Math-UA.233: Theory of Probability
Lecture 2

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

First words

Almost every aspect of life contains an element of chance.

In probability theory we learn to reason *quantitatively* about chance. Nowadays, probability theory is essential to

> *STATISTICS, THE SCIENCES, MEDICINE, ENGINEERING, FINANCE, INSURANCE, ...*

We will mention a few example applications in this course.

But this course is about the core of the theory, which is common to *all* the applications. Not the more specialized aspects of any one of them.

In fact, the easiest setting to introduce the basic theory is...

GAMBLING

This is because gambling and games of chance follow simple, artificial rules, and have only limited possible outcomes.

Historically, rigorous mathematical probability began with the analysis of certain gambling questions, such as the ‘Problem of the Points’. We’ll meet this one later in the course.

The Rules of Probability (Ross Chap 2)

Experiments, sample spaces and events (Ross Sec 2.2)

In probability theory, an *experiment* is any situation which has several possible outcomes, exactly one of which then happens.

This is a technical use of the term. It may include situations which we would not call ‘experiments’ in everyday speech.

For example:

- Finding out the sex of a newborn child.
- Finding out the finishing order in a 7-horse race.
- Flipping a pair of coins.
- Measuring the lifetime of a transistor.

Probability theory teaches us how to calculate with the probabilities of different things than can happen in an experiment.

The first step is to fix a mathematical description of the experiment.

The **sample space** is the set of all possible outcomes. Denoted by S (or, in other books, often Ω).

(Conversely, an **outcome** always refers to an element of S.)

Natural choices for preceding examples:

- Sex of a newborn child: $S = \{b, g\}$ (could also include some intersex variations for a more accurate description).
- 7-Horse race:

$$S = \{\text{all possible orders}\} = \{\text{all orderings of } (1, 2, \ldots, 7)\}.$$

- Flipping a pair of coins: $S = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$.
- Transistor lifetime: $S = \{\text{real numbers } x \geq 0\} = [0, \infty)$.

The sample space is *not the same thing* as the experiment. An experiment is something that happens in the real world. Then a sample space is a choice that we make about how to model that experiment. There can be more than one valid choice. The important thing is to make a good choice and then *stick to it*.

Example

Suppose that in a horse race, we care only about who wins, not the order of the other horses. Then we could use

$$S = \{\textit{all orderings of } (1, 2, \ldots, 7)\},$$

as before, or we could use

$$S = \{1, 2, \ldots, 7\} \quad (\textit{possible numbers of the horse that wins}).$$

The second is simpler. It is adequate to describe the outcomes if we don't care who comes in second, third, etc.

Having chosen the sample space, we will need to discuss different events that can occur.

To do so formally and carefully, observe that specifying an event is equivalent to specifying *the outcomes for which that event occurs*. Thus:

An **event** is a subset of the sample space S.

IMPORTANT CONSEQUENCE:

> *Two events are 'the same' if they consist of the same outcomes, even if they are described in two different ways.*

Examples:

- Sex of a newborn child: $E = \{g\} = \{\text{child is a girl}\}$.
- 7-Horse race:

$$E = \{\text{all orderings of } (1, 2, \ldots, 7) \text{ starting with } 3\} \\ = \{\text{horse 3 wins}\}.$$

- Flipping a pair of coins:
  $E = \{\mathrm{HH}, \mathrm{HT}\} = \{\text{first coin lands heads}\}$.
- Transistor lifetime:

$$E = [0, 5] = \{\text{transistor does not last longer than 5 hours}\}.$$

Be careful!

Non-example: if flipping a pair of coins, then

$$\{\mathrm{H}\}, \quad \{\mathrm{H}, \mathrm{T}\}, \quad \{\text{first coin}\}.$$

are *not* events of the sample space. They all have something to do with the experiment, but the information is of the wrong kind to describe an event.

In the first two examples, we see only possible results from a single coin, whereas the outcomes of the experiment depend on a *pair* of coins.

In the third, it seems we are choosing one of the coins itself, which isn't the same as flipping the coins.

Logic and operations with events

Thinking about events as *sets* clarifies various relations between then.

The **union** of events E and F is the event in which either E occurs, F occurs, or both. Denoted by $E \cup F$. As a set, it is the set of all outcomes which are in either E or F or both.

The **intersection** of events E and F is the event in which both E and F occur. Denoted by $E \cap F$ or just EF. As a set, it is the set of all outcomes which are in both E and F.

If E and F are events, then F **contains** E, or E is a **subset** of F, if every outcome in E is also in F. Denoted $E \subset F$. This means that if E occurs, then F necessarily also occurs.

Example

Flipping two coins. Let

$$E = \{\mathrm{HH}, \mathrm{HT}\} = \{\textit{first coins lands heads}\}.$$

$$F = \{\mathrm{HH}, \mathrm{TH}\} = \{\textit{second coins lands heads}\}.$$

Then

$$E \cup F = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\} = \{\textit{at least one of the coins lands heads}\}$$

and

$$E \cap F = \{\mathrm{HH}\} = \{\textit{both coins land heads}\}.$$

There are a few important relations between unions, intersections and complements, similar to the rules of algebra:

Theorem

- *Commutative laws:*

$$E \cup F = F \cup E, \quad E \cap F = F \cap E$$

(compare $a + b = b + a$, $ab = ba$ for real numbers).

- *Associative laws:*

$$(E \cup F) \cup G = E \cup (F \cup G), \quad (E \cap F) \cap G = E \cap (F \cap G).$$

- *Distributive laws:*

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G), \quad (E \cap F) \cup G = (E \cup G) \cap (F \cup G).$$

IDEA: Remember what it means for two events to be the same!

The **null event**, denoted $\emptyset$, is the event that does not contain any outcomes. Also called the **empty set**. Roughly, "the null event is to sample spaces as 0 is to counting".

Two events E and F are **mutually exclusive** if $E \cap F = \emptyset$: that is, if there are no outcomes in both. More informally: if E and F cannot both occur. Also called **disjoint**.

Several events $E_1, E_2, \ldots$ are **mutually exclusive** or **disjoint** if every pair of them is mutually exclusive: that is, if *at most one* of them can occur.

If E is an event, then its **complement** is the event “E does not occur”: that is, it consists of exactly those outcomes which are not in E. Denoted by E^{c}.

Example

$$S^{\mathrm{c}} = \emptyset \quad \textit{and} \quad \emptyset^{\mathrm{c}} = S.$$

Example

What is $[0,7]^{\mathrm{c}}$*?*

ANS: This is a trick question. In this example, we haven't specified the whole sample space S. We need to know this to define complements.

If, for instance, we've already chosen $S = \mathbb{R}$, then

$$[0,7]^{\mathrm{c}} = (-\infty, 0) \cup (7, \infty).$$

If instead $S = [0, \infty)$, then $[0,7]^{\mathrm{c}} = (7, \infty)$.

We also define unions and intersections of larger collections of events. If $E_1, E_2, \ldots$ is a sequence of events, then

$$\begin{aligned} \text{Union}: \quad E_1 \cup E_2 \cup \cdots &= \bigcup_{n=1}^{\infty} E_n \\ &= \{\text{all outcomes which lie in } E_n \text{ for at least one } n\} \end{aligned}$$

$$\begin{aligned} \text{Intersection}: \quad E_1 \cap E_2 \cap \cdots &= \bigcap_{n=1}^{\infty} E_n \\ &= \{\text{all outcomes which lie in } E_n \text{ for every } n\}. \end{aligned}$$

(Note that we don't need to write brackets, e.g.

$$(\cdots((E_1 \cup E_2) \cup E_3) \cup \cdots),$$

because of the associative laws.)

One more vital relation between unions, intersections and complements:

Theorem (De Morgan's Laws)

If $E_1, E_2, \ldots$ is a (finite or infinite) sequence of events, then

$$\Big(E_1 \cup E_2 \cup \cdots\Big)^{c} = E_1^{c} \cap E_2^{c} \cap \cdots$$

and

$$\Big(E_1 \cap E_2 \cap \cdots\Big)^{c} = E_1^{c} \cup E_2^{c} \cup \cdots .$$

IDEA: Remember what it means for two events to be the same!

Basic modeling of probabilities (slight deviation from Ross)

Now consider an experiment, and let S be its sample space. In probability theory, we assume that each event $E \subset S$ has a **probability value** $P(E)$.

Formally, P is a "**set function**":

$$P : \{\text{subsets of } S\} \longrightarrow \mathbb{R}.$$

The value $P(E)$ is supposed to put a number on how ‘likely’ the event E is. Although the idea of ‘likelihood’ is very intuitive, it is very hard to say exactly what it means.

The modern approach to probability theory is to *avoid this question*. Instead, we just agree on some basic *rules* that these numbers should follow, and derive the theory from those.

The interpretation can be decided later, and may vary with the context. *It doesn’t matter for building the theory as long as we agree on the rules.*

The general rules are called the *axioms of probability*. Ross puts these right at the start of probability, in Sec 2.3.

But here we defer the axioms a bit, and start with a special case which is simpler:

> *Now assume that* S *is finite, say* $\{s_1, \ldots, s_n\}$*. This means the experiment has only finitely many possible outcomes.*

Examples:

- Order in a horse race: YES
- Transistor lifetime: NO

A **probability distribution** on $S = \{s_1, \dots, s_n\}$: real values

$$p_1,\ p_2,\ \dots,\ p_n \geq 0$$

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

Intuitively, p_i is the "probability of getting outcome s_i in the experiment". We don't need to know exactly what this means to start calculating. What we *do* need is the following rule:

In the setting above, for any other event $E \subset S$, its **probability** is defined to be

$$P(E) = \sum_{i \text{ such that } s_i \in E} p_i .$$

CONVENTION FOR THE NULL EVENT: $P(\emptyset) = 0$.

Example (Ross 2.3a)

Flipping a coin for which heads and tails are equally likely:

$$S = \{\mathrm{H}, \mathrm{T}\}, \quad p_{\mathrm{H}} = p_{\mathrm{T}} = 1/2.$$

What if the coin is biased?

Example (Ross 2.3b)

Rolling a fair, six-sided die:

$$S = \{1, 2, 3, 4, 5, 6\}, \quad p_1 = p_2 = \cdots = p_6 = 1/6$$

What is

$$P(\textit{rolling an even number})?$$

Later we will do probability on sample spaces that are not finite.

Then specifying the probabilities values $P(E)$ for events $E \subset S$ is more complicated, and we will need a different approach.

But let us defer that a bit longer.

Equally likely outcomes

In general, the rules of probability theory do *not* tell us how to choose the values $p_1, \ldots, p_n$, only how to compute once we have them.

The actual choice of a probability distribution depends on the experiment being studied. Normally this is a modeling choice that we make at the same time as choosing the sample space.

The simplest possible choice is the following.

Definition

Again let $S = \{s_1, s_2, \ldots, s_n\}$. The outcomes are ***equally likely*** *if*

$$p_1 = p_2 = \cdots = p_n = 1/n.$$

This is also called the ***uniform distribution on*** *S.*

In many simple cases this is a natural assumption, especially in gambling: think of picking from a well-shuffled pack of cards. In other applications, one needs a different choice of probability values.

(Clearly nothing like this will work if S is infinite.)

Assuming equally likely outcomes, we get an easy formula for the probabilities of other events.

Proposition

If S is finite and the outcomes are equally likely, then for any event E,

$$\boxed{P(E) = \frac{\textit{number of outcomes in } E}{\textit{number of outcomes in } S} = \underbrace{\frac{\#E}{\#S} \quad \textit{or} \quad \frac{|E|}{|S|}}_{\text{notation}}}.$$

Example (Ross 2.5a)

If two dice are rolled, what is the probability that the sum of the upturned faces will be equal to 7*?*