In the game *Let's Make a Deal*, you are given a choice between three doors A, B, and C. One of the doors hides a prize. Behind the other two are goats. You pick a door, say door A. The host opens another door, say door C, to reveal a goat. You are then given a choice: You can stay with door A or switch to door B. Should you stay or should you switch?

Since you found out that door C hides a goat, there are two possibilities: Either the prize is behind door A and the goat is behind door B, or vice versa. So it shouldn't matter what you decide; you have a 50-50 chance at getting the prize.

Is this reasoning correct?

1 Probability models

Probability is the branch of mathematics that is concerned with these sorts of questions. In order to apply it, we need to come up with a model that precisely describes the situation at hand. A (finite) probability model consists of the following three elements:

- A *sample space*, which specifies the set of all possible outcomes of an experiment;
- Sets of outcomes of interest, called *events*;
- An assignment of *probabilities* to the different outcomes.

The sample space In *Let's Make a Deal*, the outcome is specified by all the choices that were made before you get to decide whether to stay to switch. These include (1) The door x concealing the prize; (2) The door y you initially choose; and (3) The door z that the host opens to reveal a goat. So the outcomes can be described by triples of the form (x, y, z), where x, y, and z refer to one of the three doors.

For example, (B, A, C) describes the outcome in which the prize is behind door B, you initially choose door A, then the host reveals door C. In this particular case, switching (to door B) would win you the prize, while staying with door A would reveal another goat.

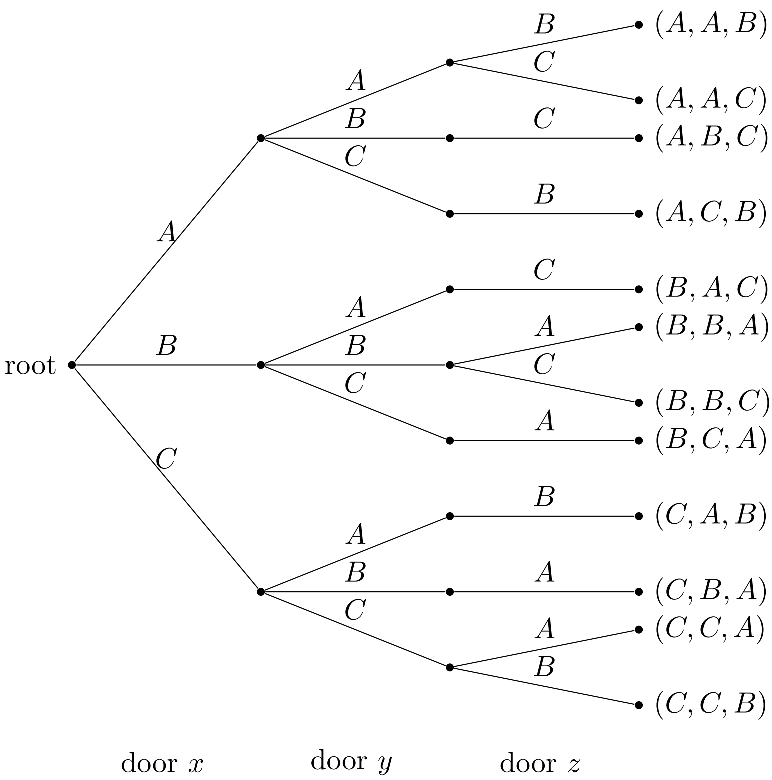
In our model of *Let's Make a Deal*, not all possible triples of As, Bs, and Cs are valid outcomes. For example, (A, B, B) is not valid because the host is not supposed to open the door that you picked; the introductory paragraph says that the host opens *another* door. We will also assume that outcomes like (A, B, A) are not valid; the host doesn't want to give away the prize for free! Ruling out such devious plays, we can model the sample space of *Let's Make a Deal* as follows:

> The sample space consists of those outcomes (x, y, z) in which x can be any door, y can be any door, and z can be any door different from x and y.

In general, the sample space of a probabilistic experiment can be very large, but in this example it is small enough that we can list them all:

$$\begin{aligned} S = \{&(A, A, B), (A, A, C), (A, B, C), (A, C, B), (B, A, C), (B, B, A), \\ &\qquad (B, B, C), (B, C, A), (C, A, B), (C, B, A), (C, C, A), (C, C, B)\}. \end{aligned}$$

In this example it is useful to represent the sample space by a *tree diagram* in which the vertices represent the possible states at different stages of the experiment, and the edges represent the choices available at a given stage:



Events An *event* is a subset of the sample space. Events are usually described by a predicate that the outcomes must satisfy. For example, the event "the prize is behind door A" (i.e., "$x = A$") consists of the outcomes

$$\{(A, A, B), (A, A, C), (A, B, C), (A, C, B)\},$$

while the event "the host opened door C" (i.e., $z = C$) is

$$\{(A, A, C), (A, B, C), (B, A, C), (B, B, C)\}.$$

We are interested in the event E described by the predicate "contestant wins by switching". This consists of those outcomes (x, y, z) for which the door different from both y and z contains the prize, namely the outcomes for which x, y, and z are all different:

$$E = \{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}. \tag{1}$$

Since the sample space contains 12 outcomes and the contestant wins by switching in 6 of those 12, it may be tempting to conclude that if the player decides to switch, the probability that they win is 50%. This is incorrect because *not all outcomes are equally likely.* To explain why we need to talk about probabilities.

Probabilities In order to complete our model, we need to assign a probability to each possible outcome. The probabilities are non-negative numbers that add up to one. There are many possible ways of doing so. Which is the correct one?

Assigning probabilities to outcomes in a reasonable manner can be tricky. It is usually a good idea to break up the experiment into simple "components" and reason about them separately. In the case of *Let's Make a Deal*, let's try to reason separately about what happens in the different stages of the game.

In the very beginning, the game host has to choose the door x that hides the prize. What are the probabilities that we should assign to the different doors? In the absence of any additional information, it seems reasonable to assume that all three doors are equally likely to hide the prize, namely that the choices $x = A$, $x = B$, and $x = C$ all have probability $1/3$.

Now that the prize has been safely hidden behind door x, the contestant is about to choose door y. How should we assign probabilities to the different choices? Again, it seems reasonable to assume (in the absence of additional information, once x is fixed, the choices $y = A$, $y = B$, and $y = C$ all have probability $1/3$.

Finally the host has to choose which door z to open. Looking at the tree diagram, you notice that for certain settings of x and y, once these have been fixed the choice of z is forced. For example, if $x = A$ and $y = B$, the host has no choice but to reveal $z = C$, so this choice must be made with probability one. In other settings, for example when $x = y = B$, the host has two possible choices for z. In such cases, it is reasonable to assume that these are made with equal probabilities, namely the choices $z = A$ and $z = C$ each occur with probability $1/2$.

We can summarize our probability model by the following rule:

> At any node in the tree diagram, all outgoing edges are taken with equal probability.

This rule allows us to calculate the probability of all outcomes: For example, to reach the outcome (B, B, A), there are 3 choices for x, 3 choices for y (given that $x = B$) and 2 choices for z (given that $x = B$ and $y = B$), so the outcome (B, B, A) has probability $\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}$. On the other hand, the outcome (B, C, A) has probability $\frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}$. Thus not all outcomes are equally likely! In fact,

$$\text{Outcome } (x, y, z) \text{ has probability } \begin{cases} \frac{1}{18}, & \text{if } x = y, \\ \frac{1}{9}, & \text{if } x \neq y. \end{cases}$$

The probability of an event is the sum of the probabilities of all the outcomes in it. The probability that the contestant wins after switching is therefore the sum of the probabilities of the six outcomes in (1). Since each of them has probability $\frac{1}{9}$, we obtain that

$$\Pr[E] = 6 \cdot \frac{1}{9} = \frac{2}{3}$$

so switching doubles your chances of winning the prize over staying!

Here is a more conceptual explanation as to why switching pays off in *Let's Make a Deal.* If the contestant happens to initially pick the door that hides the prize (i.e., $x = y$), then switching never wins. However, if he picks either one of the other two doors (i.e., $x \neq y$), then switching always wins. In other words, the contestant wins by switching *if and only if* doors x and y are different, which happens with probability $2/3$.

This example illustrates a general template for reasoning about probabilities in experiments with finitely many outcomes. First, identify and describe mathematically the sample space, which comprises *all possible* outcomes. Second, identify one or more *events* that you are interested in. Third, assign probabilities to each possible outcome. Be careful; different outcomes might have different probabilities! Finally, calculate the probability of the event(s) of interest by adding up the probabilities of the constituent outcomes.

2 Balls and bins

You toss five balls into three bins at random. What is the probability that none of the bins are empty?

The first task is to describe the sample space. There are several possible ways to specify the outcomes mathematically:

Option A: as an arrangement of five stars and two bars like in Lecture 10; for example, $\star\star|\star\star\star|$ would describe 2 balls in the first bin, 3 in the second bin, and zero in the third bin.

Option B: as a sequence (x, y, z) of three numbers that add to five, specifying the occupancies of the three bins; the above configuration would be described by the sequence $(2, 3, 0)$.

Option C: as a sequence of 1s, 2s, and 3s of length 5 describing the bin that the first, second, and so on balls land in; for example $(2, 3, 1, 1, 2)$ would describe the outcome in which the first ball lands in bin 2, the second ball lands in bin 3, and so on.

Which of these is the correct description? All three can describe the outcome of the experiment unambigously, but one is better than the other two for working out probabilities. This will become apparent shortly.

Next, we need to specify the event of interest, namely the set E of outcomes in which none of the bins are empty. Under option A, E is the set of star-bar sequences in which there is at least one star to the left and to the right of every bar. Under option B, E is the set of triples (x, y, z) in which all of x, y, and z are nonzero. Under option C, E is the set of sequences in which every number 1, 2, 3 occurs at least once.

We now come to the assignment of probabilities to outcomes. As there is no precise description of how the balls are assigned to the bins, we have to use common sense to come up with a probability model. One reasonable way to visualize the assignment is to imagine that first, a ball is tossed randomly into one of the three bins and is equally likely to land into each; once this is done, a second ball is tossed, and this one is also equally likely to land into each bin; and so on until the fifth ball is tossed. The corresponding tree diagram is a full ternary tree of depth five.

Under these assumptions, it should be evident that option C is the most natural way to represent outcomes. The probability of outcome $(2, 3, 1, 1, 2)$ is then $\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = 1/3^5$. So is the probability of outcome $(1, 1, 1, 1, 1)$. In fact, all outcomes have the same probability.

To summarize our discussion so far, our probability model is the following:

Sample space: $S = \{1, 2, 3\}^5$ (all sequences of 1s, 2s, and 3s of length 5)

Event of interest: $E = \{x \in S \colon x \text{ contains a 1, a 2, and a 3}\}$.

Probabilities: All $x \in S$ have the same probability $1/3^5$.

A probability space S in which all outcomes have equal probabilities is said to have *equally likely outcomes*. In general, each outcome must then occur with probability $1/|S|$. The probability of any event E is then given by the formula

$$\Pr[E] = \frac{|E|}{|S|} \qquad \text{assuming equally likely outcomes.}$$

Going back to balls and bins, all that remains is to calculate the size of E, namely to count the number of sequences that contain a 1, a 2, and a 3. The complement event $\overline{E}$ can be written as a union of three sets $N_1 \cup N_2 \cup N_3$, where N_i is the set of those sequences that do not contain any instances of symbol i. By the inclusion-exclusion formula,

$$|\overline{E}| = |N_1| + |N_2| + |N_3| - |N_1 \cap N_2| - |N_1 \cap N_3| - |N_2 \cap N_3| + |N_1 \cap N_2 \cap N_3|.$$

By the product rule, $|N_1| = |N_2| = |N_3| = 2^5$, $|N_1 \cap N_2| = |N_2 \cap N_3| = |N_1 \cap N_3| = 1$, and $|N_1 \cap N_2 \cap N_3| = 0$, so $|\overline{E}| = 3 \cdot 2^5 - 3$. Therefore,

$$\Pr[E] = \frac{|E|}{|S|} = \frac{|S| - |\overline{E}|}{|S|} = \frac{3^5 - 3 \cdot 2^5 + 3}{3^5} = \frac{50}{81},$$

which is about 62%.

The assumption of equally likely outcomes was very handy as it allowed us to convert our probability question into a counting question. This assumption must always be justified carefully; if used uncritically it can easily lead to incorrect answers. To illustrate this point, suppose that we use the option A representation, and we wrongly assume that all outcomes (i.e., all sequences of five stars and two bars) are equally likely. In this model, the event E has probability

$$\Pr_{\text{option A}}[E] = \frac{|E|_{\text{option A}}}{|S|_{\text{option A}}} = \frac{\binom{5-1}{3-1}}{\binom{5+3-1}{5}} = \frac{\binom{4}{2}}{\binom{7}{5}} = \frac{2}{7}$$

which is less than 29%.

3 Birthdays

There are about 100 students in class today. What is the probability that some pair of students have the same birthday?

Here is a plausible probability model for this question. We can think of the 365 days of the year as bins, the $k = 100$ students as balls, and the assignment of birthdays to students as a balls-into-bins experiment. In other words, the sample space is the set $S = \{1, 2, \ldots, 365\}^k$ of sequences of numbers between 1 and 365 of length 100. We assume equally likely outcomes.

Is this a reasonable model? In our modeling we did make some simplifying assumptions about the world. Our choice of sample space does not take leap years (i.e. February 29 birthdays) into account. Another assumption we made in our assignment of probabilities is that all days any given person is equally likely to be born on any day of the year. Certain studies indicate that this is not true; summer birthdays are in fact more favored. Yet another possible complication is that there may be relationships between the students that may affect their birthdays. For instance if the class has twins then the outcome is predetermined.

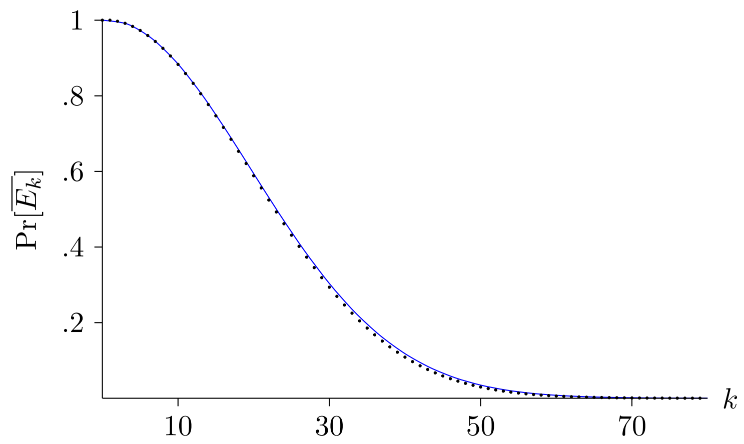
Despite these possible shortcomings, let us stick with our simple balls-and-bins models for birthdays and calculate what it predicts. The event E_k of interest consists of those sequences $(x_1, x_2, \ldots, x_k)$ in S in which at least two of the entries are the same, namely there exist indices $i \neq j$ such that $x_i = x_j$. It will be easier to work with the complement $\overline{E_k}$: This is the event that all people in class have distinct birthday, or the above sequence has no repeated entries.

Since all events are equally likely, $\Pr[\overline{E_k}] = |\overline{E_k}|/|S|$. The size of S is 365^k. The size of $\overline{E_k}$ can be calculated by the generalized product rule: There are 365 choices for x_1, 364 remaining choices for x_2, and so on. Therefore $|\overline{E_k}| = 365 \cdot 364 \cdots (365 - k + 1)$, and

$$\Pr[\overline{E_k}] = \frac{|\overline{E_k}|}{|S|} = \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k}.$$

When there are $k = 100$ people, the probability that all birthdays are distinct is $3 \cdot 10^{-7}$. The probability of a same birthday is overwhelmingly large!

The following plot shows the probability of all k people having distinct birthdays as k ranges from 1 to 80. The probability drops below half at $k = 23$. So in a room with 23 people it is already more likely that two people have the same birthday than not.



More generally, if there are n bins and k balls, the probability that all balls fall in distinct bins (under equally likely outcomes) equals

$$\Pr[\overline{E_k}] = \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} = 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right).$$

Using the inequality $1 - x \le e^{-x}$, we obtain that

$$\Pr[\overline{E_k}] \le e^0 \cdot e^{-1/n} \cdot e^{-2/n} \cdots e^{-(k-1)/n} = e^{(1+2+\cdots+(k-1))/n} = e^{-\binom{k}{2}/n}.$$

The blue line in the above plot represents this function. For example, when $n = 365$ and $k = 100$, the probability of all birthdays being distinct is at most $e^{-\binom{100}{2}/365}$, which is about $12 \cdot 10^{-7}$.

4 Intransitive dice

There are two dice on the table with possibly different face values. Alice chooses a die and Bob takes the other one. They toss their dice larger value wins. Which die should Alice choose?

To be concrete, suppose these are 3-sided dice with face values $X = \{2, 6, 7\}$ and $Y = \{1, 5, 9\}$, respectively. (If you prefer 6-sided dice just pretend each face value occurs twice.) Alice wants to choose the die that has the larger probability of winning the game. The experiment here consists of tossing the two dice, so the sample space is the product set

$$X \times Y = \{(2,1), (2,5), (2,9), (6,1), (6,5), (6,9), (7,1), (7,5), (7,9)\}.$$

The event E_{XY} of interest consists of those outcomes in which die X beats die Y, namely

$$E_{XY} = \{(x, y) \in X \times Y : x > y\} = \{(2,1), (6,1), (6,5), (7,1), (7,5)\}.$$

Given no information to the contrary, we may assume that all outcomes are equally likely and so

$$\Pr[E_{XY}] = \frac{|E_{XY}|}{|X \times Y|} = \frac{5}{9}$$

and Alice should choose die X, which gives her a 5/9 chance of winning.

Now Bob brings in a third die with face values $Z = \{3, 4, 8\}$. Now Alice can choose any one of the three die and Bob gets to pick among the remaining two. How does this additional choice affect Alice's strategy?

To answer this question let us figure out the winning probabilities for the other pairs of dice. If Alice were to play die Y against Bob's die Z, the sample space would be

$$Y \times Z = \{(1,3),(1,4),(1,8),(5,3),(5,4),(5,8),(9,3),(9,4),(9,8)\}$$

and, assuming equally likely outcomes, the probability of the event $E_{YZ} = \{(y, z) \in Y \times Z : y > z\}$ is $\Pr[E_{YZ}] = |E_{YZ}|/|Y \times Z|$, which is also 5/9. If Alice were to play die Z against Bob's die X then

$$Z \times X = \{(3,2),(3,6),(3,7),(4,2),(4,6),(4,7),(8,2),(8,6),(8,7)\}$$

and the probability of the event E_{ZX} that die Z beats die X is again 5/9.

In conclusion, no matter which die Alice picks, Bob can always beat her with probability 5/9. The tables have now turned and Bob has the advantage!

5 Axioms of probability

After all these examples we can give a precise mathematical definition of probability. A *probability space* consists of a set S called the *sample space* and a function Pr that assigns a non-negative number $\Pr[E]$ to subsets E of S called *events*. The function Pr must satisfy the following two axioms:

Axiom 1 $\Pr[S] = 1$

Axiom 2 If $E_1, E_2, E_3, \ldots$ are disjoint, then $\Pr[E_1 \cup E_2 \cup E_3 \cup \ldots] = \Pr[E_1] + \Pr[E_2] + \Pr[E_3] + \cdots$.

Axioms 1 and 2 are good enough to specify the probabilities of any event we could conceivably be interested in. For example, for any event A,

$$\begin{aligned}
\Pr[A] + \Pr[\overline{A}] &= \Pr[A \cup \overline{A}] && \text{by Axiom 2} \\
&= \Pr[S] \\
&= 1 && \text{by Axiom 1.}
\end{aligned}$$

Therefore, $\Pr[\overline{A}] = 1 - \Pr[A]$, or in words the probability of the event not happening is one minus the probability of it happening.

Similarly, if $A \subseteq B$, then

$$\begin{aligned}
\Pr[A] &= \Pr[A \cup (B - A)] \\
&= \Pr[A] + \Pr[B - A] && \text{by Axiom 2} \\
&\geq \Pr[A] && \text{because } \Pr[B - A] \text{ is non-negative.}
\end{aligned}$$

In words, larger events are more probable. You can also derive the inclusion-exclusion formula for probabilities

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

and its extension to three sets

$$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[A \cap C] - \Pr[B \cap C] + \Pr[A \cap B \cap C]$$

and so on. These mirror closely the identities about set size from Lecture 10.

One important difference is that the axioms of probability are meaningful even when the sample space S is infinite. Infinite sample spaces come up, for instance, in the following type of question.

> You toss a fair coin until it comes up heads. What is the probability that the number of tosses was even?

In principle, the coin tossing could go on for an arbitrarily long time, so the sample space that describes this experiment is infinite:

$$S = \{\mathtt{H}, \mathtt{TH}, \mathtt{TTH}, \mathtt{TTTH}, \ldots\}.$$

Since the coin is fair, these outcomes have probabilities $1/2, 1/4, 1/8, 1/16$, and so on. An outcome with n tosses has probability $1/2^n$. Therefore,

$$\Pr[S] = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots = 1$$

as required by Axiom 1.

The event E of interest consists of those outcomes with an even number of tosses, namely

$$E = \{\mathtt{TH}, \mathtt{TTTH}, \mathtt{TTTTTH}, \ldots\}.$$

By Axiom 2,

$$\begin{aligned}
\Pr[E] &= \Pr[\{\mathtt{TH}\}] + \Pr[\{\mathtt{TTTH}\}] + \Pr[\{\mathtt{TTTTTH}\}] + \cdots \\
&= \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} \\
&= \frac{1}{4}\left(1 + \frac{1}{4} + \frac{1}{4^2} + \cdots\right) \\
&= \frac{1}{4} \cdot \frac{1}{1 - 1/4} \\
&= \frac{1}{3}.
\end{aligned}$$

References

This lecture is based on Chapter 17 of the text *Mathematics for Computer Science* by E. Lehman, T. Leighton, and A. Meyer.