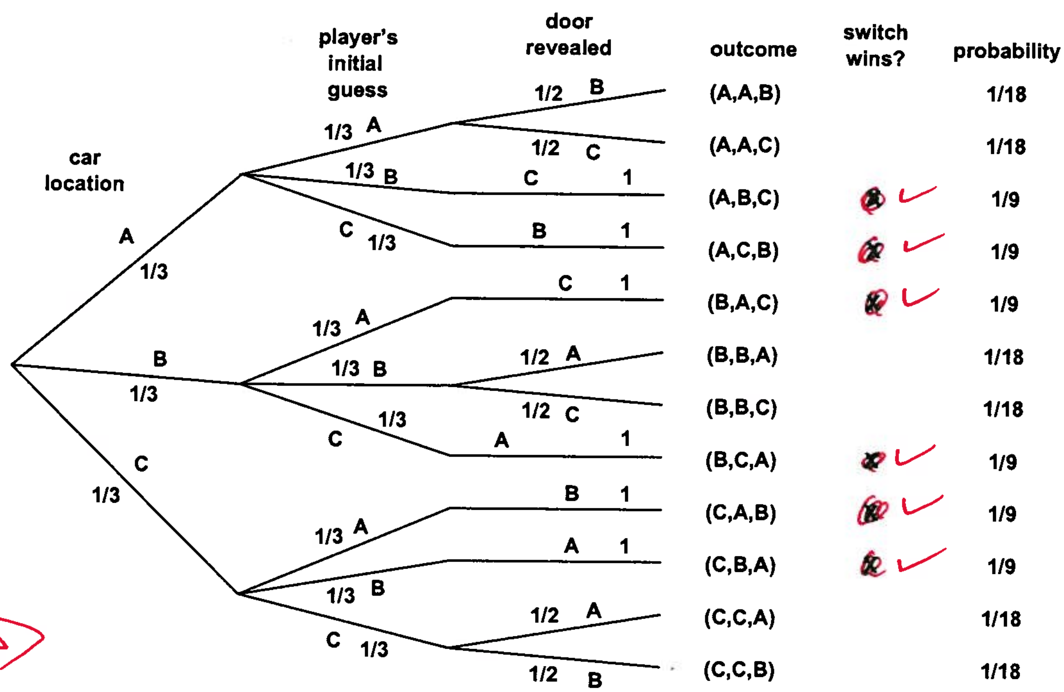
Figure A5: The tree diagram for the Monty Hall Problem where edge weights denote the probability of that branch being taken given that we are at the parent of that branch. For example, if the car is behind door A, there is a 1/3 chance that the player's initial selection is door B.

path starting at the root in our example is equally likely to go down each of the three top-level branches.

~~Now, how~~ How likely is such a ~~walk~~ path to arrive at the topmost outcome, (A, A, B)? Well, there is a 1-in-3 chance that a ~~walk~~ path 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~~ path in 18 should arrive at the (A, A, B) leaf, which is precisely the probability we assign it.

We have illustrated all of the outcome probabilities ~~Anyway, let's record all the outcome probabilities in our tree diagram.~~ In Figure A6.

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:

$$\Pr\{(A, A, B)\} = \frac{1}{18}$$

$$\Pr\{(A, A, C)\} = \frac{1}{18}$$

$$\Pr\{(A, B, C)\} = \frac{1}{9}$$

David: can we make Pr & Ex & Var be macros so we can change the bracket type later?

etc.

Figure A6: The ~~out~~ rightmost column shows ~~tree diagram for the~~ the outcome probabilities for the Monty Hall Problem. Each outcome probability is simply the product of the probabilities on the branches ~~on~~ on the path from the root to the leaf for that outcome.

14.2.4 ~~14.1.6~~ Step 4: Compute Event Probabilities

We now have a probability for each *outcome*, but we want to determine the probability of an *event*. ~~which will be~~ The probability of an event E is denoted by $\Pr[E]$ and it is the sum of the probabilities of the outcomes in ~~it~~ E. ~~The probability of an event, E, is written $\Pr\{E\}$.~~ For example, the probability of the event that the player wins by switching is:[1]

~~$\Pr[\{(A,B,C),(A,C,$~~

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

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.

---

1 ~~More formally,~~ "switching wins" is shorthand for the set of outcomes where switching wins; namely, $\{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\}$. We will frequently use such shorthand ~~inside Pr[ ] to~~ to denote events. ~~especially insi~~

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.

14.2.5

~~14.1.7~~ An Alternative Interpretation of the Monty Hall Problem

indicates that

Was Marilyn really right? Our analysis ~~suggests~~ she was. But 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

~~INSERT E goes here~~
~~(text on p910)~~

only to contestants who picked the correct door initially. In this case, switching never works!

— INSERT C goes here —

~~14.2~~ 14.4 Set Theory and Probability

Let's abstract what we've just done in this Monty Hall example into a general mathematical definition of probability. In the Monty Hall example, there were only finitely many possible outcomes. Other examples in this course will have a countably infinite number of outcomes.

General probability theory deals with uncountable sets like the set of real numbers, but we won't need these, and sticking to countable sets lets us define the probability of events using sums instead of integrals. It also lets us avoid some distracting technical problems in set theory like the Banach-Tarski "paradox" mentioned in Chapter ??.

INSERT C

14.3 Strange Dice

~~Now that you we have the 4-step method,~~

~~let's see if we can use~~

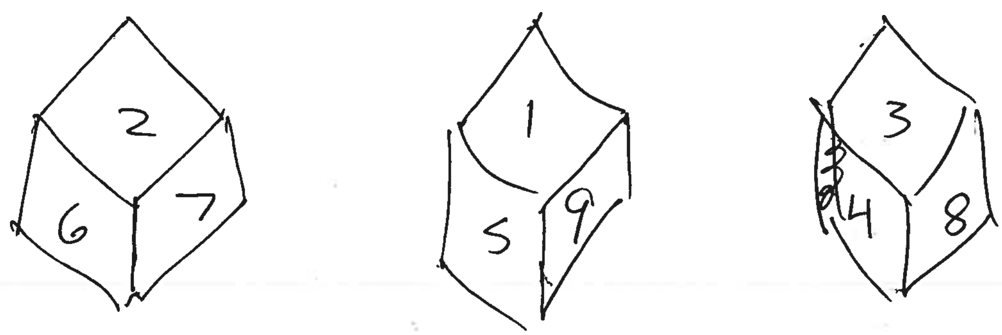
~~let's try it out on something~~

The 4-step method is surprisingly powerful. Let's get some more practice with it. Imagine, if you will, the following scenario.

It's a typical Saturday night. ~~You are~~ You're at your favorite pub ~~comfortably seated in your usual barstool at the pub, sipping some lemonade, when~~ contemplating the true meaning of infinite cardinalities, when a burly-looking biker plops down ~~in~~ on the stool next to you. ~~Just as you You try minding your own business, but then~~ Just as you are about to ~~discover~~ get your mind around $P(P(\mathbb{R}))$, biker dude slaps 3 strange-looking dice on the bar and ~~offers~~ challenges you to a ~~wager~~ \$100 wager.

The rules are simple. Each player selects one die and rolls it once. The player with the ~~higher~~ lower value ~~showing~~ pays

The ~~of~~ other ~~p~~ player $100 ~~in tears~~, ~~and in~~ ~~tears~~!" Naturally, you are skeptical. A quick inspection ~~It sounds simple, b~~ reveals that these are ~~Naturally, you are skeptical. These are~~ not ordinary dice. ~~As shown in Figure A7~~, They each have 6-sides, but the numbers on the dice are different, as shown in Figure A7.



A B C

Figure A7: The strange dice. The number on each concealed face is the same as the number on the opposite face. For example, when you roll die A, ~~you have a 1/3 chance of gettin~~ the probability of getting a 2, 6, ~~and~~ or 7 ~~dice~~ each are 1/3.

~~he assumes~~

Biker dude notices your hesitation and so he offers to let you pick a die first, and then he will choose his die from the two that are leftover. That seals the deal since you figure that ~~th~~ you now have an advantage.

But which of the dice should ~~use~~ you choose? Die B is appealing because it has a 9, which is a sure winner if it comes up. ~~Die~~ Then again, die A has two fairly large numbers, and die B has an 8 and no really-small values.

In the end, you ~~take~~ choose die B because it has the 9 and then biker dude ~~pick~~ selects die A. Let's see ~~who is more likely to win.~~ what the probability is that you will win.[1] Not surprisingly, we will use the 4-step method to compute this probability.

---

1 Of course, you probably should have done this before picking die B in the first place.

C-4

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.

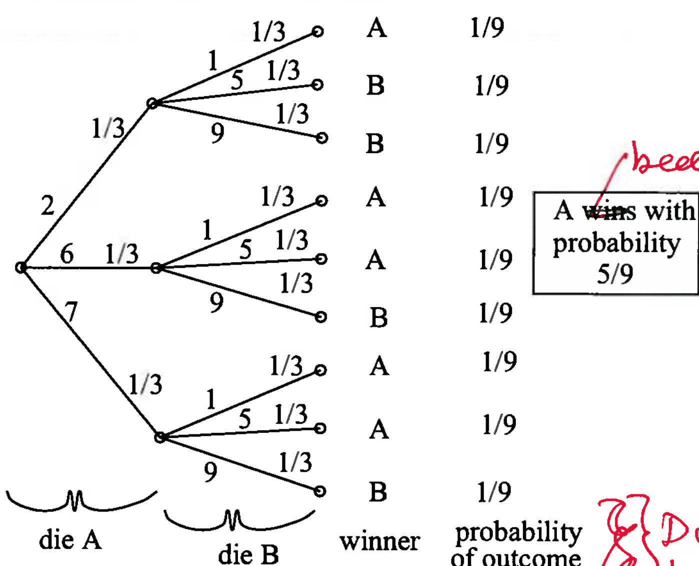
~~2.1.1~~ 14.3.1 Die A versus Die B

~~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~~ shown in Figure A8.[1] (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.)

This is footnote 1

Figure A8: The tree diagram for one roll of die A versus die B. Die A wins ~~[illegible]~~ with probability 5/9.



David: put these labels on top

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

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

C-5

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.~~ in Figure A8.

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. in Figure A8.

~~INSERT Figure here~~

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:

David: Pr[A>B]

David: Pr[(7,5)]

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

~~Bad A8.~~

~~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!~~

~~INSERT D goes here~~

~~2.1.2 Die B versus Die C~~

INSERT E goes here (text from 2.1.3 in lecture notes.)

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

DAVID: this figure goes later at the end of page C-10

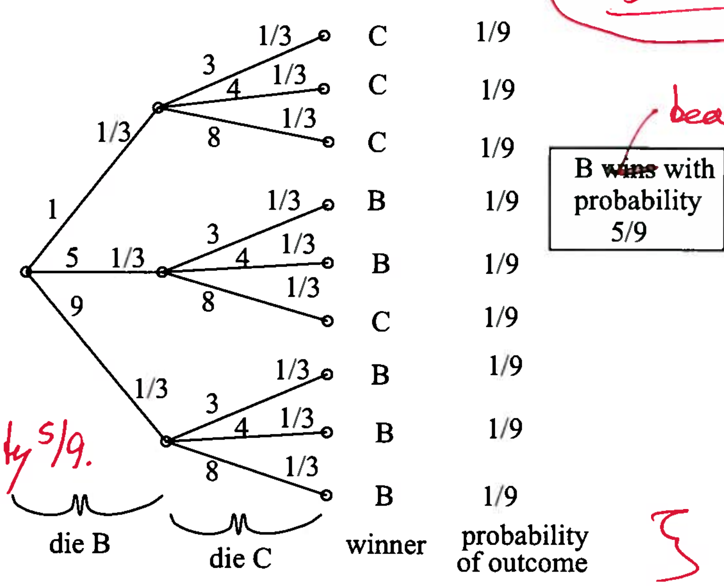


Figure A10: The tree diagram for one roll of die B versus die C. Die B wins with probability 5/9.

more labels to top

~~INSERTED~~ C-6

In this case, all the outcome probabilities are the same. In general, when the probability of every outcome is the same, we say that the ~~probab~~ sample space is *uniform*. ~~Uniform~~ Computing event probabilities for uniform sample spaces is particularly easy since you just have to compute the number of outcomes in the event. In particular, for any event E in a ~~unfi~~ uniform sample space S,

$$\Pr[E] = \frac{|E|}{|S|}. \qquad \text{(EqnF1)}$$

In this case, E is the event that die A beats die B, so $|E| = 5$, $|S| = 9$ and

$$\Pr[E] = 5/9.$$

This is is bad news for you.

~~Bad News!~~ Die A beats die B more than half the time and, not surprisingly, you just lost $100.

Biker dude consoles you on your "bad luck" ~~and offers to go double or nothing.~~ and, given that he's a ~~nice guy~~ sensitive guy beneath all that leather, he offers to go double or nothing.[1] Given that your wallet only has $25 in it, this sounds like ~~the~~ a good plan. Plus, you figure that choosing die A will give _you_ the advantage.

So you choose A and biker dude chooses C. Can you guess who is more likely to win? (HINT: it is generally not a good idea to gamble ~~w~~ with ~~bikers~~ someone you don't know in a bar, especially when you are gambling with strange dice.)

1 _Double or nothing_ is slang for doing another wager after you have lost the first. If you lose again, you will owe biker dude _double_ what you owed him before. If you win, you will now be even and you will owe him _nothing_.

14.2.2 Die A Versus Die C

~~The complete sample space~~ ~~tree diagram, com~~

We can ~~com~~ construct the tree diagram and outcome probabilities as before. ~~Th~~ The result is shown in Figure A9, and there is bad news again. Die C will beat Die A with probability 5/9, and you lose once again.

— Figure A9 goes here —

You now owe biker dude $200 and he asks for his money. ~~You reply that you are wondering need to go to the if you can outrace biker dude back to your dorm, and tell him that~~ you You reply that you need to go to the bathroom. Being a sensitive guy, biker dude nods understandingly and offers yet another wager. ~~He~~ This time, he'll let you have die C. ~~provided that~~ He'll even let you raise the wager to $200 so you can get your money back.

C-9

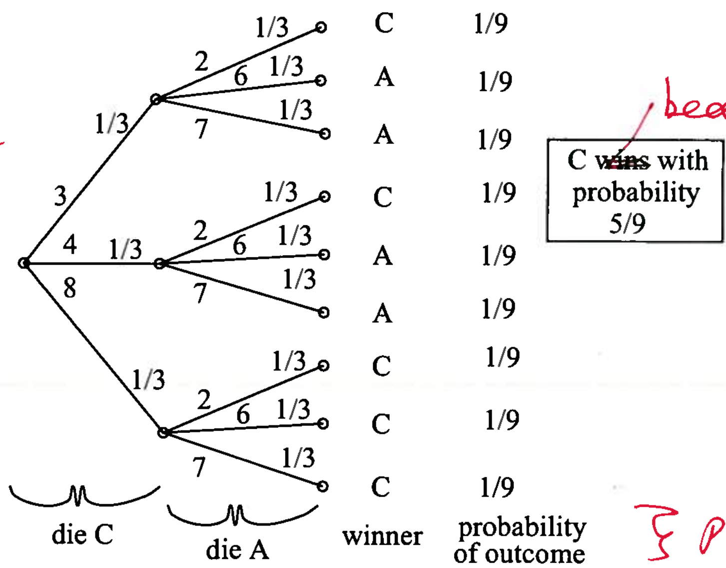
The analysis is the same as before and leads to the conclusion that die B beats die C with probability 5/9 as well. Therefore, you had beter not choose die C; if you do, I'll pick die B and most likely win!

14 ~~3.2.2~~

~~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.

Figure A9: The tree diagram for one roll of die C versus die A. Die C wins with probability 5/9.



put labels at top

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.

~~C-8~~ C-10
~~D-3~~

This is too good a deal to pass up.

Die C ~~has bea~~ is likely to beat die ~~B~~ A and die A is likely to beat die B, so the odds are surely in your favor this time. Biker dude must really be a nice guy. ~~at heart~~

So you pick C and biker dude picks ~~A~~ B. Let's use to tree method to figure out ~~your chan~~ the probability that you win.

14.2.3 Die B versus Die C

The tree diagram and outcome probabilities for B versus C are shown in Figure A10. But surely there is a mistake! The data in Figure A10 shows that die B ~~ca~~ wins with probability 5/9. ~~Not only are you $ now $400 in the hole,~~ ~~But~~ How is it possible that

~~C bea~~

~~C~~ B beats ~~A~~ C with probability 5/9,

C ~~A~~ beats ~~B~~ A with probability 5/9, and

A beats B with probability 5/9?

— Figure A10 goes here —

The problem is not with the math, but with your intuition. It seems that the ~~"Cheats" relat~~ "likely-to-beat" relation should be transitive. But ~~this~~ it is not, and ~~you~~ whatever die you pick, biker dude can pick one of the others ~~Now owe biker dude $400~~ and ~~have a~~ be likely to win. ~~So this game was rigged~~

So picking first is a big disadvantage and you now owe biker dude $400.

Just when you think matters can't get worse, biker dude offers you one final wager for $1,000. This time, ~~he will a~~ you demand to choose second. Biker dude agrees, but with the condition that instead of rolling each die once, you each roll your die twice and your score is the sum of your rolls.

Believing that you finally have a winning wager, you agree.[1] Biker dude

---

1 Did we mention that ~~gambling~~ playing strange gambling games with strangers in a bar is a bad idea?

chooses die B and, of course, you grab die A. That's because die A will beat die B with probability 5/9 on one roll and so surely, ~~it are~~ two rolls of die A are likely to ~~be~~ beat two rolls of die B, right?

Wrong!

14.2.4 Rolling Twice ~~Die A versus Die B, Revisited~~ ~~Rolled Twice~~.

If each player rolls twice, the tree diagram will have four levels and $3^4 = 81$ outcomes. ~~That~~ ~~means~~ This means that it will take ~~writing out~~ ~~down~~ a while to ~~the~~ write down the entire tree diagram. ~~is go~~ ~~you~~ ~~As~~ We can, however ~~we~~ easily write down the first two levels (as we have done in Figure A11(a)) and then notice that the remaining two levels consist of 9 identical copies of the tree in Figure A11(b).

Figure A11 goes here (on p C-14 now)

The probability of each outcome is $(1/3)^4 = 1/81$ and so, once again, we have a uniform sample space. ~~By eq~~ ~~This means that~~ By ~~we see the probability that A beats B is~~

~~The other player \$100.~~

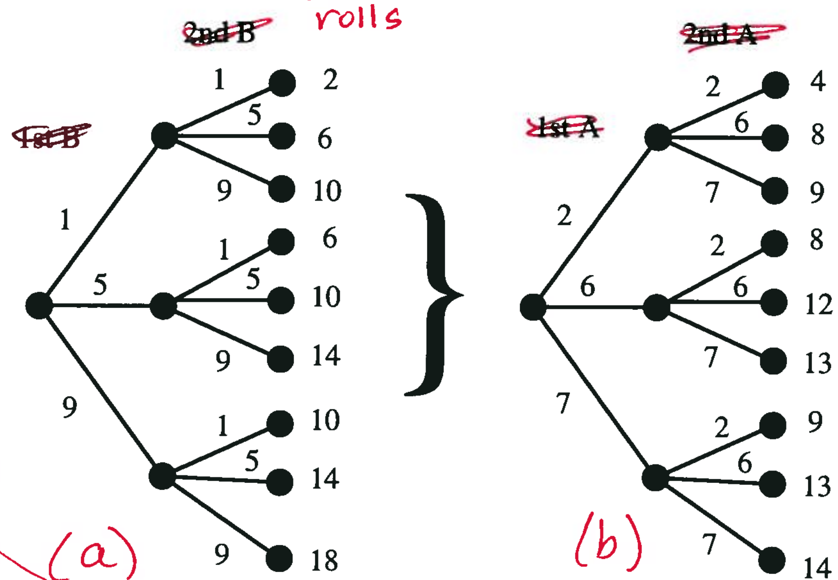
~~the~~

By Equation F1, this means that the probability that A wins is the number of outcomes where A beats B divided by 81. ~~Let's see how to~~

To compute the number of outcomes where A beats B, we observe that the

DAVID: Swap (a) & (b) below in terms of Left & Right

C-2014



In the full tree diagram, each leaf of the tree in (a) is actually the root of a copy of the tree shown in (b).

Parts of the

Figure A11: tree diagram for B versus A where each die is rolled twice. The first two levels are shown in (a). The last two levels consist of 9 copies of the tree in (b).

~~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.~~ ~~each leaf~~

~~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~~ 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 ~~$B$~~ A wins iff ~~$y > x$~~ $x > y$. If ~~$y = 2$~~ $x = 4$, there is ~~no x~~ one y (namely $y = 2$) for which ~~$y > x$~~ $x > y$. If ~~$y = 6$~~ $x = 8$, there ~~is 1 value of x, namely $x = 4$, for which $y > x$~~ are 3 values of y for which $x > y$. Continuing the count in this way, the number of pairs for which ~~$y > x$ is~~ $x > y$ is

$$1 + 3 + 3 + 3 + 3 + 6 + 6 + 6 + 6 = 37.$$

$$\cancel{0 + 1 + 1 + 5 + 5 + 5 + 8 + 8 + 9 = 42,}$$

~~while~~ A similar count shows that there are ~~only 37~~ 42 pairs for which $x > y$, and there are two pairs $((14, 14), (14, 14))$ which result in ties. This means that A loses to B with probability $42/81 > 1/2$, and ties with probability $2/81$. Die A wins with probability only $37/81$.

~~Thus, rolling die B twice is more likely to win than rolling die A twice! How can this be?~~ # How can it be ~~We say~~ showed 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 die strength reverses no matter which two die we picked. So for 1 roll, ~~we had~~

$$A \succ B \succ C \succ A,$$

but for two rolls,

$$A \prec B \prec C \prec A$$

where we have used the symbols ~~ee~~ > and < to denote which die is more likely to ~~e be~~ result in the larger value. This is surprising even to us, but at least we don't owe biker dude \$1400.

14.2.5 Even More Rolls

Now that we know that strange things can happen with strange dice, it is natural, at least for mathematicians, to ask how strange things can get. ~~En~~ It turns out that ~~a~~ things can get very strange. In fact, ~~it was~~ mathematicians[1] recently made the following discovery:

Theorem F2: For any $n \in \mathbb{N}$, there is a set of n dice $D_1, D_2, \ldots, D_n$ such that for any n-node tournament graph[2] G, there is a number of rolls k such that if each die is rolled k times,

---

1 Reference ~~to~~ Ron Graham paper

2 Recall that a tournament graph is a directed graph for which there is precisely one directed edge between any distinct nodes. In other words, for every $u, v \in V(G)$, either u beats v or v beats u but not both.

then ~~die~~ for all $i \neq j$, the ~~probability that the~~ sum of the k rolls for D_i ~~is~~ will exceed the sum ~~o~~ for D_j with probability greater than 1/2 iff $D_i \to D_j \in E(G)$.

~~The~~ ~~Under~~ ~~The~~ It will probably take a few attempts of reading theorem F2 to understand what it is saying. The idea is that for some ~~dice~~ sets of dice, by rolling them different numbers of times, ~~you get~~ the dice have ~~dif~~ varying strengths relative to each other. (this is what we observed for the dice in Figure A7.) ~~In fact, for some very~~ ~~& theorem F2 says~~ Theorem F2 says ~~strange~~ that there is a set of (very) strange dice where every possible collection of relative strengths can be observed by varying the number of rolls. ~~For example, all possible relative strengths (which corresponds to means all possible tournaments) for 3 n=3 dice are shown in Figure A13.~~ For example, the 8 possible relative strengths for $n=3$ dice are shown in Figure A13. ~~One roll of the~~

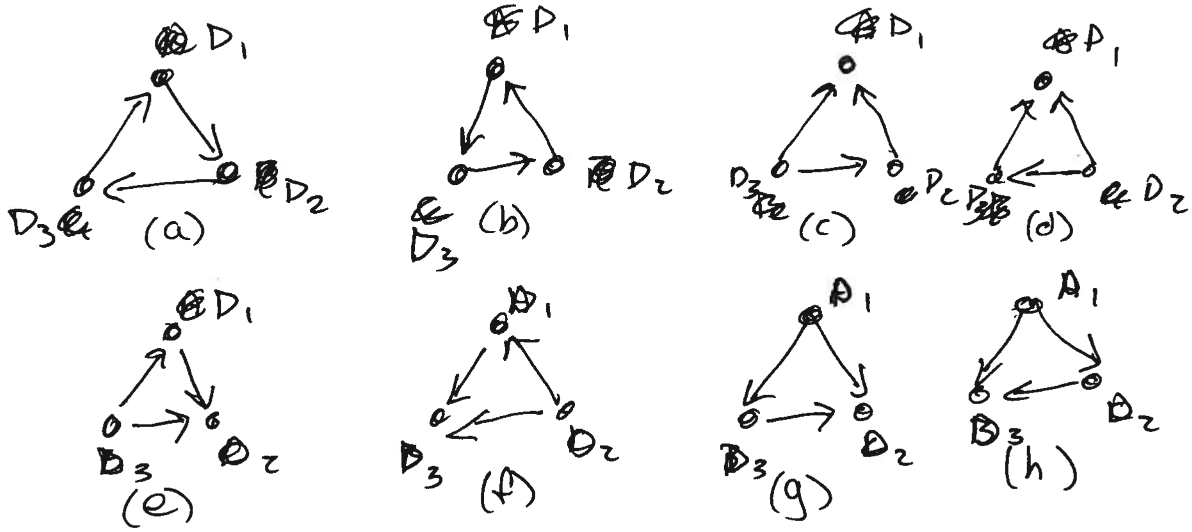


Figure A13: All possible relative strengths for 3 dice D_1, D_2, and D_3. The edge $D_i \rightarrow D_j$ denotes that the sum of rolls for D_i is likely to be ~~higher~~ greater than the sum of rolls for D_j.

~~dice from Figure A7 results~~

Our analysis ~~shows~~ for the dice in Figure A7 showed that for 1 roll, we have the relative strengths shown in Figure A13(a) and for two rolls, we have the (reverse) relative strengths shown in Figure A13(b). Can you figure out what other relative strengths are possible for these dice by using more rolls? This might be worth doing if you are prone to gambling with strangers in bars.

14.4 Set Theory and Probability

The study of probability is very closely tied to ~~se~~ set theory. That is because any set can be a sample space. ~~In other~~ Hence, sets and sample spaces are the same thing. ~~In this context, probabilities are simply weights on the item~~ This means that most of the rules and identities that we have developed for sets extend very naturally to probability. We'll cover several examples in this section, but first lets review some definitions ~~that should already be familiar.~~ that should already be familiar.

~~text for INSERT~~

~~14.2.0~~

~~14.2.1~~ 14.4.1 Probability Spaces

~~we thought it might be helpful to review the~~

Definition 14.2.1. A countable *sample space*,[1] $\mathcal{S}$, is a nonempty countable set. An element $w \in \mathcal{S}$ is called an *outcome*. A subset of $\mathcal{S}$ is called an *event*.

Definition 14.2.2. A *probability function* on a sample space, $\mathcal{S}$, is a total function $\Pr : \mathcal{S} \to \mathbb{R}$ such that

- $\Pr\{w\} \geq 0$ for all $w \in \mathcal{S}$, and
- $\sum_{w \in \mathcal{S}} \Pr\{w\} = 1$.

The sample space together with a probability function is called a *probability space*.

For any event, $E \subseteq \mathcal{S}$, the *probability of* E is defined to be the sum of the probabilities of the outcomes in E:

$$\Pr\{E\} ::= \sum_{w \in E} \Pr\{w\}.$$

1 Yes, sample spaces can be infinite. We'll see some examples shortly. If you did not read Chapter 13, don't worry. ~~as it is~~ Countable means that you can list the elements of the sample space as $w_1, w_2, w_3, \ldots$.

14.4.2 Probability Rules From Set Theory

An immediate consequence of the definition of event probability is that for *disjoint* events~~,~~ E and F,

$$\Pr\{E \cup F\} = \Pr\{E\} + \Pr\{F\}.$$

This generalizes to a countable number of events, as follows: ~~Namely, a collection of sets is *pairwise disjoint* when no element is in more than one of them —formally, $A \cap B = \emptyset$ for all sets $A \neq B$ in the collection.~~

Rule (Sum Rule). *If $\{E_0, E_1, \dots\}$ is collection of pairwise disjoint events*[2]*, then*

$$\Pr\left\{\bigcup_{n\in\mathbb{N}} E_n\right\} = \sum_{n\in\mathbb{N}} \Pr\{E_n\}.$$

The Sum Rule~~[1]~~ lets us analyze a complicated event by breaking it down into simpler

~~[1]If you think like a mathematician, you should be wondering if the infinite sum is really necessary. Namely, suppose we had only used finite sums in Definition 14.2.2 instead of sums over all natural numbers. Would this imply the result for infinite sums? It's hard to find counterexamples, but there are some: it is possible to find a pathological "probability" measure on a sample space satisfying the Sum Rule for finite unions, in~~

2 A collection of events is pairwise disjoint if they do not share any outcomes.

cases. For example, if the probability that a randomly chosen MIT student is native to the United States is 60%, to Canada is 5%, and to Mexico is 5%, then the probability that a random MIT student is native to North America is 70%.

Another consequence of the Sum Rule is that $\Pr\{A\} + \Pr\{\overline{A}\} = 1$, which follows because $\Pr\{\mathcal{S}\} = 1$ and $\mathcal{S}$ is the union of the disjoint sets A and $\overline{A}$. This equation often comes up in the form

Rule (Complement Rule).

$$\Pr\{\overline{A}\} = 1 - \Pr\{A\}.$$

Sometimes the easiest way to compute the probability of an event is to compute the

---

which the outcomes $w_0, w_1, \ldots$ each have probability zero, and the probability assigned to any event is either zero or one! So the infinite Sum Rule fails dramatically, since the whole space is of measure one, but it is a union of the outcomes of measure zero.

The construction of such weird examples is beyond the scope of this text. You can learn more about this by taking a course in Set Theory and Logic that covers the topic of "ultrafilters."

probability of its complement and then apply this formula.

Some further basic facts about probability parallel facts about cardinalities of finite sets. In particular:

$$\Pr\{B - A\} = \Pr\{B\} - \Pr\{A \cap B\}, \qquad \text{(Difference Rule)}$$

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}, \qquad \text{(Inclusion-Exclusion)}$$

$$\Pr\{A \cup B\} \leq \Pr\{A\} + \Pr\{B\}. \qquad \text{(Boole's Inequality)}$$

The Difference Rule follows from the Sum Rule because B is the union of the disjoint sets $B - A$ and $A \cap B$. Inclusion-Exclusion then follows from the Sum and Difference Rules, because $A \cup B$ is the union of the disjoint sets A and $B - A$. Boole's inequality is an immediate consequence of Inclusion-Exclusion since probabilities are nonnegative.

The two event Inclusion-Exclusion equation above generalizes to n events in the same way as the corresponding Inclusion-Exclusion rule for n sets. Boole's inequality also generalizes to

$$\Pr\{E_1 \cup \cdots \cup E_n\} \leq \Pr\{E_1\} + \cdots + \Pr\{E_n\}. \qquad \text{(Union Bound)}$$

This simple Union Bound is ~~actually~~ useful in many calculations. For example, suppose that E_i is the event that the i-th critical component in a spacecraft fails. Then $E_1 \cup \cdots \cup E_n$ is the event that *some* critical component fails. The Union Bound can give an adequate upper bound on this vital probability.

Similarly, the Difference Rule implies that

$$\text{If } A \subseteq B, \text{ then } \Pr\{A\} \leq \Pr\{B\}. \qquad \text{(Monotonicity)}$$

INSERT F goes here

~~14.2.2 An Infinite Sample Space~~

~~Suppose~~ two players take turns flipping a fair coin. Whoever flips heads first is declared the winner. What is the probability that the first player wins? A tree diagram for this

INSERT F

14.4.3 Uniform ~~Sample~~ Probability Spaces

~~A finite samp~~

~~As we saw in Section 14.3, a finite sample probabi~~

~~space is said to be uniform if every the proba~~

<u>Definition</u>: A finite probability space $\mathcal{S}$, $\Pr$ is said to be <u>uniform</u> if $\Pr[w]$ is the same for every outcome $w \in \mathcal{S}$.

As we saw ~~in Secti~~ for the ~~stra~~ strange dice problem, uniform sample spaces are particularly easy to work with. That's because for any event $E \subseteq \mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \qquad \text{(Eqn 62)}$$

This means that once we know the cardinality of E and $\mathcal{S}$, we can immediately obtain $\Pr[E]$. That's great news because we developed lots of tools for computing the cardinality of a set in ~~chapters 9~~ Part III.

For example, suppose that you select 5 cards at random from a standard deck of 52 cards. What is the probability of having ~~[illegible]~~ a full house? ~~Given~~

Normally, this question would ~~be~~ take some effort to answer. But from the ~~a~~ analysis in Section 11.9.2, we know that

$$|S| = \binom{13}{5}$$

and

$$|E| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$

where E is the event that we have a full house. Since every 5-card hand is equally likely, we can apply Equation 6.2 to find that

$$\Pr[E] = \frac{13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}}{\binom{13}{5}}$$

$$= \frac{13 \cdot 12 \cdot 4 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}$$

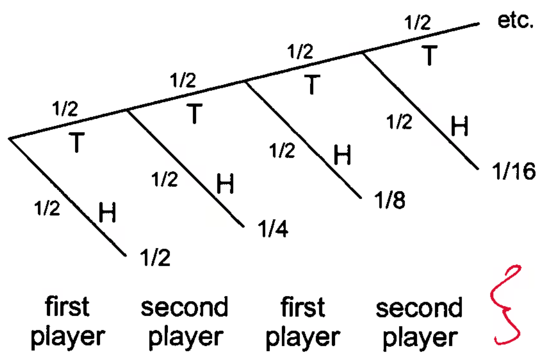
$$= \frac{18}{12495}$$

$$\approx 1/694.$$

14.4.4 Infinite Probability Spaces

General probability theory deals with uncountable sets like $\mathbb{R}$, but ~~for our purp~~ in computer science, it is ~~gusua~~ sufficient to restrict our attention to countable probability spaces. It's also alot easier - infinite sample spaces are hard enough to work with without having to deal with uncountable spaces.

~~In~~ ~~Countable~~ infinite probability spaces ~~arise~~ are ~~all the t~~ fairly common. For example, ~~consider~~ suppose problem is shown ~~below~~ in Figure A15.:



put labels on top

Figure A15: The tree diagram for the game where players take turns flipping a fair coin. The first player to ~~ge~~ flip heads wins.

The event that the first player wins contains an infinite number of outcomes, but we can still sum their probabilities:

$$
\begin{aligned}
\Pr[\text{first player wins}] &= \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128} + \cdots \\
&= \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^n \\
&= \frac{1}{2}\left(\frac{1}{1 - 1/4}\right) = \frac{2}{3}.
\end{aligned}
$$

Similarly, we can compute the probability that the second player wins:

$$\Pr[\text{second player wins}] = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \cdots$$
$$= \frac{1}{3}.$$

In this case, the ~~To be formal about this,~~ sample space is the infinite set

$$\mathcal{S} ::= \{\mathtt{T}^n\mathtt{H} \mid n \in \mathbb{N}\}$$

where $\mathtt{T}^n$ stands for a length n string of T's. The probability function is

$$\Pr[\mathtt{T}^n\mathtt{H}] ::= \frac{1}{2^{n+1}}.$$

~~Since this function is obviously nonnegative,~~ To verify that this is a probability space, we just have to check that all the probabilities are nonnegative and that they sum to 1. ~~But this follows directly from~~ Nonnegativity is obvious and applying the formula for the sum of a geometric series, we find that

$$\sum_{\cancel{\mathtt{T}^n\mathtt{H} \in \mathcal{S}}\ n \in \mathbb{N}} \Pr[\mathtt{T}^n\mathtt{H}] = \sum_{n \in \mathbb{N}} \frac{1}{2^{n+1}} \cancel{= \frac{1}{2}\sum_{n\in\mathbb{N}}\frac{1}{2^n} = 1}$$
$$= 1.$$

Notice that this model does not have an outcome corresponding to the possibility that both players keep flipping tails forever —in the diagram, flipping forever corresponds to following the infinite path in the tree without ever reaching a leaf/outcome. If leaving this possibility out of the model bothers you, you're welcome to fix it by adding another outcome, w_{forever}, to indicate that that's what happened. Of course since the probabililities of the other outcomes already sum to 1, you have to define the probability of w_{forever} to be 0. ~~Now~~ Outcomes with probability zero will have no impact on our calculations, so there's no harm in adding it in if it makes you happier. On the other hand, there's also no harm in simply leaving it out as we did, since it has no impact.

The mathematical machinery we've developed is adequate to model and analyze many interesting probability problems with infinite sample spaces. However, some intricate infinite processes require uncountable sample spaces along with more powerful (and more complex) measure-theoretic notions of probability. For example, if we

~~David – please include the rest of~~
~~the material from the next page~~ as is until
~~you get 14.3 the old "14.3 Conditional Probability"!~~
~~That page is now~~ will become ~~the beginning of Ch 15.~~

generate an infinite sequence of random bits $b_1, b_2, b_3, \ldots$, then what is the probability that

$$\frac{b_1}{2^1} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \cdots$$

is a rational number? Fortunately, we won't have any need to worry about such things.

END OF CH14

14.3 Conditional Probability

Suppose that we pick a random person in the world. Everyone has an equal chance of being selected. Let A be the event that the person is an MIT student, and let B be the event that the person lives in Cambridge. What are the probabilities of these events? Intuitively, we're picking a random point in the big ellipse shown below and asking how likely that point is to fall into region A or B:

This will be part of CH15