CIS 2033 Lecture 1, Spring 2017[1]

Instructor: David Dobor

January 2017

[1] Readings from the textbook: Chapter 1, Sec 1

We discuss administrativia and dive right into the material.

Introduction

A Probabilistic Model

A probabilistic model is a quantitative description of a phenomenon or an experiment whose outcome is uncertain. Putting together such a model involves two key steps. First, we need to describe the possible outcomes of the experiment. This is done by specifying a so-called sample space. And then by specifying a probability law, which assigns probabilities to outcomes or to collections of outcomes.

The probability law tells us, for example, whether one outcome is much more likely than some other outcome. Probabilities have to satisfy certain basic properties in order to be meaningful. For example, probabilities cannot be negative. These are the axioms of probability theory.

Interestingly, there will be very few axioms, but they are powerful, and we will see that they imply many other properties that were not part of the axioms.

We will then go through a couple of very simple examples involving models with either discrete or continuous outcomes. As you will see many times in this class, discrete models are conceptually much easier. Continuous models involve some more sophisticated concepts, and we will point out some of the subtle issues that arise.

Finally, we will talk a little bit about the big picture, about the role of probability theory, and its relation with the real world.

Building the Probabilistic Model

Putting together a probabilistic model– that is, a model of a random phenomenon or a random experiment– involves two steps:

1. In the first step we describe the possible outcomes of the phenomenon or experiment of interest.

2. In the second step we describe our beliefs about the likelihood of the different possible outcomes by specifying a probability law.

Step1: Describing Outcomes

So let's start with the description of the possible outcomes of an experiment. We carry out some experiment: for example, we flip a coin, or maybe we flip five coins simultaneously, or maybe we roll a die. Whatever that experiment is, it has a number of possible outcomes, and we start by making a list of the possible outcomes.

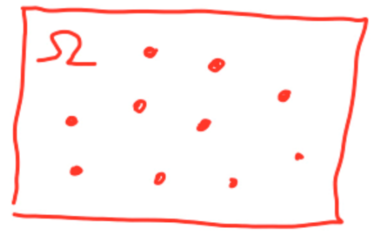


Figure 1: The sample space Ω can be represented by a drawing like this one. Each individual dot represents an outcome. The outcomes are mutually exclusive and make up the entire space Ω.

We create a set that we usually denote by capital omega: Ω. That set is called the sample space and is the set of all possible outcomes of our experiment. The elements of that set should have certain properties. Namely, the elements *should be mutually exclusive and collectively exhaustive.*

What does that mean? Mutually exclusive means that, if at the end of the experiment, I tell you that a particular outcome happened (a red dot), then it should not be possible for another outcome (some other red dot) also to have happened. At the end of the experiment, there can only be one of the outcomes that has happened.

Being collectively exhaustive means something else – that, together, all of these elements of the set exhaust all the possibilities. So no matter what, at the end, you will be able to point to one of the outcomes and say, that's the one that occurred.

To summarize– this set should be such that, at the end of the experiment, you should be always able to point to one, and exactly one, of the possible outcomes and say that this is the outcome that occurred.

Physically different outcomes should be distinguished in the sample space and correspond to distinct points. But when we say physically different outcomes, what do we mean? We really mean different in all relevant aspects but perhaps not different in irrelevant aspects.

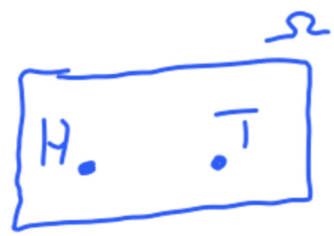


Figure 2: A sample space Ω with two outcomes.

Example 1. Let's make more precise what I mean by that by looking at a very simple, and maybe silly, example, which is the following. Suppose that you flip a coin and you see whether it resulted in heads or tails. So you have a perfectly legitimate sample space for this experiment which consists of just two points – one head and one tail.

Together these two outcomes exhaust all possibilities. And the two outcomes are mutually exclusive. So this is a legitimate sample space for this experiment. Now suppose that while you were flipping the coin, you also looked outside the window to check the weather. And then you could say that my sample space is really heads, and it's raining. Another possible outcome is heads and no rain. Another possible outcome is tails, and it's raining, and, finally, another possible outcome is tails and no rain. This set consisting of four elements (see Fig 3), is also a perfectly legitimate sample space for the experiment of flipping a coin. The elements of this sample space are mutually exclusive and collectively exhaustive. Exactly one of these outcomes will have materialized at the end of the experiment.

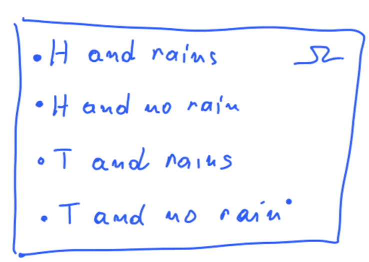


Figure 3: A sample space Ω with four outcomes.

So which sample space is the correct one? This sample space, the second one, involves some irrelevant details. So the preferred sample space for describing the flipping of a coin is the simpler one, the first one, which is sort of at the right granularity, given what we're interested in.

Ultimately, the question of which one is the right sample space depends on what kind of questions you want to answer. For example, if you have a theory that the weather affects the behavior of coins, then, in order to play with that theory, or maybe check it out, then you might want to work with the second sample space. This is a common feature in all of science. Whenever you put together a model, you need to decide how detailed you want your model to be. And the right level of detail is the one that captures those aspects that are relevant and of interest to you.

Check Your Understanding: For the experiment of flipping a coin, and for each one of the following choices, determine whether we have a legitimate sample space:

$$\Omega = \{\text{Heads and it is raining, Heads and it is not raining, Tails}\} \quad (1)$$

$$\Omega = \{\text{Heads and it is raining, Tails and it is not raining, Tails}\} \quad (2)$$

LET US now look at some more examples of sample spaces. Sample spaces are sets, and a set can be discrete, finite, infinite, continuous, and so on.

EXAMPLE 2: Let us start with a simpler case in which we have a sample space that is discrete and finite. The particular experiment we will be looking at is the following. We take a very special die, a tetrahedral die. So it's a die that has four faces numbered from 1 up 4. We roll it once. And then we roll it again. Were not dealing here with two probabilistic experiments. We're dealing with a single probabilistic experiment that involves two rolls of the die within that experiment. What is the sample space of that experiment?

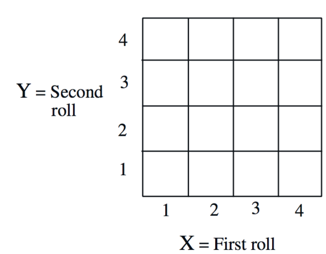


Figure 4: A representation for two rolls of a 4-sided die.

Well, one possible representation is the following. We take note of the result of the first roll. And then we take note of the result of the second roll. And this gives us a pair of numbers. Each one of the possible pairs of numbers corresponds to one of the little squares in this diagram. For example, if the first roll is 1 and the second is also 1, then this particular outcome can be represented by the lowest leftmost square. If the first roll is it 2 and the second is a 3, then this particular outcome can be represented by the square we might call "(2, 3)" (See Fig 5). If the first roll is a 3 and then the next one is a 2, then this particular can be represented by the square we might call "(3, 2)", and so on.

Notice that the outcomes "(2, 3)" and "(3, 2)" are pretty closely related. In both cases, we observe a 2 and we observe a 3. But we distinguish those two outcomes because in those two outcomes, the 2 and the 3 happen in different order, and the order in which they appear may be a detail which is of interest to us. So we make this distinction in the sample space and we keep the (3, 2) and the (2, 3) as separate outcomes.

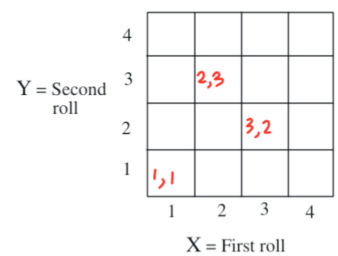


Figure 5: We distinguish between outcomes (2, 3) and (3, 2).

Now, this is a case of a model in which the probabilistic experiment can be described in phases or stages. We could think about rolling the die once and then going ahead with the second roll. So we have two stages.

A very useful way of describing the sample space of experiments– whenever we have an experiment with several stages, either real stages or imagined stages. A very useful way of describing it is by providing a sequential description in terms of a tree. We call a diagram of the kind shown in Figure 6 a tree. You can think of the leftmost point on this picture as the *root* of the tree from which you start. The endpoints of the tree are usually called *leaves*.

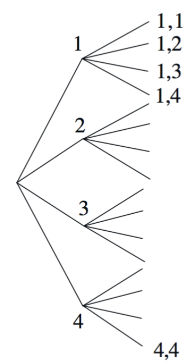


Figure 6: A different representation for the same experiment.

So here's how we read this tree. The experiment starts. We carry out the first phase, which in this case is the first roll, and we see what happens. So maybe we get a 2 in the first roll. If that's the case, we follow the branch from the root that leads to the "2". We then continue the experiment by rolling the die again, and we take note of what happened in the second roll. Maybe the result was a 3 this time. We then follow that branch from "2" that reads "(2, 3)", and so on.

Notice that in this tree we once more have a distinction. The outcome 2 followed by a 3 is different from the outcome 3 followed by a 2. In both cases, we have 16 possible outcomes. 4 times 4 makes 16. And similarly, if you count, the number of leaves is equal to 16.

The previous example involves a sample space that is *discrete and finite*: there are only 16 possible outcomes. But sample spaces can also be infinite; they can also be continuous sets.

EXAMPLE 3: Here's an example of an experiment that involves a continuous sample space. Here we have a rectangular target which is the unit square. You throw a dart on that target. Suppose that you are so skilled that no matter what, when you throw the dart, it always falls inside the target.

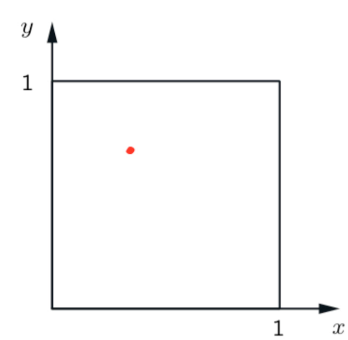
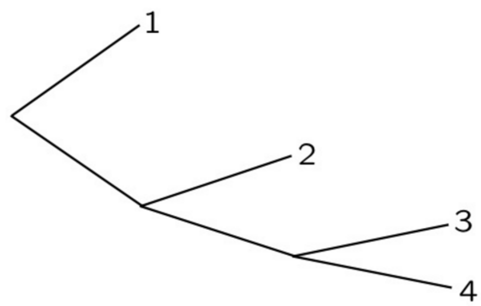


Figure 7: An example of a coninuous outcome space - throw darts into the square (x, y) such that $0 \leq x, y \leq 1$.

Once the dart hits the target, you record the coordinates x and y of the particular point that resulted from your throw. You record x and y with infinite precision. So x and y are real numbers. In this experiment, the sample space is just the set of x, y pairs that lie between 0 and 1.

Check Your Understanding: Paul checks the weather forecast. If the forecast is good, Paul will go out for a walk. If the forecast is bad, then Paul will either stay home or go out. If he goes out, he might either remember or forget his umbrella. In the tree diagram below, identify the leaf that corresponds to the event that the forecast is bad and Paul stays home.



Step 2: Describing Our Beliefs about the Outcomes

We now come to the second and much more interesting part. We need to specify which outcomes are more likely to occur and which ones are less likely to occur, and we will do that by assigning probabilities to the different outcomes.

However, as we try to do this assignment, we run into a difficulty, which is the following. Remember the previous experiment involving a continuous sample space, which was the unit square and in which we throw a dart at random and record the point that occurred?

In this experiment, what do you think is the probability of a particular point? Let's say what is the probability that my dart hits exactly the center of this square. Well, this probability would be essentially 0. Hitting the center exactly with infinite precision should be 0.

And so it's natural that in such a continuous model any individual point should have a 0 probability. For this reason instead of assigning probabilities to individual points, we will instead assign probabilities to whole sets, that is, to subsets of the sample space.

So here we have our sample space, which is some abstract set Ω.

Figure 8: In the continuous case, we assign probabilities to whole whole sets, not to individual points.

Here is a subset of the sample space. Call it A. We're going to assign a probability to that subset A, which we're going to denote with $P(A)$, which we read as the probability of set A. So probabilities will be assigned to subsets.

And these will not cause us difficulties in the continuous case because even though individual points would have 0 probability, if you ask me what are the odds that my dart falls in the upper half, let's say, of this diagram, then that should be a reasonable positive number. So even though individual outcomes may have probabilities equal to zero, sets of outcomes in general would be expected to have positive probabilities.

$P(A)$

Figure 9: Probability assigned to set A.

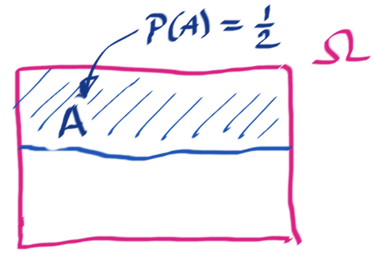


Figure 10: This upper half of the samples space gets $P(A)$ of $\frac{1}{2}$ assigned to it.

So, to repeat, we're going to assign probabilities to the various subsets of the sample space. And here comes a piece of terminology:

A subset of the sample space is called an event.

Why is it called an event? Because once we carry out the experiment and we observe the outcome of the experiment, either that outcome is inside the set A – and in that case we say that event A has occurred – or the outcome falls outside the set A in which case we say that event A did not occur.

A subset of the sample space is called an event.

Axioms of Probability

Now we want to move on and describe certain rules - "rules of the game" in probabilistic models, which are basically the rules that these probabilities should satisfy. They shouldn't be completely arbitrary. First, by convention, probabilities are always given in the range between 0 and 1. Intuitively, 0 probability means that we believe that something practically cannot happen. And probability of 1 means that we're practically certain that an event of interest is going to happen.

These rules that any probabilistic model should satisfy are called the axioms of probability theory.

OUR FIRST AXIOM is a nonnegativity axiom. Namely, probabilities will always be non-negative numbers. It's a reasonable rule.

$\mathrm{P}(A) \geq 0$

Figure 11: Axiom 1: Nonnegativity.

THE SECOND RULE is that if the subset that we're looking at is actually not a subset but is the entire sample space Ω, the probability of it should always be equal to 1.

$\mathrm{P}(\Omega) = 1$

Figure 12: Axiom 2: Normalization.

What does that mean? We know that the outcome is going to be an element of the sample space. This is the definition of the sample space. So we have absolute certainty that our outcome is going to be in Ω. Or, using a different language, we have absolute certainty that event Ω is going to occur. And we capture this certainty by saying that $P(\Omega) = 1$.

These two axioms are pretty simple and very intuitive. The more interesting axiom is the next one that says something a little more complicated. Before we discuss that particular axiom, a quick reminder about set theoretic notation. If we have two sets, let's say a set A, and another set, another set B, we use the notation $A \cap B$ to refer to the collection of elements that belong to both A and B. So in Figure 13, the *intersection* of A and B is the shaded set. We use the notation $A \cup B$ to refer to the set of elements that belong to A or to B or to both. So in Figure 14, the union of the two sets would be the set represented by the blue shaded area.

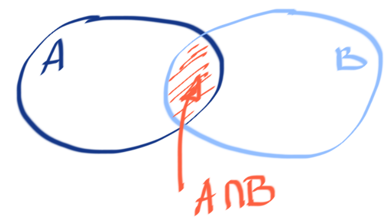


Figure 13: Intersection of Events.

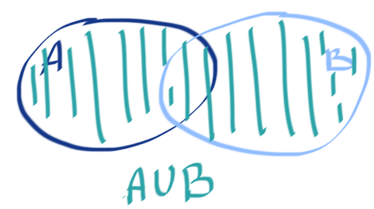


Figure 14: Union of Events.

NOW LET US LOOK AT THE THIRD AXIOM. It says that if we have two events, i.e. two subsets of the sample space, which are disjoint, then the probability that the outcome of the experiment falls in the union of A and B is equal to the sum of the probabilities of these two sets. (As shown in Figure 15, two sets being disjoint means that their intersection has no elements, i.e. their intersection is the empty set, and we use this symbol $\emptyset$ to denote the empty set) *This is called the additivity axiom.* It says that we can add probabilities of different sets when those two sets are disjoint.

In some sense we can think of probability as being one pound of some substance (of course your are free to think of ice-cream here) which is spread over our sample space and $P(A)$ is how much of that substance is sitting on top of a set A. What this axiom is saying is that whenever the sets A and B are disjoint from each other, the total amount of that substance sitting on top of A and B is how much is sitting on top of A plus how much is sitting on top of B.

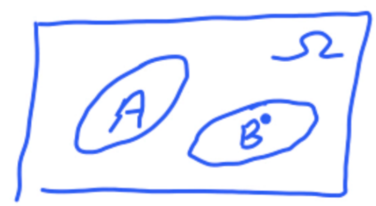


Figure 15: Non Intersecting of Events.

AND ONE FINAL NOTE: The additivity axiom needs to be refined a bit. We will talk about that a little later. Other than this refinement, these three axioms are the only requirements in order to have a legitimate probability model. At this point you may ask, shouldn't there be more requirements? Shouldn't we, for example, say that probabilities cannot be greater than 1? Yes and no. We do not want probabilities to be larger than 1, but we do not need to say it. As we will see in the next segment, such a requirement follows from what we have already said. And the same is true for several other natural properties of probabilities.

If $A \cap B = \emptyset$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$

Figure 16: Axiom 3: Additivity.

Check Your Understanding: Let A and B be events, with $P(A) = 0.6$ and $P(B) = 0.7$. Can these two events be disjoint?