Probabilities of outcomes and events

Outcomes and events

Probabilities are functions of **events**, where events can arise from one or more **outcomes** of an experiment or observation. To use an example from DNA sequences, if we pick a random location of DNA, the possible outcomes to observe are the letters `A`, `C`, `G`, `T`. The set of all possible outcomes is $S = \{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}\}$. The set of all possible outcomes is also called the **sample space**. Events of interest might be $E_1 =$ an `A` is observed, or $E_2 =$ a purine is observed. We can think of an event as a subset of the set of all possible outcomes. For example, E_2 can also be described as $\{\texttt{A}, \texttt{G}\}$.

A more traditional example to use is with dice. For one roll of a die, the possible outcomes are $S = \{1, 2, 3, 4, 5, 6\}$. Events can be any subset of these. For example, the following are events for this sample space:

- $E_1 =$ a 1 is rolled
- $E_2 =$ a 2 is rolled
- $E_i =$ an i is rolled, where i is a particular value in {1,2,3,4,5,6}
- $E =$ an even number is rolled $= \{2, 4, 6\}$
- $D =$ an odd number is rolled $= \{1, 3, 5\}$
- $F =$ a prime number is rolled $= \{2, 3, 5\}$
- $G = \{1, 5\}$

where event G is the event that either a 1 or a 5 is rolled. Often events have natural descriptions, like "an odd number is rolled", but events can also be arbitrary subsets like G.

Sample spaces and events can also involve more complicated descriptions. For example, we might be interested in rolling two dice at once. In this case the sample space is

$$S = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\} = \{(1, 1), (1, 2), (1, 3), \ldots, (6, 6)\}$$

The sample space can be organized more neatly like this

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6).

A similar problem that arises for DNA is a word consisting of three consecutive DNA letters. Here the sample space consists of triples of three letters, $S = \{(i, j, k) : i, j, k \in \{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}\}$. Usually these are written without parentheses. For example elements of the sample space include the three-letter combinations `AAG`, `TAG`, etc.

Three-letter combinations of DNA letters, also called **codons** are particularly important becuase in coding regions of genes, they code for amino acids, which get chained together to make proteins. For example, `AAG` codes for Lysine, and `TAG` is a stop codon.

The number of codons is $4 \times 4 \times 4 = 64$ (4 choices for each position), but there are only 20 amino acids. In most cases, more than one codon codes for the same amino acid. For example, `AAA` and `AAG` both code for lysine. The amino acid leucine has 6 codons. In thinking about events and outcomes, the event that a particular amino acid is coded for can be realized by multiple outcomes (codons).

Operations on events

Often we are interested in combining events to construct new events of interest. For example, in the game of Craps, you roll two six-sided dice and add the result. You win immediately if you roll a 7 or 11. You loose immediately if you roll a 2, 3, or 12. Any other result requires rolling the dice again. Thus the events or winning and loosing on the first roll can be constructed from other events whose probabilities might already be known. Let E_i be the event that two dice have a sum of i, $i \in \{2, 3, \dots, 12\}$. Then the event of winning on the first toss can be written as $W = E_7$ or E_{11}. The event of loosing on the first toss can be written as $L = E_2$ or E_3 or E_{12}.

Instead of writting "or", the notation $\cup$, called **union** is often used. So we write $W = E_7 \cup E_{11}$, and $L = E_2 \cup E_3 \cup E_{12}$. The union notation is common in other areas of mathematics. In general, the set $A \cup B$ can be defined by

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

For example, the function $f(x) = 1/x$ is defined on $(-\infty, 0) \cup (0, \infty)$. The meaning is the same in probability, but the values of interest are points in the sample space. The union of two sets uses an inclusive sense of "or". For example, if E is the event that the first number rolled is 6, and F is the event that the second number rolled is 6, then

$$E \cup F = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6), (1,6), (2,6), (3,6), (4,6), (5,6)\}$$

In addition to unions, we often want to consider what two sets, or events, have in common. This is done with **intersections**, notated by $\cap$. Here

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

In this case, $E \cap F = \{(6,6)\}$, a set consisting of only one element.

The idea of intersections and unions can be extended to infintely many sets. For example, let A_i be the event that there are exactly i car accidents in Albuquerque this year. Let A be the event that there is at least one accident in Albuquerque this year. Then

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots$$

Such a (countably) infinite union can also be written using notation similar to the summation notation Σ:

$$A = \bigcup_{i=1}^{\infty} A_i$$

Large numbers of intersections can also be presented this way. Let W_i be the event that a team wins the ith round in a tournament. Then the event that the team wins rounds $1, \dots, n$ can be represented as

$$\bigcap_{i=1}^{n} W_i.$$

We might also describe the **complement** of an event as the set of outcomes not included in the event. In set notation, we can write $E^c = S \setminus E$. For example, if $S = \{1, 2, 3, 4, 5, 6\}$ and $E = \{2, 3, 5\}$, then $E^c = \{1, 4, 6\}$.

An important rule for complements is **DeMorgan's Law**:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$$

which you can convince yourself of using Venn Diagrams.

As an example, if A is the event that it rains today and B is the event that it snows today, then $(A \cup B)^c$ means that it is not the case that there is either rain OR snow today, and this is equivalent to there not being rain AND there not being snow today.

DeMorgan's Law extends to larger unions and intersections as well:

$$\left(\bigcup A_i\right)^c = \bigcap A_i^c \quad \left(\bigcap A_i\right)^c = \bigcup A_i^c$$

Probabilities

Probabilities are numbers between 0 and 1 assigned to events. Mathematically, we can think of a probability as a function from the set of events that maps onto real numbers between 0 and 1. Different probability models can assign different probabilities to the same set of events.

It is common to present probability functions as obeying the following axioms:

- For any event E, $0 \leq P(E) \leq 1$
- For the sample space S, $P(S) = 1$
- If $E_1, E_2, \ldots$ are mutually exclusive, then $P(\cup E_i) = \sum P(E_i)$

The third axiom also applies to finite numbers of events, for example if E_1 and E_2 are mutually exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Some helpful properties of probabilities are

1. $P(A^c) = 1 - P(A)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(AB)$.
3. $P(A \cup B) \leq P(A) + P(B)$
4. $P(A \cup B) \geq P(A) + P(B) - 1$
5. A and B are **independent** if and only if $P(AB) = P(A)P(B)$
6. If A and B are **mutually exclusive**, then $P(A \cup B) = P(A) + P(B)$
7. $P((A \cup B)^c) = 1 - P(A^c B^c)$

A common approach to simple probability problems involving dice and other highly symmetrical situations is to assume that different outcomes are equally likely. Then the probability of an event is the number of outcomes in the event divided by the number of possible, equally likely outcomes. For example, the probability of rolling an even number on a 6-sided die is $3/6 = 1/2$.

As an example of applying some of these rules, there is the birthday example. This is a famous example saying that if there are more than n people in a room, chances are at least two people have the same birthday. We'll try that for this class.

To simplify the problem, suppose that no one is born on February 29th, and that everyone has the same chance of being born on any of the other 365 days. Also assume that everyone's birthday is independent (there are no twins or close siblings in the room). The first person's birthday can be anything. Then the probability that the second person's birthday is distinct from the first person's birthday is 364/365. If the first two people have distinct birthdays, there are 363 choices for the third person's birthday that lead to three distinct birthdays. Continuing in this way, the probability of n distinct birthdays is

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365 - n + 1}{365}$$

We have to be a little careful that we counted exactly n people here, so that the final term is $365 - n + 1$ rather than $365 - n$ or $365 - n - 1$. If $n = 4$, then the formula says the final term is $(365 - 4 + 1)/365 = 362/365$, which is correct.

A calculator or program (in say, MATLAB or R) can be used to calculate the probability for specific values of n.

Conditional probability

Often we have partial information about an event which can affect it's probability. For example, suppose you roll a die, and you know that it isn't a 1. What is the probability that it is a 6? Let E_i be the event that an i occurs. Then we write this conditional probability as $P(E_6|E_1^c)$ where the vertical bar is read "given that".

In this case, assuming that the die is fair, you can guess that there are 5 possible die rolls, $2, \ldots, 6$, and since each is equally likely, each should have a probability of 1/5 since the five probabilities should be equal and add up to 1. Thus $P(E_6|E_1^c) = 1/5$.

This is correct, but difficult to apply for more complicated cases. Suppose that two dice are rolled, then what is the probability that the second die is a 6 given that at least one of the dice is a 6? This is a bit trickier. There are three ways that at least one of the dice could be a 6: the first die is a 6 but not the second, the second one is but not the first, and both are 6.

Recall the sample space $S =$

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	**(1,6)**
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	**(2,6)**
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	**(3,6)**
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	**(4,6)**
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	**(5,6)**
(6,1)	**(6,2)**	**(6,3)**	**(6,4)**	**(6,5)**	**(6,6)**

I have bolded all of the cases where at least one of the die rolls is a 6. We can think of the conditional probability as a probability on this reduced sample space. That is, let the reduced sample space be

$$S' = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

There are 11 outcomes in this reduced sample space. We still would expect all of these outcomes to be equally likely. The probability that the second die is a 6 given that at least one roll is a 6 is the number of favorable outcomes in the reduced space divided by the total number of outcomes in the reduced space. Let E be the event that there is at least one 6 and F the event that the second roll is a 6. We have This is

$$P(F|E) = P_{S'}(F) = \frac{6}{11}$$

where $P_{S'}$ is the name for the probability function on the reduced sample space. We can have more than one probability function floating around just like a math problem can have multiple functions, f_1, f_2, etc. We can also think of this in the following way

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{6/36}{11/36}$$

and this is the usual formula for conditional probabilities, and it holds for any events E and F where $P(E) > 0$. We take this to be the definition

For any events, E and F with $P(F) > 0, P(F|E) = P(E \cap F)/P(F)$

The definition holds even for cases where the underlying sample space does not have equally likely outcomes.

Example. Suppose you are playing craps at a casino. What is the probability that you rolled a 7 given that you won on the first toss?

Solution. Let A be the event that you rolled a 7 on the first toss and B the event that you rolled an 11. It helps to know that $P(A) = 6/36$, $P(B) = 2/36$. Then

$$P(A|A \cup B) = \frac{P(A \cap (A \cup B))}{P(A \cup B)} = \frac{P(A)}{P(A) + P(B)} = \frac{6/36}{6/36 + 2/36} = 6/8 = 3/4$$

To get this result, I used the fact that A and B are mutually exclusive so that $P(A \cup B) = P(A) + P(B)$. I also used the fact that $A \cap (A \cup B) = A$. Why is this true? The intuitive way of seeing it is that, in the context of the problem, if you rolled a 7 AND you rolled a 7 or an 11, then this tells you nothing more than that you rolled a 7.

To show more formally that two sets (events) are equal, you show that if x belongs to one set, then it belongs to the other. In this case, if $x \in A$, then $x \in A \cup B$ since if $x \in A$, then it is true that either $x \in A$ or $x \in B$. On the other hand, if $x \in A \cap (A \cup B)$, then $x \in A$ and $x \in A\ cupB$, so $x \in A$. Thus, if x belongs to the left hand side, then it belongs to the right-hand side, and vice versa, so A and $A \cap (A \cup B)$ have the same elements.

The set manipulation is similar to "disjunction introduction" in symbolic logic. According to this principle, if P is true, then the statement "P or Q" is also true. Similarly here, if A occurs, then $A \cup B$ occurs, and so does $A \cap (A \cup B)$, so you could write

$$A \Rightarrow A \cup B \Rightarrow A \cap (A \cup B) \text{ and } A \cap (A \cup B) \Rightarrow A$$

Example. A famously confusing problem is the following: A king has one sibling. What is the probability that his sibling is a sister?

In the problem, assume that males and females are equally likely in the population (and in royal families), and don't assume that the king was born first (or was more likely to be born first).

If you didn't know anything about the family except that there were two kids, the sample space would be $S = \{bb, bg, gb, gg\}$, where gg means that two girls were born, and bg means that a boy was born first followed by a girl. In this sample space, all outcomes are equally probable. We are given the information that the family has at least one boy. Thus the reduced sample space is $S' = \{bb, bg, gb\}$. In the reduced sample space, two out of three outcomes result in the boy having a sister, so the answer is 2/3. Another way to do the problem is let A be the event that the first child is a boy and B the event that the second child is a boy. Let C be the event that the king has a sister.

$$P(C|A \cup B) = \frac{P(C \cap (A \cup B))}{P(A \cup B)} = \frac{P((C \cap A) \cup (C \cap B))}{3/4} = \frac{P(\{bg\} \cup \{gb\})}{3/4} = \frac{2/4}{3/4} = \frac{2}{3}$$

It is tempting to say that the answer is 1/2. This would be correct if we knew that the first child is a boy and that there are two children. But knowing that at least one child is boy is different information than knowing that the first child is a boy, and results in a different probability.

Total Probability. An important formula for manipulating conditional probabilities is called the Total Probability formula. The idea is that if you partition a sample space into events $A_1, A_2, \ldots, A_n$ such that the A_is are mutually exclusive and $\cup_i A_i = S$, then for any event B,

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

This formula is very useful for dividing a difficult probability problem into separate cases that are easier to handle. A special case occurs when $n = 2$. In this case, the space is partitioned into A and A^c, resulting in the following

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Example As an example, suppose that if I take bus 1, then I am late with probability 0.1. If I take bus 2, I am late with probability 0.2. The probability that I take bus 1 is 0.4, and the probability that I take bus 2 is 0.6. What is the probability that I am late?

Solution.

$$P(L) = P(L|\text{ bus } 1)P(\text{bus } 1) + P(L|\text{bus } 2)P(\text{bus } 2) = (0.1)(0.6) + (0.2)(0.4) = 0.06 + 0.08 = 0.14$$

Bayes Formula Suppose in the previous example, I was late. What was the probability that I took bus 2? Bayes formula is useful for this kind of problem. We were given information about the probability of being late given that I took bus 2, but now the question flips the events around and asks us the probability of taking bus 2 given that I was late. This is sometimes called "inverse probability", but it is just a consequence of standard probability. The formula is derived as follows:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(BA)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

From the first term to the fourth, notice that we have inverted which event gets conditioned from $A|B$ to $B|A$. Often this is written in the following form, which replaces $P(B)$ in the denominator with the total probability formula:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_i P(B|A_i)P(A_i)}$$

In the bus example we have

$$P(\text{bus } 2|L) = \frac{P(L|\text{bus } 2)P(\text{bus } 2)}{P(L)} = \frac{(0.2)(0.4)}{0.14} = \frac{4}{7} > \frac{1}{2}$$

Notice that $P(\text{bus } 2) = 0.4 < 1/2$, but knowing that I'm late increases the conditional probability of having taken bus 2 to a value greater than 1/2. This makes sense: knowling that I'm late gives some evidence that I took the bus that is more likely to result in me being late. This use of Bayes Formula is called **updating**.

Example Consider a game in which you and a friend take turns throwing a single fair die. The first player to roll a 6 wins. What is the probability that the first player to roll wins the game?

Solution. This can be solved with conditional probability. We divide the game into three cases: Player 1 wins on the first toss, Player 2 wins on the second toss, and neither player wins on the first or second toss. Call these events A_1, A_2, and A_3, and let W be the event that Player 1 wins. Then

$$\begin{aligned}
P(W) &= P(W|A_1)P(A_1) + P(W|A_2)P(A_2) + P(W|A_3)P(A_3) \\
&= 1 \cdot \frac{1}{6} + 0 \cdot \frac{5}{6} \cdot \frac{1}{6} + P(W) \cdot \frac{5}{6} \cdot \frac{5}{6} \\
&= \frac{1}{6} + \frac{25}{36} \cdot P(W) \\
&\Rightarrow P(W) - \frac{25}{36} P(W) = \frac{1}{6} \\
&\Rightarrow \frac{11}{36} P(W) = \frac{1}{6} \\
&\Rightarrow P(W) = \frac{6}{11}
\end{aligned}$$

We conclude that the first player has some advantage (55% versus 45% chance of winning). This problem involved putting $P(W)$ on both sides of the equation and then solving for $P(W)$, which is often a useful technique. Another way of solving is to use infinite series. Let W_i be the event that Player 1 wins on the ith turn. For $i > 1$, this means that there were $i - 1$ turns where no one won, and that Player 1 wins on the

ith turn. The probability of no one winning on the ith turn is $(5/6)^2 = (25/36)$. Therefore we have

$$\begin{aligned}
P(W) &= P(\cup_{i=1}^{\infty} W_i) \\
&= \sum_{i=1}^{\infty} P(W_i) \\
&= \sum_{i=1}^{\infty} (25/36)^{i-1}(1/6) \\
&= (1/6)\sum_{i=0}^{\infty} (25/36)^i \\
&= \frac{(1/6)}{1 - 25/36} \\
&= \frac{6}{11}
\end{aligned}$$

using the fact that we have a geometric series $\sum_i r^i$ with $r = 25/36 < 1$.

Random variables

The story you hear in many textbooks is that random variables are functions that map outcomes into real numbers. In many cases, we think of random variables as numerical, but often in biology we don't. Instead we might think of random variables as being DNA letters, amino acids, or evolutionary trees, and it is seems artificial to convert these into real numbers. Random variables can also be random vectors, random matrices, and other more complicated objects.

In many cases, a random variable is a way of describing some family of events. Instead of using E_i for the event that an i is rolled, we might let X be the random variable denoting the roll of a die and let $X = 1, 2, 3, 4, 5$, or 6. We then might be interested in $P(X > 4)$, for instance, and this is the same thing as $P(E_5 \cup E_6)$. We can also go in another direction, by letting $X = 1$ if event E occurs and $X = 0$ otherwise. Then $P(X = 1) = P(E)$. We can go back and forth between events and random variables depending on what is convenient.

When random variables are numeric, we can think of their average (mean) value, how much they vary, or other mathematical properties. In many cases, random variables are also used for continuous quantities. We might model the high temperature today as a random variable X and be intersted in $P(X < 100)$, for instance, or $P(99 < X < 100)$, the probability that X is in a certain interval.

A convention for random variables is to use upper case roman italics, the same font as for events, but letters near the end of the alphabet, such as X, Y, and Z are more likely to be used.

Random variables are often classified as being discrete (in which case they take a finite or countably infinite number of possible values, for example integers), or continuous, in which case they are typically real numbers. Occasionally a random variable is a mixture of both. For example, a random variable T be might an evolutionary tree with both a topology and a vector of branch lengths. Another example might be a random pair (X, Z) where X is a person's age and Z is their gender.

Random variables are often used in probability models. A typical example of a random variable is to let X be the sum of two 6-sided dice. We might be interested in $P(X = i)$ or $P(X \in \{7, 11\})$, for example. We also work with multiple random variables. For example, if X_1 is the first die and X_2 is the second die, then we might want $P(\max(X_1, X_2) = i)$ for different values of i.

For a discrete random variable, the list (or formula) of all possible values of the variable is its **probability mass function**. If X_1 is the value of a six-sided die, we have

$$P(X = i) = \begin{cases} 1/6 & i \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases}$$

As another example, if $Z = \max(X_1, X_2)$, then the mass function for Z is

$$P(Z = i) = \begin{cases} 1/36 & i = 1 \\ 3/36 & i = 2 \\ 5/36 & i = 3 \\ 7/36 & i = 4 \\ 9/36 & i = 5 \\ 11/36 & i = 6 \\ 0 & \text{otherwise} \end{cases}$$

It is traditional to specify probability distributions for all real numbers, and the "otherwise" case is put in for that reason.

A few discrete distributions come up a lot. One is the binomial distribution. This occurs when you have n trials, each independent and each with a probability p of a particular outcome, often called "success". For example, if you roll a die 5 times and let X be the number of 6s that appear, then assuming that the die is fair and the rolls are independent, then X has a binomial distribution with $n = 4$ and $p = 1/6$. If the die is not fair but the rolls are still independent, then X is binomial with $n = 4$ and some $p \neq 1/6$.

We can try to determine the distribution of X by hand. Denote the rolls of the die by either `6` or `-` in the case of any non 6 roll. Then here are the possible sequences of rolls

```
----
---6
--6-
-6--
6---
--66
-6-6
6--6
-66-
6-6-
66--
-666
6-66
66-6
666-
6666
```

How many results were possible with this way of counting? Answer: 16. How do we know this? We can think of each sequence as 0s and 1s instead of `6` and `-`, and we are considering all binary sequences of length 4. There are $2^4 = 16$ such sequences, so we can be sure we got them all. Is each sequence equally likely? No, but the probability of each sequence only depends on the number of 6s. Note that there is 1 sequence with no 6s, 4 with one 6, 6 with two 6s, 4 with three 6s, and 1 with all 6s. These numbers are **binomial coefficients**.

Because there are 6 sequences with two 6s, the probability that $X = 2$ is $6 \cdot (5/6)^2(1/6)^2$. For example the sequence –66 has probability $(5/6)^2(1/6)^2$ and the sequence -6-6 has probability $(5/6)(1/6)(5/6)(1/6) = (5/6)^2(1/6)^2$. Since there are six sequences with two 6s, we can multiply the probability of any particular sequence by 6 to get the total probability of two 6s.

For larger problems, we want a way to count how many sequences have a certain number of successes instead of enumerating them. This is done with binomial coefficients. If there are n trials, then the number of ways to have k successes is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \times (n-1) \times \cdots \times (n-k+1)}{k!}$$

I think of this like pizza toppings. If a pizza place has only four toppings, pepperoni, mushroom, olives, and green chile, how many ways can you order a two-topping pizza? The answer is $\binom{4}{2} = 6$ assuming that we don't care about the order, e.g., that a pepperoni and green chile pizza is the same as green chile and pepperoni pizza. If a pizza place has 10 toppings, how many ways can you order a 3-topping pizza?

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 10 \times 3 \times 4 = 120.$$

Getting back to the probability problem. If you have a binomial random variable with n trials, each with probability p of success, then

$$P(X = i) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i} & i \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

The term $p^i(1-p)^{n-i}$ is the probability that say, the first i trials are a success, and the next $n-i$ trials are failures. The term $\binom{n}{i}$ counts how many sequences of successes and failures there are with exactly i successes.

Birthday Problem revisited. Now we have the tools to do the birthday problem allowing for the possibility that someone is born on February 29th. Here we assume that all birthdays are equally likely within a 4-year span of time (which includes one February 29th and that all birthdays are independent. Again we ask what is the probability that at least two people have the same birthday.

Solution. First, let the probability that someone is born on a particular day other than February 29th be $a = 4/(365 * 4 + 1) = 4/1461$, and let the probability that someone is born on February 29th be $b = 1/1461$. We consider the following events: $A =$ No one is born on February 29th, $B =$ exactly one person is born on February 29th, $C =$ two or more people are born on February 29th. Then let E be the event that two or more people are born on the same day. The total probability formula gives us

$$P(E^c) = P(E^c|A)P(A) + P(E^c|B)P(B) + P(E^c|C)P(C)$$

Assuming that no one is born on February 29th, everyone is equally likely to be born on any of the other 365, so this reduces to the original, simplified birthday problem. Let the probability that everyone has a distinct birthday in the original birthday problem be denoted $p(n)$. The probability that noone is born on February 29th is $P(A) = (1-b)^n$. Assuming that only one person is born on February 29th, the person does not have a tied birthday with anyone else, and so $P(E|B)$ reduces to the original birthday problem with 1 less person. Thus, $P(E|B) = p(n-1)$. The probability that exactly one person has a birthday on February 29th is the probability of a binomial random variable $X = 1$ with n trials and $b = 1/1461$. Thus $P(B) = P(X = 1) = nb(1-b)^{n-1}$. The probability $P(E^c|C)$ is equal to 0 since if C occurs, then it is impossible for everyone to have a distinct birthday. This way, we avoid having to compute $P(C)$. Combining everything we have

$$P(E^c) = p(n)(1-b)^n + p(n-1)nb(1-b)^{n-1}$$

and $P(E) = 1 - P(E^c)$.

How much of a difference does this improved, more realistic model make?

Here is some R code to test assuming that `bd(n)` is the probability that no two people have the same birthday in the original birthday model:

```
> bd2
function(n) {
prob <- 1
b <- 1/1461
prob <- bd(n)*(1-b)^n + bd(n-1)*n*b*(1-b)^(n-1)
return(prob)
```

```
}
> bd3
function(n) {
return((1-bd(n))/(1-bd2(n)))
}
> sapply(2:100,bd3)
> sapply(2:100,bd3)
 [1] 1.001199 1.001197 1.001194 1.001189 1.001183 1.001175 1.001165 1.001154
 [9] 1.001141 1.001127 1.001111 1.001094 1.001075 1.001055 1.001033 1.001011
[17] 1.000987 1.000962 1.000936 1.000909 1.000881 1.000853 1.000823 1.000794
[25] 1.000763 1.000733 1.000702 1.000670 1.000639 1.000608 1.000577 1.000546
[33] 1.000516 1.000486 1.000456 1.000427 1.000399 1.000372 1.000345 1.000320
[41] 1.000296 1.000272 1.000250 1.000228 1.000208 1.000189 1.000172 1.000155
[49] 1.000140 1.000125 1.000112 1.000100 1.000089 1.000078 1.000069 1.000061
[57] 1.000053 1.000046 1.000040 1.000035 1.000030 1.000026 1.000022 1.000019
[65] 1.000016 1.000014 1.000011 1.000010 1.000008 1.000007 1.000006 1.000005
[73] 1.000004 1.000003 1.000003 1.000002 1.000002 1.000001 1.000001 1.000001
[81] 1.000001 1.000001 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
[89] 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
[97] 1.000000 1.000000 1.000000
```

Here `bd3()` computes the ratio in probabilities of at least two people with the same birthday under the two models.

In other words, it makes very little difference, less than 1% for small n, whether we worry about the February 29th case or not. The numbers do suggest that not worrying about February 29th makes the probabilities slightly higher that two or more people will have the same birthday. However, for 23 people, this probability is 0.5072972 for the model that ignores February 29th and 0.506865 for the model that accounts for February 29th.

The take home story: Is the extra realism in modeling worth the extra effort? Probably not in this case. Other assumptions might be more important, like assuming that people are equally likely to be born at different times of year. If people are both

Probabilities as definite integrals

For continuous random variables, a probability can be represented by an integral of the density function. If a continuous random variable has density $f_X(x)$, then $P(a < X < b) = \int_a^b f_X(x)dx$. Note that there is a convention of using an upper case letter for a random variable and a lower case letter for particular values of a random variable. Thus we may write $P(X > x)$ for the probability that a random variable X is larger than the value x, and this convention is often used for both discrete and continuous random variables.

As an example of using integrals for probabilities, let

$$f(x) = \begin{cases} 3e^{-3x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

A random variable X with this density is an exponential random variable with rate 3. Then

$$P(X > 1) = \int_1^\infty 3e^{-3x}dx = -e^{-3x}\big|_1^\infty = e^{-3}$$

In some cases, you have two random variables to consider simultaneously. Let X have density $f_X(x)$ and let Y have density $f_Y(y)$. Suppose X and Y are independent. Then the joint density is $f_{X,Y} = f_X(x)f_Y(y)$. Probabilities involving joint densities can be obtained by integrating the joint density over both x and y.

Example. One lightbulb has a lifespan that is exponential with mean 500 hrs. Another lightbulb costs twice as much as has a lifespan with mean 1000 hrs. This means that the rate at which the cheaper lightbult burns out is 2 (per 1000 hrs of use), and the more expensive lightbulb has a rate of 1 (per 1000 hrs of use). What is the probability that the more expensive lightbult lasts longer?

Solution. Let X be the lifespan of the cheaper lightbult. Then $X \sim exp(2)$, and $Y \sim exp(1)$. Thus

$$\begin{aligned} P(Y > X) &= \int_0^\infty \int_x^\infty 2e^{-2x} e^{-y} \, dy \, dx \\ &= \int_0^\infty 2e^{-2x} (-e^{-y}) \Big|_x^\infty \, dx \\ &= \int_0^\infty 2e^{-3x} \, dx \\ &= -\frac{2}{3} e^{-3x} \Big|_0^\infty \\ &= \frac{2}{3} \end{aligned}$$

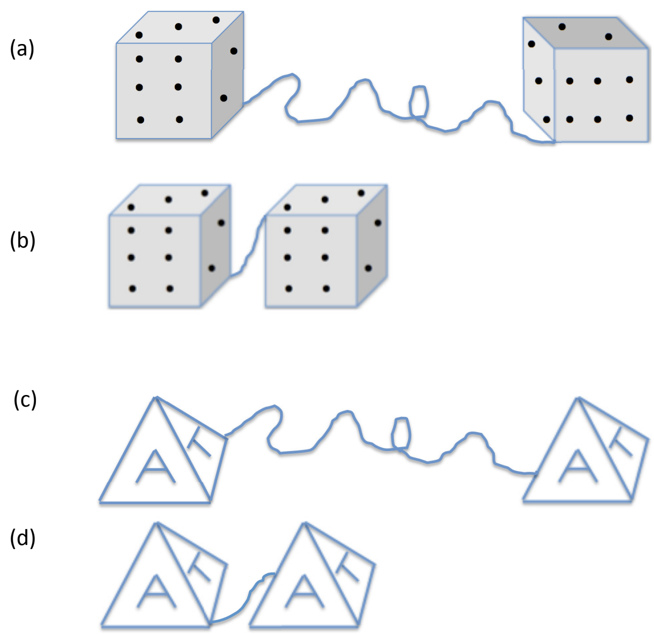

Figure 1: Non-independence for dice. The pair of dice I(a) are connected by a long string which might slightly constrain their joint values. The pair of dice in (b) is connected by a shorter string, making them less independent. In this example, it might be physically possible but very unlikely for a roll of the bottom pair to result in two 3s. By imagining the string getting longer or shorter, we can think of independence as a matter of degree. (c) and (d) illustrate weak and strong dependence for DNA dice.