1 Logic

Logic is about good and bad reasoning. In order to talk clearly about reasoning, logicians have given precise meanings to some ordinary words. This chapter is a review of their language.

ARGUMENTS

Logicians attach a special sense to the word *argument.* In ordinary language, it usually takes two to argue. One dictionary defines an argument as:

1 A quarrel.
2 A discussion in which reasons are put forward in support of and against a proposition, proposal, or case.
3 A point or series of reasons presented to support a proposition which is the conclusion of the argument.

Definition (3) is what logicians mean by an argument.

Reasoning is stated or written out in arguments. So logicians study arguments (in sense 3).

An argument thus divides up into:

A point or series of reasons which are called *premises,*
and a *conclusion.*

Premises and conclusion are *propositions,* statements that can be either true or false. Propositions are "true-or-false."

GOING WRONG

The premises are supposed to be reasons for the conclusion. Logic tries to understand the idea of a good reason.

We find arguments convincing when we know that the premises are true, and when we see that they give a good reason for the conclusion.

So two things can go wrong with an argument:

- the premises may be false.
- the premises may not provide a good reason for the conclusion.

Here is an argument:

(*J) If James wants a job, then he will get a haircut tomorrow.
James will get a haircut tomorrow.
So:
James wants a job.

The first two propositions are the premises. The third proposition is the conclusion.

Someone might offer this argument, thinking the premises give a conclusive reason for the conclusion. They do not. The premises could be true and the conclusion false, for any number of reasons. For example:

James has a date with a girl who likes tidy men, and his hair is a mess.
He has to go home to his family, who would be disgusted by how he looks.
It is the third Monday of the month, and he always gets a haircut then.
No way does he want a job! Of course, if he did want a job, he'd get a haircut tomorrow.

Argument (*J), if offered as a conclusive argument, commits an error—a common error. That is why we labeled it with a "star" in front, as a warning that it is a bad argument.

Argument (*J) commits a *fallacy*. A fallacy is an error in reasoning that is so common that logicians have noted it. Sometimes they give it a name. Argument (*J) commits the fallacy called "affirming the consequent." The first premise in the argument is of the form:

If A, then C.

A is called the *antecedent* of this "if-then" proposition, and C is called the *consequent*.

The second premise of (*J) is of the form "C." So in stating this premise, we "affirm the consequent."

The conclusion is of the form "A." It is a fallacy to infer the antecedent A from the consequent C. That is the fallacy of affirming the consequent.

TWO WAYS TO CRITICIZE

Here is a conclusive argument that only looks a little like (*J):

(J) If James wants a job, then he will get a haircut tomorrow.
James wants a job.
So:
James will get a haircut tomorrow.

Here the premises do provide a conclusive reason for the conclusion. If the premises are true, then the conclusion must be true too.

But you might question the premises.

You might question the first premise if you knew that James wants a job as a rock musician. The last thing he wants is a haircut.

You might also question the second premise. Does James really want a job?

There are two basic ways to criticize an argument:

- Challenge the premises—show that at least one is false.
- Challenge the reasoning—show that the premises are not a good reason for the conclusion.

The two basic types of criticism apply to any kind of argument whatsoever. But logic is concerned only with reasoning. It cannot in general tell whether premises are true or false. It can only tell whether the reasoning is good or bad.

VALIDITY

Here is another conclusive argument:

(K) Every automobile sold by Queen Street Motors is rust-proofed.
Barbara's car was sold by Queen Street Motors.
Therefore:
Barbara's car is rust-proofed.

If the two premises of (K) are true, then the conclusion must be true too. The same goes for (J) above. But not for (*J)!

This idea defines a valid argument. It is logically impossible for the conclusion to be false given that the premises are true.

Validity is best explained in terms of logical form. The logical form of arguments (J) and (K) is:

1. If *A*, then *C*.	4. Every *F* is *G*.
2. *A*.	5. *b* is *F*.
So:	Therefore:
3. *C*.	6. *b* is *G*.

Whenever an argument of one of these forms has true premises, then the conclusion is also true. That is a definition of a valid argument form.

Valid is a technical term of deductive logic. The opposite of *valid* is *invalid*. In ordinary life, we talk about a valid driver's license. We say someone is making a valid point if there is a basis for it, or even if it is true. But we will stick to the special, logicians' meaning of the word. Arguments are valid or invalid.

Argument (*J) above was invalid. Here is another invalid argument:

(*K) Every automobile sold by Queen Street Motors is rust-proofed.
Barbara's car is rust-proofed.
Therefore:
Barbara's car was sold by Queen Street Motors.

This is invalid because the conclusion could be false, even when the premises are true. Many companies sell rust-proofed cars, so Barbara need not have bought hers at Queen Street Motors.

TRUE VERSUS VALID

Be careful about *true* and *valid*. In logic:

> Propositions are true or false.
> Arguments are valid or invalid.

You should also distinguish the argument (K) about Barbara's car from an "if-then" or conditional proposition like this:

If Barbara's car was sold by Queen Street Motors, and if every automobile sold by Queen Street Motors is rust-proofed, then Barbara's car is rust-proofed.

This is a true proposition of the form,

If p and if q, then r.

Or, in finer detail,

If b is F, and if every F is G, then b is G.

Argument (K), on the other hand, is of the form:

4. p.	Or, in finer detail,	4. Every F is G.
5. q.		5. b is F.
So:		Therefore:
6. r.		6. b is G.

To every argument there is a corresponding conditional proposition "if-then." An argument is valid if and only if the corresponding conditional proposition is a truth of logic.

METAPHORS

There are many ways to suggest the idea of validity:

The conclusion follows from the premises.
Whenever the premises are true, the conclusion *must* be true too.
The conclusion is a logical consequence of the premises.
The conclusion is implicitly contained in the premises.
Valid argument forms are truth-preserving.

"Truth-preserving" means that whenever you start out with true premises, you will end up with a true conclusion.

When you reason from true premises using a valid argument, you never risk drawing a false conclusion. When your premises are true, there is no risk that the conclusion will be false.

Textbooks on deductive logic make precise sense of these metaphors. For the purposes of this book, one metaphor says best what matters for validity:

> Valid arguments are risk-free arguments.

SOUND

A *valid* argument never takes you from true premises to a false conclusion.

But, of course, the argument might have a false premise.

We say an argument is *sound* when:

- all the premises are true, and
- the argument is valid.

Thus an argument may be unsound because:

- A premise is false.
- The argument is invalid.

Validity has to do with the logical connection between premises and conclusion, and *not* with the truth of the premises or the conclusion.

Soundness for deductive logic has to do with *both* validity *and* the truth of the premises.

LIKE BUILDING A HOUSE

Making a deductive argument is like building a house.

- It may be built on sand, and so fall down, because the foundations are not solid. That is like having a false premise.

- Or it may be badly built. That is like having an invalid argument.
- And, of course, a house built on sand with bad design may still stay up. That is like an invalid argument with false premises and a true conclusion.

There are two ways to criticize a contractor who built a house. "The foundations are no good!" Or, "The house is badly built!" Likewise, if someone shows you a deduction you can make two kinds of criticism. "One of your premises is false." Or, "The argument is invalid." Or both, of course.

VALIDITY IS NOT TRUTH!

A valid argument can have a *false premise* but a *true conclusion*. Example:

(R) Every famous philosopher who lived to be over ninety was a mathematical logician.
Bertrand Russell was a famous philosopher who lived to be over ninety.
So:
Bertrand Russell was a mathematical logician.

This argument is valid. The conclusion is true.

But the first premise is false. Thomas Hobbes, the famous political philosopher, lived to be over ninety, but he was not a mathematical logician.

Likewise an argument with *false premises* and a *false conclusion* could be valid. Validity is about the connection between premises and conclusion, not about truth or falsehood.

INVALIDITY IS NOT FALSEHOOD!

An invalid argument can have *true premises* and a *true conclusion*. Example:

(*R) Some philosophers now dead were witty and wrote many books.
Bertrand Russell was a philosopher, now dead.
So:
Bertrand Russell was witty and wrote many books.

Both premises are true. The conclusion is true. But the argument is invalid.

TWO WAYS TO CRITICIZE A DEDUCTION

Both (R) and (*R) are unsound, but for quite different reasons.

You can tell that (*R) is unsound because it is invalid. You can tell it is invalid without knowing *anything* about Bertrand Russell (except that "Bertrand Russell" was someone's name).

Likewise, you can tell that (R) is valid without knowing anything about Bertrand Russell.

But to know whether the premises are true, you have to know something about the world, about history, about philosophers, about Bertrand Russell and others.

Maybe you did not know that Bertrand Russell was witty or that Thomas Hobbes was a famous political philosopher who lived to be over ninety. Now you do.

You need not know anything special about the world to know whether an argument is valid or invalid. But you need to know some facts to know whether a premise is true or false.

There are two ways to criticize a deduction:

- A premise is false.
- The argument is invalid.

So there is a *division of labor*.

Who is an expert on the truth of premises?

Detectives, nurses, surgeons, pollsters, historians, astrologers, zoologists, investigative reporters, you and me.

Who is an expert on validity?

A logician.

Logicians study the relations between premises and conclusions, but, as logicians, are not especially qualified to tell whether the premises are true or false.

EXERCISES

1 *Propositions.* The premises and conclusion of an argument are propositions. Propositions are expressed by statements that can be either true or false. For brevity, we say that propositions are true-or-false.

The headline of a newspaper story is:

SEIZED SERPENTS MAKE STRANGE OFFICE-FELLOWS
SHIPPING ERROR LANDS OFFICIAL WITH PYTHONS

There was a bizarre mix-up. A man who runs a tropical fish store in Windsor, Ontario, was delivered a box of ball pythons from a dealer in California. The newspaper tells us that:

> The ball python is a central African ground dweller that can grow to more than a meter on a diet of small mammals.

(a) Is that true-or-false?
(b) Do you know whether it is true?
(c) Is it what logicians call a proposition? [You should give the same answer to (c) as to (a).]

The newspaper goes on to tell us that:

> The ball python is named for its tendency to curl up into a ball.

(d) Is that true-or-false?
(e) Do you know whether it is true?

The story continues:

> The shipment of tropical fish intended for Windsor went to a snake dealer in Ohio.

(f) Is that a proposition?

In logic, propositions express matters of fact that can be either true or false. Judgments of personal taste, such as "avocados are delicious," are not strictly matters of fact. Avocados taste good to some people and taste slimy and disgusting to others. The proposition that avocados are delicious is not strictly speaking true-or-false. But if I say "avocados taste delicious to me," I am stating something about me, which happens to be true.

Joe, the man who owns the fish store, is quoted as saying:

> Ball pythons are very attractive animals.

(g) Is that true-or-false? Is it a proposition?

Suppose that he had said,

> I think ball pythons are very attractive animals.

(h) Is that true-or-false? Is it a proposition?

The newspaper begins the story by saying "It is not so nice to share your office with a box of snakes for two months." Then it adds, as a full paragraph:

> Especially when it was all a result of being soft-hearted.

(i) Is that a proposition?

Joe has to feed the snakes a lot of live mice. According to the reporter, Joe said,

> I'm not really too thrilled to hear baby mice squeaking and screaming behind me while I'm on the telephone.

(j) Is that a proposition?

Then Joe said,

> Thank God they don't eat every day!

(k) Is that a proposition?

He next asked,

> Do you know any zoos or schools who might want these snakes?

(l) Is that a proposition?

Joe phoned Federal Express, the shipper who had mixed up the deliveries, saying:

> You owe me for my expenses, my trouble, and your mistake.

(m) Is that a proposition?

The story ended happily:

> On Wednesday Federal Express bargained a $1000 payment to Joe.

(n) Is that a proposition?

2 *False all over.* State two arguments—they can be silly ones—in which the premises and conclusion are all false, and such that one argument is (a) valid and the other is (b) invalid.

3 *Unsound.* Is either of your answers to question 2 a sound argument?

4 *Combinations.* Only one of the following eight combinations is impossible. Which one?

(a) All premises true. Conclusion true. Valid.
(b) All premises true. Conclusion false. Valid.
(c) One premise false. Conclusion true. Valid.
(d) One premise false. Conclusion false. Valid.
(e) All premises true. Conclusion true. Invalid.
(f) All premises true. Conclusion false. Invalid.
(g) One premise false. Conclusion true. Invalid.
(h) One premise false. Conclusion false. Invalid.

5 *Soundness.* Which of the combinations just listed are sound arguments?

6 *Conditional propositions.* Which of the following is true-or-false? Which is valid-or-invalid? Which is an argument? Which is a conditional proposition?

(a) Tom, Dick, and Harry died.
So:
All men are mortal.

(b) If Tom, Dick, and Harry died, then all men are mortal.

7 *Chewing tobacco.* Which of these arguments are valid?

(a) I follow three major league teams. Most of their top hitters chew tobacco at the plate.
So:
Chewing tobacco improves batting average.

(b) The top six hitters in the National League chew tobacco at the plate.
So:
Chewing tobacco improves batting average.

(c) A study, by the American Dental Association, of 158 players on seven major league teams during the 1988 season, showed that the mean batting average for chewers was .238, compared to .248 for non users. Abstainers also had a higher fielding average.
So:
Chewing tobacco does not improve batting average.

(d) In 1921, every major league pitcher who chewed tobacco when up to bat had a higher batting average than any major league pitcher who did not.
So:
Chewing tobacco improves the batting average of pitchers.

8 *Inductive baseball.* None of the arguments (7a)–(7d) is valid. Invalid arguments are not conclusive. But some non-conclusive arguments are better than others. They are risky arguments. Each of the arguments (a)–(d) is risky. We have not done any inductive logic yet, but you probably think some of (7a)–(7d) are better arguments than others. Which is best? Which is worst?

KEY WORDS FOR REVIEW

Argument	Conclusion
Proposition	Valid
True-or-false	Sound
Premise	Conditional

2 What Is Inductive Logic?

Inductive logic is about risky arguments. It analyses inductive arguments using probability. There are other kinds of risky arguments. There is inference to the best explanation, and there are arguments based on testimony.

Valid arguments are risk-free. Inductive logic studies risky arguments. A risky argument can be a very good one, and yet its conclusion can be false, even when the premises are true. Most of our arguments are risky.

Begin with the big picture. The Big Bang theory of the origin of our universe is well supported by present evidence, but it could be wrong. That is a risk.

We now have very strong evidence that smoking causes lung cancer. But the reasoning from all that evidence to the conclusion "smoking causes lung cancer" is still risky. It might just turn out that people predisposed to nicotine addiction are also predisposed to lung cancer, in which case our inference, that smoking causes lung cancer, would be in question after all.

After a lot of research, a company concludes that it can make a profit by marketing a special left-handed mouse for personal computers. It is taking a risk.

You want to be in the same class as your friend Jan. You reason that Jan likes mathematics, and so will take another logic class. You sign up for inductive logic. You have made a risky argument.

ORANGES

Here are some everyday examples of risky arguments.

A small grocer sells her old fruit at half-price. I want a box of oranges, cheap. But I want them to be good, sweet, and not rotten. The grocer takes an orange from the top of a box, cuts it open, and shows it to me. Her argument is:

(A) This orange is good.
So:
All (or almost all) the oranges in the box are good.

The premise is evidence for the conclusion: but not very good evidence. Most of the oranges in the box may be rotten.

Argument (A) is not a valid argument. Even if the premise is true, the conclusion may be false. This is a risky argument.

If I buy the box at half-price on the strength of this argument, I am taking a big risk. So I reach into the box, pick an orange at random, and pull it out. It is good too. I buy the box. My reasoning is:

(B) This orange that I chose at random is good.
So:
All (or almost all) the oranges in the box are good.

This argument is also risky. But it is not as risky as (A).

Julia takes six oranges at random. One, but only one, is squishy. She buys the box at half-price. Her argument is:

(C) Of these six oranges that I chose at random, five are good and one is rotten.
So:
Most (but not all) of the oranges in the box are good.

Argument (C) is based on more data than (B). But it is not a valid argument. Even though five out of six oranges that Julia picked at random are fine, she may just have been lucky. Perhaps most of the remaining oranges are rotten.

SAMPLES AND POPULATIONS

There are many forms of risky argument. Arguments (A)–(C) all have this basic form:

Statement about a sample drawn from a given population.
So:
Statement about the population as a whole.

We may also go the other way around. I might know that almost all the oranges in this box are good. I pick four oranges at random to squeeze a big glass of orange juice. I reason:

All or almost all the oranges in this box are good.
These four oranges are taken at random from this box.
So:
These four oranges are good.

This too is a risky argument. I might pick a rotten orange, even if most of the oranges in the box are fine. The form of my argument is:

Statement about a population.
So:
Statement about a sample.

We can also go from sample to sample:

These four oranges that I chose at random are good.
So:
The next four oranges that I draw at random will also be good.

The basic form of this argument is:

Statement about a sample.
So:
Statement about a new sample.

PROPORTIONS

We can try to be more exact about our arguments. These are small juice oranges, 60 to the box. A cautious person might express "almost all" by "90%," and then the argument would look like this:

These four oranges, that I chose at random from a box of 60 oranges, are good.
So:
At least 90% (or 54) of the oranges in the box are good.

At least 90% (or 54) of the oranges in this box are good. These four oranges are taken at random from this box.
So:
These four oranges are good.

PROBABILITY

Most of us are happy putting a "probably" into these arguments:

These four oranges, that I chose at random from a box of 60 oranges, are good.
So, probably:
At least 90% (or 54) of the oranges in the box are good.

At least 90% (or 54) of the oranges in this box are good.
These four oranges are taken at random from this box.
So, probably:
These four oranges are good.

These four oranges, that I chose at random from a box of 60 oranges, are good.
So, probably:
The next four oranges that I draw at random will also be good.

Can we put in numerical probability values? That would be one way of telling which arguments are riskier than others. We will use ideas of probability to study risk.

Probability is a fundamental tool for inductive logic.

We will only do enough probability calculations to make ideas clear. *The focus in this book is on the ideas, not on the numbers.*

DEDUCING PROBABILITIES

Inductive logic uses probabilities. *But not all arguments using probabilities are inductive.* Not all arguments where you see the word "probability" are risky. Probability can be made into a rigorous mathematical idea. Mathematics is a deductive science. We make deductions using probability. In chapter 6 we state basic laws, or axioms, of probability. We *deduce* other facts about probability from these axioms.

Here is a simple deduction about probabilities:

This die has six faces, labeled 1, 2, 3, 4, 5, 6.
Each face is equally probable. (Each face is as likely as any other to turn up on a roll of the die.)
So,
The probability of rolling a 4 is 1/6.

This argument is valid. You already know this. Even if you have never studied probability, you make probabilities add up to 1.

You intuitively know that when the events are *mutually exclusive*—the die can land only one face up on any roll—and *exhaustive*—the die must land with one of the six faces up—then the probabilities add up to 1.

Why is the argument valid? Given the basic laws of probability, whenever the premises of an argument of this form are true, then the conclusion must be true too.

Here is another valid argument about probability.

This die has six faces, labeled 1, 2, 3, 4, 5, 6.
Each face is equally probable.
So:
The probability of rolling a 3 or a 4 is 1/3.

Even if you have never studied probability, you know that probabilities add up. If two events are *mutually exclusive*—one or the other can happen, but not both at the same time—then the probability that one or the other happens is the sum of their probabilities.

Given the basic laws of probability, whenever the premises of an argument of this form are true, then the conclusion must be true too. So the argument is valid.

The two arguments just stated are both valid. Notice how they differ from this one:

This die has six faces, labeled 1, 2, 3, 4, 5, 6.
In a sequence of 227 rolls, a 4 was rolled just 38 times.
So:
The probability of rolling a 4 with this die is about 1/6.

That is a risky argument. The conclusion might be false, even with true premises. The die might be somewhat biased against 4. The probability of rolling a 4 might be 1/8. Yet, by chance, in the last 227 rolls we managed to roll 4 almost exactly 1/6 of the time.

ANOTHER KIND OF RISKY ARGUMENT

Probability is a fundamental tool for inductive logic. But we have just seen that:

- There are also deductively valid arguments about probability.

Likewise:

- Many kinds of risky argument need not involve probability.

There may be more to a risky argument than inductive logic. Inductive logic does study risky arguments—but maybe not every kind of risky argument. Here is a new kind of risky argument. It begins with somebody noticing that:

It is very unusual in our university for most of the students in a large elementary class to get As. But in one class they did.

That is odd. It is something to be explained. One explanation is that the instructor is an easy marker.

Almost all the students in that class got As.
So:
The instructor must be a really easy marker.

Here we are *not* inferring from a sample to a population, or from a population to a sample.

We are offering a *hypothesis* to explain the observed facts. There might be other explanations. Almost all the students in that class got As,

So:
That was a very gifted class.

So:
The instructor is a marvelous teacher.
So:
The material in that course is far too easy for well-prepared students.

Each of these arguments ends with a *plausible explanation* of the curious fact that almost everyone in the class got an A grade.

Remember argument (*J) on page 2:

(*J) If James wants a job, then he will get a haircut tomorrow.
James will get a haircut tomorrow.
So:
James wants a job.

This is an invalid argument. It is still an argument, a risky argument. Let us have some more details. James gets his hair cut once in a blue moon. He is broke. You hear he is going to the barber tomorrow. Why on earth? Because he wants a job. The conclusion is a *plausible explanation*.

INFERENCE TO THE BEST EXPLANATION

Each of the arguments we've just looked at is an *inference to a plausible explanation.*

If one explanation is much more plausible than any other, it is an *inference to the best explanation.*

Many pieces of reasoning in science are like that. Some philosophers think that whenever we reach a theoretical conclusion, we are arguing to the best explanation. For example, cosmology was changed radically around 1967, when the Big Bang theory of the universe became widely accepted. The Big Bang theory says that our universe came into existence with a gigantic "explosion" at a definite date in the past. Why did people reach this amazing conclusion? Because two radio astronomers discovered that a certain low "background radiation" seems to be uniformly distributed everywhere in space that can be checked with a radio telescope. The best explanation, then and now, is that this background radiation is the result of a "Big Bang."

"ABDUCTION"

One philosopher who thought deeply about probability was Charles Sanders Peirce (1839–1914). Notice that it is spelled PEIrce. His name is not "Pierce." Worse still, his name is correctly pronounced "purse"! He came from an old New England family that spelled their name "Pers" or "Perse."

Peirce liked things to come in groups of three. He thought that there are three types of good argument: deduction, induction, and inference to the best explanation. Since he liked symmetries, he invented a new name for inference to the best explanation. He called it *abduction*. So his picture of logic is this:

	Deduction
Logic	Induction
	Abduction

Induction and abduction are, in his theory, two distinct types of risky argument.

Some philosophers believe that probability is a very useful tool in analyzing arguments to the best explanation. Other philosophers, like Peirce, do not think so. There is a debate about that. We leave that debate to philosophers of science. The issues are very interesting, but this book will not discuss inference to the best explanation.

TESTIMONY

Most of what you believe, you believe because someone told you so.

How reliable are your parents? Your psychology instructor? The evening news? Believing what they say involves risky arguments.

I know I was born on February 14, because my mother told me so.
So:
I was born on February 14.

My psychology instructor says that Freud was a fraud, and is a worthless guide to human psychology.
So:
Freud is a worthless guide to human psychology.

According to the evening news, the mayor is meeting with out-of-town officials to discuss the effect of the flood.
So:
The mayor is meeting with out-of-town officials to discuss the effect of the flood.

These are risky arguments. The evening news may be misinformed. Your psychology instructor may hate Freud, and be a very biased informant.

The argument about your birthday is the least risky. It is still risky. How do you know that your parents are telling the truth?

> You look at your birth certificate. You can't doubt that! Well, maybe your parents lied by a day, so they could benefit from a new law about child benefits that took effect the day after you were born. Or maybe you were born on Friday the thirteenth, and they thought it would be better if you thought you were born on Valentine's Day. Or maybe you were born on a taxi ride to the hospital, and in the excitement no one noticed whether you were born before or after midnight . . .

All the examples are arguments based on the *testimony* of someone else: your family, your instructor, the evening news.

Some kinds of testimony can be analyzed using probability, but there are a lot of problems. Inductive logic touches on testimony, but there is a lot more to testimony than probability.

In this book we will *not* discuss inference to the best explanation, and we will *not* discuss testimony. But if you really want to understand risky arguments, you should think about testimony, and inference to the best explanation. In this book we study only one side of probability.

ROUGH DEFINITION OF INDUCTIVE LOGIC

> Inductive logic analyzes risky arguments using probability ideas.

DECISION THEORY

There is a whole other side to reasoning: *decision*. We don't just reason about what to believe.

We reason about what to do.

The probability theory of practical reasoning is called *decision theory*, and it is very close to inductive logic.

We decide what to do on the basis of two ingredients:

- What we think will probably happen (*beliefs*).
- What we want (*values*).

Decision theory involves both probabilities and values. We measure values by what are called *utilities*.

ROUGH DEFINITION OF DECISION THEORY

> Decision theory analyzes risky decision -making using ideas of probability and utility.

EXERCISES

1 *Fees.* With a budgetary crisis, administrators at Memorial University state that they must either increase fees by 35% or increase class sizes and limit course offerings. Students are asked which option they prefer. There is a sharp difference of opinion.

Which of these risky arguments is from sample to population? From population to sample? From sample to sample?

(a) The student body as a whole is strongly opposed to a major fee increase.
65 students will be asked about the fee increase.
So:
Most of the 65 students will say that they oppose a major fee increase.

(b) A questionnaire was given to 40 students from all subjects and years.
32 said they were opposed to a major fee increase.
So:
Most students are opposed to a major fee increase.

(c) The student body as a whole is strongly opposed to a major fee increase.
So (probably):
The next student we ask will oppose a major fee increase.

(d) A questionnaire was a given to 40 students from all subjects and years.
32 said they were opposed to a major fee increase.
So (probably):
The next student we ask will oppose a major fee increase.

2 *More fees.* Which of these is an inference to a plausible explanation? Which is an inference based on testimony?

(a) The student body as a whole is strongly opposed to a major fee increase.
So:
They prefer to save money rather than get a quality education.

(b) The student body as a whole is strongly opposed to a major fee increase.
So:
Many students are so poor, and loans are so hard to get, that many students would have to drop out of school if fees went up.

(c) Duodecimal Research Corporation polled the students and found that 46% are living below the official government poverty line.
So:
The students at Memorial cannot afford a major fee increase.

3 Look back at the Odd Questions on pages xv–xvii. Each question will be discussed later on. But regardless of which answer is correct, we can see that any answer you give involves an argument.

3.1 *Boys and girls.* Someone argues:

About as many boys as girls are born in hospitals.
Many babies are born every week at City General.
In Cornwall, a country town, there is a small hospital where only a few babies are born every week.
An unusual week at a hospital is one where more than 55% of the babies are girls, or more than 55% are boys.
An unusual week occurred at either Cornwall or City General Hospital last week.
So:
The unusual week occurred at Cornwall Hospital.

Explain why this is a risky argument.

3.2 *Pia.* The premises are as stated in Odd Question 2.
Which is the riskier conclusion, given those premises?
(a) Pia is an active feminist.
(e) Pia is a bank teller and an active feminist who takes yoga classes.

3.3 *Lotteries.* Your aunt offers you as a present one of two free Lotto 6/49 tickets for next week's drawing. They are:

A. 1, 2, 3, 4, 5, and 6.
B. 39, 36, 32, 21, 14, and 3.

(a) Construct an argument for choosing (A). If you think it is stupid to prefer (A) over (B), then you can produce a bad or weak argument! But try to make it plausible.
(b) You decide to take (A). Is this a risky decision?

3.4 *Dice.*

Two dice are fair: each face falls as often as any other, and the number that falls uppermost with one die has no effect on the number that falls uppermost with the other die.

So:

It is more probable that 7 occurs on a throw of these two dice, than 6.

Is this a risky argument?

3.5 *Taxicabs.* Amos and Daniel are both jurors at a trial. They both hear the same information as evidence, namely the information stated in Odd Question 5. In the end, they have to make a judgment about what happened.

Amos concludes: So, the sideswiper was blue.
Daniel concludes: So, the sideswiper was green.

(a) Are these risky arguments?
(b) Could you think of them as risky decisions?

3.6 *Strep throat.* The physician has the information reported in Odd Question 6. She concludes that the results are worthless, and sends out for more tests. Explain why that is a risky decision.

4 *Ludwig van Beethoven.*
(a) What kind of argument is this? How good is it?

Beethoven was in tremendous pain during some of his most creative periods—pain produced by cirrhosis of the liver, chronic kidney stones (passing a stone is excruciatingly painful), and bouts of nonstop diarrhea.
Yet his compositions are profound and often joyous.
So:
He took both pain killers and alcohol, and these drugs produced states of elation when he did his composing.

(b) Give an example of a new piece of information which, when added to the premises, strengthens the argument.

Books on "critical thinking" teach you how to analyze real-life complicated arguments. Among other things, they teach you how to read, listen, and think critically about the things that people actually say and write. This is not a book for critical

thinking, but it is worth looking at a few real-life arguments. All are taken from a daily newspaper.

5 *The slender oarfish.*

A rare deep-sea creature, the slender oarfish, is helping Japanese scientists predict major earthquakes. In Japanese folklore, if an oarfish, which normally lives at depths of more than 200 meters, is landed in nets, then major tremors are not far behind.

Two slender oarfish were caught in fixed nets recently only days before a series of earthquakes shook Japan. This reminds us that one of these fish was caught two days before a major earthquake hit Nijima Island, near Tokyo, in 1963. Moreover, when shock waves hit Uwajima Bay in 1968, the same type of rare fish was caught.

The oarfish has a unique elongated shape, which could make it susceptible to underwater shock waves. It may be stunned and then float to the surface. Or the real reason could be that poisonous gases are released from the Earth's crust during seismic activity. At any rate, whenever an oarfish is netted, a geological upheaval is in progress or about to occur.

And, having just caught some slender oarfish, Japanese seismologists are afraid that another disaster is imminent.

(a) In the first paragraph, there is a statement based on testimony. What is it? On what testimony is it based?
(b) The third paragraph states one conclusion of the entire discussion. What is the conclusion?
(c) The second paragraph states some evidence for this conclusion. Would you say that the argument to the conclusion (b) is more like an argument from population to sample, or from sample to population?
(d) The third paragraph offers two plausible explanations for the facts stated in the second paragraph. What are they?
(e) There are several distinct arguments leading to the final conclusion in the fourth paragraph. Describe how the arguments fit together.

6 *Women engineers.*

Since 1986, only 11% of engineering school graduates have been women. That showing is particularly poor considering that in other formerly male-dominated fields there are signs of real progress. Some examples from 1986: law, 48%; commerce, 44%; medicine, 45%; and in the biological sciences, nearly 50% of the graduates are women.

(a) What is the conclusion? (b) What kind of argument is it? Valid? Inductive and risky? Inference to a plausible explanation?

7 *Plastic surgery.*

In her private counseling service for women, Martha Laurence, a professor of social work, tries to get behind the reasons women give for wanting plastic surgery. "Usually it is because they have a lack of confidence in who they are, the way they are," she said. "There is no simple answer, but the real problem is one of equity and of women's control over the self."

Her conclusion is that "the real problem is one of equity and of women's control over the self." What type of argument does she have for this conclusion?

8 *Manitoba marijuana.*

Basement operations are sprouting up in rural Manitoba to supply hydroponically grown marijuana for the Winnipeg market, police say. As Constable Duane Rhone of the rural Selkirk community of Winnipeg said in a recent interview, "It's cheap, it's easy to set up and there is a high return on investment. You can produce more marijuana of a better quality in a small amount of space," he said, adding that the necessary equipment is readily available. "It's become the thing to do. We've been seeing a lot more of this hydroponics marijuana in the last little while. There must be plenty more of these operators that we don't know about."

Conclusion: There are many as-yet undiscovered marijuana growers in rural Manitoba.

What kind of argument is Constable Rhone offering?

KEY WORDS FOR REVIEW

Population	Sample
Inference to the best explanation	Testimony
Inductive logic	Decision Theory

3 The Gambler's Fallacy

Most of the main ideas about probability come up right at the beginning. Two major ones are **independence** and **randomness**. Even more important for clear thinking is the notion of a **probability model**.

ROULETTE

A gambler is betting on what he thinks is a *fair* roulette wheel. The wheel is divided into 38 segments, of which:

- 18 segments are black.
- 18 segments are red.
- 2 segments are green, and marked with zeroes.

If you bet \$10 on red, and the wheel stops at red, you win \$20. Likewise if you bet \$10 on black and it stops at black, you win \$20. Otherwise you lose. The house always wins when the wheel stops at zero.

Now imagine that there has been a long run—a dozen spins—in which the wheel stopped at black. The gambler decides to bet on red, because he thinks:

> The wheel must come up red soon.
> This wheel is fair, so it stops on red as often as it stops on black.
> Since it has not stopped on red recently, it must stop there soon. I'll bet on red.

The argument is a risky one. The conclusion is, "The wheel must stop on red in the next few spins." The argument leads to a risky decision. The gambler decides to bet on red. There you have it, an argument and a decision. Do you agree with the gambler?

Since this chapter is called "the gambler's fallacy" there must be something wrong with the gambler's argument. Can you say what?

We will spend some time explaining one way to talk about the argument—and about probability in general. Do not expect formal definitions yet. Try to get some fairly clear ideas.

FAIR

Arguments have premises. The gambler's main premise was that the roulette wheel is fair. Fair? What is fair? The word has quite a few meanings.

A judge may be fair.
A fair wage.
A fair settlement.
A fair grade in this course.
A fair game.
A fair coin.
One greedy child cuts a cake in half; another greedy child chooses which half to take. "That's fair."
"Affirmative action" to help minorities or women in the workplace: is that fair?

What is the opposite of fair? Something is unfair if it favors one party over another. A judge who is fair is not biased in favor of one party or the other. We use the same word—*biased*—for gambling devices. A gambling setup like a coin or a roulette wheel is unfair if it is biased.

BIASED

If a coin tends to come up heads more often than tails, it is biased.

If spins of a roulette wheel tend more often to be red than black, the wheel is biased.

A biased coin tends to come up heads more often than tails, or vice versa. What do we mean by "tends"? That is a hard question. The coin comes up heads more often than tails. Always? In any sequence of tosses with this setup? No, it must be "on average" or "in the long run." And what does that mean? In the long run we're all dead.

Yet we do seem to have a rough intuitive idea of averages in the long run. We have to start somewhere. We think of these setups as somehow involving "chance." Call them *chance setups*. We can in principle make repeated *trials* on a chance setup: tosses, spins, draws, samples. Trials on a setup have a definite set of possible *outcomes*:

- With a coin: heads, tails.
- With a die: 1, 2, 3, 4, 5, 6.
- With a roulette wheel: each of the 38 segments.

We have the idea of how frequently different outcomes occur on repeated trials. If each outcome occurs as often as every other, we say the setup is unbiased.

A chance setup is unbiased if and only if the relative frequency in the long run of each outcome is equal to that of any other.

Since this is a philosophy book, we will come back and worry about what we mean by "long run," "tends," and "relative frequency." We do seem to have some intuitions about what these ideas mean, and we start with those intuitions. Later on, philosophical analysis tries to get clearer about what they mean.

INDEPENDENCE

There are many ways to be unfair. A coin tossing device is "unfair" if it regularly gives heads more often than tails. It is biased. But that is not the only way to be unfair.

You can very easily learn to toss a coin, flipping it with your thumb and catching it on your wrist, so that it almost always appears heads if you last tossed tails, and tails if it last fell heads. (You may not believe this, but practice for five minutes and then you can amaze your friends. You are on your way to becoming a magician.) When you get the knack, heads and tails come up equally often:

H T H T H T H T H T H T

Your system is unbiased. But would you count that as "fair" tossing? No. Something is fishy about this kind of tossing. A gambler would have a field day. He does not have to wait until he sees 12 tails in a row (he never will). He sees one tails—and then bets heads. He is sure to win!

Hence lack of bias in the system does not guarantee that a chance setup is fair. We need more than that. The idea of a fair tossing device seems to involve there being *no regularity* in the outcomes. Or they should be *random*. Randomness is a very hard idea. Outcomes from a chance setup are random (we think) if outcomes are not *influenced* by the outcomes of previous trials.

The setup should not have a *"memory."* A fair setup does not know, at any trial, what happened on previous trials.

There are even more ways to think about this. For example, if a gambler knew that two heads in a row are usually followed by tails, then he could place bets so as to make a profit. But trials on that setup would not be independent. Randomness is sometimes defined as: *no successful gambling system is possible.*

The idea of *complexity* is also used. Random sequences are so complex that we cannot predict them. The complexity of a sequence can be measured by the length of the shortest computer program needed to generate the sequence. A sequence is called random, relative to some computational system, if the shortest program needed to generate it is as long as the sequence itself.

Here we have a family of related ideas:

random	no influence from previous trials
no regularity	no memory of previous trials
complexity	impossibility of a gambling system

Some students like the metaphor of "no memory." Others focus on randomness. People happy with computers like complexity. Each of these ideas gets at a central notion. We say that the outcome of each trial should be *independent* of the outcome of other trials on the setup. A more precise account of independence is given in chapter 6. For the present this will suffice:

> Trials on a chance setup are independent if and only if the probabilities of the outcomes of a trial are not influenced by the outcomes of previous trials.

This is not a definition. It is an explanation of an idea. A chance setup is fair if and only if:

- It is unbiased, and
- outcomes are independent of each other.

TWO WAYS TO BE UNFAIR

Hence a chance setup can be "unfair" in two different ways. It can be biased. For example, heads tends to come up more often than tails. But there could also be some regularity in the sequence of outcomes. Trials may not be independent of each other. Since there are two ways to be unfair, there are four possible combinations:

Fair: unbiased, independent.
Unfair: unbiased, not independent.
Unfair: biased, independent.
Unfair: biased, not independent.

Here are examples of each combination.

UNBIASED AND INDEPENDENT

A favorite model for probability is a big container of balls, an urn, from which we draw balls "at random." Imagine an urn with 50 balls in it, numbered 1 to 50, but otherwise as round and smooth and as similar as could be—the same circumference, the same weight. There is lots of room to spare. A trial consists of shaking the urn very well, drawing a ball, noting the number, and putting it back. This is called *sampling with replacement.* We expect each number to be drawn as often as every other. We imagine that the draws are *unbiased* and *independent.*

BIASED BONES, INDEPENDENT OUTCOMES

People gambled a long time ago. In the beginning they did not have dice. They used bones. The heelbones of a running animal like a deer or a horse can land in only four ways. They are natural randomizers. These bones are sometimes called "knucklebones." They are ancestors of our dice. Some gamblers still talk about "rolling the bones."

Of course, every knucklebone is different. One class tossed a knucklebone from Turkey, about 6,000 years old. Three outcomes were labeled with colored spots: red, black, blue. The fourth outcome was left unmarked.

The bone was tossed 300 times. Here is a summary of results:

unmarked: 110	blue: 88
red: 50	black: 52

Or in percentages, rounding off:

unmarked: 37%	blue: 29%
red: 17%	black: 17%

Although red and black seem to occur almost equally often, unmarked comes up more often than red and black combined. These bones are clearly *biased.*

But we could not detect any regularity in the rolls. It did not look as if the outcome of any roll depended on previous rolls. Rolls of this bone seem to be *independent.*

UNBIASED DRAWS, DEPENDENT OUTCOMES

Imagine an urn with an equal number of red and green balls. At the first draw, we might think that there will be no bias for red or green. But suppose we sample *without replacement.* That is, once a ball is drawn, we draw another without replacing, until the urn is empty; then, if we want, we restore all the balls to the urn and start again. If we draw a green ball the first time, then at the next draw there is one more red ball in the urn than there are green balls. So we would expect a better chance of getting a red ball than a green one, the next time around.

Hence the result of the second draw is *not* independent of the result of the first draw. Yet overall, green and red balls are drawn equally often. The setup is *unbiased* (red and green come up equally often), but trials are influenced by previous outcomes. They are *not independent.*

BIAS AND DEPENDENCE

Now imagine that 90% of the balls are red, and 10% are green. We sample without replacement. Then there is a *bias* in favor of red. And trials are *not independent.*

THE GAMBLER'S FALLACY

We have just seen that a chance setup can be unfair in two different ways. It can be biased. And trials may not be independent. The difference matters. *Fallacious Gambler* said:

> I think this is a fair roulette wheel.
> I have just seen twelve black spins in a row.
> Since the wheel is fair, black and red come up equally often.

Hence:

> Red has to come up pretty soon.
> I'd better start betting red.
> Maybe red won't come up the very next time, but a lot of reds have to come up soon.

This is called "the gambler's fallacy." We commit many fallacies in inductive reasoning.

What is the gambler's fallacy? The fallacy does not involve *bias*. It involves *independence.*

The gambler thinks that a sequence of twelve blacks makes it more likely that the wheel will stop at red next time. *If so, a past sequence affects future outcomes.* So trials on the device would not be independent, and the device would not be fair after all.

Thus the gambler is being *inconsistent.* His premises are:

- The setup is fair, and
- there have been twelve black spins in a row.

He infers:

- Some reds must turn up soon.

That conclusion would follow only if outcomes from the setup were not independent. That would be inconsistent with the gambler's first premise.

IMPOSSIBILITY OF A SUCCESSFUL GAMBLING SYSTEM

There are many ways to think about randomness and independence. One definition says that *outcomes are random if and only if a successful gambling system is impossible.* That does not mean that you can't win. Someone has to win. A gambling system is impossible if no betting system is guaranteed to win.

For example, suppose a coin-tossing setup has a "memory." It is made so that every time there has been heads followed by tails, the coin falls tails. Every time there has been tails followed by heads, the coin falls heads. It can never produce the sequences HTH or THT. So you might see this sequence:

H T T T T H H T T H H T

But not this one:

H T T T T H T T T H H T

If the setup never allows THT or HTH, a very profitable gambling system is possible: when you see HT, bet on T next; when you see TH, bet on H next; otherwise don't bet. So for the first sequence above you bet like this:

HT (bet T) TTTH (bet H) HT (bet T) TH (bet H) HT

You win every bet you make.

Our fallacious gambler dreams that a profitable gambling system is possible. His system is, in part, "When you see 12 blacks in a row, bet red." This system would work only if the spins of the roulette wheel were not independent. And, of course, they may not be! But if the gambler's premise is that the roulette wheel is fair, then he should think the spins are independent of each other. So his fallacy is thinking both that the roulette wheel is fair, and that a gambling system is possible.

COMPOUND OUTCOMES

There is yet another way to think of a fair setup. A coin is unbiased if on average the two outcomes heads and tails come up equally often. But we can also think of all outcomes of two trials, namely,

HH HT TH TT

If trials are independent, then each of these four compound outcomes will occur on average equally often. Likewise, each sequence of 13 outcomes from the roulette wheel, made up of just B and R, will come up equally often. Thus on average,

B B B B B B B B B B B B R

occurs neither more nor less frequently than

B B B B B B B B B B B B B.

Half the time a sequence of 12 blacks is followed by red. Half the time a sequence of 12 blacks is followed by black.

ODD QUESTION 3

> In Lotto 6/49, a standard government-run lottery, you choose 6 out of 49 numbers (1 through 49). You win the biggest prize—maybe millions of dollars—if these 6 are drawn. (The prize money is divided between all those who choose the lucky numbers. If no one wins, then most of the prize money is put back into next week's lottery.)

Suppose your aunt offers you, *free*, a choice between two tickets in the lottery, with numbers as shown:

A. You win if 1, 2, 3, 4, 5, and 6 are drawn.
B. You win if 39, 36, 32, 21, 14, and 3 are drawn.

Do you prefer A, B, or are you indifferent between the two?

If the lottery is fair, then any sequence of outcomes will be as probable as any other. Some people might prefer ticket A because it is simpler to check the ticket to see if you have won. That's practical.

Some people might prefer ticket B because they have a kid brother who is fourteen years old and a little sister who is three. That is superstition.

If you choose your tickets *simply by the probability*, then you should have no preference between A and B.

But! There might be a real advantage in choosing A over B!

The very large prizes are split between all the people who chose the lucky numbers that win. It may be that most people like irregular-looking outcomes like B. They cannot believe that a regular sequence like A would occur. So fewer people choose sequence A than choose sequence B. Hence, if sequence A is drawn, the prize may be bigger, for each winner, than if sequence B is drawn.

But! Maybe enough people know this so that they try to outwit the herd, and become a little herd themselves. When government lotteries were finally introduced into Great Britain late in the twentieth century, a lot of people chose 1, 2, 3, 4, 5, 6 because they thought that no one else would. If it had come up (it did not) the payoff would have been quite small, because the prize would have been split among so many players.

ODD QUESTION 7

> "Imitate" a coin. That is, write down a sequence of 100 H (for heads) and T (for tails) without tossing a coin—but a sequence that you think will fool everyone into thinking it is the report of tossing a fair coin.

That sounds easy, and it is. But most people try to build in *too much* irregularity to make the sequence look random. It is the same instinct that inclines most of us to think, for a moment, that B is a better choice for a lottery ticket than A.

A *run* is a sequence of identical outcomes—like the run of 12 blacks at roulette. Most people think that a sequence of 100 tosses will fluctuate a good deal between heads and tails.

Hardly anyone making up a sequence of 10 tosses puts in a run of seven heads in a row. It is true that the chance of getting 7 heads in a row, with a fair coin, is only 1/64 (½ × ½ × ½ × ½ × ½ × ½ × ½). But in tossing a coin 100 times, you have at least 93 chances to start tossing 7 heads in a row, because each of the first 93 tosses could begin a run of 7.

It is more probable than not, in 100 tosses, that you will get 7 heads in a row. It is certainly more probable than not, that you will get at least 6 heads in a row. Yet almost no one writes down a pretend sequence, in which there are even 6 heads in a row.

This example may help some people with the gambler's fallacy. We have this feeling that if there have been 12 blacks in a row, then the roulette wheel had better hurry up and stop at a red! Not so. There is, so to speak, all the time in the world for things to even out.

THE ALERT LEARNER

We began with a person we call *Fallacious Gambler*. He meets *Stodgy Logic*. *Stodgy Logic* says to *Fallacious Gambler*:

> Your premise was that the setup is fair. But now you reason as if the setup is not fair! You think the only ingredient in unfairness is bias. You forgot about independence. That's your fallacy.

They meet a third party, *Alert Learner*. She reasons:

> We've been spinning black with this wheel all too frequently. I thought that the wheel was unbiased. But that must be wrong. The wheel must be biased toward black! So I'll start betting on black.

She has made an inductive argument: to the risky conclusion that the wheel is biased. *Alert Learner* bets black. *Fallacious Gambler* bets red. *Stodgy Logic* says there is no point in betting either way. Who is right?

Alert Learner might be right. Maybe the wheel *is* biased. Pure logic cannot tell you. You have to know more about this casino, how it gets its roulette wheels, and so on. In a real-life casino we would guess that the wheel is not loaded for black. There would be no point. People would catch on quickly, and all the alert learners in this world would be getting rich.

I expect that *Alert Learner* is wrong. I think so because I am worldly-wise. My belief that she is wrong has nothing to do with inductive logic. It depends on what I know about casinos, wheels, and gamblers. I may be wrong. Another risky argument.

RISKY AIRPLANES

Vincent and Gina have been visiting their grandparents and have to fly home. Two companies fly their route, Alpha Air and Gamma Goways. Gamma has just crashed one of its planes.

> VINCENT: Let's take Gamma—they crash only one flight in a million, and they've just had their crash!
> GINA: Don't be crazy. This accident shows that Gamma is negligent; we had better take Alpha.

Who is right? Gina is thinking like *Alert Learner*. But her grandmother may say: "Gamma will be extra careful and check all its planes for just the fault that caused the crash. So Gamma is the one to fly." Gina's boyfriend says: "No way! Gamma has a long record of safety violations. Moreover, its pilots are poorly trained. Their planes are old, cheap, and poorly maintained. Their cabin staff are underpaid. There is no way they can fix things in a week—or in a year."

Vincent, Gina, grandma, and the boyfriend all argue differently. Each brings in different premises. A premise may be false. Logic cannot tell which premises are true. Logic can only tell when premises are good reasons for a conclusion. All logic can tell us is that *if* Vincent is reasoning like the fallacious gambler, *then* he is wrong. And it seems that he is.

MODELS

In reasoning using probability, we often turn to *simple and artificial models* of complex situations. Real life is almost always complicated, except when we have deliberately made an apparatus and rules—as in a gambling situation. People made the roulette wheel to have some symmetries, so that each segment turns up as often as every other. Real life is not like that. But artificial games may still be useful *probability models* that help us to think about the real world.

A basic strategy of probability thinking: make simple and artificial models. Compare the models to a real-life situation. We can apply precise and often mathematical or logical concepts to the models. That is why they help us to think clearly. But we also have to see how applicable they are to a real-life problem.

For example: it is very hard to find a place to park on my street. The fine for illegal parking is $20. My guest wants to park illegally. How likely is it that she will get a ticket this evening? The parking inspector comes around only one evening a week. So I make a model using a lottery, in which one red card and six green cards are put in a hat, and one is drawn at random. I compare a red card to getting a fine, and a green one to a free parking place.

Another guest says: "What *is* the right model of this situation? I got a ticket parking outside last night." We believe the inspectors almost never come around two days running. Now we make a new model of the situation, in which trials are not independent.

The question for Vincent and Gina is: what is the right model of the situation with Gamma Goways?

TWO WAYS TO GO WRONG WITH A MODEL

Serious thinking about risks, which uses probability models, can go wrong in two very different ways.

- The model may not represent reality well. That is a mistake about the real world.
- We can draw wrong conclusions from the model. That is a logical error.

Alert Learner said, "Bet on black." This is because she thought we were using the wrong model, and were making a mistake about the real world.

Fallacious Gambler said, "Bet on red." He made a logical error. He forgot about independence.

These two kinds of error recall the two ways to criticize an argument.

- Challenge the premises—show that at least one is false.
- Challenge the reasoning—show that the premises are not a good reason for the conclusion.

Criticizing the model is like challenging the premises. Criticizing the analysis of the model is like challenging the reasoning.

EXERCISES

1 *Roulette wheels.* At North American casinos, roulette wheels have two "zeros" where the house collects all. In Europe the rules are similar, but there is only one zero.
(a) If you had to bet, would you rather play roulette in Europe or North America?
(b) Does the difference in odds make any difference to the gambler's fallacy?

2 *Shuffling.* In card games such as bridge, cards are shuffled before each deal. The deal usually has a "memory" of previous games, because shuffling is imperfect. After a deal, cards stay in an order that still reflects the order present before the deal. Even professional dealers are not perfect. A pack must be shuffled at least seven times by a professional to eliminate all traces of its previous order.
Suppose some good friends are playing, and they shuffle the pack only twice before each deal.
(a) In repeated playing, are the hands that are dealt biased?
(b) Are they independent?
(c) Is the setup fair, according to our definition of a fair setup?

3 *Lotto.* Two mathematicians from Stanford University analyzed data from the Canadian Lotto 6/49. Using a model of choices people made, they calculated that the least popular ticket was 20, 30, 39, 40, 41, 48. If you are going to buy a Lotto ticket, would that be a good number to choose?

4 *Numerals.* We can produce a sequence of numerals by rolling a die. The *trials* consist of rolling the die. The possible *outcomes* are 1 through 6. What about the numerals produced by the following setups? Are these setups unbiased? Are trials independent?

(a) *Birth weights.* Newborn babies at a large hospital are weighed to the nearest gram. Take just the last digit in the birth weight in grams to generate a sequence of numerals from 0 through 9. A trial consists of a birth, weighing, and recording. The possible outcomes are 0 through 9. What if the babies were weighed only to the nearest pound, and the last digit were used?

(b) *Telephone poles.* A telephone line runs alongside a long straight road in the prairies. The poles are numbered consecutively. Trials, in order, consist of noting the last digit on successive poles.

(c) *Books.* A book has exactly 600 pages, numbered 1–600. You open the pages at random. Each time, you note the last digit of the left-hand page.

(d) *Cruise ships.* The captain on a cruise ship reports, at noon each day, the number of nautical miles the ship has sailed in the past twenty-four hours. Each day, passengers bet on the last digit in the distance. Twenty percent of the stakes go to charity. The rest is divided among those who bet on the right digit. (Might you have any reason for betting on one digit rather than another?)

5 *Fallacious Gambler strikes back.* "I've been reading this old book, *A Treatise on Probability*, published in 1921 by the famous economist John Maynard Keynes. He says that a German psychologist, Dr. Marbe, studied 80,000 spins of roulette at Monte Carlo, and found that long runs do not occur nearly as often as theory predicts. A run of 13 blacks is more improbable than anyone believes, so having seen 12 blacks in a row, I am going to bet red!" (a) Is our friend still a fallacious gambler? (b) How would you explain Dr. Marbe's data?

6 *Counting.* In varieties of poker known as blackjack, "twenty-one," etc., when played at a casino, the odds favor the house. But not by much. Hands of blackjack used to be dealt until the pack was exhausted. That was like sampling *without replacement*. Explain why a *gambling system* should be possible when a game is played according to these rules.

7 *The American draft.* During the Vietnam War, there was much criticism of the way in which young American men were drafted into the army. The poorer you were, the more likely you were to get conscripted. Blacks and Hispanics were more likely to be drafted than other Americans. So the old system was replaced by a lottery: 365 balls were placed in a large rotating urn. The balls were labeled, one for each day of the year (an extra ball was added for leap years).

The draw was done in public, on television. To show the public that every ball went in, no cheating, the balls were put in the urn in order, starting with 365 (it was not a leap year) and going down to 1. Pretty young women then reached into the urn and drew one ball at a time.

The first men to be drafted were those whose birthdays were drawn early in the draw. The army drafted men who had the first birthday drawn, then the second, and so on, until it had as many soldiers as it wanted.

The first few balls that were drawn did seem to come from anywhere in the year, but after a while, there was a pronounced tendency for the balls to be from the early months. So if you were born in February, you had a good chance of being drafted. If you were born in December, you had little chance of being drafted.

(a) Was this sampling with, or without, replacement?

(b) Do you think the draws were biased?

(c) If so, the set-up was not "fair" in the sense of this chapter. Was it unfair to the men of draft age?

(d) Suggest at least two explanations of what was happening.

8 *X rays* are potentially harmful. They may damage a cell that then becomes cancerous.

Sara says to her dentist: "Please do not take any X rays. I had a terrible skiing accident three years ago and broke many leg bones, and so had many X rays then. I don't want any more, because that would make cancer even more likely."

Dentist: "Nonsense. Your past X rays make no difference to the danger of future X rays. To think otherwise is to commit the gambler's fallacy!"

Continue the argument.

9 *A dry August.* A man started a farm in arid but fertile land. Weather records show that during each month from March through August there is a fifty-fifty chance of one good rainfall, enough to let him irrigate for the whole season. A completely dry growing season occurs less than once in sixty years.

It is now mid-August. There has been no rain since a big storm at the end of last year. The farmer has run out of water. His crops are dying. He is optimistic, because he thinks:

> It will almost certainly rain soon. The statistics prove it!

Has he committed the gambler's fallacy? Are there any errors in the argument which are different from the gambler's fallacy?

10 *The inverse gambler's fallacy.* Albert enjoys gambling with dice.

(a) If Albert rolls four fair dice simultaneously, what is the probability that on a single roll he gets 6, 6, 6, 6?

(b) Trapper: "Albert was rolling four dice last night, and got four sixes." Nelson: "I bet he rolled the dice many times to get four sixes!" Is Nelson's conclusion reasonable?

(c) Lucie visits Albert. As she enters, he rolls four dice and he shouts "Hooray!" for he has just rolled four sixes. Lucie: "I bet you've been rolling the dice for a long time tonight to get that result!" Now Lucie may have many reasons for saying this—perhaps Albert is a lunatic dice-roller. But simply on the evidence that he has just rolled four sixes, is her conclusion reasonable?

11 *Lucky aunt.* (a) "Did you know your aunt won a medium prize in the lottery last year?"—"Oh, I suppose she played every week." Is that a good inference? (b) The lottery numbers have just been announced on TV. Phone rings, it is your aunt. "I just won the lottery!"—"Oh, I suppose you've been playing every week." Is that a good inference?

12 *The argument from design.* There is a famous argument for the existence of a Creator. If you found a watch in the middle of a desert, where "no man has been before," you would still infer that this well-designed timekeeping instrument had been made by a craftsman or factory. Here we are in the midst of an extraordinarily organized and well-designed universe, where there is an extraordinary adjustment of causes, and of means to ends. Therefore, our universe too must have a Creator.

A common objection: The emergence of a well-organized universe just by chance would be astonishing. But if we imagine that matter in motion has been around, if not forever, at least for a very long time, then of course sooner or later we would arrive at a well-organized universe, just by chance. Hence the argument from design is defective.

Is this a sound objection?

KEY WORDS FOR REVIEW

Relative frequency	Gambling system
Chance setup	Complexity
Biased/unbiased	Independence
Random	Probability model

4 Elementary Probability Ideas

This chapter explains the usual notation for talking about probability, and then reminds you how to add and multiply with probabilities.

WHAT HAS A PROBABILITY?

Suppose you want to take out car insurance. The insurance company will want to know your age, sex, driving experience, make of car, and so forth. They do so because they have a question in mind:

What is the probability that you will have an automobile accident next year?

That asks about a *proposition* (statement, assertion, conjecture, etc.):

"You will have an automobile accident next year."

The company wants to know: *What is the probability that this proposition is true?*

The insurers could ask the same question in a different way:

What is the probability of your having an automobile accident next year?

This asks about an *event* (something of a certain sort happening). Will there be

"an automobile accident next year, in which you are driving one of the cars involved"?

The company wants to know: *What is the probability of this event occurring?*

Obviously these are two different ways of asking the same question.

PROPOSITIONS AND EVENTS

Logicians are interested in arguments from premises to conclusions. Premises and conclusions are propositions. So inductive logic textbooks usually talk about the probability of propositions.

Most statisticians and most textbooks of probability talk about the probability of events.

So there are two languages of probability, propositions and events.

Propositions are true or false.
Events occur or do not occur.

Most of what we say in terms of propositions can be translated into event-language, and most of what we say in terms of events can be translated into proposition-language.

To begin with, we will sometimes talk one way, sometimes the other.

The distinction between propositions and events is not an important one now. It matters only in the second half of this book.

WHY LEARN TWO LANGUAGES WHEN ONE WILL DO?

Because some students will already talk the event language, and others will talk the proposition language.

Because some students will go on to learn more statistics, and talk the event language. Other students will follow logic, and talk the proposition language.

The important thing is to be able to understand anyone who has something useful to say.

There is a general moral here. Be very careful and very clear about what you say. But do not be dogmatic about your own language. Be prepared to express any careful thought in the language your audience will understand. And be prepared to learn from someone who talks a language with which you are not familiar.

NOTATION: LOGIC

Propositions or events are represented by capital letters: A, B, C . . .

Logical compounds will be represented as follows, no matter whether we have propositions or events in mind:

Disjunction (or): AvB for (A, or B, or both). We read this "A or B."

Conjunction (and): A&B for (A and B).

Negation (not): ~A for (not A).

Example in the proposition language:

Let Z be the proposition that the roulette wheel stops at a zero. Let B be the proposition that the wheel stops at black.

ZvB is the proposition that the wheel stops at a zero or black.

Example:

Let Z: the wheel stops at a zero.
Let B: the wheel stops at black.

Then: ZvB = one or the other of those events occurs = the wheel stops at black or a zero.

Let R: the wheel stops at red.

In roulette, the wheel stops at a zero or black (ZvB) if and only if it does not stop at red (~R). So,

~R is equivalent to (ZvB).
R is equivalent to ~(ZvB).

NOTATION: SETS

Statisticians usually do not talk about propositions. They talk about events in terms of set theory. Here is a rough translation of proposition language into event language.

The **disjunction** of two propositions, AvB, corresponds to the **union** of two sets of events, A∪B.

The **conjunction** of two propositions, A&B, corresponds to the **intersection** of two sets of events, A∩B.

The **negation** of a proposition, ~A, corresponds to the **complement** of a set of events, often written A'.

NOTATION: PROBABILITY

In courses on probability and statistics, textbooks usually write P( ) for probability. But our notation for probability will be:

Pr ( ).

In the roulette example (Z for zero, R for red, B for black), all these are probabilities:

Pr(Z) Pr(ZvB) Pr(~(ZvB)) Pr(R)

Earlier on this page we said that ~R is equivalent to (ZvB). So:

Pr(~R) = Pr(ZvB).

That is, the probability of not stopping at a red segment is the same as the probability of stopping at a zero or a black segment.

TWO CONVENTIONS

All of us—whether we were ever taught any probability theory or not—have got into the habit of expressing probabilities by percentages or fractions. That is:

> Probabilities lie between 0 and 1.

In symbols, for any A,

$$0 \leq \Pr(A) \leq 1$$

At the extremes we have 0 and 1.

In the language of propositions, what is certainly true has probability 1.

In the language of events, what is bound to happen has probability 1.

In probability textbooks, the sure event or a proposition that is certainly true is often represented by the last letter in the Greek alphabet, omega, written as a capital letter: Ω. So our convention is written:

$$\Pr(\Omega) = 1$$

> The probability of a proposition that is certainly true, or of an event that is sure to happen, is 1.

MUTUALLY EXCLUSIVE

Two propositions are called *mutually exclusive* if they can't both be true at once. An ordinary roulette wheel cannot both stop at a zero (the house wins) and, on the same spin, stop at red. Hence these two propositions cannot both be true. They are mutually exclusive:

The wheel will stop at a zero on the next spin.
The wheel will stop at red on the next spin.

Likewise, two events which cannot both occur at once are called *mutually exclusive*. They are also called *disjoint*.

ADDING PROBABILITIES

There are some things about probability that "everybody" in college seems to know. For the moment we will just use this common knowledge. "Everybody" knows how to add probabilities.

More carefully: *the probabilities of mutually exclusive propositions or events add up.*

If A and B are mutually exclusive, $\Pr(A\text{v}B) = \Pr(A) + \Pr(B)$.

Thus if the probability of zero, in roulette, is 1/19, and the probability of red is 9/19, the probability that one or the other happens is:

$$\Pr(Z\text{v}R) = \Pr(Z) + \Pr(R) = 1/19 + 9/19 = 10/19.$$

Example: Take a *fair* die.
Let: E = the die falls with an even number up.

E = (the die falls 2, 4, or 6). $\Pr(E) = ½$

Why? Because $\Pr(2) = 1/6$. $\Pr(4) = 1/6$. $\Pr(6) = 1/6$.

The events 2, 4, and 6 are mutually exclusive.
Add $(1/6) + (1/6) + (1/6)$. You get ½.

People who roll dice call the one-spot on a die the *ace*.
Let M = the die falls either ace up, or with a prime number up.

M = (the die falls 1, 2, 3, or 5). $\Pr(M) = 4/6 = 2/3$

But you cannot add Pr(E) to Pr(M) to get

** $\Pr(E\text{v}M) = 7/6$. (*WRONG*)

(We already know probabilities lie between 0 and 1, so 7/6 is impossible.)

Why can't we add them up? Because E and M *overlap*: 2 is in both E and M. E and M are not mutually exclusive.

In fact, (EvM) = (the die falls 1, 2, 3, 4, 5, or 6), so that,

$$\Pr(E\text{v}M) = 1.$$

You cannot add if the events or propositions "overlap."

> Adding probabilities is for mutually exclusive events or propositions.

INDEPENDENCE

Intuitively:

Two events are independent when the occurrence of one does not influence the probability of the occurrence of the other.

Two propositions are independent when the truth of one does not make the truth of the other any more or less probable.

Many people—like *Fallacious Gambler*—don't understand independence very well. All the same, "everybody" seems to know that probabilities can be multiplied. More carefully: *the probabilities of independent events or propositions can be multiplied.*

MULTIPLYING

If A and B are independent, Pr(A&B) = Pr(A) × Pr(B).

We are rolling two fair dice. The outcome of tossing one die is independent of the outcome of tossing the other, so the probability of getting

a five on the first toss ($Five_1$)
and a six on the second toss (Six_2) is:

$Pr(Five_1 \& Six_2) = Pr(Five_1) \times Pr(Six_2) = 1/6 \times 1/6 = 1/36.$

Independence matters! Here is a mistake:

The probability of getting an even number (E) with a fair die is 1/2.
We found that the probability of M, of getting either an ace or a prime number, is 2/3.
What is the probability that on a single toss a die comes up both E and M?
We cannot reason:

** Pr(E&M) = Pr(E) × Pr(M) = 1/2 × 2/3 = 1/3. (*WRONG*)

The two events are not independent. In fact, only one outcome is both even and prime, namely 2. Hence:

Pr(E&M) = Pr(2) = 1/6.

Sometimes the fallacy is not obvious. Suppose you decide that the probability of the Toronto Blue Jays playing in the next World Series is 0.3 [Pr(J)], and that the probability of the Los Angeles Dodgers playing in the next world series is 0.4 [Pr(D)]. You cannot conclude that

** Pr(D&J) = Pr(D) × Pr(J) = 0.4 × 0.3 = 0.12. (*WRONG*)

This is because the two events may not be independent. Maybe they are. But maybe, because of various player trades and so on, the Dodgers will do well only if they trade some players with the Jays, in which case the Jays won't do so well.

Multiplying probabilities is for independent events or propositions.

SIXES AND SEVENS: ODD QUESTION 4

Odd Question 4 went like this:

To throw a total of 7 with a pair of dice, you have to get a 1 and a 6, or a 2 and a 5, or a 3 and a 4.
To throw a total of 6 with a pair of dice, you have to get a 1 and a 5, or a 2 and a 4, or a 3 and another 3.
With two fair dice, you would expect:
______(a) To throw 7 more frequently than 6.
______(b) To throw 6 more frequently than 7.
______(c) To throw 6 and 7 equally often.

Many people think that 6 and 7 are equally probable. In fact, 7 is more probable than 6.

Look closely at what can happen in one roll of two dice, *X* and *Y*. We assume tosses are independent. There are 36 possible outcomes. In this table, (3,5), for example, means that X fell 3, while Y fell 5.

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Circle the outcomes that add up to 6. How many?
Put a square around the outcomes that add up to 7. How many?
We can get a sum of seven in six ways: (1,6) or (2,5) or (3,4) or (4,3) or (5,2) or (6,1).
Each of these outcomes has probability 1/36. (*Independent tosses*) So,

Pr(7 with 2 dice) = 6/36 = 1/6. (*Mutually exclusive outcomes*)

But we can get a sum of six in only five ways: (1,5) or (2,4) or (3,3) or (4,2) or (5,1). So,

Pr(6 with 2 dice) = 5/36.

COMPOUNDING

Throwing a 6 with one die is a single event. Throwing a sum of seven with two dice is a *compound* event. It involves two distinct outcomes, which are combined in the event "the sum of the dice equals 7."

A lot of simple probability reasoning involves compound events. Imagine a fair coin, and two urns, Urn 1 and Urn 2, made up as follows:

Urn 1: 3 red balls, 1 green one.
Urn 2: 1 red ball, 3 green ones.

In a fair drawing from Urn 1, the probability of getting a Red ball is $Pr(R_1) = 3/4$.

With Urn 2, it is $Pr(R_2) = 1/4$.

Now suppose we pick an urn by tossing a fair coin. If we get heads, we draw from Urn 1; if tails, from Urn 2. Assume independence, that is, that the toss of the coin has no effect on the urns.

What is the probability that we toss a coin and then draw a red ball from Urn 1? We first have to toss heads, and then draw a red ball from Urn 1 (R_1).

$$Pr(H\&R_1) = 1/2 \times 3/4 = 3/8.$$

The probability of tossing a coin and then drawing a red ball from Urn 2 (R_2) is:

$$Pr(T\&R_2) = 1/2 \times 1/4 = 1/8.$$

What is the probability of getting a red ball, using this setup? That is a compound event. We can get a red ball by getting heads with the coin, and then drawing red from Urn 1, ($H\&R_1$), or by getting tails, and then drawing red from Urn 2 ($T\&R_2$).

These are mutually exclusive events, and so can be added.

$$Pr(Red) = Pr(H\&R_1) + Pr(T\&R_2) = 3/8 + 1/8 = 1/2.$$

So the probability of drawing red, in this set-up, is 1/2.

A TRICK QUESTION

Suppose we select one of those two urns by tossing a coin, and then make *two* draws from *that* urn with replacement. What is the probability of drawing two reds in a row in this setup?

We know the probability of getting one red is 1/2.

Is the probability of two reds $(1/2)(1/2) = 1/4$? NO!

The reason is that we can get two reds in a row in two different ways, which we'll call X and Y:

X: By tossing heads (an event of probability ½), and then getting red from Urn 1 (R_1, an event of probability 3/4) followed by replacing the ball and again drawing red from Urn 1 (another event of probability 3/4).

Y: By tossing tails (probability 1/2), and then getting red from Urn 2 (R_2), followed by replacing the ball and again drawing red from Urn 2 (another R_2).

The probabilities are:

- $Pr(X) = (1/2)(3/4)(3/4) = 9/32$.
- $Pr(Y) = (1/2)(1/4)(1/4) = 1/32$.

Hence,

Pr(first ball drawn is red & second ball drawn is red)
$= Pr(X) + Pr(Y) = 10/32 = 5/16$.

UNDERSTANDING THE TRICK QUESTION

Did you think that the probability of two reds would be 1/4? Here is one way to understand why not. Think of doing two different experiments over and over again.

Experiment 1. Choose an urn by tossing a coin, and then draw a ball.

Results 1. You draw a red ball about half the time.

Experiment 2. Choose an urn by tossing a coin, and then draw two balls with replacement. (After you have drawn a ball, you put it back in the urn.)

Results 2. You get two red balls about 5/16 of the time, two green balls about 5/16 of the time, and a mix of one red and one green about $6/16 = 3/8$ of the time.

Explanation. Once you have picked an urn with a "bias" for a given color, it is more probable that both balls will be of that color, than that you will get one majority and one minority color.

LAPLACE

This example was a great favorite with P. S. de Laplace (1749–1827), a truly major figure in the development of probability theory. He wrote the very first introductory college textbook about probability, *A Philosophical Essay on Probabilities*. He wrote this text for a class he taught at the polytechnic school in Paris in 1795, between the French Revolution and the rule of Napoleon.

Laplace was one of the finest mathematicians of his day. His *Analytic Theory of Probabilities* is still a rich source of ideas. His *Celestial Mechanics*—the mathematics of gravitation and astronomy—was equally important. He was very popular with the army, because he used mathematics to improve the French artillery. He used to go to Napoleon's vegetarian lunches, where he gave informal talks about probability theory.

EXERCISES

1 *Galileo.* Don't feel bad if you gave the wrong answer to Odd Question 4, about rolling dice. A long time ago someone asked a similar question, about throwing three dice. Galileo (1564–1642), one of the greatest astronomers and physicists ever, took the time to explain the right and wrong answers.

Explain why you might (wrongly) expect three fair dice to yield a sum of 9 as often as they yield a sum of 10. Why is it wrong to think that 9 is as probable as 10?

2 *One card.* A card is drawn from a standard deck of fifty-two cards which have been well shuffled seven times. What is the probability that the card is:

(a) Either a face card (jack, queen, king) or a ten?
(b) Either a spade or a face card?

3 *Two cards.* When two cards are drawn in succession from a standard pack of cards, what are the probabilities of drawing:
(a) Two hearts in a row, with replacement, and (b) without replacement.
(c) Two cards, neither of which is a heart, with replacement, and (d) without replacement.

4 *Archery.* An archer's target has four concentric circles around a bull's-eye. For a certain archer, the probabilities of scoring are as follows:

Pr(hit the bull's-eye) = 0.1
Pr(hit first circle, but not bull's-eye) = 0.3.
Pr(hit second circle, but no better) = 0.2.
Pr(hit third circle, but no better) = 0.2.
Pr(hit fourth circle, but no better) = 0.1.

Her shots are independent.
(a) What is the probability that in two shots she scores a bull's-eye on the first shot, and the third circle on the second shot?
(b) What is the probability that in two shots she hits the bull's-eye once, and the third circle once?
(c) What is the probability that on any one shot she misses the target entirely?

5 *Polio from diapers* (a news story).

> *Southampton, England*: A man contracted polio from the soiled diaper of his niece, who had been vaccinated against the disease just days before, doctors said yesterday. "The probability of a person contracting polio from soiled diapers is literally one in three million," said consultant Martin Wale. What did Dr. Wale mean?

6 *Languages.* We distinguish between a "proposition-language" and an "event-language" for probability. Which language was used in:
(a) Question 2. (b) Question 3. (c) Question 4. (d) Dr. Wale's statement in question 5?

KEY WORDS FOR REVIEW

Events	Addition
Propositions	Independence
Mutually exclusive	Multiplication

5 Conditional Probability

The most important new idea about probability is the probability that something happens, on condition that something else happens. This is called conditional probability.

CATEGORICAL AND CONDITIONAL

We express probabilities in numbers. Here is a story I read in the newspaper. The old tennis pro Ivan was discussing the probability that the rising young star Stefan would beat the established player Boris in the semifinals. Ivan was set to play Pete in the other semifinals match. He said,

> The probability that Stefan will beat Boris is 40%.

Or he could have said,

> The chance of Stefan's winning is 0.4.

These are *categorical* statements, no ifs and buts about them. Ivan might also have this opinion:

> Of course I'm going to win my semifinal match, but if I were to lose, then Stefan would not be so scared of meeting me in the finals, and he would play better; there would then be a 50–50 chance that Stefan would beat Boris.

This is the probability of Stefan's winning in his semifinal match, *conditional* on Ivan losing the other semifinal. We call it the *conditional probability*. Here are other examples:

> *Categorical*: The probability that there will be a bumper grain crop on the prairies next summer.

Conditional: The probability that there will be a bumper grain crop next summer, given that there has been very heavy snowfall the previous winter.

Categorical: The probability of dealing an ace as the second card from a standard pack of well-shuffled cards (regardless of what card is dealt first).

There are 4 aces and 52 cards, any one of which may come up as the second card. So the probability of getting an ace as the second card should be 4/52 = 1/13.

Conditional: The probability of dealing an ace as the second card *on condition that* the first card dealt was a king.

If a king was dealt first, there are 51 cards remaining. There are 4 aces still in the pack, so the *conditional probability* is 4/51.

Conditional: The probability of dealing an ace as the second card on condition that the first card dealt was also an ace.

When an ace is dealt first, there are 51 cards remaining, but only 3 aces, so the conditional probability is 3/51.

NOTATION

Categorical probability is represented:

Pr( )

Conditional probability is represented:

Pr(/).

Examples of categorical probability:

Pr(S wins the final) = 0.4.
Pr(second card dealt is an ace) = 1/13.

Examples of conditional probability:

Pr(S wins his semifinal/I loses his semifinal) = 0.5.
Pr(second card dealt is an ace/first card dealt is a king) = 4/51.

BINGO

Bingo players know about conditional probability.

In a game of bingo, you have a 5×5 card with 25 squares. Each square is marked with a different number from 1 to 99. The master of ceremonies draws numbered balls from a bag. Each time a number on your board is drawn, you fill in the corresponding square. You win (BINGO!) when you fill in a complete column, row, or diagonal.

Bingo players are fairly relaxed when they start the game. The probability that they will soon complete a line is small. But as they begin to fill in a line they get very excited, because the conditional probability of their winning is not so small.

PARKING TICKETS

If you park overnight near my home, and don't live on the block, you may be ticketed for not having a permit for overnight parking. The fine will be $20. But the street is only patrolled on average about once a week.

What is the probability of being fined?

Apparently the street is never patrolled on two consecutive nights. What is the probability of being ticketed tonight, *conditional* on having been ticketed on this street last night?

DEFINITION OF CONDITIONAL PROBABILITY

There is a very handy definition of conditional probability. We first state it, and then illustrate how it works.

When Pr(B) > 0
Pr(A/B) = Pr(A&B)/Pr(B)

Pr(B) must be a positive number, because we cannot divide by zero. But why is the rest of this definition sensible? Some examples will suggest why.

CONDITIONAL DICE

Think of a fair die. We say the outcome of a toss is even if it falls 2, 4, or 6 face up.

Here is conditional probability:

Pr(6/even)

In ordinary English:

The probability that we roll a 6, on condition that we rolled an even number.
The conditional probability of sixes, given evens.

With a fair die, we roll 2, 4, and 6 equally often. So 6 comes up a third of the time that we get an even outcome.

Pr(6/even) = 1/3.

This fits our definition, because,

Pr(6 & even) = Pr(6) = 1/6.
Pr(even) = 1/2.
Pr(6/even) = (1/6)/(1/2) = 1/3.

OVERLAPS

Now ask a more complicated question, which involves *overlapping* outcomes. Let M mean that the die either falls 1 up or falls with a prime number up (2, 3, 5). Thus M happens when the die falls 1, 2, 3, or 5 uppermost. What is

Pr(even/M)?

The only prime even number is 2. There are 4 ways to throw M (1, 2, 3, 5). Hence, if the die is fair,

Pr(even/M) = 1/4.

This fits our definition, because

Pr(even & M) = 1/6.
Pr(M) = 4/6.
Pr(even/M) = (1/6)/(4/6) = 1/4.

WELL-SHUFFLED CARDS

Think of a well-shuffled standard pack of 52 cards, from which the dealer deals the top card. He tells you that it is *either* red, *or* clubs. But not which. Call this information RvC.

Clubs are black. There are 13 clubs in the pack, and 26 other cards that are red. We were told that the first card is RvC. What is the probability that it is an ace? What is Pr(A/RvC)? A & (RvC) is equivalent to ace of clubs, or a red ace, diamonds or hearts. For a total of 3. Hence 3 cards out of the 39 RvC cards are aces.

Hence the conditional probability is:

Pr[A/(RvC)] = 1/13.

This agrees with our definition:

Pr[A & (RvC)] = 3/52.
Pr(RvC) = 39/52.

$$\Pr[(A/(RvC)] = \frac{\Pr[A \,\&\, (RvC)]}{\Pr(RvC)} = 3/39 = 1/13.$$

URNS

Imagine two urns, each containing red and green balls. Urn A has 80% red balls, 20% green, and Urn B has 60% green, 40% red. You pick an urn at random. Is it A or B? Let's draw balls from the urn and use this information to guess which urn it is. After each draw, the ball drawn *is replaced*. Hence for any draw, the probability of getting red from urn A is 0.8, and from urn B, the probability of getting red is 0.4.

Pr(R/A) = 0.8
Pr(R/B) = 0.4
Pr(A) = Pr(B) = 0.5

You draw a red ball. If you are like *Alert Learner*, that may lead you to suspect that this is urn A (which has more red balls than green ones). That is just a hunch. Let's be more exact.

We want to find Pr(A/R), which is [Pr(A&R)]/[Pr(R)].

You can get a red ball from either urn A or urn B. You get a red ball either when the event A&R happens, or when the event B&R happens. Event R is thus identical to (A&R)v(B&R).

The two alternatives (A&R) and (B&R) are mutually exclusive, so we can add up the probabilities.

Pr(R) = Pr(A&R) + Pr(B&R) [1]

The probability of getting urn B is 0.5; the probability of getting a red ball from it is 0.4, so that the probability of both happening is

Pr(B&R) = Pr(R&B) = Pr(R/B)Pr(B) = 0.4 × 0.5 = 0.2.

Likewise,

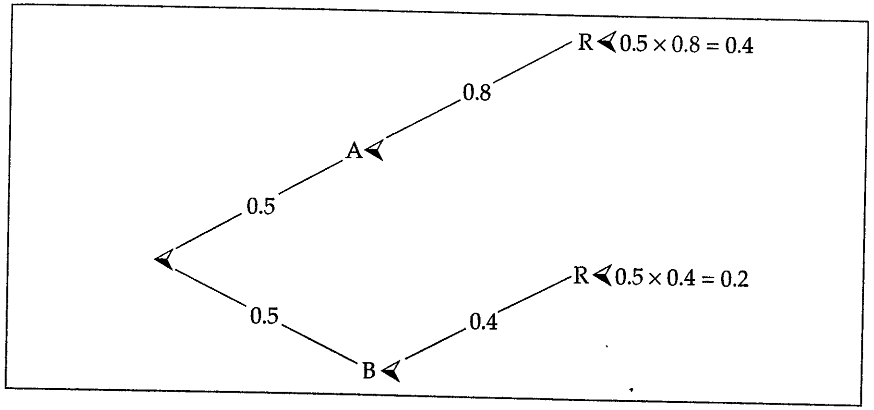
Pr(A&R) = 0.8 × 0.5 = 0.4.

Putting these into formula [1] above,

Pr(R) = Pr(A&R) + Pr(B&R) = 0.4 + 0.2 = 0.6.
Hence Pr(A/R) = Pr(A&R)/Pr(R) = (0.4)/(0.6) = 2/3.

DRAWING THE CALCULATION TO CHECK IT

You may find it helpful to visualize the calculation as a branching tree. We start out with our coin and the two urns. How can we get to a red ball? There are two routes. We can toss a heads (probability 0.5), giving us urn A. Then we can draw a red ball (probability 0.8). That is the route shown here on the top branch.



We can also get an R by tossing tails, going to urn B, and then drawing a red ball, as shown on the bottom branch.

We get to R on one of the two branches. So the total probability of ending up with R is the sum of the probabilities at the end of each branch. Here it is 0.4 + 0.2 = 0.6.

The probability of getting to an R following an A branch is 0.4.

Thus that part of the probability that gets you to R by A, namely Pr(A/R), is .4/.6 = 2/3.

MODELS

All our examples up to now have been dice, cards, urns. Now we turn to more interesting cases, more like real life. In each we make a model of a situation, and say that the real-life story is modeled by a standard ball-and-urn example.

SHOCK ABSORBERS

An automobile plant contracted to buy shock absorbers from two local suppliers, Bolt & Co. and Acme Inc. Bolt supplies 40% and Acme 60%. All shocks are subject to quality control. The ones that pass are called reliable.

Of Acme's shocks, 96% test reliable. But Bolt has been having some problems on the line, and recently only 72% of Bolt's shock absorbers have tested reliable.

What is the probability that a randomly chosen shock absorber will test reliable?

Intuitive guess: the probability will be lower than 0.96, because Acme's product is diluted by a proportion of shock absorbers from Bolt. The probability must be between 0.96 and 0.72, and nearer to 0.96. But by how much?

Solution

Let A = The shock chosen at random was made by Acme.
Let B = The shock chosen at random was made by Bolt.
Let R = The shock chosen at random is reliable.

Pr(A) = 0.6	Pr(R/A) = 0.96	So, Pr(R&A) = 0.576.
Pr(B) = 0.4	Pr(R/B) = 0.72	So, Pr(R&B) = 0.288.
R = (R&A)v(R&B)		

Answer: Pr(R) = (.6×.96) + (.4×.72) = 0.576 + 0.288 = 0.864.

We can ask a more interesting question.

What is the conditional probability that a randomly chosen shock absorber, which is tested and found to be reliable, is made by Bolt?

Intuitive guess: look at the numbers. The automobile plant buys more shocks from Acme than Bolt. And Bolt's shocks are much less reliable than Acme's. Both these pieces of evidence count against a reliable shock, chosen at random, being made by Bolt. We expect that the probability that the shock is from Bolt is less that 0.4. But by how much?

Solution

We require Pr(B/R).

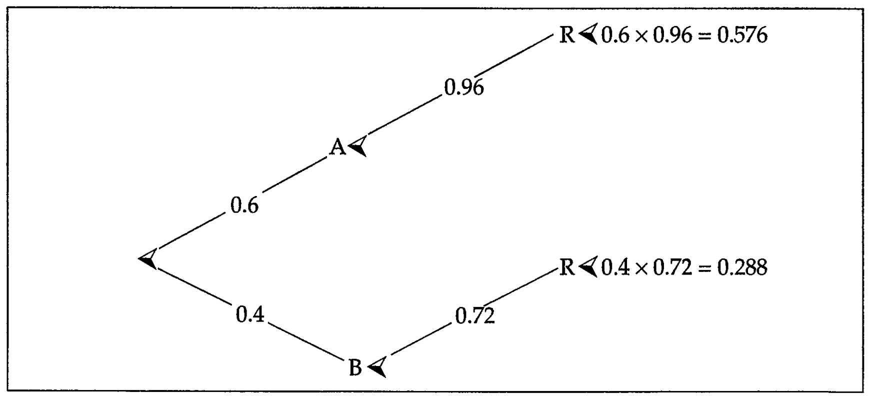
By definition, Pr(B/R) = Pr(B&R)/Pr(R) = 0.288/0.864 = 1/3.

Actually, you may like to do this without any multiplying, because almost all the numbers cancel:

$$= \frac{0.4 \times 0.72}{(.4\times.72) + (.6\times.96)} = 1/3$$

Answer: Pr(B/R) = 1/3.

DRAWING TO CHECK



Pr(R) = 0.576 + 0.288 = 0.864. Pr(B/R) = 1/3.

WEIGHTLIFTERS

You learn that a certain country has two teams of weightlifters, either of which it may send to an international competition. Of the members of one team (the Steroid team), 80% have been regularly using steroids, but only 20% of the members of the other team are regular users (the Cleaner team). The head coach flips a fair coin to decide which team will be sent.

One member of the competing team is tested at random. He has been using steroids.

What is the conditional probability that the team in competition is the Steroid team, given that a member was found by a urine test to be using steroids? That is, what is Pr(S/U)?

Solution

Let S = The coach sent the Steroid team.
Let C = The coach sent the Cleaner team.
Let U = A member selected at random uses steroids.

Pr(S) = 0.5	Pr(U/S) = 0.8	Pr(U&S) = 0.4
Pr(C) = 0.5	Pr(U/C) = 0.2	Pr(U&C) = 0.1
U = (U&S)v(U&C)	Pr(U) = 0.4 + 0.1 = 0.5	

Pr(S/U) = [Pr(S&U)]/[Pr(U)] = 0.4/0.5 = 0.8

Answer: Pr(Steroid team/selected member uses steroids) = 0.8.

So the fact that we randomly selected a team member who uses steroids, is pretty good evidence that this is the Steroid team.

TWO IN A ROW: WITH REPLACEMENT

Back to the urns on page 51. Suppose you pick an urn at random, and make two draws, *with replacement*. You get a red, and then a red again. What is $P(A/R_1\&R_2)$?

Let R_1 be the event that the first ball drawn is red, and R_2 the event that the second ball drawn is red. Then you can work out $P(A/R_1\&R_2)$ as:

$$\frac{Pr(A\&R_1\&R_2)}{Pr(R_1\&R_2)}$$

Now we know $Pr(A\&R_1\&R_2) = Pr(R_2/A\&R_1)Pr(A\&R_1) = 0.8 \times 0.4 = 0.32$.
Likewise, $Pr(B\&R_1\&R_2) = 0.08$.
$Pr(R_1\&R_2) = Pr(A\&R_1\&R_2) + Pr(B\&R_1\&R_2) = 0.32 + 0.08 = 0.4$
$Pr(A/R_1\&R_2) = 0.32/0.4 = 4/5 = 0.8$

Conditional probability of urn A, given that we:
draw one red ball, is 2/3
draw a second ball after replacement, also red, is 0.8

Thus a second red ball "increases the conditional probability" that this is urn A. The extra red ball may be taken as more *evidence*.

This suggests how we *learn by experience* by obtaining more evidence.

THE GAMBLER'S FALLACY ONCE AGAIN

Fallacious Gambler thought that he could "learn from experience" when he saw that a fair (unbiased, independent trials) roulette wheel came up 12 blacks in a row. That is, he thought that:

Pr(red on 13th trial/12 blacks in a row) > ½.

But if trials are independent, this probability is

$$\frac{Pr(BBBBBBBBBBBBR)}{Pr(BBBBBBBBBBBB)} = \frac{(½)^{13}}{(½)^{12}} = ½.$$

This is a new way to understand the gambler's fallacy.

TWO WEIGHTLIFTERS: WITHOUT REPLACEMENT

Back to the weightlifters. Suppose we test two weightlifters chosen at random from a team that the coach selected by tossing a fair coin. We think: if both weightlifters test positive, that is pretty strong evidence that this is the Steroid team. Probability confirms this hunch.

We are sampling the team without replacement. So say there are ten members to a team. We randomly test two members.

Let S = The coach sent the Steroid team.
Let C = The coach sent the Cleaner team.
Let U_1 = The first member selected at random uses steroids.
Let U_2 = The second member selected at random uses steroids.

If we have the Steroid team, the probability that the first person tested uses steroids is 0.8 (on page 54 we had Pr(U/S) = 0.8). What is the probability of selecting two users?

There is a 4/5 probability of selecting one user. After the first person is selected, and turns out to be a user, there are 9 team members left, 7 of whom use steroids. So there is a 7/9 probability of getting a user for the next test. Hence the probability that the first two persons chosen from the Steroid team use steroids is $4/5 \times 7/9 = 28/45$.

Likewise, the probability that the first two persons chosen from the Cleaner team use steroids is $1/5 \times 1/9 = 1/45$.

The probability that the coach sent the Steroid team, when both team members selected at random are users, is,

$\Pr(S\&U_1\&U_2) = 0.5(28/45) = 28/90.$

Likewise, $\Pr(C\&U_1\&U_2) = 0.5(1/45) = 1/90.$

$\Pr(U_1\&U_2) = \Pr(S\&U_1\&U_2) + \Pr(C\&U_1\&U_2) = 29/90.$

$$\Pr(S/U_1\&U_2) = \frac{\Pr(S\&U_1\&U_2)}{\Pr(U_1\&U_2)} = 28/29 > 0.96$$

Conditional probability that this is the Steroid team, given that we randomly selected:
one weightlifter who was a user, is 0.8
a second weightlifter who was also a user, is > 0.96

Getting two members who use steroids seems to be powerful *evidence* that the coach picked the Steroid team.

EXERCISES

1 *Phony precision about tennis.* In real life, the newspaper story about tennis quoted Ivan as stating a probability not of "40%" but:

The probability that Stefan will beat Boris in the semifinals is only 37.325%.

Can you make any sense out of this precise fraction?

2 *Heat lamps.* Three percent of production batches of *Tropicana* heat lamps fall below quality standards. Six percent of the batches of *Florida* heat lamps are below quality standards. A hardware store buys 40% of its heat lamps from *Tropicana,* and 60% from *Florida.*
(a) What is the probability that a lamp taken at random in the store is made by *Tropicana* and is below quality standards?
(b) What is the probability that a lamp taken at random in the store is below quality standards?
(c) What is the probability that a lamp from this store, and found to be below quality standards, is made by *Tropicana*?

3 *The Triangle.* An unhealthy triangular-shaped region in an old industrial city once had a lot of chemical industry. Two percent of the children in the city live in the triangle. Fourteen percent of these test positive for excessive presence of toxic metals in the tissues. The rate of positive tests for children in the city, not living in the triangle, is only 1%.
(a) What is the probability that a child who lives in the city, and who is chosen at random, both lives in the Triangle and tests positive?
(b) What is the probability that a child living in the city, chosen at random, tests positive?
(c) What is the probability that a child chosen at random, who tests positive, lives in the Triangle?

4 *Taxicabs.* Draw a tree diagram for the taxicab problem, Odd Question 5.

5 *Laplace's trick question.* Look back at Laplace's question (page 44). An experiment consists of tossing a coin to select an urn, then drawing a ball, noting its color, replacing it, and drawing another ball and noting its color. Find, Pr(second ball drawn is red/first ball drawn is red).

6 *Understanding the question.* On page 45 we ended Chapter 4 with a way to understand Laplace's trick question. How does your answer to question 6 help with understanding Laplace's question?

KEY WORDS FOR REVIEW

Categorical
Conditional
Definition of conditional probability
Calculating conditional probabilities
Models
Learning from experience

6 The Basic Rules of Probability

This chapter summarizes the rules you have been using for adding and multiplying probabilities, and for using conditional probability. It also gives a pictorial way to understand the rules.

The rules that follow are informal versions of standard axioms for elementary probability theory.

ASSUMPTIONS

The rules stated here take some things for granted:

- The rules are for finite groups of propositions (or events).
- If A and B are propositions (or events), then so are AvB, A&B, and ~A.
- Elementary deductive logic (or elementary set theory) is taken for granted.
- If A and B are *logically equivalent*, then Pr(A) = Pr(B). [Or, in set theory, if A and B are events which are provably the same sets of events, Pr(A) = Pr(B).]

NORMALITY

The probability of any proposition or event A lies between 0 and 1.

(1) $0 \leq \Pr(A) \leq 1$

Why the name "normality"? A measure is said to be *normalized* if it is put on a scale between 0 and 1.

CERTAINTY

An event that is sure to happen has probability 1. A proposition that is certainly true has probability 1.

(2) Pr(certain proposition) = 1
Pr(sure event) = 1

Often the Greek letter Ω is used to represent certainty: $\Pr(\Omega) = 1$.

ADDITIVITY

If two events or propositions A and B are mutually exclusive (disjoint, incompatible), the probability that one or the other happens (or is true) is the sum of their probabilities.

(3) If A and B are mutually exclusive, then
$\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

OVERLAP

When A and B are not mutually exclusive, we have to subtract the probability of their overlap. In a moment we will *deduce* this from rules (1)–(3).

(4) $\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A\&B)$

CONDITIONAL PROBABILITY

The only basic rules are (1)–(3). Now comes a *definition*.

(5) If $\Pr(B) > 0$, then $\Pr(A/B) = \dfrac{\Pr(A\&B)}{\Pr(B)}$

MULTIPLICATION

The definition of conditional probability implies that:

(6) If $\Pr(B) > 0$, $\Pr(A\&B) = \Pr(A/B)\Pr(B)$.

TOTAL PROBABILITY

Another consequence of the definition of conditional probability:

(7) If $0 < \Pr(B) < 1$, $\Pr(A) = \Pr(B)\Pr(A/B) + \Pr(\sim B)\Pr(A/\sim B)$.

In practice this is a very useful rule. What is the probability that you will get a grade of A in this course? Maybe there are just two possibilities: you study hard, or you do not study hard. Then:

Pr(A) = Pr(study hard)Pr(A/study hard) + Pr(don't study)Pr(A/don't study).

Try putting in some numbers that describe yourself.

LOGICAL CONSEQUENCE

When B logically entails A, then

Pr(B) ≤ Pr(A).

This is because, when B entails A, B is logically equivalent to A&B. Since

Pr(A) = Pr(A&B) + Pr(A&~B) = Pr(B) + Pr(A&~B),

Pr(A) will be bigger than Pr(B) except when Pr(A&~B) = 0.

STATISTICAL INDEPENDENCE

Thus far we have been very informal when talking about independence. Now we state a *definition* of one concept, often called statistical independence.

(8) If 0 < Pr(A) and 0 < Pr(B), then,
A and B are statistically independent if and only if:
Pr(A/B) = Pr(A).

PROOF OF THE RULE FOR OVERLAP

(4) Pr(AvB) = Pr(A) + Pr(B) − Pr(A&B).

This rule follows from rules (1)–(3), and the logical assumption on page 58, that logically equivalent propositions have the same probability.

AvB is logically equivalent to: (A&B) v (A&~B) v (~A&B) (*)

Why? Those familiar with "truth tables" can check it out. But you can see it directly. A is logically equivalent to (A&B) v (A&~B). B is logically equivalent to (A&B) v (~A&B).

Now the three components (A&B), (A&~B), and (~A&B) are mutually exclusive. (Why?) Hence we can add their probabilities, using (*):

Pr(AvB) = Pr(A&B) + Pr(A&~B) + Pr(~A&B) (**)
A is logically equivalent to [(A&B)v(A&~B)], and
B is logically equivalent to [(A&B)v(~A&B)].

So,

Pr(A) = Pr(A&B) + Pr(A&~B).
Pr(B) = Pr(A&B) + Pr(~A&B).

Since it makes no difference to add and then subtract something in (**):

Pr(AvB) = Pr(A&B) + Pr(A&~B) + Pr(~A&B) + Pr(A&B) − Pr(A&B)

Hence,

Pr(AvB) = Pr(A) + Pr(B) − Pr(A&B).

CONDITIONALIZING THE RULES

It is easy to check that the basic rules (1)–(3), and (5), the definition of conditional probability, all hold in conditional form. That is, the rules hold if we replace Pr(A), Pr(B), Pr(A/B), and so on, by Pr(A/E), Pr(B/E), P(A/B&E), and so on.

Normality

(1C) 0 ≤ Pr(A/E) ≤ 1

Certainty

We need to check that for E, such that Pr(E) > 0,

(2C) Pr([sure event]/E) = 1.

Now E is logically equivalent to the occurrence of E with something that is sure to happen. Hence,

Pr([sure event] & E) = Pr(E).
Pr([sure event/E]) = [Pr(E)]/[Pr(E)] = 1.

Additivity

Let Pr(E) > 0. If A and B are mutually exclusive, then

Pr[(AvB)/E] = Pr[(AvB)&E]/Pr(E) = Pr(A&E)/Pr(E) + Pr(B&E)/Pr(E).
(3C) Pr[(AvB)/E] = Pr(A/E) + Pr(B/E).

Conditional probability

This is the only case you should examine carefully. The conditionalized form of (5) is:

(5C) If Pr(E) > 0 and Pr(B/E) > 0, then

$$\Pr[A/(B\&E)] = \frac{\Pr[(A\&B)/E]}{\Pr(B/E)}.$$

We prove this starting from (5),

$$\Pr[A/(B\&E)] = \frac{\Pr(A\&B\&E)}{\Pr(B\&E)}.$$

The numerator (on top of the fraction) is Pr(A&B&E) = Pr[(A&B)/E] × Pr(E).
The denominator (bottom of the fraction) is Pr(B&E) = Pr(B/E) × Pr(E).
Dividing the numerator by the denominator, we get (5C).

Many philosophers and inductive logicians take conditional probability, rather than categorical probability, as the primitive idea. Their basic rules are, then, versions of (1C), (2C), (3C), and (5C). Formally, the end results are in all essential respects identical to our approach that begins with categorical probability and then defines conditional probability. But when we start to ask about various meanings of these rules, we find that a conditional probability approach sometimes makes more sense.

STATISTICAL INDEPENDENCE AGAIN

Our first intuitive explanation of independence (page 25) said that trials on a chance setup are independent if and only if the probabilities of the outcomes of a trial are not influenced by the outcomes of previous trials. But this left open what "influenced" really means. We also spoke of *randomness*, of trials having no *memory*, and of *the impossibility of a gambling system*. These are all valuable metaphors.

The idea of conditional probability makes one exact definition possible. The probability of A should be no different from the probability of A given B, Pr(A/B).

Naturally, independence should be a symmetric relation: A is independent of B if and only if B is independent of A.

In other words, when 0 < Pr(A) and 0 < Pr(B), we expect that:

If Pr(A/B) = Pr(A), then Pr(B/A) = Pr(B) (and vice versa).

This is proved from definition (8) on page 60.

Suppose that Pr(A/B) = Pr(A).
By (5), Pr(A) = [Pr(A&B)]/[Pr(B)].
And so Pr(B) = [Pr(A&B)]/[Pr(A)].
So, since A&B is logically equivalent to B&A,
Pr(B) = Pr(B&A)/Pr(A) = Pr(B/A).

MULTIPLE INDEPENDENCE

Definition (8) defines the statistical independence of a pair of propositions. That is called "pairwise" independence. But a whole group of events or propositions could be mutually independent. This idea is easily defined.

It follows from (6) and (8) that when A and B are statistically independent:

Pr(A&B) = Pr(A)Pr(B)

(See exercise 3.) This can be generalized to the statistical independence of any number of events. For example A, B, and C are statistically independent if and only if A, B, and C are pairwise independent, and

Pr(A&B&C) = Pr(A)Pr(B)Pr(C).

VENN DIAGRAMS

John Venn (1824–1923) was an English logician who in 1866 published the first systematic theory of probabilities explained in terms of relative frequencies. Most people remember him only for "Venn diagrams" in deductive logic. Venn diagrams are used to represent *deductive* arguments involving the quantifiers *all*, *some*, and *no*.

You can also use Venn diagrams to represent probability relations. These drawings help some people who think spatially or pictorially.

Imagine that you have eight musicians:

Four of them are singers, with no other musical abilities.
Three of them can whistle but cannot sing.
One can both whistle and sing.

A Venn diagram can picture this group, using a set of circles. One circle is used for each class. Circles overlap when the classes overlap. Our diagram looks like this:

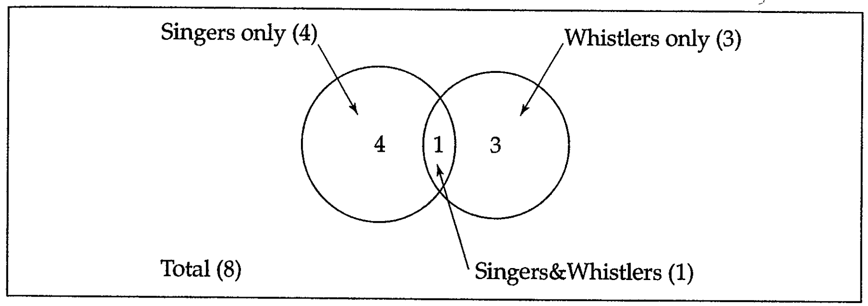


FIGURE 6.1

The circle representing the singers contains five units (four singers plus one singer&whistler), while the circle representing the whistlers has four units (three whistlers plus one singer&whistler). The overlapping region has an area of one unit, since only one of the eight people fits into both categories. We will think of the area of each segment as proportional to the number of people in that segment.

Now say we are interested in the probability of selecting, at random, a singer from the group of eight people. Since there are five singers in the group of eight people, the answer is 5/8.

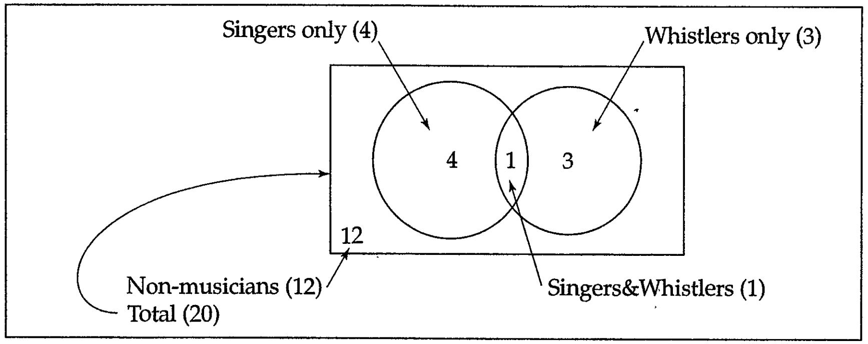
What is the probability that a singer is chosen, on condition that the person chosen is also a whistler? Since you know the person selected is a whistler, this limits the group to the whistlers' circle. It contains four people. Only one of the four is in the singers' circle. Hence only one of the four possible choices is a singer. Hence, the probability that a singer is chosen, given that the singer is also a whistler, is 1/4.

Now let us generalize the example. Put our 8 musicians in a room with 12 nonmusical people, resulting in a group of 20 people. Imagine we were interested in these two events:

Event A = a singer is selected at random from the whole group.
Event B = a whistler is selected at random from the whole group.

Here is a Venn diagram of the situation, where the entire box represents the room full of twenty people.

FIGURE 6.2



Notice the major change from the previous diagram: Figure 6.2 now has its circles enclosed in a rectangle. By convention, the area of the rectangle is set to 1. The areas of each of the circles correspond to the probability of occurrence of an event of the type that it represents: the area of circle A is 5/20, or 0.25, since there are 5 singers among 20 people. Likewise, the area of circle B is 4/20, or 0.2. The area of the region of overlap between A & B is 1/20, or 0.05.

These drawings can be used to illustrate the basic rules of probability.

(1) *Normality*: $0 \leq \Pr(A) \leq 1$.

This corresponds to the rectangle having an area of 1 unit: since all circles must lie within the rectangle, no circle, and hence no event can have a probability of greater than 1.

(2) *Certainty*: Pr(sure event) = 1. Pr(certain proposition) = 1.

With Venn diagrams, an event that is sure to happen, or a proposition that is certain, corresponds to a "circle" that fills the entire rectangle, which by convention has unit area 1.

(3) *Additivity*: If A and B are mutually exclusive, then:

$\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

If two groups are mutually exclusive they do not overlap, and the area covering members of either group is just the sum of the areas of each.

(4) *Overlap*:

To calculate the probability of AvB, determine how much of the rectangle is covered by circles A and B. This will be all the area in A, plus the area that

appears *only* in B. The area only in B is the areas in B, less the area of overlap with A.

$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A\&B)$

(5) *Conditional*:

Given that event B has happened, what is the probability that event A will also happen? Look at Figure 6.2. If B has happened, you know that the person selected is a whistler. So we want the proportion of the area of B, that includes A. That is, the area of A&B divided by the area of B.

$\Pr(A/B) = \Pr(A\&B) \div \Pr(B)$, so long as $\Pr(B) > 0$.

So, in our numerical example, $\Pr(A/B) = 1/4$.
Conversely, $\Pr(B/A) = \Pr(A \& B)/\Pr(A) = 1/5 = 0.2$.

ODD QUESTION 2

Recall the Odd Question about Pia:

2. Pia is thirty-one years old, single, outspoken, and smart. She was a philosophy major. When a student, she was an ardent supporter of Native American rights, and she picketed a department store that had no facilities for nursing mothers. Rank the following statements in order of probability from 1 (most probable) to 6 (least probable). (Ties are allowed.)

______(a) Pia is an active feminist.
______(b) Pia is a bank teller.
______(c) Pia works in a small bookstore.
______(d) Pia is a bank teller and an active feminist.
______(e) Pia is a bank teller and an active feminist who takes yoga classes.
______(f) Pia works in a small bookstore and is an active feminist who takes yoga classes.

This is a famous example, first studied empirically by the psychologists Amos Tversky and Daniel Kahneman. They found that very many people think that, given the whole story:

The most probable description is (f) Pia works in a small bookstore and is an active feminist who takes yoga classes.

In fact, they rank the possibilities something like this, from most probable to least probable:

(f), (e), (d), (a), (c), (b).

But just look at the logical consequence rule on page 60. Since, for example, (f) logically entails (a) and (b), (a) and (b) must be more probable than (f).

In general:

$$\Pr(A\&B) \leq \Pr(B).$$

It follows that the probability rankings given by many people, with (f) most probable, are completely wrong. There are many ways of ranking (a)–(f), but any ranking should obey these inequalities:

$\Pr(a) \geq \Pr(d) \geq \Pr(e)$.
$\Pr(b) \geq \Pr(d) \geq \Pr(e)$.
$\Pr(a) \geq \Pr(f)$.
$\Pr(c) \geq \Pr(f)$.

ARE PEOPLE STUPID?

Some readers of Tversky and Kahneman conclude that we human beings are irrational, because so many of us come up with the wrong probability orderings. But perhaps people are merely *careless*!

Perhaps most of us do not attend closely to the exact wording of the question, "Which of statements (a)–(f) are more probable, that is have the highest probability."

Instead we think, "Which is the most useful, instructive, and likely to be true thing to say about Pia?"

When we are asked a question, most of us want to be informative, useful, or interesting. We don't necessarily want simply to say what is most probable, in the strict sense of having the highest probability.

For example, suppose I ask you whether you think the rate of inflation next year will be (a) less than 3%, (b) between 3% and 4%, or (c) greater than 4%.

You could reply, (a)-or-(b)-or-(c). You would certainly be right! That would be the answer with the highest probability. But it would be totally uninformative.

You could reply, (b)-or-(c). That is more probable than simply (b), or simply (c), assuming that both are possible (thanks to additivity). But that is a less interesting and less useful answer than (c), or (b), by itself.

Perhaps what many people do, when they look at Odd Question 2, is to form a character analysis of Pia, and then make an interesting guess about what she is doing nowadays.

If that is what is happening, then people who said it was most probable that Pia works in a small bookstore and is an active feminist who takes yoga classes, are not irrational.

They are just answering the wrong question—but maybe answering a more useful question than the one that was asked.

AXIOMS: HUYGENS

Probability can be axiomatized in many ways. The first axioms, or basic rules, were published in 1657 by the Dutch physicist Christiaan Huygens (1629–1695), famous for his wave theory of light. Strictly speaking, Huygens did not use the idea of probability at all. Instead, he used the idea of the fair price of something like a lottery ticket, or what we today would call the expected value of an event or proposition. We can still do that today. In fact, almost all approaches take probability as the idea to be axiomatized. But a few authors still take expected value as the primitive idea, in terms of which they define probability.

AXIOMS: KOLMOGOROV

The definitive axioms for probability theory were published in 1933 by the immensely influential Russian mathematician A. N. Kolmogorov (1903–1987). This theory is much more developed than our basic rules, for it applies to infinite sets and employs the full differential and integral calculus, as part of what is called measure theory.

EXERCISES

1 *Venn Diagrams.*

Let L: A person contracts a lung disease.
Let S: That person smokes.

Write each of the following probabilities using the Pr notation, and then explain it using a Venn diagram.
(a) The probability that a person either smokes or contracts lung disease (or both).
(b) The probability that a person contracts lung disease, given that he or she smokes.
(c) The probability that a person smokes, given that she or he contracts lung disease.

2 *Total probability.* Prove from the basic rules that $\Pr(A) + \Pr(\sim A) = 1$.

3 *Multiplying.* Prove from the definition of statistical independence that if $0 < \Pr(A)$, and $0 < \Pr(B)$, and A and B are statistically independent,

$$\Pr(A\&B) = \Pr(A)\Pr(B).$$

4 *Conventions.* In Chapter 4, page 40, we said that the rules for normality and certainty are just conventions. Can you think of any other plausible conventions for representing probability by numbers?

5 *Terrorists.* This is a story about a philosopher, the late Max Black.

One of Black's students was to go overseas to do some research on Kant. She was afraid that a terrorist would put a bomb on the plane. Black could not convince her that the risk was negligible. So he argued as follows:

BLACK: Well, at least you agree that it is almost impossible that *two* people should take bombs on your plane?
STUDENT: Sure.
BLACK: Then you should take a bomb on the plane. The risk that there would be another bomb on your plane is negligible.

What's the joke?

KEY WORDS FOR REVIEW

Normality	Conditional probability	Total probability
Certainty	Venn diagrams	Logical consequence
Additivity	Multiplication	Statistical independence

7 Bayes' Rule

One of the most useful consequences of the basic rules helps us understand how to make use of new evidence. Bayes' Rule is one key to "learning from experience."

Chapter 5 ended with several examples of the same form: urns, shock absorbers, weightlifters. The numbers were changed a bit, but the problems in each case were identical.

For example, on page 51 there were two urns A and B, each containing a known proportion of red and green balls. An urn was picked at random. So we knew:

Pr(A) and Pr(B).

Then there was another event R, such as drawing a red ball from an urn. The probability of getting red from urn A was 0.8. The probability of getting red from urn B was 0.4. So we knew:

Pr(R/A) and Pr(R/B).

Then we asked, what is the probability that the urn drawn was A, *conditional* on drawing a red ball? We asked for:

Pr(A/R) = ? Pr(B/R) = ?

Chapter 5 solved these problems directly from the definition of conditional probability. There is an easy rule for solving problems like that. It is called *Bayes' Rule.*

In the urn problem we ask which of two *hypotheses* is true: Urn A is selected, or Urn B is selected. In general we will represent hypotheses by the letter H.

We perform an *experiment* or get some *evidence*: we draw at random and observe a red ball. In general we represent evidence by the letter E.

Let's start with the simplest case, where there are only two hypotheses, H and ~H. By definition these are mutually exclusive, and exhaustive.

Let E be a proposition such that Pr(E) > 0. Then:

$$Pr(H/E) = \frac{Pr(H)Pr(E/H)}{Pr(H)Pr(E/H) + Pr(\sim H)Pr(E/\sim H)}$$

This is called *Bayes' Rule* for the case of two hypotheses.

PROOF OF BAYES' RULE

$$Pr(H\&E) = Pr(E\&H)$$

$$\frac{Pr(H\&E)Pr(E)}{Pr(E)} = \frac{Pr(E\&H)Pr(H)}{Pr(H)}$$

Using the definition of conditional probability,

$$Pr(H/E)Pr(E) = Pr(E/H)Pr(H).$$

$$Pr(H/E) = \frac{Pr(H)Pr(E/H)}{Pr(E)}$$

Since H and (~H) are mutually exclusive and exhaustive, then, by the rule of total probability on page 59,

$$Pr(E) = Pr(H)Pr(E/H) + Pr(\sim H)Pr(E/\sim H).$$

Which gives us Bayes' Rule:

$$(1) \quad Pr(H/E) = \frac{Pr(H)Pr(E/H)}{Pr(H)Pr(E/H) + Pr(\sim H)Pr(E/\sim H)}$$

GENERALIZATION

The same formula holds for any number of *mutually exclusive* and *jointly exhaustive* hypotheses:

$H_1, H_2, H_3, H_4, \ldots, H_k$, such that for each i, $Pr(H_i) > 0$.

Mutually exclusive means that only one of the hypotheses can be true. *Jointly exhaustive* means that at least one must be true.

By extending the above argument, if Pr(E) > 0, and for every i, $Pr(H_i) > 0$, we get for any hypothesis H_k,

$$(2) \quad Pr(H_j/E) = \frac{Pr(H_j)\,Pr(E/H_j)}{\Sigma[Pr(H_i)\,Pr(E/H_i)]}$$

Here the Σ (the Greek capital letter sigma, or S in Greek) stands for the *sum* of the terms with subscript i. Add all the terms $[Pr(H_i)Pr(E/H_i)]$ for i = 1, i = 2, up to i = k.

Formula (1) and its generalization (2) are known as **Bayes' Rule.**

The rule is just a way to combine a couple of basic rules, namely conditional and total probability. Bayes' Rule is trivial, but it is very tidy. It has a major role in some theories about inductive logic, explained in Chapters 13–15 and 21.

URNS

Here is the urn problem from page 51:

Imagine two urns, each containing red and green balls. Urn A has 80% red balls, 20% green, and Urn B has 60% green, 40% red. You pick an urn at random, and then can draw balls from the urn in order to guess which urn it is. After each draw, the ball drawn *is replaced.* Hence for any draw, the probability of getting red from urn A is 0.8, and from urn B it is 0.4.

$$Pr(R/A) = 0.8 \qquad Pr(R/B) = 0.4 \qquad Pr(A) = Pr(B) = 0.5$$

You draw a red ball. What is P(A/R)?

Solution by Bayes' Rule:

$$Pr(A/R) = \frac{Pr(A)Pr(R/A)}{Pr(A)Pr(R/A) + Pr(B)Pr(R/B)}$$

$$= (0.5 \times 0.8)/[(0.5 \times 0.8) + (0.5 \times 0.4)] = 2/3.$$

This is the same answer as was obtained on page 51.

SPIDERS

A tarantula is a large, fierce-looking, and somewhat poisonous tropical spider.

Once upon a time, 3% of consignments of bananas from Honduras were found to have tarantulas on them, and 6% of the consignments from Guatemala had tarantulas.

40% of the consignments came from Honduras. 60% came from Guatemala.

A tarantula was found on a randomly selected lot of bananas. What is the probability that this lot came from Guatemala?

Solution

Let G = The lot came from Guatemala. Pr(G) = 0.6.
Let H = The lot came from Honduras. Pr(H) = 0.4.
Let T = The lot had a tarantula on it. Pr(T/G) = 0.06. Pr(T/H) = 0.03.

$$\Pr(G/T) = \frac{\Pr(G)\Pr(T/G)}{\Pr(G)\Pr(T/G) + \Pr(H)\Pr(T/H)}$$

Answer: Pr(G/T) = (.6 × .06) / [(.6 × .06) + (.4 × .03)] = 3/4

TAXICABS: ODD QUESTION 5

Here is Odd Question 5.

You have been called to jury duty in a town where there are two taxi companies, Green Cabs Ltd. and Blue Taxi Inc. Blue Taxi uses cars painted blue; Green Cabs uses green cars.

Green Cabs dominates the market, with 85% of the taxis on the road.

On a misty winter night a taxi sideswiped another car and drove off. A witness says it was a blue cab.

The witness is tested under conditions like those on the night of the accident, and 80% of the time she correctly reports the color of the cab that is seen. That is, regardless of whether she is shown a blue or a green cab in misty evening light, she gets the color right 80% of the time.

You conclude, on the basis of this information:

______(a) The probability that the sideswiper was blue is 0.8.
______(b) It is more likely that the sideswiper was blue, but the probability is less than 0.8.
______(c) It is just as probable that the sideswiper was green as that it was blue.
______(d) It is more likely than not that the sideswiper was green.

This question, like Odd Question 2, was invented by Amos Tversky and Daniel Kahneman. They have done very extensive psychological testing on this question, and found that many people think that (a) or (b) is correct. Very few think that (d) is correct. Yet (d) is, in the natural probability model, the right answer! Here is how Bayes' Rule answers the question.

Solution

Let G = A taxi selected at random is green. Pr(G) = 0.85.
Let B = A taxi selected at random is blue. Pr(B) = 0.15.
Let W_b = The witness states that the taxi is blue.
$\Pr(W_b/B) = 0.8$.

Moreover, $\Pr(W_b/G) = 0.2$, because the witness gives a *wrong* answer 20% of the time, so the probability that she says "blue" when the cab was green is 20%.

We require $\Pr(B/W_b)$ and $\Pr(G/W_b)$.

$$\Pr(B/W_b) = \frac{\Pr(B)\Pr(W_b/B)}{\Pr(B)\Pr(W_b/B) + \Pr(G)\Pr(W_b/G)}$$

$$\Pr(B/W_b) = (.15 \times .8) \,/\, [(.15 \times .8) + (.85 \times .2)] = 12/29 \approx 0.41$$

Answer:

$\Pr(B/W_b) \approx 0.41$.
$\Pr(G/W_b) \approx 1 - 0.41 = 0.59$.
It is more likely that the sideswiper was green.

BASE RATES

Why do so few people feel, intuitively, that (d) is the right answer? Tversky and Kahneman argue that people tend to ignore the *base rate* or background information. We focus on the fact that the witness is right 80% of the time. We ignore the fact that most of the cabs in town are green.

Suppose that we made a great many experiments with the witness, randomly selecting cabs and showing them to her on a misty night. If 100 cabs were picked at random, then we'd expect something like this:

The witness sees about 85 green cabs. She correctly identifies 80% of these as green: about 68.
She incorrectly identifies 20% as blue: about 17.
She sees about 15 blue cabs. She correctly identifies 80% of these as blue: about 12.
She incorrectly identifies 20% as green: about 3.

So the witness identifies about 29 cabs as blue, but only 12 of these are blue! In fact, the more we think of the problem as one about frequencies, the clearer the Bayesian answer becomes.

Some critics say that the taxicab problem does not show that we make mistakes easily. The question is asked in the wrong way. If we had been asked just about frequencies, say the critics, we would have given pretty much the right answer straightaway!

RELIABILITY

Our witness was pretty reliable: right 80% of the time. How can a reliable witness not be trustworthy? Because of the base rates. We tend to confuse two different ideas of "reliability."

Idea 1: $\Pr(W_b/B)$: How reliable is she at identifying a cab as blue, given that it is in fact blue? This is a characteristic of the witness and her perceptual acumen.

Idea 2: $\Pr(B/W_b)$: How well can what the witness said be relied on, given that she said the cab is blue? This is a characteristic of the witness and the base rate.

FALSE POSITIVES

Base rates are very striking with medical diagnoses. Suppose I am tested for a terrible disease. I am told that the test is 99% right. If I have the disease, the test says YES with probability 99%. If I do not have the disease, it says NO with probability 99%.

I am tested for the disease. The test says YES. I am terrified.

But suppose the disease is very rare. In the general population, only one person in 10,000 has this disease.

Then among one million people, only 100 have the disease.

In testing a million people at random, our excellent test will answer YES for about 1% of the population, that is, 10,000 people. But as we see by a simple calculation in the next section, *at most 100 of these people actually have the disease*! I am relieved, unless I am in a population especially at risk.

I was terrified by a result YES, plus the test "reliability" (*Idea 1*):

Pr(YES/I'm sick).

But I am relieved once I find out about the "reliability" of a test result (*Idea 2*):

Pr(I'm sick/YES).

A test result of YES, when the correct answer is NO, is called a *false positive*. In our example, about 9,900 of the YES results were false positives.

Thus even a very "reliable" test may be quite misleading, if the base rate for the disease is very low. Exactly this argument was used against universal testing for the HIV virus in the entire population. Even a quite reliable test would give far too many false positives. Even a reliable test can be trusted only when applied to a population "at risk," that is, where the base rate for the disease is substantial.

PROBABILITY OF A FALSE POSITIVE

The result of testing an individual for a condition D is *positive* when according to the test the individual has the condition D.

The result of testing an individual for a condition D is a *false positive* when the individual does not have condition D, and yet the test result is nevertheless positive.

How much can we rely on a test result? This is *Idea 2* about reliability. The probability of a false positive is a good indicator of the extent to which you should rely on (or doubt) a test result.

Let D be the hypothesis that an individual has condition D.
Let Y be YES, a positive test result for an individual.
A false positive occurs when an individual does not have condition D, even though the test result is Y.
The probability of a false positive is $Pr(\sim D/Y)$.

In our example of the rare disease:

The base rate is $Pr(D) = 1/10{,}000$. Hence $Pr(\sim D) = 9{,}999/10{,}000$.
The test's "reliability" (*Idea 1*) is $Pr(Y/D) = 0.99$.
And $Pr(Y/\sim D) = 0.01$.

Applying Bayes' Rule,

$$Pr(-D/Y) = \frac{Pr(-D)Pr(Y/-D)}{Pr(\sim D)Pr(Y/\sim D) + Pr(D)Pr(Y/D)} = 9999/(9999 + 99) \approx 0.99.$$

STREP THROAT: ODD QUESTION 6

6. You are a physician. You think it is quite likely that one of your patients has strep throat, but you aren't sure. You take some swabs from the throat and send them to a lab for testing. The test is (like nearly all lab tests) not perfect.

If the patient has strep throat, then 70% of the time the lab says YES. But 30% of the time it says NO.
If the patient does **not** have strep throat, then 90% of the time the lab says NO. But 10% of the time it says YES.

You send five successive swabs to the lab, from the same patient. You get back these results, in order:

YES, NO, YES, NO, YES

You conclude:
______(a) These results are worthless.
______(b) It is likely that the patient does **not** have strep throat.
______(c) It is **slightly** more likely than not, that the patient **does** have strep throat.
______(d) It is **very much more** likely than not, that the patient **does** have strep throat.

In my experience almost no one finds the correct answer very obvious. It looks as if the yes-no-yes-no-yes does not add up to much. In fact, it is very good evidence that your patient has strep throat.

Let S = the patient has strep throat.
Let ~S = the patient does not have strep throat.
Let Y = a test result is positive.
Let N = a test result is negative

You think it likely that the patient has strep throat. Let us, to get a sense of the problem, put a number to this, a probability of 90%, that the patient has strep throat. $Pr(S) = 0.9$.

Solution
We know the conditional probabilities, and we assume that test outcomes are independent.

$\Pr(Y/S) = 0.7 \qquad \Pr(N/S) = 0.3$
$\Pr(Y/{\sim}S) = 0.1 \qquad \Pr(N/{\sim}S) = 0.9$

We need to find Pr(S/YNYNY).

$\Pr(YNYNY/S) = 0.7 \times 0.3 \times 0.7 \times 0.3 \times 0.7 = 0.03087$
$\Pr(YNYNY/{\sim}S) = 0.1 \times 0.9 \times 0.1 \times 0.9 \times 0.1 = 0.00081$

$$\Pr(S/YNYNY) = \frac{\Pr(S)\Pr(YNYNY/S)}{\Pr(S)\Pr(YNYNY/S) + \Pr({\sim}S)\Pr(YNYNY/{\sim}S)}$$

$$\Pr(S/YNYNY) = \frac{0.9 \times 0.03087}{(0.9 \times 0.03087) + (0.1 \times 0.00081)} = 0.997$$

Or you can do the calculation with the original figures, most of which cancel, to give Pr(S/YNYNY) = 343/344. Starting with a prior assumption that Pr(S) = 0.9, we have found that Pr(S/YNYNY) is almost 1!

Answer: So (d) is correct: *It is very much more* likely than not, that the patient *does* have strep throat.

SHEER IGNORANCE

But you are not a physician. You cannot read the signs well. You might just as well toss a coin to decide whether your friend has strep throat. You would model your ignorance as tossing a coin:

Pr(S) = 0.5.

Then you learn of the test results. Should they impress you, or are they meaningless? You require Pr(S/YNYNY).

Solution
Using the same formula as before, but with Pr(S) =0.5,
$\Pr(S/YNYNY) = (.5 \times .03087)/[(.5 \times .03087) + (.5 \times .00081)] \approx 0.974.$
Or, exactly, 343/352.

Answer: This result shows once again that the test results YNYNY are *powerful* evidence that your friend has strep throat.

REV. THOMAS BAYES

Bayes' Rule is named after Thomas Bayes (1702–1761), an English minister who was interested in probability and induction. He probably disagreed strongly with the Scottish philosopher David Hume about evidence. Chapter 21 explains how one might evade Hume's philosophical problem about induction by using Bayesian ideas.

Bayes wrote an essay that was published in 1763 (after his death). It contains the solution to a sophisticated problem like the examples given above. He imagines that a ball is thrown onto a billiard table. The table is "so made and leveled" that a ball is as likely to land on any spot as on any other. A line is drawn through the ball, parallel to the ends of the table. This divides the table into two parts, A and B, with A at a distance of *a* inches from one end.

Now suppose you do not know the value of *a*. The ball has been thrown behind your back, and removed by another player.

Then the ball is thrown *n* times. You are told that on *k* tosses the ball falls in segment A of the table, and in $n-k$ tosses it falls in segment B. Can you make a guess, on the basis of this information, about the value of *a*? Obviously, if most of the balls fell in A, then *a* must cover most of the length of the table; if it is about 50:50 A and B, then *a* should be about half the length of the table.

Thomas Bayes shows how to solve this problem exactly, finding, for any distance x, and any interval ε, the probability that the unknown *a* lies between (x−ε) and (x+ε).

The idea he used is the same as in our examples, but the mathematics is hard. What is now called *Bayes' Rule* (or, misleadingly, Bayes' Theorem) is a trivial simplification of Bayes' work. In fact, as we saw in Chapter 4, all the work we do with Bayes' Rule can be done from first principles, starting with the definition of conditional probability.

EXERCISES

1 *Lamps and triangles.* Use Bayes' Rule to solve 2(c), and 3(c) in the exercises for Chapter 5, page 56.

2 *Double dipping.*
Contents of urn A: 60 red, 40 green balls.
Contents of urn B: 10 red, 90 green balls.
An urn is chosen by flipping a fair coin.
(a) Two balls are drawn from this urn with replacement. Both are red. What is the probability that we have urn A?
(b) Two balls are drawn from this urn without replacement. Both are red. What is the probability that we have urn A?

3 *Tests.* A professor gives a true-false examination consisting of thirty T-F questions. The questions whose answers are "true" are randomly distributed among the thirty questions. The professor thinks that ¾ of the class are serious, and have correctly mastered the material, and that the probability of a correct answer on any question from such students is 75%. The remaining students will answer at random. She glances at a couple of questions from a test picked haphazardly. Both questions are answered correctly. What is the probability that this is the test of a serious student?

4 *Weightlifters.* Recall the coach that sent one of two teams for competition (page 54 above). Each team has ten members. Eight members of the Steroid team (S) use steroids (U). Two members of the Cleaner team (C) use steroids. The coach chooses which team to send for competition by tossing a fair coin.

One athletics committee tests for steroids in the urine of only one randomly chosen member of the team that has been sent. The test is 100% effective. If this team member is a user, the team is rejected.
(a) What would be a false positive rejection of the entire team?
(b) What is the probability of a false positive?
(c) Another committee is more rigorous. It randomly chooses two different members. What is the probability of a false positive?

5 *Three hypotheses.* (a) State Bayes' Rule for the conditional probability Pr(F/E) with three mutually exclusive and exhaustive hypotheses, F, G, H. (b) Prove it.

6 *Computer crashes.* A small company has just bought three software packages to solve an accounting problem. They are called Fog, Golem, and Hotshot. On first trials, Fog crashes 10% of the time, Golem 20% of the time, and Hotshot 30% of the time.

Of ten employees, six are assigned Fog, three are assigned Golem, and one is assigned Hotshot. Sophia was assigned a program at random. It crashed on the first trial. What is the probability that she was assigned Hotshot?

7 *Deterring burglars.* This example is based on a letter that a sociologist wrote to the daily newspaper. He thinks that it is a good idea for people to have handguns at home, in order to deter burglars. He states the following (amazing) information:

> The rate with which a home in the United States is burgled at least once per year is 10%. The rate for Canada is 40%, and for Great Britain is 60%. These rates have been stable for the past decade.

Don't believe everything a professor says, especially when he writes to the newspaper! Suppose, however, that the information is correct as stated, and that:

> Jenny Park, Larry Chen, and Ali Sami were trainee investment bankers for a multinational company. During the last calendar year Jenny had a home in the United States, Larry in Great Britain, and Ali in Canada.

One of the trainees is picked at random. This person was burgled last year. What is the probability that this person was Ali?

KEY WORDS FOR REVIEW

Bayes' Rule
Base rates
False positives

8 Expected Value

Inductive logic is risky. We need it when we are uncertain. Not just uncertain about what will happen, or what is true, but also when we are uncertain about what to do. Decisions need more than probability. They are based on the value of possible outcomes of our actions. The technical name for value is **utility**. This chapter shows how to combine probability and utility. But it ends with a famous paradox.

ACTS

Should you open a small business?
Should you take an umbrella?
Should you buy a Lotto ticket?
Should you move in with someone you love?

In each case you settle on an *act*. Doing nothing at all counts as an act.

Acts have *consequences*.

You go broke (or maybe found a great company).
You stay dry when everyone else is sopping wet (or you mislay your umbrella).
You waste a dollar (or perhaps win a fortune).
You live happily ever after (or split up a week later).
You do absolutely nothing active at all: that counts as an act, too.

Some consequences are desirable. Some are not. Suppose you can represent the cost or benefit of a possible consequence by a number—so many dollars, perhaps. Call that number the *utility* of the consequence.

Suppose you can also represent the *probability* of each possible consequence of an act by a number.

In making a decision, we want to assess the relative merits of each possible