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

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

leftover from CH14 (ignore

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

CHAPTER 15

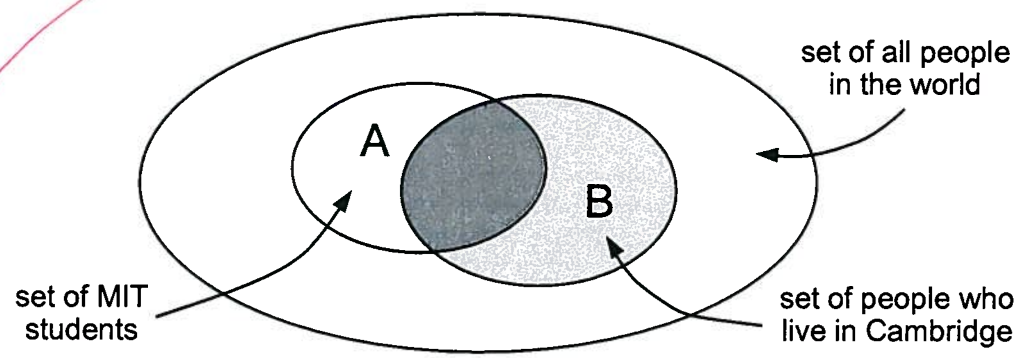
~~14.3~~ Conditional Probability

15.1 Definitions

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~~ in Figure B1 and asking how likely that point is to fall into region A or B.

~~Subsets of the sample space of~~

Figure B1: ~~A is the event that a~~ Selecting a random person. A is the event that the person is an MIT student. B is the event that the person lives in Cambridge.



The vast majority of people in the world neither live in Cambridge nor are MIT students, so events A and B both have low probability. But what is the probability that a person is an MIT student, *given* that the person lives in Cambridge? This should be much greater— but what is it exactly?

What we're asking for is called a *conditional probability*; that is, the probability that one event happens, given that some other event definitely happens. Questions about conditional probabilities come up all the time:

- What is the probability that it will rain this afternoon, given that it is cloudy this morning?
- What is the probability that two rolled dice sum to 10, given that both are odd?
- What is the probability that I'll get four-of-a-kind in Texas No Limit Hold 'Em Poker, given that I'm initially dealt two queens?

David: pls check we have a macro for all occurrences of Pr

There is a special notation for conditional probabilities. In general, $\Pr\{A \mid B\}$ denotes the probability of event A, given that event B happens. So, in our example, $\Pr\{A \mid B\}$ is the probability that a random person is an MIT student, given that he or she is a Cambridge resident.

How do we compute $\Pr\{A \mid B\}$? Since we are *given* that the person lives in Cambridge, we can forget about everyone in the world who does not. Thus, all outcomes outside event B are irrelevant. So, intuitively, $\Pr\{A \mid B\}$ should be the fraction of Cambridge residents that are also MIT students; that is, the answer should be the probability that the person is in set $A \cap B$ (the darkly shaded region in Figure B 1) divided by the probability that the person is in set B (the lightly shaded region). This motivates the definition of conditional probability:

Definition 14.3.1.

$$\Pr\{A \mid B\} ::= \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

If $\Pr\{B\} = 0$, then the conditional probability $\Pr\{A \mid B\}$ is undefined.

Pure probability is often counterintuitive, but conditional probability is worse! Conditioning can subtly alter probabilities and produce unexpected results in randomized algorithms and computer systems as well as in betting games. Yet, the mathematical definition of conditional probability given above is very simple and should give you no trouble— provided you rely on formal reasoning and not intuition. The ~~tree~~ for step method will also be very helpful as we will see in the next ~~ee~~ examples.

15.2 Using the ~~Tree~~ Four-step method to Determine ~~a~~ Conditional Probability

15.2

15.2.1 ~~14.3.1~~ The "Halting Problem"

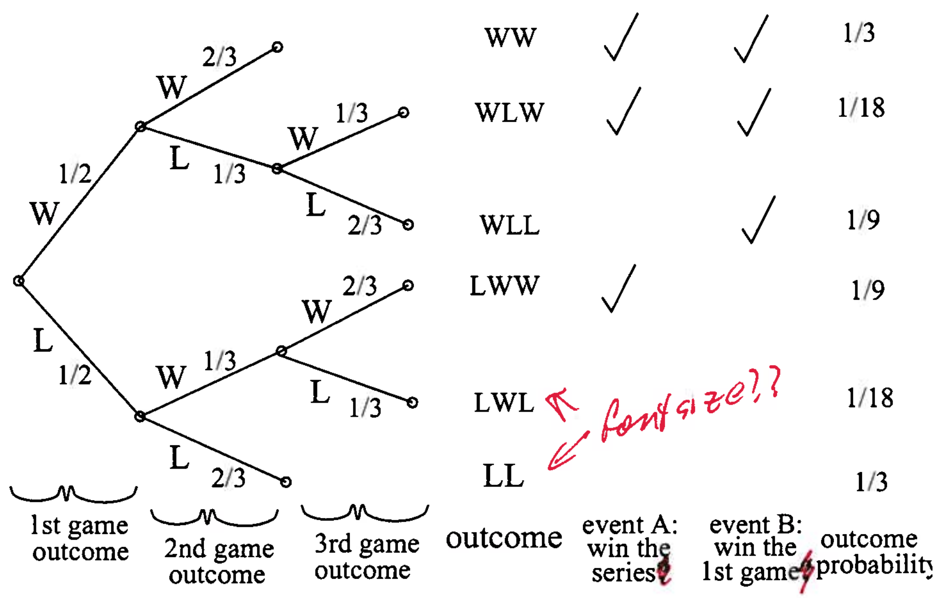
The *Halting Problem* was the first example of a property that could not be tested by any program. It was introduced by Alan Turing in his seminal 1936 paper. The problem is to determine whether a Turing machine halts on a given . . . yadda yadda yadda . . . what's much *more important*, it was the name of the MIT EECS department's famed C-league hockey team.

In a best-of-three tournament, the Halting Problem wins the first game with probability 1/2. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the Halting Problem won the previous game, then they are invigorated by victory and win the current game with probability 2/3. If they lost the previous game, then they are demoralized by defeat and win the current game with probablity only 1/3. What is the probability that the Halting Problem wins the tournament, given that they win the first game?

This is a question about a conditional probability. Let A be the event that the Halting Problem wins the tournament, and let B be the event that they win the first game. Our goal is then to determine the conditional probability $\Pr\{A \mid B\}$.

We can tackle conditional probability questions just like ordinary probability problems: using a tree diagram and the four step method. A complete tree diagram is shown in Figure 32. ~~below, followed by an explanation of its construction and use~~:

Figure B2: The tree diagram for computing the probability that the "Halting Problem" wins ~~the~~ two out of three games given that they won the first game.



font size??

← move to top

15.2.1 Step 1: Find the Sample Space

Each internal vertex in the tree diagram has two children, one corresponding to a win for the Halting Problem (labeled W) and one corresponding to a loss (labeled L). The complete sample space is:

$$S = \{WW,\ WLW,\ WLL,\ LWW,\ LWL,\ LL\}$$

15.2.2

Step 2: Define Events of Interest

The event that the Halting Problem wins the whole tournament is:

$$T = \{WW,\ WLW,\ LWW\}$$

And the event that the Halting Problem wins the first game is:

$$F = \{WW, WLW, WLL\}$$

The outcomes in these events are indicated with checkmarks in the tree diagram in Figure B2.

15.2.3

Step 3: Determine Outcome Probabilities

Next, we must assign a probability to each outcome. We begin by labeling edges as specified in the problem statement. Specifically, The Halting Problem has a $1/2$ chance of winning the first game, so the two edges leaving the root are each assigned probability $1/2$. Other edges are labeled $1/3$ or $2/3$ based on the outcome of the preceding game.

We then find the probability of each outcome by multiplying all probabilities along the corresponding root-to-leaf path. For example, the probability of outcome WLL is:

$$\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{9}$$

Step 4: Compute Event Probabilities

We can now compute the probability that The Halting Problem wins the tournament, given that they win the first game:

$$\begin{aligned}
\Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\
&= \frac{\Pr\{\{WW, WLW\}\}}{\Pr\{\{WW, WLW, WLL\}\}} \\
&= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\
&= \frac{7}{9}
\end{aligned}$$

We're done! If the Halting Problem wins the first game, then they win the whole tournament with probability 7/9.

~~15.3~~ 15.2.2

~~14.3~~.2 Why Tree Diagrams Work

We've now settled into a routine of solving probability problems using tree diagrams. But we've left a big question unaddressed: what is the mathematical justification behind those funny little pictures? Why do they work?

The answer involves conditional probabilities. In fact, the probabilities that we've been recording on the edges of tree diagrams *are* conditional probabilities. For example, consider the uppermost path in the tree diagram for the Halting Problem, which corresponds to the outcome WW. The first edge is labeled $1/2$, which is the probability that the Halting Problem wins the first game. The second edge is labeled $2/3$, which is the probability that the Halting Problem wins the second game, *given* that they won the first— that's a conditional probability! More generally, on each edge of a tree diagram, we record the probability that the experiment proceeds along that path, given that it reaches the parent vertex.

So we've been using conditional probabilities all along. But why can we multiply edge probabilities to get outcome probabilities? For example, we concluded that:

$$\begin{aligned}\Pr\{WW\} &= \frac{1}{2} \cdot \frac{2}{3} \\ &= \frac{1}{3}\end{aligned}$$

Why is this correct?

The answer goes back to Definition 14.3.1 of conditional probability which could be written in a form called the *Product Rule* for probabilities:

Rule (Product Rule for 2 Events). *If* $\Pr\{E_1\} \neq 0$, *then:*

$$\Pr\{E_1 \cap E_2\} = \Pr\{E_1\} \cdot \Pr\{E_2 \mid E_1\}$$

Multiplying edge probabilities in a tree diagram amounts to evaluating the right side of this equation. For example:

$$\begin{aligned}
&\Pr\{\text{win first game} \cap \text{win second game}\} \\
&\qquad = \Pr\{\text{win first game}\} \cdot \Pr\{\text{win second game} \mid \text{win first game}\} \\
&\qquad = \frac{1}{2} \cdot \frac{2}{3}
\end{aligned}$$

So the Product Rule is the formal justification for multiplying edge probabilities to get outcome probabilities! Of course to justify multiplying edge probabilities along longer paths, we need a Product Rule for n events. ~~The pattern of the n event rule should be apparent from~~

Rule (Product Rule for ~~3~~ n Events).

$$\Pr\{E_1 \cap E_2 \cancel{\cap E_3} \cap \cdots \cap E_n\} = \Pr\{E_1\} \cdot \Pr\{E_2 \mid E_1\} \cdot \Pr\{E_3 \mid E_1 \cap E_2\} \cdots \Pr\{E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}\}$$

providing $\Pr\{E_1 \cap E_2 \cap \cdots \cap E_{n-1}\} \neq 0.$

This rule follows from the definition of conditional probability and ~~the trivial identity.~~ induction on n.

$$\cancel{\Pr\{E_1 \cap E_2 \cap E_3\} = \Pr\{E_1\} \cdot \frac{\Pr\{E_2 \cap E_1\}}{\Pr\{E_1\}} \cdot \frac{\Pr\{E_3 \cap E_2 \cap E_1\}}{\Pr\{E_2 \cap E_1\}}}$$

This is INSERT C and goes to p950

15.4 Conditional Identities

~~14.3.3~~ 15.4.1 The Law of Total Probability

Breaking a probability calculation into cases simplifies many problems. The idea is to calculate the probability of an event A by splitting into two cases based on whether or not another event E occurs. That is, calculate the probability of $A \cap E$ and $A \cap \overline{E}$. By the Sum Rule, the sum of these probabilities equals $\Pr\{A\}$. Expressing the intersection probabilities as conditional probabilities yields

Rule (Total Probability).

$$\Pr\{A\} = \Pr\{A \mid E\} \cdot \Pr\{E\} + \Pr\{A \mid \overline{E}\} \cdot \Pr\{\overline{E}\}.$$

For example, suppose we conduct the following experiment. First, we flip a coin. If heads comes up, then we roll one die and take the result. If tails comes up, then we roll two dice and take the sum of the two results. What is the probability that this process yields a 2? Let E be the event that the coin comes up heads, and let A be the event that we get a 2 overall. Assuming that the coin is fair, $\Pr\{E\} = \Pr\{\overline{E}\} = 1/2$. There are now two cases. If we flip heads, then we roll a 2 on a single die with probabilty $\Pr\{A \mid E\} = 1/6$. On the other hand, if we flip tails, then we get a sum of 2 on two dice with probability $\Pr\{A \mid \overline{E}\} = 1/36$. Therefore, the probability that the whole process yields a 2 is

$$\Pr\{A\} = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{36} = \frac{7}{72}.$$

There is also a form of the rule to handle more than two cases.

Rule (Multicase Total Probability). *If $E_1, \ldots, E_n$ are pairwise disjoint events whose union* end of insert C

is the whole sample space, then:

$$\Pr\{A\} = \sum_{i=1}^{n} \Pr\{A \mid E_i\} \cdot \Pr\{E_i\}.$$

~~EDITING NOTE:~~

This is insert D → go to p 950

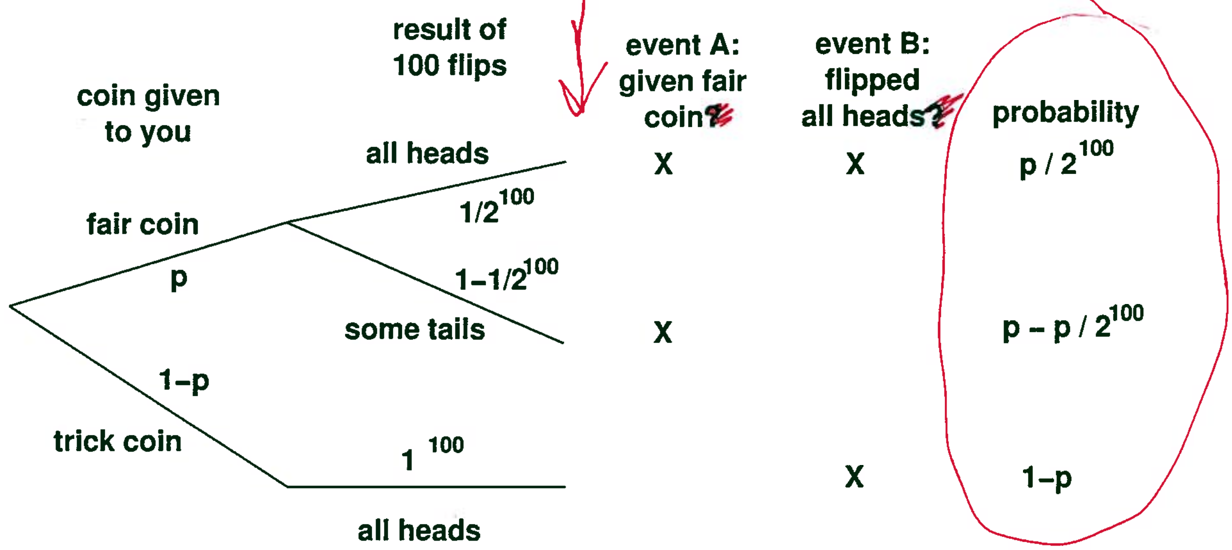
15.3.2 A Coin Problem

~~Now for a problem that even bothers us.~~

suppose that ~~S~~someone hands you either a fair coin or a trick coin with heads on both sides. You flip the coin 100 times and see heads every time. What can you say about the probability that you flipped the fair coin? Remarkably— nothing!

In order to make sense out of this outrageous claim, let's formalize the problem. The sample space is worked out in the tree diagram ~~below~~ shown in Figure C2. We do not know the probability that you were handed the fair coin initially— you were just given one coin or the other— so let's call that p.

Figure C2: The tree diagram for the coin flipping problem.



Let A be the event that you were handed the fair coin, and let B be the event that you flipped 100 straight heads. Now, we're looking for $\Pr\{A \mid B\}$, the probability that you were handed the fair coin, given that you flipped 100 heads. The outcome probabilities are worked out in ~~the tree diagram~~ Figure C2. Plugging the results into the definition of conditional p is not close to 1 and hence that you are very likely to have flipped the trick coin.

probability gives:

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

$$= \frac{p/2^{100}}{1 - p + p/2^{100}}$$

$$= \frac{p}{2^{100}(1-p) + p}$$

This expression is very small for moderate values of p because of the 2^{100} term in the denominator. For example, if $p = 1/2$, then the probability that you were given the fair coin is essentially zero.

But we *do not know* the probability p that you were given the fair coin. And perhaps the value of p is *not* moderate; in fact, maybe $p = 1 - 2^{-100}$. Then there is nearly an even chance that you have the fair coin, given that you flipped 100 heads. In fact, maybe you were handed the fair coin with probability $p = 1$. Then the probability that you were given the fair coin is, well, 1!

Of course, it is extremely unlikely that you would ~~see~~ flip 100 straight heads, but in this case, that is a given ~~we are asking ab~~ from the assumption of the conditional probability. And so if you really did see 100 straight heads, it would be very tempting to ~~ass~~ also assume that ~~p~~

15.3.3 Polling

A similar problem arises in polling before an election. A pollster picks a random American and asks his or her party affiliation. If this process is repeated many times, what can be said about the population as a whole? To clarify the analogy, suppose that the country contains only two people. There is either one Republican and one Democrat (like the fair coin), or there are two Republicans (like the trick coin). The pollster picks a random citizen 100 times, which is analogous to flipping the coin 100 times. Suppose that he picks a Republican every single time. However, even given this polling data, the probability that there is one citizen in each party could still be anywhere between 0 and 1!

What the pollster *can* say is that either:

1. Something earth-shatteringly unlikely happened during the poll.

2. There are two Republicans.

end of insert D

This is as far as probability theory can take us; from here, you must draw your own conclusions. Based on life experience, many people would consider the second possibility more plausible. However, if you are just *convinced* that the country isn't entirely Republican (say, because you're a citizen and a Democrat), then you might believe that the first possibility is actually more likely.

we will talk alot more about polling in chapter 16.

15.2.3 ~~14.3.4~~ Medical Testing

There is an unpleasant condition called *BO* suffered by 10% of the population. There are no prior symptoms; victims just suddenly start to stink. Fortunately, there is a test for latent *BO* before things start to smell. The test is not perfect, however:

- If you have the condition, there is a 10% chance that the test will say you do not. (These are called "false negatives".)

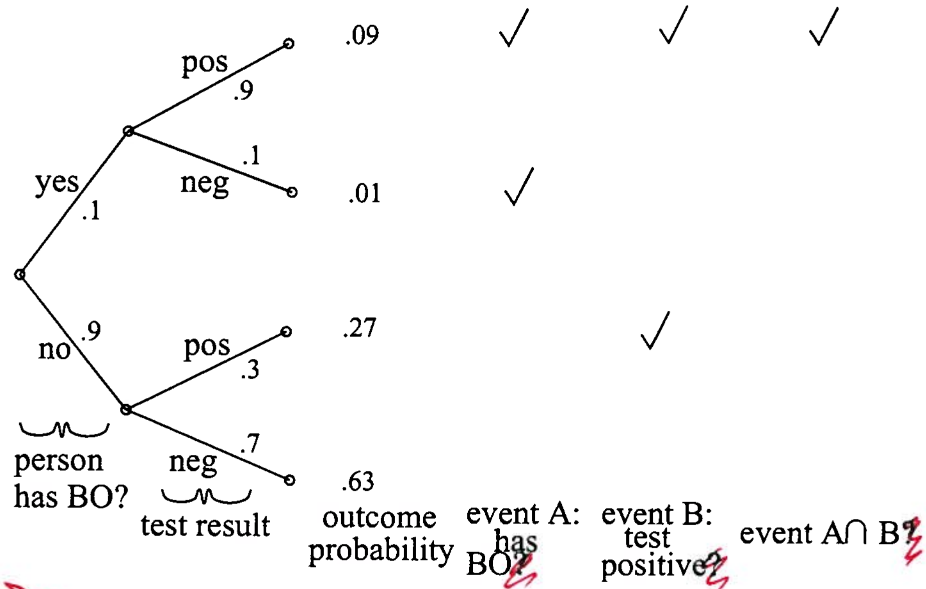
- If you do not have the condition, there is a 30% chance that the test will say you do. ~~(~~These are "false positives". ~~)~~

Suppose a random person is tested for latent *BO*. If the test is positive, then what is the probability that the person has the condition?

Step 1: Find the Sample Space

The sample space is found with the tree diagram ~~below.~~ shown in Figure C1.

Figure C1: The tree diagram for ~~the testing latent~~ the BO problem,



← fix labels & put on top

Step 2: Define Events of Interest

Let A be the event that the person has BO. Let B be the event that the test was positive.

The outcomes in each event are marked in the tree diagram. We want to find $\Pr\{A \mid B\}$, the probability that a person has BO, given that the test was positive.

Step 3: Find Outcome Probabilities

First, we assign probabilities to edges. These probabilities are drawn directly from the problem statement. By the Product Rule, the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path. All probabilities are shown in ~~the~~ Figure. C.1.

Step 4: Compute Event Probabilities

~~P~~ From Definition 14.3.1, we have:

$$\begin{aligned} \Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \\ &= \frac{0.09}{0.09 + 0.27} \\ &= \frac{1}{4} \end{aligned}$$

So, if ~~If~~ you test positive, then there is only a 25% chance that you have the condition!

This answer is initially surprising, but makes sense on reflection. There are two ways you could test positive. First, it could be that you are sick and the test is correct. Second, it could be that you are healthy and the test is incorrect. The problem is that almost everyone is healthy; therefore, most of the positive results arise from incorrect tests of healthy people!

We can also compute the probability that the test is correct for a random person. This event consists of two outcomes. The person could be sick and the test positive (probability 0.09), or the person could be healthy and the test negative (probability 0.63). Therefore, the test is correct with probability $0.09 + 0.63 = 0.72$. This is a relief; the test is correct almost three-quarters of the time.

But wait! There is a simple way to make the test correct 90% of the time: always return a negative result! This "test" gives the right answer for all healthy people and the wrong answer only for the 10% that actually have the condition. ~~The best strategy~~ So a better strategy by this measure is to completely ignore the test result!

There is a similar paradox in weather forecasting. During winter, almost all days in Boston are wet and overcast. Predicting miserable weather every day may be more accurate than really trying to get it right!

It goes to p950

This is Insert E

15.4.2 Conditioning on a Single Event ~~14.3.5 Conditional Identities~~

The probability rules ~~above~~ that we derived in chapter 14 extend to probabilities conditioned on the same event. For example, the Inclusion-Exclusion formula for two sets holds when all probabilities are conditioned on an event C:

$$\Pr\{A \cup B \mid C\} = \Pr\{A \mid C\} + \Pr\{B \mid C\} - \Pr\{A \cap B \mid C\}.$$

This follows from the fact that if $\Pr\{C\} \neq 0$ ~~and we define~~ then

$$\cancel{\Pr_C\{A\} ::= \Pr\{A \mid C\}}$$

$$\begin{aligned}
\Pr[A \cup B \mid C] &= \frac{\Pr[(A \cup B) \cap C]}{\Pr[C]} \\
&= \frac{\Pr[(A \cap C) \cup (B \cap C)]}{\Pr[C]} \\
&= \frac{\Pr[A \cap C] + \Pr[B \cap C] - \Pr[A \cap B \cap C]}{\Pr[C]} \\
&= \Pr[A \mid C] + \Pr[B \mid C] - \Pr[A \cap B \mid C].
\end{aligned}$$

~~end of insert E~~

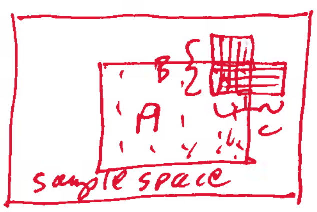
~~then $\Pr_C\{\}$ satisfies the definition of being probabilty function.~~ Insert

It is important not to mix up events before and after the conditioning bar. For example, the following is *not* a valid identity:

False Claim.

$$\Pr\{A \mid B \cup C\} = \Pr\{A \mid B\} + \Pr\{A \mid C\} - \Pr\{A \mid B \cap C\}. \qquad (14.1)$$

A counterexample is shown ~~below~~ in Figure D2. In this case, $\Pr\{A \mid B\} = \cancel{1}\ 1/2$, $\Pr\{A \mid C\} = \cancel{1}\ 1/2$, ~~and~~ $\Pr(A \mid B \cap C) = 1$ and $\Pr\{A \mid B \cup C\} = \cancel{1}\ 1/3$. However, since $\cancel{1 \neq 1 + 1}$ $1/3 \neq \frac{1}{2} + \frac{1}{2} - 1$, ~~the equation above~~ Equation 14.1 does not hold.



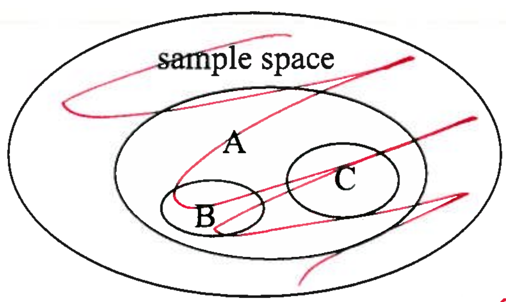


Figure D2: A counterexample to Equation 14.1. Event A is the gray ~~region~~ rectangle, event B ~~has~~ is the rectangle with the vertical stripes, and event C ~~has~~ is the rectangle with the horizontal stripes. ~~The intersection of B and~~ $B \cap C$ lies entirely within A while $B - C$ and $C - B$ are entirely outside of A.

So you're convinced that this equation is false in general, right? Let's see if you *really* believe that.

~~this is~~

15.4.3 ~~14.3.6~~ Discrimination Lawsuit

Several years ago there was a sex discrimination lawsuit against ~~Berkeley.~~ a famous university. A female math professor was denied tenure, allegedly because she was a woman. She argued that in every one of ~~Berkeley~~ the university's 22 departments, the percentage of male applicants accepted was greater than the percentage of female applicants accepted. This sounds very suspicious!

However, ~~Berkeley's~~ the university's lawyers argued that across ~~the whole~~ university as a whole the percentage of male ~~tenure~~ applicants accepted was actually *lower* than the percentage of female applicants accepted. This suggests that if there was any sex discrimination, then it was against men! Surely, at least one party in the dispute must be lying.

Let's simplify the problem and express both arguments in terms of conditional probabilities. To simplify matters, ~~S~~suppose that there are only two departments, EE and CS, and consider the experiment where we pick a random applicant. Define the following events:

~~I we have with~~

- Let A be the event that the applicant is accepted.
- Let F_{EE} the event that the applicant is a female applying to EE.
- Let F_{CS} the event that the applicant is a female applying to CS.
- Let M_{EE} the event that the applicant is a male applying to EE.
- Let M_{CS} the event that the applicant is a male applying to CS.

Assume that all applicants are either male or female, and that no applicant applied to both departments. That is, the events F_{EE}, F_{CS}, M_{EE}, and M_{CS} are all disjoint.

In these terms, the plaintiff is make the following argument:

$$\Pr\{A \mid F_{EE}\} < \Pr\{A \mid M_{EE}\}$$

$$\Pr\{A \mid F_{CS}\} < \Pr\{A \mid M_{CS}\}$$

That is, in both departments, the probability that a woman is accepted for tenure is less than the probability that a man is accepted. The university retorts that overall a woman applicant is *more* likely to be accepted than a man:

$$\Pr\{A \mid F_{EE} \cup F_{CS}\} > \Pr\{A \mid M_{EE} \cup M_{CS}\}$$

It is easy to believe that these two positions are contradictory. In fact, we might even try to prove this by adding the plaintiff's two inequalities and then arguing as follows:

$$\Pr\{A \mid F_{EE}\} + \Pr\{A \mid F_{CS}\} < \Pr\{A \mid M_{EE}\} + \Pr\{A \mid M_{CS}\}$$

$$\Rightarrow \qquad \Pr\{A \mid F_{EE} \cup F_{CS}\} < \Pr\{A \mid M_{EE} \cup M_{CS}\}$$

The second line exactly contradicts the university's position! But there is a big problem with this argument; the second inequality follows from the first only if we accept the false identity (14.1). This argument is bogus! Maybe the two parties do not hold contradictory positions after all!

In fact, the table ~~below~~ in Figure D3 shows a set of application statistics for which the assertions of end of insert @ E

Figure D3: A scenario where ~~In each depart~~ females are less likely to be admitted than males in each department, but more likely to be admitted overall

both the plaintiff and the university holds

CS	0 females accepted, 1 applied	0%
	50 males accepted, 100 applied	50%
EE	70 females accepted, 100 applied	70%
	1 male accepted, 1 applied	100%
Overall	70 females accepted, 101 applied	≈ 70%
	51 males accepted, 101 applied	≈ 51%

In this case, a higher percentage of males were accepted in both departments, but overall a higher percentage of females were accepted! Bizarre! ~~Perhaps this is why~~

15.3 ~~14.3.7~~ *A Posteriori* Probabilities ← section

— INSERT H goes here —

Suppose that we turn the hockey question around: what is the probability that the Halting Problem won their first game, given that they won the series?

This seems like an absurd question! After all, if the Halting Problem won the series, then the winner of the first game has already been determined. Therefore, who won the first game is a question of fact, not a question of probability. However, our mathematical theory of probability contains no notion of one event preceding another— there is no

~~Mark Twain once said~~

~~The me~~
If you think about it too much, the medical testing problem we just considered could ~~be~~ ~~the issue resolves~~ ~~the particular,~~ start to trouble you. ~~For~~ The concern would ~~be~~ ~~when~~ be that ~~either you~~ by the time you take the test, you either have the BO condition or you don't — ~~it is not~~ you just don't know which it is. So you may wonder if a statement like "if you tested positive, then you have the condition with probability 25%" makes sense.

In fact, such a statement does make sense. It means that 25% of people who test positive actually have the condition. ~~Any~~ It is true that any particular person has it or they don't, ~~but a randomly selected~~ but a randomly selected ~~individual~~ person ~~with a posit~~ among those who test positive will have the ~~co~~ condition with probability 25%.

~~Anyw~~ ~~Anyway, whether~~

~~It~~ ~~At~~ Anyway, if ~~me~~ the medical testing example ~~bot~~ bothers you, you will ~~be even more bothere~~ definitely be worried by the following ~~of~~ examples, which go even farther down this path.

15.3.1 The "Halting Problem," in Reverse ~~Revisited~~

notion of time at all. Therefore, from a mathematical perspective, this is a perfectly valid question. And this is also a meaningful question from a practical perspective. Suppose that you're told that the Halting Problem won the series, but not told the results of individual games. Then, from your perspective, it makes perfect sense to wonder how likely it is that The Halting Problem won the first game.

A conditional probability $\Pr\{B \mid A\}$ is called *a posteriori* if event B precedes event A in time. Here are some other examples of a posteriori probabilities:

- The probability it was cloudy this morning, given that it rained in the afternoon.
- The probability that I was initially dealt two queens in Texas No Limit Hold 'Em poker, given that I eventually got four-of-a-kind.

Mathematically, a posteriori probabilities are *no different* from ordinary probabilities; the distinction is only at a higher, philosophical level. Our only reason for drawing attention to them is to say, "Don't let them rattle you."

Let's return to the original problem. The probability that the Halting Problem won their first game, given that they won the series is $\Pr\{B \mid A\}$. We can compute this using the definition of conditional probability and ~~our earlier~~ the tree diagram in Figure B2:

$$\begin{aligned}\Pr\{B \mid A\} &= \frac{\Pr\{B \cap A\}}{\Pr\{A\}} \\ &= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\ &= \frac{7}{9}\end{aligned}$$

This answer is suspicious! In the preceding section, we showed that $\Pr\{A \mid B\}$ was also $7/9$. Could it be true that $\Pr\{A \mid B\} = \Pr\{B \mid A\}$ in general? Some reflection suggests this is unlikely. For example, the probability that I feel uneasy, given that I was abducted by aliens, is pretty large. But the probability that I was abducted by aliens, given that I feel uneasy, is rather small.

Let's work out the general conditions under which $\Pr\{A \mid B\} = \Pr\{B \mid A\}$. By the definition of conditional probability, this equation holds if an only if:

$$\frac{\Pr\{A \cap B\}}{\Pr\{B\}} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}}$$

This equation, in turn, holds only if the denominators are equal or the numerator is 0:

$$\Pr\{B\} = \Pr\{A\} \qquad \text{or} \qquad \Pr\{A \cap B\} = 0$$

The former condition holds in the hockey example; the probability that the Halting Problem wins the series (event A) is equal to the probability that it wins the first game (event B). In fact, both probabilities are $1/2$.

In general, such

~~Such~~ pairs of probabilities are related by Bayes' Rule:

Theorem 14.3.2 (Bayes' Rule). *If* $\Pr\{A\}$ *and* $\Pr\{B\}$ *are nonzero, then:*

$$\frac{\Pr\{A \mid B\} \cdot \Pr\{B\}}{\Pr\{A\}} = \Pr\{B \mid A\} \tag{14.2}$$

reverse left & right

around =

Proof. When $\Pr\{A\}$ and $\Pr\{B\}$ are nonzero, we have

$$\Pr\{A \mid B\} \cdot \Pr\{B\} = \Pr\{A \cap B\} = \Pr\{B \mid A\} \cdot \Pr\{A\}$$

by definition of conditional probability. Dividing by $\Pr\{A\}$ gives (14.2). ■

¶ Next, let's look at a problem that even bothers us.

~~In the hockey problem, the probability that the Halting Problem wins the first game is 1/2 and so is the probability that the Halting Problem wins the series. Therefore, $\Pr\{A\} = \Pr\{B\} = 1/2$. This, together with Bayes' Rule, explains why $\Pr\{A \mid B\}$ and $\Pr\{B \mid A\}$ turned out to be equal in the hockey example.~~

INSERT D goes here (text on pp 932-934)

INSERT C goes here (text on pp 930-932)

INSERT E goes here (text on pp 941-946)

~~14.4~~ 15.5 Independence

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The

15.5.1 Definition

mathematical concept that captures this intuition is called *independence*:

X1

Definition. Events A and B are independent if ~~and only if:~~ $Pr(B) = 0$ or if

$$Pr(A \mid B) = Pr(A) \qquad (14.3)$$

$$\cancel{\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}}$$

A and B are independent if

In other words, knowing that B happens does not alter the probability that A happens, ~~as is~~ as is the case with flipping two coins on ~~the case with flipping two coins~~ ~~on different~~ opposite sides of a room. ~~sides of the room.~~

Equivalently, A and B are independent if and only if

$$Pr(A \cap B) = Pr(A) \cdot Pr(B). \qquad \text{(Eqn D3)}$$

This follows from the definition of independence and ~~Equation~~ Definition 14.3.1.

Generally, independence is something you *assume* in modeling a phenomenon— or wish you could realistically assume. Many useful probability formulas only hold if certain events are independent, so a dash of independence can greatly simplify the analysis of a system.

For example, ~~let's~~ consider

~~**14.4.1 Examples**~~

~~Let's return to~~ the experiment of flipping two fair coins. Let A be the event that the first coin comes up heads, and let B be the event that the second coin is heads. If we assume that A and B are independent, then the probability that both coins come up heads is:

$$\begin{aligned}\Pr\{A \cap B\} &= \Pr\{A\} \cdot \Pr\{B\} \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4}\end{aligned}$$

On the other hand, let C be the event that tomorrow is cloudy and R be the event that tomorrow is rainy. Perhaps $\Pr\{C\} = 1/5$ and $\Pr\{R\} = 1/10$ ~~around here~~ in Boston. If these events were independent, then we could conclude that the probability of a rainy, cloudy day was quite small:

$$\begin{aligned}\Pr\{R \cap C\} &= \Pr\{R\} \cdot \Pr\{C\} \\ &= \frac{1}{5} \cdot \frac{1}{10} \\ &= \frac{1}{50}\end{aligned}$$

Unfortunately, these events are definitely not independent; in particular, every rainy day is cloudy. Thus, the probability of a rainy, cloudy day is actually $1/10$.

14.4.2 Working with Independence

There is another way to think about independence that you may find more intuitive. According to the definition, events A and B are independent if and only if $\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}$. This equation holds even if $\Pr\{B\} = 0$, but assuming it is not, we can divide both sides by $\Pr\{B\}$ and use the definition of conditional probability to obtain an alternative formulation of independence.

Proposition. *If* $\Pr\{B\} \neq 0$, *then events A and B are independent if and only if*

$$\Pr\{A \mid B\} = \Pr\{A\}. \tag{14.3}$$

Equation (14.3) says that events A and B are independent if the probability of A is unaffected by the fact that B happens. In these terms, the two coin tosses of the previous section were independent, because the probability that one coin comes up heads is unaffected by the fact that the other came up heads. Turning to our other example, the probability of clouds in the sky is strongly affected by the fact that it is raining. So, as we noted before, these events are not independent.

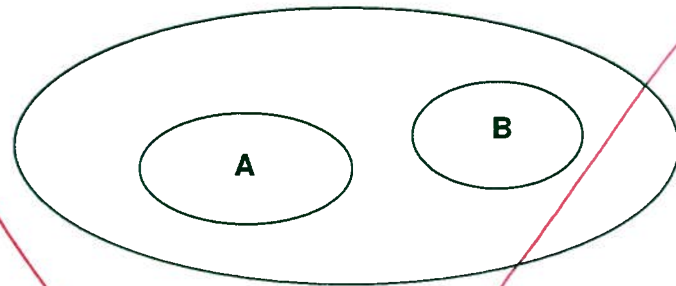
Potential Pitfall.

~~**Warning:**~~ Students sometimes get the idea that disjoint events are independent. The *opposite* is true: if $A \cap B = \emptyset$, then knowing that A happens means you know that B does not happen. So disjoint events are *never* independent —unless one of them has probability zero.

EDITING NOTE:

Some Intuition

Suppose that A and B are disjoint events, as shown in the figure below.



Are these events independent? Let's check. On one hand, we know

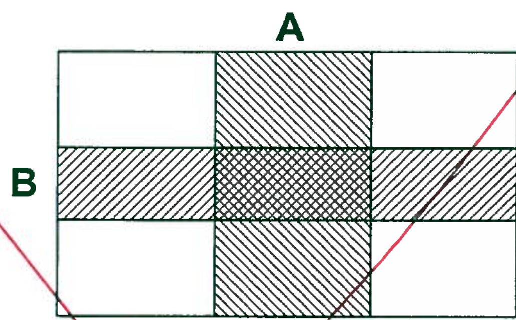
$$\Pr\{A \cap B\} = 0$$

because $A \cap B$ contains no outcomes. On the other hand, we have

$$\Pr\{A\} \cdot \Pr\{B\} > 0$$

except in degenerate cases where A or B has zero probability. Thus, *disjointness and independence are very different ideas.*

Here's a better mental picture of what independent events look like.



The sample space is the whole rectangle. Event A is a vertical stripe, and event B is a horizontal stripe. Assume that the probability of each event is proportional to its area in the diagram. Now if A covers an α-fraction of the sample space, and B covers a β-fraction, then the area of the intersection region is $\alpha \cdot \beta$. In terms of probability:

$$\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}$$

■

15.5.2

~~14.4.3~~ Mutual Independence

We have defined what it means for two events to be independent. But what if ~~how can we talk about independence when~~ there are more than two events? For example, how can we say that the flips ~~orientations~~ of n coins are all independent of one another?

Events $E_1, \ldots, E_n$ are *mutually independent* if and only if *for every subset* of the events, the probability of the intersection is the product of the probabilities. In other words, all footnote: There is an equivalent definition based on conditional probabilities, as in Definition X1, ~~but the definit~~ but it is more complicated for large values of n.

of the following equations must hold:[1]

$$\Pr\{E_i \cap E_j\} = \Pr\{E_i\} \cdot \Pr\{E_j\} \qquad \text{for all distinct } i, j$$

$$\Pr\{E_i \cap E_j \cap E_k\} = \Pr\{E_i\} \cdot \Pr\{E_j\} \cdot \Pr\{E_k\} \qquad \text{for all distinct } i, j, k$$

$$\Pr\{E_i \cap E_j \cap E_k \cap E_l\} = \Pr\{E_i\} \cdot \Pr\{E_j\} \cdot \Pr\{E_k\} \cdot \Pr\{E_l\} \quad \text{for all distinct } i, j, k, l$$

$$\cdots$$

$$\Pr\{E_1 \cap \cdots \cap E_n\} = \Pr\{E_1\} \cdots \Pr\{E_n\}$$

As an example, if we toss 100 fair coins and let E_i be the event that the ith coin lands heads, then we might reasonably assume that $E_1, \ldots, E_{100}$ are mutually independent.

~~EDITING NOTE:~~

15.5.3 DNA Testing

Assumptions about indepence are routinely made in practice. Frequently, such assumptions are quite reasonable. Sometimes, however, the reasonableness of ~~the~~ an independence assumption is not so clear, and the consequences of a faulty assumption can be severe.

For example, consider the following

~~This is~~ testimony from the O. J. Simpson murder trial on May 15, 1995:

remove box & put in text

MR. CLARKE: When you make these estimations of frequency— and I believe you touched a little bit on a concept called independence?

DR. COTTON: Yes, I did.

MR. CLARKE: And what is that again?

DR. COTTON: It means whether or not you inherit one allele that you have is not— does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn't mean you'll automatically or with some probability inherit one at 6,000. What you inherit from one parent is what you inherit from the other. *(Got that? – EAL)*

MR. CLARKE: Why is that important?

DR. COTTON: Mathematically that's important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

MR. CLARKE: How do you— well, first of all, are these markers independent that

The jury was told that genetic markers in blood found at the crime scene matched Simpson's. Furthermore, they were told that the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker A.
- 1 person in 50 marker B.
- 1 person in 40 has marker C.
- 1 person in 5 has marker D.
- 1 person in 170 has marker E.

Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

$$\begin{aligned}
\Pr\{A \cap B \cap C \cap D \cap E\} &= \Pr\{A\} \cdot \Pr\{B\} \cdot \Pr\{C\} \cdot \Pr\{D\} \cdot \Pr\{E\} \\
&= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} \\
&= \frac{1}{170,000,000}
\end{aligned}$$

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples.

The jury was widely mocked for failing to "understand" the DNA evidence. If you were a juror, would *you* accept the 1 in 170 million calculation?

15.5.4

~~14.4.4~~ Pairwise Independence

The definition of mutual independence seems awfully complicated— there are so many conditions! Here's an example that illustrates the subtlety of independence when more than two events are involved and the need for all those conditions. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- A_1 is the event that coin 1 matches coin 2.
- A_2 is the event that coin 2 matches coin 3.
- A_3 is the event that coin 3 matches coin 1.

Are A_1, A_2, A_3 mutually independent?

The sample space for this experiment is:

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Every outcome has probability $(1/2)^3 = 1/8$ by our assumption that the coins are mutually independent.

To see if events A_1, A_2, and A_3 are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event A_i:

$$\begin{aligned}
\Pr\{A_1\} &= \Pr\{HHH\} + \Pr\{HHT\} + \Pr\{TTH\} + \Pr\{TTT\} \\
&= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{2}
\end{aligned}$$

By symmetry, $\Pr\{A_2\} = \Pr\{A_3\} = 1/2$ as well. Now we can begin checking all the equalities required for mutual independence.

$$\begin{aligned}
\Pr\{A_1 \cap A_2\} &= \Pr\{HHH\} + \Pr\{TTT\} \\
&= \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{4} \\
&= \frac{1}{2} \cdot \frac{1}{2} \\
&= \Pr\{A_1\} \Pr\{A_2\}
\end{aligned}$$

By symmetry, $\Pr\{A_1 \cap A_3\} = \Pr\{A_1\} \cdot \Pr\{A_3\}$ and $\Pr\{A_2 \cap A_3\} = \Pr\{A_2\} \cdot \Pr\{A_3\}$ must hold also. Finally, we must check one last condition:

$$\begin{aligned}
\Pr\{A_1 \cap A_2 \cap A_3\} &= \Pr\{HHH\} + \Pr\{TTT\} \\
&= \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{4} \\
&\neq \Pr\{A_1\}\Pr\{A_2\}\Pr\{A_3\} = \frac{1}{8}
\end{aligned}$$

The three events A_1, A_2, and A_3 are not mutually independent even though any two of them are independent! This not-quite mutual independence seems weird at first, but it happens. It even generalizes:

Definition 14.4.1. A set $A_1, A_2, \ldots$ of events is *k-way independent* iff every set of k of these events is mutually independent. The set is *pairwise independent* iff it is 2-way independent.

So the sets A_1, A_2, A_3 above are pairwise independent, but not mutually independent.

Pairwise independence is a much weaker property than mutual independence, ~~but it's~~ it is sometimes good enough in practice, as we will see ~~all that's needed to justify a standard approach to making probabilistic estimates that~~ later. ~~will come up later.~~

~~EDITING NOTE:~~

For example, suppose that the prosecutors in the O. J. Simpson trial were wrong and markers A, B, C, D, and E appear only *pairwise* independently. Then the probability that a randomly-selected person has all five markers is no more than:

$$
\begin{aligned}
\Pr\{A \cap B \cap C \cap D \cap E\} &\le \Pr\{A \cap E\} \\
&= \Pr\{A\} \cdot \Pr\{E\} \\
&= \frac{1}{100} \cdot \frac{1}{170} \\
&= \frac{1}{17{,}000}
\end{aligned}
$$

~~# On the other hand, it is a lot better than we can say if there was no independence at all between the markers. In this case the best we can say is that the probability of a match is at most 1 in 170.~~

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the A and E markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000— a far cry from 1 in 170 million! And this is the strongest conclusion we can reach assuming only pairwise independence.

— INSERT Q goes here —

~~Is a probability of 1 in 17,000 good enough to be certain beyond a reasonable doubt?~~

15.5.5 ~~14.5~~ The Birthday ~~Principle~~ Paradox ← subsection

Suppose that there are 100 ~~There are 85~~ students in a class. What is the probability that some birthday is shared by two people? Comparing 100 ~~85~~ students to the 365 possible birthdays, you might guess the probability lies somewhere around 1/3 ~~1/4~~—but you'd be wrong: the probability that there will be two people in the class with matching birthdays is actually ~~more than 0.9999~~ .99999692... . STET ~~STET~~

In other words, the probability that all 100 birthdays are different is less than 1 in 3,000,000 ~~100,000~~.

Why is this probability so small? The answer involves a phenomenon known as the ~~Birthday Principle or~~ Birthday Paradox (or the Birthday Principle), which is surprisingly important in computer science, as we'll see later.

Before delving into the analysis, we'll need to

— INSERT P goes here —

On the other hand, ~~that~~ the 1 in 17,000 bound that we get ~~from~~ by assuming pairwise independence is a lot better than the bound that we would have if there were no independence at all. For example, if the markers ~~are not~~ pairwise ~~independent~~, then it is possible that are dependent,

everyone with marker E has marker A,
everyone with marker A has marker B,
everyone with marker B has marker C, and
everyone with marker C has marker D.

In such a scenario, ~~the best that we can say is that~~ the probability of a match is

$$\Pr[E] = 1/170.$$

So ~~the~~ a stronger ~~the~~ independence assumption ~~the less likely it is for a match to occur.~~ ~~makes for a tighter~~ ~~smaller~~ ~~probabilit~~ leads to a smaller bound on

~~the~~ the probability of a match. The trick is to figure out what independence assumption is reasonable. ~~E~~ Assuming that the markers are <u>mutually</u> independent may well <u>not</u> be reasonable unless you have examined hundreds of millions of blood ~~ty~~ samples. Otherwise, how would you know, ~~that the p~~ for example, ~~the the probabili~~ that marker ~~E~~ D does not show up more frequently whenever the other four markers are simultaneously present?

We will conclude our discussion of independence with one final, ~~ex~~ and somewhat famous, example known as the Birthday ~~Principle~~ Paradox.

INSERT P

P-1

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the A and E markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000— a far ~~cry from 1 in 170 million!~~ And this is the strongest conclusion we can reach assuming only pairwise independence.

3 The Birthday Paradox

After all, ~~in a hash~~ whether or not two items collide in a hash table really has nothing to do with human reproductive preferences.

~~Suppose that there are 100 students in a lecture hall.~~ There are 365 possible birthdays, ~~ignoring February 29. What is the probability that two students have the same birthday? 50%? 90%? 99%? Let's~~ make some modeling assumptions:

START HERE

- For each student, all possible birthdays are equally likely. The idea underlying this assumption is that each student's birthday is determined by a random process involving parents, fate, and, um, some issues that we discussed earlier in the context of graph theory. ~~Our~~ The assumption is not completely accurate, however; a disproportionate number of babies are born in August and September, for example. ~~(Counting back nine months explains the reason why!)~~ ~~the principle in computer~~ For applications of computer science, the assumption is more ~~reasonable; whether or not two items collide~~
- Birthdays are mutually independent. This isn't perfectly accurate either. For example, if there are twins in the ~~lecture hall~~ class, then their birthdays are surely not independent.

We'll stick with these assumptions, despite their limitations. Part of the reason is to simplify the analysis. But the bigger reason is that our conclusions will apply to many situations in computer science where twins, leap days, and romantic holidays are not considerations. Also in pursuit of generality, let's switch from specific numbers to variables. Let m be the number of people in the room, and let N be the number of days in a year.

~~3.1 The Four-Step Method~~

We can solve this problem using the standard four-step method. However, a tree diagram will be of little value because the sample space is so enormous. This time we'll have to proceed without the visual aid!

Step 1: Find the Sample Space

Let's number the people in the room from 1 to m. An outcome of the experiment is a sequence $(b_1, \ldots, b_m)$ where b_i is the birthday of the ith person. The sample space is the set of all such sequences:

$$S = \{(b_1, \ldots, b_m) \mid b_i \in \{1, \ldots, N\}\}$$

P-2

Step 2: Define Events of Interest

Our goal is to determine the probability of the event A, in which some two people have the same birthday. This event is a little awkward to study directly, however. So we'll use a common trick, which is to analyze the *complementary* event $\overline{A}$, in which all m people have different birthdays:

$$\overline{A} = \{(b_1, \ldots, b_m) \in S \mid \text{all } b_i \text{ are distinct}\}$$

If we can compute $\Pr\left(\overline{A}\right)$, then we can compute what we really want, $\Pr(A)$, using the ~~relation:~~ ~~formula:~~ identity:

$$\Pr(A) + \Pr\left(\overline{A}\right) = 1$$

Step 3: Assign Outcome Probabilities

We need to compute the probability that m people have a particular combination of birthdays $(b_1, \ldots, b_m)$. There are N possible birthdays and all of them are equally likely for each student. Therefore, the probability that the ith person was born on day b_i is $1/N$. Since we're assuming that birthdays are mutually independent, we can multiply probabilities. Therefore, the probability that the first person was born on day b_1, the second on day b_2, and so forth is $(1/N)^m$. This is the probability of every outcome in the sample space, which means that the sample space is uniform. That's good news, because as we have seen, it means that the analysis will be simpler.

Step 4: Compute Event Probabilities

~~Now we're~~ We're interested in the probability of the event $\overline{A}$ in which everyone has a different birthday:

$$\overline{A} = \{(b_1, \ldots, b_m) \in S \mid \text{all } b_i \text{ are distinct}\}$$

This is a gigantic set. In fact, there are N choices for b_1, $N-1$ choices for b_2, and so forth. Therefore, by the Generalized Product Rule:

$$\left|\overline{A}\right| = N(N-1)(N-2)\ldots(N-m+1) = N!/(N-m)!$$

Since the sample space is uniform, we can ~~de~~ conclude that ~~The probability of the event $\overline{A}$ is the sum of the probabilities of all these outcomes. Happily, this sum is easy~~ to compute, owing to ~~the fact that every outcome has the same probability~~:

$$\begin{aligned}
\Pr\left(\overline{A}\right) &= \cancel{\sum_{w \in A} \Pr(w)} \\
&= \frac{\left|\overline{A}\right|}{N^m} \\
&\cancel{= \frac{N(N-1)(N-2)\ldots(N-m+1)}{N^m}} \\
&\cancel{= \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\ldots\left(1-\frac{m-1}{N}\right)} \\
&= \frac{N!}{N^m (N-m)!} \qquad \text{(Eqn E4)}
\end{aligned}$$

We're done!

or are we? While correct, it would certainly be nicer to have a closed form expression for Equation E4. That means finding an approximation for N! and (N-m)! But this is what we learned how to do in Chapter 9. In fact, ~~by using the bounds~~ from Theorem 9.6.1, we

$$\cancel{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)} \le n! \le \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}}$$

~~where~~

~~from Theorem 9.6.1, to bound N! and (N-m)!~~

~~in Equation E4, and then doing~~

~~form~~ know that

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{a_n} \qquad (\text{eqn } E7)$$

where

$$\frac{1}{12n+1} \le a_n \le \frac{1}{12n}.$$

~~Plugging~~ N! and (N-m)! ~~form in Equation E7, order for N! and (N-m)!~~

~~and grinding through a~~ bunch of algebra, ~~we~~

we ~~found that~~

Plugging Equation E7 in for $N!$ and $(N-m)!$ in Equation E4a and simplifying yields a closed form expression for the probability that all m birthdays are different:

$$P[\bar{A}] = \frac{N!}{N^m (N-m)!}$$

$$= \frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N e^{a_N}}{N^m \sqrt{2\pi(N-m)}\left(\frac{N-m}{e}\right)^{N-m} e^{a_{N-m}}}$$

$$= \sqrt{\frac{N}{N-m}}\ \frac{e^{N\ln(N) - N + a_N}}{e^{m\ln(N)}\ e^{(N-m)\ln(N-m) - (N-m) + a_{N-m}}}$$

$$= \sqrt{\frac{N}{N-m}}\ e^{(N-m)\ln(N) - (N-m)\ln(N-m) - m + a_N - a_{N-m}}$$

$$= \sqrt{\frac{N}{N-m}}\ e^{(N-m)\ln\left(\frac{N}{N-m}\right) - m + a_N - a_{N-m}} \quad \text{Eqn E9}$$

We can now evaluate Equation E9 for $m=100$ and $N=365$ to find the probability that all 100 birthdays are different is[1]

$$3.07\ldots \cdot 10^{-7}.$$

1 The contribution of a_N and a_{N-m} is so small that it is lost in the ... after 3.07.

~~The~~ we can also plug in other values of m to find the number of people needed for the probability of a matching birthday to be about 1/2. In particular, for $m = 23$ and $N = 365$, ~~we find~~ Equation E9 reveals that the probability that all the birthdays differ is 0.493.... So if you are in a room with 22 other ~~[illegible]~~

~~The [illegible]~~
~~The issue of when the first match~~
~~In computer science applications~~

people, the probability that some pair of people share a birthday will be a little better than 1/2. ~~[illegible]~~ ~~Because that~~ It is because 23 seems to be a small number of people for a match that the phenomenon ~~the Birthday Principle~~ is ~~often~~ called the <u>Birthday Paradox</u>.

15.5.6 Applications to Hashing

Hashing is frequently used in computer science to map ~~a subset of a large set into a small set.~~ large strings of data into short strings of data. In a typical scenario, you ~~For example suppose that~~ have a set of m ~~1000 messages~~ items ~~(keys such as messages, key)~~ ~~(key messages such as messages, addresses, variables, and etc and so on,~~ and ~~that~~ you would like to assign each ~~message~~ item to a number from 1 to N where ~~N is small~~ no pair of items is assigned to the same number and N is as small as possible. ~~The~~ For example, the items might be messages, addresses, or variables. The numbers might represent storage locations, devices, indices, or digital signatures.

If two items are assigned to the same number, then a collision is said to occur. Collisions are generally

For example, collisions

bad. ~~since they~~ can correspond to two ~~messages items~~ variable being ~~put~~ stored in the same place or two messages being assigned the same digital signature. ~~The~~ In fact, ~~when~~ finding collisions is a common technique in breaking cryptographic codes; ~~and they~~ Such ~~attacks~~ techniques are often referred to as Birthday attacks. ~~A referen~~

~~How big does N need to be for the probability of a collision to be low?~~

~~INSERT Q goes here~~

~~How large can m be~~

~~Suppose the value of N is fixed. How many items can~~ ~~In practice, the assignment of a number to an item is done randomly.~~

~~Given N, how~~

~~The problem now is, given N, how large can m be before a collision is likely to occur. For example,~~

For efficiency purposes, it is generally desirable to make N ~~the size of the hash table~~ as small as possible to accomodate the hashing of m items without collisions.

INSERT Q

~~Ideally~~

In practice, the assignment of a number to an item is done using a hash function

$$h : S \to [1, N]$$

where S is the set of items and $n = |S|$.

Typically, ~~it is~~ the values of h(s) are assumed to ~~produce random~~ ~~values that~~ ~~values are~~ be uniformly selected from [1, N] and ~~that are~~ to be mutually independent.

~~For e~~

Ideally, ~~m ≈~~ N would be only a little larger than m. Unfortunately, this is not possible for ~~truly~~ random hash functions. ~~In fact~~ To see why, lets take a closer look at Equation E9.

~~As N grows~~

~~If $m = o(N^{2/3})$~~

~~We already know that if~~

~~As m~~

~~If m is near N, then the value of $P[\bar{A}]$ is small. In fact, even if $m = o(N^{2/3})$,~~

$$\cancel{\Pr[A] \approx} \; e^{\cancel{(N-m)\ln\left(\frac{N}{N-m}\right)}}$$

~~As N gets large,~~

$$\cancel{(N-m)\ln\left(\frac{N}{N-m}\right) - m} = \cancel{-(N-m)\ln\left(\frac{N-m}{N}\right)} - m$$

$$= \cancel{-(N-m)\ln\left(1-\frac{m}{N}\right)} - m$$

$$= -(N-m)\cancel{\left(-\frac{m}{N} - \frac{m^2}{2N^2} - \frac{m^3}{3N^3} - \cdots\right)} - m$$

Using the Taylor Series expansion for

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots$$

in Equation E9, we find that

$$\cancel{\Pr[\bar{A}] = \sqrt{\frac{N}{N-m}}\; e^{(N-m)\ln\left(\frac{N}{N-m}\right) - m + a_N - a_{N-m}}}$$

$$\cancel{= \sqrt{1 + \frac{m}{N-m}}\; e}$$

$$(N-m)\ln\left(\frac{N}{N-m}\right) - m = -(N-m)\ln\left(\frac{N-m}{N}\right) - m$$

$$= -(N-m)\ln\left(1 - \frac{m}{N}\right) - m$$

$$= -(N-m)\left(-\frac{m}{N} - \frac{m^2}{2N^2} - \frac{m^3}{3N^2} - \cdots\right) - m$$

$$= \left(m + \frac{m^2}{2N} + \frac{m^3}{3N^2} + \cdots\right) - \left(\frac{m^2}{N} + \frac{m^3}{2N^2} + \frac{m^4}{3N^3} + \cdots\right) - m$$

$$= -\frac{m^2}{2N} - \frac{m^3}{6N^2} - \frac{m^4}{12N^3} - \cdots .$$

Hence, for $m = o(N^{2/3})$,

$$\Pr[\bar{A}] = \sqrt{\frac{N}{N-m}}\; e^{-\frac{m^2}{2N} - \frac{m^3}{6N^2} - \frac{m^4}{12N^3} - \cdots - a_N - a_{N-m}}$$

$$\sim e^{-m^2/2N} .$$

This means that if

$$m = \sqrt{2\ln(2)}\,\sqrt{N}$$

$$= 1.177\ldots\,\sqrt{N},$$

then ~~and~~ ~~then~~ there ~~is about~~ $\Pr[A] \approx 1/2$ will be a collision with probability near 1/2. ~~Th~~

In other words, N needs to ~~be~~ grow quadratically with m in order to avoid collisions. This unfortunate fact ~~is~~ is known as the <u>Birthday Principle</u> ~~e~~ ~~and it means that hash either in order~~ ~~and it as~~ and it ~~governs~~ limits the efficiency of hashing ~~app~~ in practice — either N ~~needs~~ is quadratic in ~~the~~ the number of items being hashed or you need to be able to deal with collisions.

15.7 Problems

To work this out, we'll assume that the probability that a randomly chosen student has a given birthday is $1/d$, where $d = 365$ in this case. We'll also assume that a class is composed of n randomly and mutually independently selected students, with $n = 85$ in this case. These randomness assumptions are not really true, since more babies are born at certain times of year, and students' class selections are typically not independent of each other, but simplifying in this way gives us a start on analyzing the problem. More importantly, these assumptions are justifiable in important computer science applications of birthday matching. For example, the birthday matching is a good model for collisions between items randomly inserted into a hash table. So we won't worry about things like Spring procreation preferences that make January birthdays more common, or about twins' preferences to take classes together (or not).

EDITING NOTE: or that fact that a student can't be selected twice in making up a

~~class list.~~

Selecting a sequence of n students for a class yields a sequence of n birthdays. Under

mutual independence

the assumptions above, the d^n possible birthday sequences are equally likely outcomes. Let's examine the consequences of this probability model by focussing on the ith and jth elements in a birthday sequence, where $1 \leq i \neq j \leq n$. It makes for a better story if we refer to the ith birthday as "Alice's" and the jth as "Bob's."

Now since Bob's birthday is assumed to be independent of Alice's, it follows that whichever of the d birthdays Alice's happens to be, the probability that Bob has the same birthday $1/d$. Next, If we look at two other birthdays —call them "Carol's" and "Don's" —then whether Alice and Bob have matching birthdays has nothing to do with whether Carol and Don have matching birthdays. That is, the event that Alice and Bob have matching birthdays is independent of the event that Carol and Don have matching birthdays. In fact, for any set of non-overlapping couples, the events that a couple has matching birthdays are mutually independent.

In fact, it's pretty clear that the probability that Alice and Bob have matching birthdays remains $1/d$ whether or not Carol and *Alice* have matching birthdays. That is, the event that Alice and Bob match is also independent of Alice and Carol matching. In short, the set of all events in which a couple has macthing birthdays is *pairwise* independent, despite the overlapping couples. This will be important in Chapter 17 because pairwise independence will be enough to justify some conclusions about the expected number of matches. However, it's obvious that these matching birthday events are *not* mutually independent, not even 3-way independent: if Alice and Bob match and also Alice and Carol match, then Bob and Carol will match.

We could justify all these assertions of independence routinely using the four step method, but it's pretty boring, and we'll skip it.

It turns out that as long as the number of students is noticeably smaller than the number of possible birthdays, we can get a pretty good estimate of the birthday matching probabilities by *pretending* that the matching events are mutually independent. (An intuitive justification for this is that with only a small number of matching pairs, it's likely that none of the pairs overlap.) Then the probability of *no* matching birthdays would be the same as rth power of the probability that a couple does *not* have matching birthdays, where $r ::= \binom{n}{2}$ is the number of couples. That is, the probability of no matching birthdays would be

$$(1 - 1/d)^{\binom{n}{2}}. \tag{14.4}$$

Using the fact that $e^x > 1 + x$ for all x,[2] we would conclude that the probability of no

[2]This approximation is obtained by truncating the Taylor series $e^{-x} = 1 - x + x^2/2! - x^3/3! + \cdots$. The approximation $e^{-x} \approx 1 - x$ is pretty accurate when x is small.

matching birthdays is at most

$$e^{-\frac{\binom{n}{2}}{d}}. \tag{14.5}$$

The matching birthday problem fits in here so far as a nice example illustrating pair-wise and mutual independence. But it's actually not hard to justify the bound (14.5) without any pretence or any explicit consideration of independence. Namely, there are $d(d-1)(d-2)\cdots(d-(n-1))$ length n sequences of distinct birthdays. So the probability that everyone has a different birthday is:

$$\frac{d(d-1)(d-2)\cdots(d-(n-1))}{d^n}$$

$$= \frac{d}{d}\cdot\frac{d-1}{d}\cdot\frac{d-2}{d}\cdots\frac{d-(n-1)}{d}$$

$$= \left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right)$$

$$< e^0\cdot e^{-1/d}\cdot e^{-2/d}\cdots e^{-(n-1)/d} \qquad \text{(since } 1+x<e^x\text{)}$$

$$= e^{-\left(\sum_{i=1}^{n-1} i/d\right)}$$

$$= e^{-(n(n-1)/2d)}$$

$$= \text{the bound (14.5).}$$

For $n = 85$ and $d = 365$, (14.5) is less than $1/17,000$, which means the probability of having some pair of matching birthdays actually is more than $1 - 1/17,000 > 0.9999$. So it would be pretty astonishing if there were no pair of students in the class with matching birthdays.

For $d \leq n^2/2$, the probability of no match turns out to be asymptotically equal to the upper bound (14.5). For $d = n^2/2$ in particular, the probability of no match is asymptotically equal to $1/e$. This leads to a rule of thumb which is useful in many contexts in computer science:

> **The Birthday Principle**
>
> If there are d days in a year and $\sqrt{2d}$ people in a room, then the probability that two share a birthday is about $1 - 1/e \approx 0.632$.

For example, the Birthday Principle says that if you have $\sqrt{2 \cdot 365} \approx 27$ people in a room, then the probability that two share a birthday is about 0.632. The actual probability is about 0.626, so the approximation is quite good.

Among other applications, the Birthday Principle famously comes into play as the basis of "birthday attacks" that crack certain cryptographic systems.

Class Problems

Homework Problems

Class Problems

Practice Problems

Class Problems

Homework Problems

Class Problems