Math-UA.233: Theory of Probability
Lecture 3

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

From last time... 1

To model an experiment, we choose:

- A sample space S, containing all possible outcomes (according to some description).
- A probability value $P(E)$ for each event $E \subset S$, quantifying how 'likely' that event is.

From last time... 2

Here's a simple template for obtaining probability values:

If $S = \{s_1, \ldots, s_n\}$, then a **probability distribution** on S specifies real numbers $p_1, \ldots, p_n \geq 0$ which satisfy

$$p_1 + \cdots + p_n = 1.$$

Then for any event $E \subset S$ we define its probability value by

$$\boxed{P(E) = \sum_{s_i \in E} p_i}.$$

(WARNING: We'll be taking a different approach by the end of this class.)

From last time... 3

In many simple cases, we assume 'equally likely outcomes': S is finite, and

$$\boxed{P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}}.$$

We also call this choice of P the 'uniform distribution on S'.

Warm-up examples

Example (Ross 2.5b)

Three balls are ***randomly drawn*** *from a bowl containing 6 white and 5 black balls. What is the probability that one of the balls is white and the other two are black?*

NOTEWORTHY FEATURES:

1. The phrase ‘randomly drawn’ in this problem really means ‘*assume equally likely outcomes*’.
2. There are two natural approaches, depending whether we order the balls or not. They give different calculations but the same final answer. *Either is correct*.

Example (Ross 2.5c)

A committee of 5 is to be selected from a group of 6 men and 9 women. If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

Again, two correct approaches, but one ends up being much easier!

Beyond equally likely outcomes

Some experiments clearly require a different choice of probabilities than ‘equally likely outcomes’.

Suppose we flip a coin, so $S = \{\mathrm{H}, \mathrm{T}\}$. Then we must have

$$p_{\mathrm{H}} = p \quad \text{and} \quad p_{\mathrm{T}} = 1 - p$$

for some $0 \leq p \leq 1$. (Because these two single-outcome probabilities must be non-negative and sum to 1).

If the coin is fair, that means $p = 1/2$, i.e. the outcomes are equally likely.

But if the coin is biased, then $p \neq 1/2$. This basic example will come up repeatedly later. It is called the ‘p-biased coin’.

Even if we start under an assumption of equally likely outcomes, this can change if we switch to a different description of the experiment.

(We already saw an example last time when we discussed rolling two dice in case we can't tell which die is which.)

Example

Alice and Bob play a game. They roll a fair die, and if it gives 5 or 6 then Alice wins, otherwise Bob wins.

- *'Complete' description of the game:* $S = \{1, 2, 3, 4, 5, 6\}$ *and outcomes equally likely so* $p_1 = \cdots = p_6 = 1/6$.

But if we only care who wins, then we could switch to:

- *'Reduced' description:* $S' = \{\mathrm{A}, \mathrm{B}\}$, *where* A *means 'Alice wins'. Now we treat this as a single outcome, and ignore the exact value shown by the die. Here the correct probability distribution is*

  $$p'_{\mathrm{A}} = p_5 + p_6 = 1/3, \quad p'_{\mathrm{B}} = p_1 + \cdots + p_4 = 2/3.$$

  This reduced description is mathematically the same as the $1/3$*-biased coin.*

Similar but more complicated:

Example

We roll two fair dice, and record the sum of the values shown.

- *'Complete' description:*

$$S = \{(i,j) : 1 \le i,j \le 6\}, \quad \text{distribution} = \text{uniform}.$$

- *'Reduced' description:* $S' = \{2, 3, \ldots, 12\}$ *and*

$$p'_2 = \frac{1}{36}, \quad p'_3 = \frac{1}{18}, \quad p'_4 = \frac{1}{12}, \quad \ldots, \quad p'_{12} = \frac{1}{36}.$$

A variant of this idea appears with experiments in which we are *waiting for something to happen*.

Example

An urn contains 2 indistinguishable red and 2 indistinguishable blue balls. They are withdrawn one-by-one at random (and not replaced) until a red ball is obtained.

Possible sample space: $S = \{1, 2, 3\}$*, indicating how many balls are withdrawn up to and including the first red.*

Probability distribution:

$$p_1 = \frac{2}{4} = \frac{1}{2}, \quad p_2 = \frac{2 \cdot 2}{4 \cdot 3} = \frac{1}{3}, \quad p_3 = \frac{2 \cdot 1 \cdot 2}{4 \cdot 3 \cdot 2} = \frac{1}{6}.$$

WHY THESE VALUES? Each is obtained by considering the withdrawal of a *fixed* number of balls (1, 2, or 3, respectively).

On the other hand, sometimes you start with equally likely outcomes, change the description, and still end up with equally likely outcomes.

Example (Ross 2.5e)

Suppose that $n + m$ *distinguishable balls, of which n are red and m are blue, are arranged in a linear order in such a way that all* $(n + m)!$ *possible orderings are equally likely [think: a thorough shuffling of n red cards and m blue cards].*

If we record the result of this experiment by listing only the colours of the successive balls, show that all the possible results remain equally likely.

The axioms of probability

Suppose $S = \{s_1, \ldots, s_n\}$ and $p_1, \ldots, p_n$ is a probability distribution.

Here are some simple consequences of how we define the values $P(E)$.

Proposition

1. *Any event $E \subset S$ satisfies $0 \leq P(E) \leq 1$;*
2. $P(S) = 1$*;*
3. *If $E_1, E_2, \ldots, E_n \subset S$ are disjoint (= mutually exclusive), then*
$$P(E_1 \cup \cdots \cup E_n) = P(E_1) + \cdots + P(E_n).$$

Another approach to developing probability theory: take the three properties from the previous slide as the *definition* of how probability values $P(E)$ for $E \subset S$ should behave. This gives:

The **axioms of probability**:

Axiom 1: Any event E satisfies $0 \leq P(E) \leq 1$.

Axiom 2: $P(S) = 1$.

Axiom 3: If the sequence of events E_1, E_2, ... are disjoint, then

$$P\Big(E_1 \cup E_2 \cup \cdots\Big) = P(E_1) + P(E_2) + \cdots .$$

(Axiom 3 often called **additivity**.)

Like many books, Ross simply starts with the axioms, and deduces everything from there. Why?

1. Often you begin a problem with info *not* about single-outcome probabilities, but about probability values for certain other events. Then it's crucial to know how to compute starting from those instead. These calculations are always based on the axioms.
2. Even if you know the single-outcome probabilities, it can be too difficult to compute $P(E)$ for some E by just adding up over all outcomes in E. Often much better methods involve describing E in stages, and knowing how probability values can be computed along those stages. These methods are based on the axioms.
3. If S is *not finite*, then we simply can't define everything in terms of single-outcome probabilities $p_1, p_2, \ldots, p_n$. *But the axioms still make sense*. They become the basis for the theory.

AN IMPORTANT CONFESSION

I cheated a bit. Look at axiom 3 again:

$$E_1, E_2, \ldots \text{ disjoint } \implies P\Big(E_1 \cup E_2 \cup \cdots\Big) = P(E_1) + P(E_2) + \cdots \,.$$

IMPORTANT: in this axiom, the sequence of events can be *finite or infinite*. The earlier proposition allowed only finite sequences.

Working with infinite sequences of events is more difficult, because then the right-hand side is a *convergent series*, not a finite sum. But it turns out that this extra strength in the axiom is crucial once we start working with infinite sample spaces.

For this reason it's sometimes called **countable additivity**.

See Ross p27 and Sec 2.6 for more discussion.

A QUICK SANITY CHECK

Let's prove that *if* S is finite, then starting from the axioms leads to *the same theory* as our simpler notion of a probability distribution.

Proposition

Suppose $S = \{s_1, \ldots, s_n\}$*, a finite set, and that the values* $P(E)$ *for* $E \subset S$ *satisfy the three axioms. Let*

$$p_1 = P(\{s_1\}),\ p_2 = P(\{s_2\}),\ \ldots,\ p_n = P(\{s_n\}).$$

Then

- $p_i \geq 0$ *for every* i *and* $p_1 + p_2 + \cdots + p_n = 1$*, and*
- *for any* $E = \{s_{i_1}, \ldots, s_{i_m}\} \subset S$*, we have*

$$P(E) = p_{i_1} + \cdots + p_{i_m}.$$

Using the axioms

From here on, I will do everything in terms of the axioms, and often I won't even mention the values $p_1, \ldots, p_n$. This matches Ross' book and most others.

Example (Ross 2.5h)

In the game of bridge, the entire deck of 52 cards is dealt out to 4 players. What is the probability that

(a) *one of the players receives all 13 spades?*

(b) *each player receives 1 ace?*

NOTEWORTHY FEATURES:

1. IDEA for part (a): break up this event into four disjoint smaller events and apply axiom 3. There's an art to finding such methods. *It comes with practice.*
2. Part (b), however, is just a 'counting' problem, doesn't use any more abstract ideas.

(For more gambling examples, see 2.5f, 2.5g and 2.5j.)

Here are two simple facts that we can deduce from the axioms.

Proposition

$$\boxed{P(E^c) = 1 - P(E)} \quad \textit{and} \quad \boxed{P(\emptyset) = 0}.$$

The first of these is incredibly useful, because of this CONSEQUENCE:

> *If $P(E)$ seems hard to compute, try finding $P(E^c)$ instead!*

Keep it in mind!

Example (Ross 2.5d)

An urn contains n balls, one of which is special. If k of these balls are withdrawn one at a time, with each selection being equally likely to be any of the balls that remains at that time, what is the probability that the special ball is chosen?

NOTEWORTHY FEATURES:

1. Two possible approaches (may order balls or not).
2. Easier to start by compute P(special ball *not* chosen).

Example (Ross Prob 2.41)

If a die is rolled 4 times, what is the probability that 6 comes up at least once?