CHAPTER 1

Logic

1. Certainty & Uncertainty

There are some things we know with *certainty*. These things are true or false given some evidence or just because they are obviously true or false. There are many more things about which we are *uncertain*. These things are more or less probable given some evidence. And there are still more things of which nobody can ever quantify the uncertainty. These things are nonsensical or paradoxical.

First I want to prove to you there are things that are true, but which cannot be proved to be true, and which are true based on *no* evidence. Suppose some statement A is true (A might be shorthand for "I am a citizen of Planet Earth"; writing just 'A' is easier than writing the entire statement; the statement is everything between the quotation marks). Also suppose some statement B is true (B might be "Some people are frightfully boring"). Then this statement: "A and B are true", is true, right? But also true is the statement "B and A are true". We were allowed to reverse the letters A and B and the joint statement stayed true. Why? Why doesn't switching make the new statement false? Nobody knows. It is just assumed that switching the letters is valid and does not change the truth of the statement. The operation of switching does not change the truth of statements like this, but nobody will ever be able to prove or explain why switching has this property. If you like, you can say we take it on faith.

That there are certain statements which are assumed true based on no evidence will not be surprising to you if you have ever studied mathematics. The basis of all mathematics rests on beliefs which are assumed to be true but cannot be proved to be true. These beliefs are called *axioms*. Axioms are the base;

theorems, lemmas, and proofs are the bricks which build upon the base using rules (like the switching statements rule) that are also assumed true. The axioms and basic rules cannot, and can never, be proved to be true. Another way to say this is, "We hold these truths to be self-evident."

Here is one of the axioms of arithmetic: For all natural numbers x and y, if $x = y$, then $y = x$. Obviously true, right? It is just like our switching statements rule above. There is no way to prove this axiom is valid. From this axiom and a couple of others, plus acceptance of some manipulation rules, all of mathematics arises. There are other axioms—two, actually—that define probability. Here, due to Cox (1961), is one of those axioms: The probability of a statement on given evidence determines the probability of its contradictory on the same evidence. I'll explain these terms as we go.

It is the job of logic, probability, and statistics to quantify the amount of certainty any given statement has. An example of a statement which might interest us: "This new drug improves memory in Alzheimer patients by at least ten percent." How probable is it that that statement is true given some specific evidence, perhaps in the form of a clinical trial? Another statement: "This stock will increase in price by at least two dollars within the next thirty days." Another: "Marketing campaign B will result in more sales than campaign A." In order to specify how probable these statements are, we need evidence, which usually comes in the form of *data*. Manipulating data to provide coherent evidence is why we need statistics.

Manipulating data, while extremely important, is in some sense only mechanical. We must always keep in mind that our goal is to make sense of the world and to quantify the uncertainty we have in given problems. So we will hold off on playing with data for several chapters until we understand exactly what probability really means.

2. Logic

We start with simple logic. Here is a classical logical argument, slightly reworked:

All statistics books are boring.

Stats 101 is a statistics book.

Therefore, Stats 101 is boring.

The structure of this argument can be broken down as follows. The two statements above the horizontal line are called *premises*; they are our *evidence* for the statement below the line, which is the *conclusion*. We can use the words "premises" and "evidence" interchangeably. We want to know the probability that the conclusion is true given these two premises. Given the evidence listed, it is 1 (probability is a number between, and including, 0 and 1). The conclusion is true given these premises. Another way to say this is the conclusion is *entailed* by the premises (or evidence).

You are no doubt tempted to say that the probability of the conclusion is not 1, that is, that the conclusion is *not* certain, because, you say to yourself, statistics is nothing if not fun. But that would be missing the point. You are not free to add to the evidence (premises) given. You *must* assess the probability of the conclusion given *only* the evidence provided.

This argument is important because it shows you that there are things we can know to be true *given* certain evidence. Another way to say this, which is commonly used in statistics, is that the conclusion is true *conditional* on certain evidence.

Here is another argument, courtesy (in form, at least) of David Hume (2003):

All the reality TV shows I have observed before have been ridiculous.

This is a (new) reality show before me.

Therefore, this reality show will be ridiculous.

The conclusion here does not follow from the premises; that is, the conclusion is not certainly true, nor is it certainly false (its probability is not 1 nor 0). You may be surprised to learn this, but the universe is not set up to guarantee that all reality TV shows will be ridiculous. It may be that, for whatever unknown reason, that *this* new show will not be ridiculous. The conclusion, then, is *contingent* on certain facts (about network executives, uncontrollable weeping of contestants, viewers' habits, etc.), and any conclusion that is contingent (on certain conditions about the universe holding) is never certainly true nor certainly false. So what is the probability that the conclusion is true? Pretty high, but not 1 and not 0. We don't need to, and there is nothing in the universe that guarantees that we can, put an exact number of this probability. (Many arguments are non-numerical, see Keynes (2004); Franklin (2001b).)

Another argument:

> I will roll a die, which has six sides, only one of which will show.
>
> Just 1 side of the six is labeled "6.".
>
> ---
>
> Therefore, the side that shows will be a "6."

The conclusion here is also not certain, as will be plainly obvious to any of us. The conclusion is contingent (not certainly true or false) given just the evidence in the two premises, and the probability that the conclusion is true is 1 in 6, or about 0.17. You knew this before reading this book, but you might not have seen it written out like this before.

Here is a very difficult argument to understand, but it is important, so we will take our time with it:

T

M

T is any tautology, which is a statement that is necessarily true, or always true no matter what: an example of a tautology is T = "Either Joe is a pain in the ass, or he is not." The statement (all the stuff inside the quotation marks) T is always true, is it not? Another tautology, T = "Tomorrow it will rain or it will not." In this book, whenever we see "T" is means a true statement; thus the probability of T being true is 1. A shorthand way to say this is the probability of T is 1 (we can leave out the "being true"). There are other true statments that are not necessarily true. One appeared in the argument before this, "Just 1 side of the six is labeled '6'" is a true statement just in case there is actually a die that has one side with a 6 on it. This is an observation, and a true one, but it is not necessarily true. The die *could* have no sides with a 6 on it.

M is just some statement, which I'll leave undefined for a moment to make a point. It should be obvious that if we know nothing about M, we cannot state any probability about it. So if you let T be, for example, the tautology about Joe, and I did not tell you anything about M, then the probability of M being true is undefined. Thus, it is possible that some statements have no probability (of being true) at all.

Let's change our tautology to T = "M is true or false," which is another way of saying, T = "M will happen or it won't." These tautologies have buried within them implicit information[1] about M, which is that M can happen or not. So it must be physically possible for M to be true. If I add evidence that M is physically possible, but not certain, then we are saying some positive thing about M, it is information about M that is useful. With this new information about M (implicit) in the premises, we can then state a probability of M being true. However, the evidence is pretty weak: saying something *might* be true doesn't say much. So the

[1]Tautologies can easily contain implicit information because of the flexibility and nuances of human language. For example, T = "Either utopians are deluded or they are not." We generally ignore this kind of information.

best we can do is to say the probability of M is greater than 0 but less than 1.

Another, technical, way to say this is that M is *contingent*. Pause here, because this is a great truth. It is a reminder that all contingent statements, which we do not know the truth or falsity of, have a probability between 0 and 1. As long as the premises we have do not entail M, then we will *never* know whether M is true or false; the best we can ever do is to say M is more or less probable (Briggs, 2007).

Much later we will meet arguments that look like this:

> I have collected a bunch of data, which I will call x and y.
>
> I want to describe my uncertainty in future values of x and y using probability models M_x and M_y.
>
> ---
>
> $f(x) > f(y)$

It will be our job to quantify the uncertinty of the statement $f(x) > f(y)$ [$f()$ is any function of the data of interest to us; and if you don't know what a function is, stick around]. Our job as statisticians is to collect the evidence (the data) and create the models that make up the two premises of the argument. The job of defining the conclusion (which is some interesting question about the data) is usually given to us by the powers that be. We use probability to tie the whole process together.

3. What is probability?

For us, **probability is a measure of the logical relation between a list of premises (or observation statements, or evidence) and some conclusion.**

This is not the only interpretation of probability. There are many more, but only two of them have large followings. The older and largest is called *frequentism*, the newer and second largest is called *Bayesianism*, named after Thomans Bayes, the man who first proved a theorem we will meet in the next Chapter. The later camp is divided into two sects, the *subjective* Bayesian and the logical or *objective Bayesian*, which we follow in this book. The differences between the two flavors of Bayesianism are slight, and have to do with if probabilities are a matter of human beliefs. Subjectivists believe that all probabilities exist only in and only for human minds; objectivists can show that most, if not all, probabilities are matters of logic. But these are mere quibbles in some sense, because the math and the methods are nearly the same for both. The real difference is between the Bayesians and the frequentists. Frequentist theory arose because of its practioners' distaste in the rampant subjectivism and relativism in an earlier incarnation of Bayesianism. The classical theorists wanted to develop probability on a solid, objective ground, freed from human opinion, and this is much to their credit. However, people are starting to realize that this approached failed (Howson and Urbach, 1993; Berger and Selke, 1987; Little, 2006).

For frequentists, probability is defined to be the frequency with which an event happens in the limit of "experiments" where that event can happen; that is, given that you run a number of "experiments" that approach infinity, then the ratio of those experiments in which the event happens to the total number of experiments is *defined* to be the probability that the event will happen. This obviously cannot tell you what the probability is for your well-defined, possibly unique, event happening now, but can only give you probabilities in the limit, after an infinite amount of time has elapsed for all those experiments to take place. There are precise mathematical definitions of these terms, but this is the (confusing) gist of it. Now, once this mathematical definition is in place, it is easy to start creating and proving theorems, which are mathematical statements that are the result of the given definition. I must stress that this is a perfectly legitimate activity. The theorems that arise from frequentist probability are of interest as mathematical objects, just as are the theorems from any branch of mathematics, however esoteric. Frequentist probability statements are, therefore, true, given this definition of probability.

But this does not imply, and it is not true, that these frequentist probability statements have any connection with reality, or with any physical counterparts. We must never mistake math for reality! Sometimes, mathematics and reality match up very well. In fact, an entire branch of math has sprung up out of this idea; it's called, not surprisingly, *applied mathematics*. You already know that a lot of mathematics exists solely for itself; this branch is usually called *pure mathematics*. Applications arising from pure math are usually unintentional. Frequentist probability and statistics theory are also pure mathematics in this sense.

Very well. All these detailed explanations are to prepare you for some of the odd methods and the even odder interpretations of these methods that have arisen out of frequentist probability theory over the past $\sim$ 100 years. We will meet these methods later in this book, and you will certainly meet them when reading results produced by other people. You will be well equipped, once you finish reading this book, to understand common claims made with classical statistics, and you will be able to understand its limitations.

4. Why isn't probability subjective?

If 3 out of 4 dentists agree that using Dr Johnston's Whitening Powder makes for shiny teeth, what is the probability that *your* dentist thinks so? Given *only* the evidence (premises) that 3 out of 4 etc., then we know the probability is 0.75 that your dentist likes Dr Johnston's Whitening Powder.

But what if you learned your dentist had just attended an "informational seminar" (with free lunch) sponsored by Galaxy Pharmaceuticals, the manufacturer of Dr Johnston's Whitening Powder? This introduces new evidence, and will therefore modify the probability that your doctor would recommend Dr Johnston's.

It may suddenly seem that probability *is* a matter of belief, of subjective feeling, because different people will have different opinions on how the free lunch will affect the doctor's endorsement. Probability cannot always be a matter of free choice, however. For example, knowing only that a die has 6 sides, and knowing *nothing else* except that the outcome of the die toss is contingent, then the probability of seeing a 6 is 1 in 6, or about 0.17, regardless of what you or I or anybody thinks.

After you learn of your doc's cozying up to the pharmaceutical representative, *you* would be inclined to increase your probability that he would tout Dr Johnston's to, say, the extent of 0.95. *I* may come to a different conclusion, say, 0.76 (just slightly higher). Why? Because we are using *different sets or collections of information*, different evidence or premises, which naturally change our probability assesments. You might know more about pharmaceutical companies than I do, for example, and this causes you to be more cynical, whereas I know more about the purity and selflessness of doctors, and this causes me to be trusting.

But, if I agreed with you about the new evidence, and I felt it was as relevant as you did, then we would share the same probability. This, of course, is very unlikely to happen. Rarely will two people agree on a list of premises when the argument involves human affairs, and so it is natural that for most complex things, people will come to different probabilities that the conclusions are true. Does this remind you of politics?

Because people never agree on the set of premises—and they cannot merely agree on them, they have to agree on them *exactly*—probabilities will differ. In this sense, probabilities are subjective—rather, it is the choice of premises that is subjective. The probabilities assigned to a conclusion *given* a set of premises is not. The probability of a conclusion always follows logically from the given premises.

5. Randomness

There is a great deal of nonsense written and said about randomness. Although it's never stated directly, there is a certain mysticism about randomness which is supposed to "bless" statistical results (samples must be "random" to be statistically valid, it is often said). We will take care of these beliefs when we meet them.

When we say something is random, all we are saying is that we don't know what that thing is. If that thing is, for example, the result of a dice roll, then we are saying that we do not know in advance what the outcome will be. That is, the outcome is "random." Only this, and nothing more. Barring quantum mechanics, about which Richard Feynman said fairly that nobody understands, there is nothing spooky or mysterious behind events that are random. Not knowing what something is, is saying that the truth of that thing is unknown, or "random." Technically, then, randomness means ignorance.

6. A taste of Boolean algebra

George Boole, in his 1854 book *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*, introduced a calculus for working with statements which is now called *Boolean Algebra*. We have been using this calculus so far, although you didn't know it because the manipulations have been mostly intuitive. But there are some formal rules that we'll need before we get much further. Our only real problem will be notation. You can look in a dozen books and see thirteen different notations, all of which mean the same thing, but which you must assimilate anew each time before you can understand what the authors are saying. There are some consistencies, but very few. I try to stick with the most-used symbols, but be warned: if you read another book, you will likely see a different notation.

We use capital Latin letters to stand for statements, which may be true or false. So A = "Real estate agents never lie" is a statement which is either true or false, but we might not know whether it is true or false. If we have another statement, B, and we write AB, this means the joint statement "A *and* B". So if B = "The *New York Times* always reports stories in an absolutely unbiased manner", then AB means "Agents never lie *and* the *Times* is always unbiased." Both parts of the statement must be true for AB to be (jointly) true.

When we write $\mathrm{A} \cup \mathrm{B}$, it means "A *or* B." If either A *or* B is true, or both are true, then the statement "$\mathrm{A} \cup \mathrm{B}$" is true.

There are, unfortunately, many ways to write that a statement is false. None of them is especially beautiful. One way, the one we'll use here, is A^F. Thus, if A = "Real estate agents never lie", then A^F = "Real estate agents sometimes lie." If A = "A Head will show when I flip this coin", then A^F = "A Head will not show when I flip this coin", which is a fancy way of saying we'll see a tail. Other ways of writing A is false: A^c (the c means complement, or opposite) and ¬A.

Sometimes we need a truth, a statement which is true. We'll label these statements T. There are statements that are true just because the universe was in a certain position, meaning we observed the statement to be true. For example, it might be true that T = "you saw a dog on your lawn last week" just in case you actually did see a dog on yor lawn last week. All observed data statements are truths in this sense. For example, T = "In patient 37, I measured a systolic blood pressure of 182 mmHg." There are other statements which are necessarily true, which are true regardless of anything. Tautologies, which we met earlier, are the most common examples of these truths.

7. Mechanics

Understanding logic and interpreting probability are only a small portion of a statistics course. The meat is in the actual mechanics of the methods: how to calculate a mean and so forth. A great deal of statistics evolved before easy access to computers became usual. Therefore, many of the methods, and even most or all that are traditionally taught in introductory courses were designed to be calculated by hand and by referrals to standardized tables. This often meant that certain crude assumptions were made which would greatly simplifyy the calculations involved. This was certainly a rational thing to do at the time.

Naturally, now that computers have become so cheap that even professors can afford them, the methods you learn in statistics classes will have changed to reflect this fact, right?

Uh...no, that's wrong, actually.

You will still find legions of students churning out sums of squares of X, sums of squares of Y, total sums of squares, computing means and variances by plugging in numbers into calculators, and looking up probabilities in tables with small print in the backs of ridiculously heavy books. I can think of two reasons for this: (1) All the textbooks are written in this old-fashioned way, and all the courses are already created around these textbooks. Do you have any idea how long it takes to develop a new college course? Not just writing new textbooks, but also creating thousands of new homework questions and exams, and answers for them, too, and on and on? Just don't even ask. (2) A lot of us professors have progressed a long way toward fogey-hood, and we feel that, since we had to calculate sums of squares by hand, then, by golly, our students will too! It builds character.

8. Homework

(1) Rewrite the first logical argument (with the conclusion "Stats 101 is boring") using one or more different premisses such that the conclusion has probability 0. Rewrite it again so that it has a probability between 0 and 1.

(2) Rewrite the second argument (with the conclusion "This reality show will be ridiculous") using one or more different premisses such that the conclusion has probability 1.

(3) What is the probability of drawing the Jack of Hearts from a standard deck of playing cards? Write your argument in the same form as the dice example.

(4) Alice hands you a deck of playing cards which she says are well shuffled. Bob hands you another deck and says nothing. What is the probability of drawing the Jack of Hearts from Alice's deck and what is it from Bob's deck? Explain your answer.

(5) Charlie hands you a third deck, but as he does so, he gives you a wink. What is the probability of drawing the Jack of Hearts from Charlie's deck? Write your answer in the form of a logical argument. Be clear about your premises.

(6) There is a famous, if not tedious, statement that goes L = "This statement is false." What is the probability that L is true? Explain how you arrived at your answer.

(7) Right before I come to class, I put either a quarter or a dime in my pocket. Once I get there, I pull out the coin and conceal it from your view. What is the probability that I reveal the quarter? Write your answer in the form of a logical argument. Be clear about your premises.

(8) Bounding probabilities. Is it possible to translate the statement "Given evidence E (about the past performance, knowledge of the starting lineup, etc.), I conclude the Detroit Tigers will most likely win tomorrow's game" into a numerical value? Explain how you arrived at your answer.

(9) Create your own tautolgy T (different from those in the text).

(10) B = "The sun rises in the west." What it the probability of BT and B∪T, where T is the tautology from the previous problem.

(11) What is the probability that A = "Wearing white shoes after Labor Day is wrong"? Explain.
(12) Write out a list of premises (be explicit) for the Dr Johnston's Whitening Powder example supposing you learned your doctor did have that lunch, then give a guess for the probability of the conclusion.
(13) A premiss of moral relativism, often called upon in Postmodern literary theory and other highly-educated people, is C = "There is no truth." What is the probability C is true?
(14) If A = "Real estate agents never lie", then A^F = "Real estate agents sometimes lie." Why isn't A^F = "Real estate agents *always* lie"?
(15) EXTRA In the argument T, therefore M, where T is the tautolgy "M will happen or it won't", why isn't the probability of M 1/2?