Chapter 4 Measures

4.1 Probability Defined

We talk informally about probabilities all the time. We might say that it is more probable than not that such-and-such team will make the playoffs. Or we might say that it's very probable that a particular defendant will be convicted at his trial. Or that it isn't very probable that the next card will be the one we need to complete this royal flush.

We also talk formally about probability in mathematical contexts. Formally, a probability function is a normalised measure over a possibility space. Below we'll be saying a fair bit about what each of those terms mean. We'll start with *measure*, then say what a *normalised measure* is, and finally (over the next two days) say something about *possibility spaces*.

There is a very important philosophical question about the connection between our informal talk and our formal talk. In particular, it is a very deep question whether this particular kind of formal model is the right model to represent our informal, intuitive concept. The vast majority of philosophers, statisticians, economists and others who work on these topics think it is, though as always there are dissenters. We'll be spending a fair bit of time later in this course on this philosophical question. But before we can even answer that question we need to understand what the mathematicians are talking about when they talk about probabilities. And that requires starting with the notion of a measure.

4.2 Measures

A measure is a function from 'regions' of some space to non-negative numbers with the following property. If A is a region that divides exactly into regions B and C, then the measure of A is the sum of the measures of B and C. And more generally, if A divides exactly into regions B_1, B_2, ..., B_n, then the measure of A will be the sum of the measures of B_1, B_2, ... and B_n.

Here's a simple example of a measure: the function that takes as input any part of New York City, and returns as output the population of that part. Assume that the following numbers are the populations of New York's five boroughs. (These numbers are far from accurate.)

Borough	Population
Brooklyn	2,500,000
Queens	2,000,000
Manhattan	1,500,000
The Bronx	1,000,000
Staten Island	500,000

We can already think of this as a function, with the left hand column giving the inputs, and the right hand column the values. Now if this function is a *measure*, it should be additive in the sense described above. So consider the part of New York City that's on Long Island. That's just Brooklyn plus Queens. If the population function is a measure, the value of that function, as applied to the Long Island part of New York, should be 2,500,000 plus 2,000,000, i.e. 4,500,000. And that makes sense: the population of Brooklyn plus Queens just is the population of Brooklyn plus the population of Queens.

Not every function from regions to numbers is a measure. Consider the function that takes a region of New York City as input, and returns as output the proportion of people in that region who are New York Mets fans. We can imagine that this function has the following values.

Borough	Mets Proportion
Brooklyn	0.6
Queens	0.75
Manhattan	0.5
The Bronx	0.25
Staten Island	0.5

Now think again about the part of New York we discussed above: the Brooklyn plus Queens part. What proportion of people in that part of the city are Mets fans? We certainly can't figure that out by just looking at the Brooklyn number from the above table, 0.6, and the Queens number, 0.75, and adding them together. That would yield the absurd result that the proportion of people in that part of the city who are Mets fans is 1.35.

That's to say, the function from a region to the proportion of people in that region who are Mets fans is *not* a measure. Measures are functions that are always additive over subregions. The value of the function applied to a whole region is the sum of the values the function takes when applied to the parts. 'Counting' functions, like population, have this property.

The measure function we looked at above takes real regions, parts of New York City, as inputs. But measures can also be defined over things that are suitably analogous to regions. Imagine a family of four children, named below, who eat the following amounts of meat at dinner.

Child	Meat Consumption (g)
Alice	400
Bruce	300
Chuck	200
Daria	100

We can imagine a function that takes a group of children (possibly including just one child, or even no children) as inputs, and has as output how many grams of meat those children ate. This function will be a measure. If the 'groups' contain just the one child, the values of the function will be given by the above table. If the group contains two children, the values will be given by the addition rule. So for the group consisting of Alice and Chuck, the value of the function will be 600. That's because the amount of meat eaten by Alice and Chuck just is the amount of meat eaten by Alice, plus the amount of meat eaten by Chuck. Whenever the value of a function, as applied to a group, is the sum of the values of the function as applied to the members, we have a measure function.

4.3 Normalised Measures

A measure function is defined over some regions. Usually one of those regions will be the 'universe' of the function; that is, the region made up of all those regions the function is defined over. In the case where the regions are regions of physical space, as in our New York example, that will just be the physical space consisting of all the smaller regions that are inputs to the function. In our New York example, the universe is just New York City. In cases where the regions are somewhat more metaphorical, as in the case of the children's meat-eating, the universe will also be defined somewhat more metaphorically. In that case, it is just the group consisting of the four children.

However the universe is defined, a normalised measure is simply a measure function where the value the function gives to the universe is 1. So for every sub-region of the universe, its measure can be understood as a proportion of the universe.

We can 'normalise' any measure by simply dividing each value through by the value of the universe. If we wanted to normalise our New York City population measure, we would simply divide all values by 7,500,000. The values we would then end up with are as follows.

Borough	Population
Brooklyn	$\frac{1}{3}$
Queens	$\frac{4}{15}$
Manhattan	$\frac{1}{5}$
The Bronx	$\frac{2}{15}$
Staten Island	$\frac{1}{3}$

Some measures may not have a well-defined universe, and in those cases we cannot normalise the measure. But generally normalisation is a simple matter of dividing everything by the value the function takes when applied to the whole universe. And the benefit of doing this is that it gives us a simple way of representing proportions.

4.4 Formalities

So far I've given a fairly informal description of what measures are, and what normalised measures are. In this section we're going to go over the details more formally. If you understand the concepts well enough already, or if you aren't familiar enough with set theory to follow this section entirely, you should feel free to skip forward to the next section. Note that this is a slightly simplified, and hence slightly inaccurate, presentation; we aren't focussing on issues to do with infinity. Those will be covered later in the course.

A measure is a function m satisfying the following conditions.

1. The domain D is a set of sets.
2. The domain is closed under union, intersection and complementation with respect to the relevant universe U. That is, if $A \in D$ and $B \in D$, then $(A \cup B) \in D$ and $(A \cup B) \in D$ and $U \setminus A \in D$
3. The range is a set of non-negative real numbers
4. The function is additive in the following sense: If $A \cap B = \emptyset$, then $m(A \cup B) = m(A) + m(B)$

We can prove some important general results about measures using just these properties. Note that we the following results follow more or less immediately from additivity.

1. $m(A) = m(A \cap B) + m(A \cap (U \setminus B))$
2. $m(B) = m(A \cap B) + m(B \cap (U \setminus A))$
3. $m(A \cup B) = m(A \cap B) + m(A \cap (U \setminus B)) + m(B \cap (U \setminus A))$

The first says that the measure of A is the measure of A's intersection with B, plus the measure of A's intersection with the complement of B. The first says that the measure of B is the measure of A's intersection with B, plus the measure of B's intersection with the complement of A. In each case the point is that a set is just made up of its intersection with some other set, plus its intersection with the complement of that set. The final line relies on the fact that the union of A and B is made up of (i) their intersection, (ii) the part of A that overlaps B's complement and (iii) the part of B that overlaps A's complement. So the measure of $A \cup B$ should be the sum of the measure of those three sets.

Note that if we add up the LHS and RHS of lines 1 and 2 above, we get

$$m(A) + m(B) = m(A \cap B) + m(A \cap (U \setminus B)) + m(A \cap B) + m(A \cap (U \setminus B))$$

And subtracting $m(A \cap B)$ from each side, we get

$$m(A) + m(B) - m(A \cap B) = m(A \cap B) + m(A \cap (U \setminus B)) + m(A \cap (U \setminus B))$$

But that equation, plus line 3 above, entails that

$$m(A) + m(B) - m(A \cap B) = m(A \cup B)$$

And that identity holds whether or not $A \cap B$ is empty. If $A \cap B$ is empty, the result is just equivalent to the addition postulate, but in general it is a stronger result, and one we'll be using a fair bit in what follows.

4.5 Possibility Space

Imagine you're watching a baseball game. There are lots of ways we could get to the final result, but there are just two ways the game could win. The home team could win, call this possibility H, or the away team could win, call this possibility A.

Let's complicate the example somewhat. Imagine that you're watching one game while keeping track of what's going on in another game. Now there are four ways that the games could end. Both home teams could win. The home team could win at your game while the away team wins the other game. The away team could win at your game while the home team wins the other game. Or both away teams could win. This is a little easier to represent on a chart.

Your game	Other game
H	H
H	A
A	H
A	A

Here H stands for home team winning, and A stands for away team winning. If we start to consider a third game, there are now 8 possibilities. We started with 4 possibilities, but now each of these divides in 2: one where the home team wins the third game, and one where the away team wins. It's just about impossible to represent these verbally, so we'll just use a chart.

Game 1	Game 2	Game 3
H	H	H
H	H	A
H	A	H
H	A	A
A	H	H
A	H	A
A	A	H
A	A	A

Of course, in general we're interested in more things than just the results of baseball games. But the same structure can be applied to many more cases.

Say that there are three propositions, p, q and r that we're interested in. And assume that all we're interested in is whether each of these propositions is true or false. Then there are eight possible ways things could turn out, relative to what we're interested in. In the following table, each row is a possibility. T means the proposition at the head of that column is true, F means that it is false.

p	q	r
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

These eight possibilities are the foundation of the possibility space we'll use to build a probability function.

A measure is an additive function. So once you've set the values of the smallest parts, you've fixed the values of the whole. That's because for any larger part, you can work out its value by summing the values of its smaller parts. We can see this in the above example. Once you've fixed how much meat each child has eaten, you've fixed how much meat each group of children have eaten. The same goes for probability functions. In the cases we're interested in, once you've fixed the measure, i.e. the probability of each of the eight basic possibilities represented by the above eight rows, you've fixed the probability of all propositions that we're interested in.

For concreteness, let's say the probability of each row is given as follows.

p	q	r	
T	T	T	0.0008
T	T	F	0.008
T	F	T	0.08
T	F	F	0.8
F	T	T	0.0002
F	T	F	0.001
F	F	T	0.01
F	F	F	0.1

So the probability of the fourth row, where p is true while q and r are false, is 0.8. (Don't worry for now about where these numbers come from; we'll spend much more time on that in what follows.) Note that these numbers sum to 1. This is required; probabilities are **normalised** measures, so they must sum to 1.

Then the probability of any proposition is simply the sum of the probabilities of each row on which it is true. For instance, the probability of p is the sum of the probabilities of the first four rows. That is, it is 0.0008 + 0.008 + 0.08 + 0.8, which is 0.8888.

In the next class we'll look at how we tell which propositions are true on which rows. Once we've done that, we'll have a fairly large portion of the formalities needed to look at many decision-theoretic puzzles.