PROBABILITIES

Outcomes, probability space, and events
As you know, a die has six faces, each marked with dots representing the numbers 1, 2, 3, 4, 5, or 6. Thus, when we roll a die we get one of six possible outcomes: the numbers 1 through 6. With a fair die each of the six faces is equally likely to show up on top. Thus, when we roll a fair die the chance or probability of getting each of the numbers 1 through 6 is 1/6.

Determining probabilities depends in an essential way on specifying the possible outcomes and their likelihood, thus on assumptions such as “the die has six faces” and “the die is fair”.

Example 13.1 We roll a pair of distinguishable fair dice (see Fig. 1; one die is green, the other is purple) together. What are the chances that we get a double (1−1, 2−2, . . . , 6−6)?

Solution: After the roll we can see one of 6 possible green sides and one of 6 purple ones. By the Product Rule, that’s 6 × 6 = 36 outcomes. Since the dice are fair, and since, presumably, we roll them together without favoring one or the other, we will assume that the 36 outcomes are equally likely. 6 of the outcomes are doubles. Intuitively, the chances of getting a double are therefore 6/36 = 1/6.

Example 13.2 We roll a fair die twice. What are the chances that we get a double?

Solution: An outcome is now a sequence of two faces shown by the die. This seems like a different situation than when we roll a green-purple pair of dice. However, there are still 36 outcomes and we can assume that they are equally likely.

This is based on the important implicit assumption that the second roll does not depend on the first one (it could, for all we know; we might be forced to roll the second time by keeping the die in the same position in which it landed in the first roll and to use a weird two-finger poke). We say
that the two rolls are \textit{independent} and we shall have a lot more to say about this later. In fact, in the previous green-purple example, we have also made the implicit assumption that when rolling the dice together the green one’s roll is independent of the purple one’s.

Out of the 36 outcomes 6 are doubles so again the chances of getting a double are $\frac{1}{6}$.

\begin{example}
We roll a pair of indistinguishable beige fair dice (see Fig. 2). What are the chances that we get a double?
\end{example}

\begin{solution}
Now there are fewer than 36 outcomes. We have $\binom{6}{2} = 15$ outcomes in which the two dice show different numbers plus 6 outcomes for the doubles, a total of 21 outcomes. If these 21 outcomes would be equally likely then the chances of of getting a double would $\frac{6}{21} = \frac{2}{7} = 28.6\%$ which is quite a bit better than $\frac{1}{6} = 16.6\%$. Backgammon players love doubles so you would think that they prefer to play with dice of the same color rather than different colors?!? But they don’t care about the color of the dice and here is why: each of the 15 outcomes with different numbers has twice the chance of showing up than one of the 6 outcomes with doubles. (And still, each die is fair, of course.)

Think of it this way. $\{2, 5\}$ with beige dice is like rolling the green-purple dice but declaring that (green 2, purple 5) is the same outcome as (purple 2, green 5). For the green-purple dice each outcome had a chance of $\frac{1}{36}$. So the chance of getting $\{2, 5\}$ with beige dice seems to be $\frac{1}{36} + \frac{1}{36} = \frac{1}{18}$. And the same for all outcomes with different numbers. On the other hand for a double, (green 6, purple 6) is exactly the same as (purple 6, green 6) so the chance is just $\frac{1}{36}$. Its also $\frac{1}{36}$ for the other doubles so the chance of getting a double, any double, is $\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$.

\end{solution}

But why are we adding? What are the ground rules of computing chances, or as we will call them, probabilities?

\textbf{Space of outcomes (a.k.a. sample space)} This models the outcomes of an experiment or process that is governed by some randomness. Examples are rolling/tossing dice, where the outcomes are the numbers shown by the dice, or tossing/flippping coins, where the outcomes are the sides (heads/tails) shown by the coins. Other examples are extracting (sampling) marbles from urns, where the outcomes are the colors of the marbles, or throwing balls into bins, where the outcomes are the bins in which the balls end up. In a random experiment, or process, exactly one
outcome from a set of outcomes is sure to occur but, in general, no outcome can be predicted with certainty. In this class we will only discuss randomness with a finite number of outcomes. From the perspective of a probability theorist this is a big simplification but, for the most part, and as of 2017, it is enough for most uses of probability theory in Computer Science. Our notation for the space of outcomes is typically \( \Omega \). We always assume \( \Omega \neq \emptyset \), i.e., we have at least one outcome in a random process. The case when \( |\Omega| = 1 \) corresponds to deterministic experiments or processes. This case is, of course, uninteresting from the perspective of probability theory but it is included, as usual, as a mathematical “edge or corner case.”

When rolling the green-purple dice the outcomes are \( \Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\} \) with \( |\Omega| = 36 \).

When rolling the beige dice the outcomes are \( \Omega = \{\{1,2\}, \{1,3\}, \ldots, \{6,4\}, \{6,5\}\} \cup \{1-1, 2-2, \ldots, 6-6\} \) with \( |\Omega| = 15 + 6 = 21 \).

**Definition 13.4 (Probability Space)** This is a pair \((\Omega, \Pr)\) where \( \Omega \) is a finite non-empty set of outcomes and \( \Pr : \Omega \to [0, 1] \) is a function that associates with each outcome \( w \in \Omega \) its probability \( \Pr[w] \) which is a real number between 0 and 1 (inclusive), such that

\[
\sum_{w \in \Omega} \Pr[w] = 1
\]

We will often call \( \Pr \) a probability distribution. A probability space \((\Omega, \Pr)\) is called uniform if all the outcomes have the same probability, which must then equal \( 1/|\Omega| \). Hence, for uniform probability spaces, the distribution is a constant function.

When rolling the green-purple dice the probability space has the outcomes listed above and it is uniform with each outcome having probability \( 1/36 \).

When rolling the beige dice the probability space has the outcomes listed above, but the space is not uniform:

\[
\Pr[\{1,2\}] = \Pr[\{1,3\}] = \cdots = \Pr[\{6,4\}] = \Pr[\{6,5\}] = 1/18
\]

\[
\Pr[1-1] = \Pr[2-2] = \cdots = \Pr[6-6] = 1/36
\]

**Definition 13.5 (Event)** Let \((\Omega, \Pr)\) be a probability space. An event in this space is a subset \( E \subseteq \Omega \). We extend the probability function to events as follows (with the abuse of notation \( \Pr[\{w\}] = \Pr[w] \))

\[
\Pr[E] = \sum_{w \in E} \Pr[w]
\]

---

1. The one exception in these lecture notes is the geometric distribution whose sample space consists, essentially, of all natural numbers.
2. Although in probability theory it is called “probability mass function” and “distribution” is reserved for something else.
3. Any subset! Those of you who know something about infinite probability spaces in which not all subsets are “measurable”, beware, this issue does not arise in this class.
In general $\Pr[E] \in [0, 1]$ but $\Pr[\emptyset] = 0$.

This definition justifies the addition used in the calculations of probabilities of doubles in Example 13.3. (Go back and check!)

**Proposition 13.6** In a uniform probability space $\Pr[E] = \frac{|E|}{|\Omega|}$. Therefore, in such spaces computing probabilities is often a pure counting problem.

**Proof:** If $(\Omega, \Pr)$ is a uniform probability space then for each $w \in \Omega$ we have $\Pr[w] = 1/|\Omega|$. It follows that

$$
\Pr[E] = \sum_{w \in E} \Pr[w] = \sum_{w \in E} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}
$$

Example 13.7 Compute the probability of the event “when we roll the green-purple dice the numbers add up to an even number”.

**Solution:** As we saw before, we are in a uniform probability space with 36 outcomes. Let $(g, p) \in [1..6] \times [1..6]$ be an outcome in this space. $g + p$ is even iff $g$ and $p$ have the same parity. This happens for exactly half of the outcomes because for each die there are as many even faces as there are odd ones. Hence the probability we seek is 1/2.

Example 13.8 We roll three fair dice. What is the probability that all three show the same number?

**Solution:** There are $6 \cdot 6 \cdot 6 = 216$ outcomes and the event of interest consists of 6 of them: $(1, 1, 1), \ldots, (6, 6, 6)$. Hence the probability is $6/216 = 1/36$.

**Fair coin, biased coin** A fair coin is equally likely to show heads or tails when flipped. This is modeled by the simplest probability space of some interest: it has two outcomes $H$ and $T$, each with probability 1/2.

A biased coin shows heads with some probability $p$ that may be different from 1/2. Flipping such a coin is often called a Bernoulli trial with parameter $p$ (where, by convention, obtaining heads
is called \textit{success} and obtaining tails is called \textit{failure}). We shall attach the name of Bernoulli also to the corresponding probability space with two outcomes: success with probability $p$ and failure with probability $q = 1 - p$. Bernoulli trials become quite significant when we consider sequences of trials performed \textit{independently} (much more about independence coming up). In the following example we substitute some intuition for a formal definition of independence that we will give later.

\textbf{Example 13.9} Suppose we flip a fair coin twice. What is the probability that we obtain one tails and one heads? What if the coin is biased in a such a way that the probability of heads is $1/3$?

\textbf{Solution}: The space of outcomes is $\Omega = \{HH, HT, TH, TT\}$. Intuitively, because the coin is fair \textit{and} because the second flip does not seem be influenced by what happened in the first flip, there is no reason to assign different likelihoods to the four outcomes. Hence the space is uniform. The event of interest is $A = \{HT, TH\}$ and its probability (by Proposition 13.6) is $\Pr[A] = |A|/|\Omega| = 2/4 = 1/2$.

Now assume that the coin is biased with heads probability $1/3$. Without making direct use of the concept of independence we can imagine a different random experiment that is intuitively equivalent to flipping a biased coin. Consider an urn in which we keep three paper labels. One of the labels has $H$ written on it and the other two have $T_1$ and $T_2$ written on them. Assuming that each label is equally likely to be extracted, sampling a label from such an urn corresponds to flipping our biased coin.

In fact, consider \textit{two} such urns, $U$ and $U'$. Now our experiment consists of extracting a label from $U$ followed by extracting a label from $U'$. This experiment has $3 \times 3 = 9$ outcomes:

$$\Omega = \{HH, HT_1, HT_2, T_1H, T_1T_1, T_1T_2, T_2H, T_2T_1, T_2T_2\}$$

We assume it’s equally likely to extract each of the three labels from each urn and also that the extraction from the second urn is not influenced by what happened in the first extraction. Therefore there is no reason to assign different likelihoods to the 9 outcomes, and we have a uniform probability space. The event of interest is $B = \{HT_1, HT_2, T_1H, T_2H\}$ and its probability is $\Pr[B] = |B|/|\Omega| = 4/9$. $\blacksquare$

\textbf{Random permutations} Consider $n \geq 1$ distinct objects $a_1, \ldots, a_n$. Recall that a permutation of these objects is a sequence of length $n$ in which each of these objects appears exactly once. The number of such permutations is $n!$. A \textit{random permutation} of $a_1, \ldots, a_n$ is an element of the \textit{uniform} probability space whose outcomes are all the permutations of $a_1, \ldots, a_n$.

\textbf{Example 13.10} Let $i, j \in [1..n]$ (not necessarily distinct). Calculate the probability that $a_i$ occurs in position $j$ in a random permutation of $a_1, \ldots, a_n$. 
Solution: Let $E_{ij}$ be the event consisting of all outcomes in which $a_i$ occurs in position $j$. In the positions other than $j$ we can have any permutation of $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$. Therefore $|E_{ij}| = (n - 1)!$ and the probability we want to calculate is $(n - 1)!/n! = 1/n$. You can interpret this result as saying that each of the $n$ elements is equally likely to occur in any given position of a random permutation. However, be very careful, this does not happen independently of what occurs in the other positions, because each element occurs exactly once. We shall come back to this after we define independence formally. See also “balls into bins” for a different model that supports such independence of positions in a sequence (with repetitions). The random permutations model will appear also in further problems in recitation, review problems and homeworks, as well as in future lectures.

Properties of probability

Proposition 13.11 (Basic properties) \footnote{It is not accidental that cardinality (of finite sets) and probability have analogous properties. Both are instances of the mathematical concept of “measure”. Other examples of measure are area and volume.} Consider a probability space $(\Omega, \Pr)$ and arbitrary events $E, A, B$ in this space.

(P0) $\Pr[E] \geq 0$

(P1) $\Pr[\Omega] = 1$

(P2) $A \cap B = \emptyset \implies \Pr[A \cup B] = \Pr[A] + \Pr[B]$

(P3) $A \subseteq B \implies \Pr[A] \leq \Pr[B]$ \quad (monotonicity)

(P4) $\Pr[\overline{E}] = 1 - \Pr[E]$ \quad where $\overline{E} = \Omega \setminus E$

(P5) $\Pr[\emptyset] = 0$

(P6) $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$

(P7) $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$

Proof: We have already noted (P0). (P1) follows from the basic property of a probability distribution in our finite spaces. To prove (P2) we note that

$$\Pr[A \cup B] = \sum_{w \in A \cup B} \Pr[w] = \sum_{w \in A} \Pr[w] + \sum_{w \in B} \Pr[w] = \Pr[A] + \Pr[B]$$

where the second equality holds precisely because $A \cap B = \emptyset$. \hfill $\blacksquare$
To prove (P3) we observe that if $A \subseteq B$ then $B = A \cup (B \setminus A)$ and $A \cap (B \setminus A) = \emptyset$ so we can apply (P2):

$$\Pr[B] = \Pr[A] + \Pr[B \setminus A] \geq \Pr[A]$$

because $\Pr[B \setminus A] \geq 0$.

To prove (P4) we observe that $E \cap \overline{E} = \emptyset$ and $E \cup \overline{E} = \Omega$ so (P4) follows from (P3) and (P1).

(P5) now follow from (P4) and (P1) because $\emptyset = \overline{\Omega}$.

To prove (P6) we apply (P2) repeatedly:

\[
\begin{align*}
\Pr[A \cup B] &= \Pr[A] + \Pr[B \setminus A] \\
\Pr[A \cup B] &= \Pr[B] + \Pr[A \setminus B] \\
\Pr[A] &= \Pr[A \setminus B] + \Pr[A \cap B] \\
\Pr[B] &= \Pr[B \setminus A] + \Pr[A \cap B]
\end{align*}
\]

followed by some algebraic manipulations.

(P7) follows immediately from (P6) observing that $\Pr[A \cap B] \geq 0$.\footnote{You must have noticed that we proved only (P0), (P1) and (P2) from the definition of finite probability space. All the other properties were shown as consequence of the first three. This is not accidental. The first three properties correspond to Kolmogorov’s axiomatization of the general theory of probability, in which only “measurable” events have probability. (Actually, a more general version of (P2) with countable unions of pairwise disjoint events is needed.)}