Math-UA.233: Theory of Probability
Lecture 4

Tim Austin

tim@cims.nyu.edu
cims.nyu.edu/∼tim

Reminder: the axioms again

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

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

Axiom 3: If the sequence of events $E_1, E_2, \ldots$ are disjoint (= mutually exclusive), then

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

Example (A nice warm-up: the birthday problem. See Ross E.g. 2.5i or Pishro-Nik E.g. 2.4)

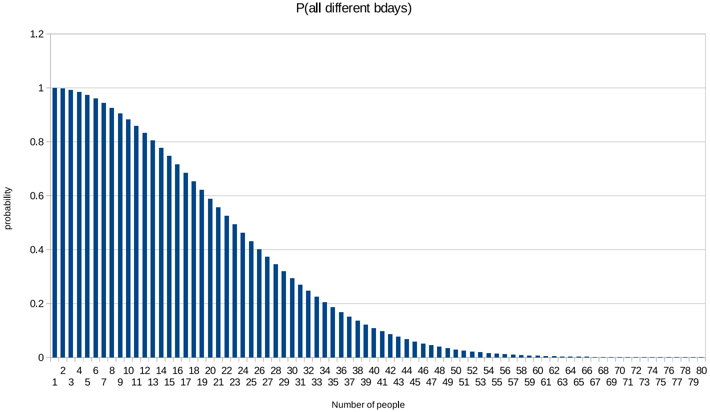
What is the probability that two members of our class have the same birthday?

SIMPLIFYING ASSUMPTIONS:

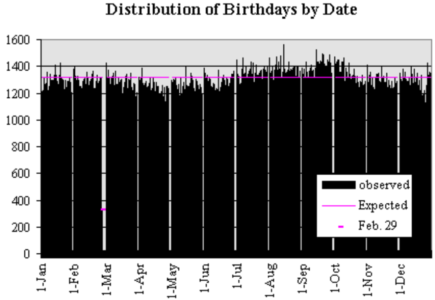
1. Ignore leap years.
2. Assume that for each person, all 365 possible birthdays are equally likely, irrespective of the other birthdays in the room.

General answer: same model but with n people,

$$P(\text{all birthdays different}) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$



Was it *really* reasonable to assume equally likely outcomes?



REMARK ON SOME COMMON TERMINOLOGY (see Pishro-Nik):

- If we draw a random sequence of n things from some collection of N things, and each thing may be sampled more than once, then we speak of sampling **with replacement.** Size of sample space: N^n. The birthday problem is an example, where n is the number of people and $N = 365$.
- If we draw such a sequence, but no thing may be chosen more than once, then we are sampling **without replacement.** Size of sample space:

$$N \times (N-1) \times \cdots \times (N-n+1) = \frac{N!}{(N-n)!}.$$

Most of the rest of today: more on how to deduce things *directly from the axioms*.

This will be essential throughout the rest of the course.

Two more general propositions

Proposition

$$\boxed{\textit{If } E \subset F \textit{ then } P(E) \leq P(F).}$$

Axiom 3 tells us how to get $P(E \cup F)$ from $P(E)$ and $P(F)$ if E and F are disjoint. But what if they aren't?

Proposition

$$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F)}.$$

Example (Ross Ex 2.4a)

Y chooses two books. She will like the first with probability 0.5*; she will like the second with probability* 0.4*; and she will like both books with probability* 0.3*. What is the probability that she will not like either book?*

A slightly more complicated calculation:

Theorem

$$
\begin{aligned}
P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) \\
- P(E \cap G) - P(F \cap G) + P(E \cap F \cap G).
\end{aligned}
$$

IDEA: Expand $P(E \cup F \cup G) = P(E \cup (F \cup G))$ in two steps.

Example (Ross 2.5l)

In a club of 100 people, 36 members play tennis, 28 play squash, and 18 play badminton. Furthermore, 22 of the members play both tennis and squash, 12 play both tennis and badminton, 9 play both squash and badminton, and 4 play all three sports. If a member is selected at random, what is the probability that they play at least one sport?

Generalization of the previous theorem:

Theorem (The inclusion-exclusion principle)

For any events $E_1, E_2, \dots, E_n$,

$$
\begin{aligned}
&P(E_1 \cup E_2 \cup \cdots \cup E_n) \\
&= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \cdots \\
&\quad + (-1)^{r-1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) + \cdots \\
&\quad + (-1)^{n-1} P(E_1 \cap E_2 \cap \cdots \cap E_n).
\end{aligned}
$$

IDEA: for instance, by induction on n (will see a different proof later).

Example (The matching problem; Ross 2.5m)

Suppose that each of n people at a party throws their hat into the center of the room. The hats are mixed up, and then everyone randomly selects a hat. What is the probability that no-one selects their own hat?

IDEA: We want $1 - P\left(\bigcup_{i=1}^{n} E_i\right)$, where

$$E_i = \{i^{\text{th}} \text{ person gets their own hat}\}.$$

Now apply inclusion-exclusion.

WHY is this is a good idea? Because in this problem, it turns out to be easier to compute *intersection* probabilities such as

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}).$$

Example (Ross 2.5n)

Ten pairs of non-identical twins are seated at random at a round table. What is the probability that no pair of twins sits together?

IDEA: We want $1 - P\left(\bigcup_{i=1}^{10} E_i\right)$, where

$$E_i = \{i^{\text{th}} \text{ twins } \underline{\text{do}} \text{ sit together}\}.$$

Now apply inclusion-exclusion.

ANS (messy):

$$1 - \left[\binom{10}{1} 2^1 \frac{18!}{19!} - \binom{10}{2} 2^2 \frac{17!}{19!} + \cdots - \binom{10}{10} 2^{10} \frac{9!}{19!}\right] \approx 0.3395.$$

What do probability values mean?

This is really a philosophical question. Roughly two schools of thought:

- *Objective* or *frequentist*: if you can repeat an experiment many times, then the 'probability' of an event is roughly the percentage of those repeats in which that event occurs. Good interpretation for, e.g., flipping a coin.
- *Subjective* or *personal*: the 'probability' of an event puts a number on 'how likely I think it is that that event will happen'. Good interpretation for, e.g., whether it rains tomorrow.

But the great insight is that *we don't have to choose* in order to use probability theory — we just have to agree on the three axioms! These are still an idealization of reality, but they cover a huge range of applications.

Example

We can imagine that someone holds the following beliefs:

- *Probability of rain today: 30%*
- *Probability of rain tomorrow: 40%*
- *Probability of rain both today and tomorrow: 20%*
- *Probability of rain either today or tomorrow: 60%.*

What's the problem here?

Beyond finite sample spaces

For more complicated experiments, it is sometimes natural to allow *infinitely many* possible outcomes. For example, some experiments could have any positive integer or even any real number as an outcome.

In these cases:

- We cannot assume equally likely outcomes (they don't make sense), and
- We may need to calculate with infinite series, using the full strength of 'countable additivity'(= axiom 3).

The next simplest setting: when the sample spaces S is **countably infinite.**

This means we can write S as an infinite list:

$$S = \{s_1, s_2, \dots\}.$$

Examples:

- How many days until the next rainfall, to the nearest whole number?

$$S = \{0, 1, 2, \dots\} \quad \text{COUNTABLE}$$

- How long *exactly* until the next rainfall?

$$S = [0, \infty) \quad \text{NOT COUNTABLE.}$$

(Countability is quite a deep concept within math, but I won't say much more about it here.)

If $S = \{s_1, s_2, \dots\}$ is countably infinite, then a probability model P can be described similarly to the finite setting: each single outcome s_i gets a probability value

$$p_i = P(\{s_i\}), \quad i = 1, 2, \dots,$$

and these satisfy:

1. $p_i \geq 0$ for every i
2. $\sum_{i=1}^{\infty} p_i = 1$, and
3. for any other $E \subset S$,

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

But now:

- These sums are *convergent series*, and
- We're using the full strength of *countable* additivity.

Some of the simplest examples: experiments in which we are waiting for something to happen, and there's no bound on how long it might take.

Example

We flip a fair coin until the first H appears. What is the probability that this occurs after an even number of flips? What is the probability that we *<u>never</u>* *get H?*

What if the coin were p-biased, where $0 \leq p \leq 1$*?*

IMPORTANT: First figure out the right model! Natural choice:

$$S = \{\text{possible numbers of flips}\} = \{1, 2, \ldots, \infty\},$$

where '∞' means 'never get H'. Now, for $i \in S$ and $i \neq \infty$, obtain p_i from the probability that, if we flip the coin exactly i times, we get $n - 1$ Ts followed by H.

Example (From Ross Prob 2.25)

A pair of dice is rolled until a sum of either 5 or 7 appears. Find the probability that a 5 occurs first.

IDEAS: First select S [many possible choices, but none finite!].

Let E_n be the event that a 5 occurs on the n^{th} roll and no 5 or 7 occurs before that. The desired probability is $\sum_{n=1}^{\infty} P(E_n)$ (WHY?). Now each term $P(E_n)$ involves rolling a *fixed* number of times, so can be found using a *finite* sample space.