Probability Examples Based on Counting

We will now look at examples of random experiments and their corresponding sample spaces, along with possible probability spaces and events. As we do so, we'll add a couple more tools to our repertoire: (1) Stirling's Approximation; and (2) How to combine multiple independent experiments into a single probability space.

Fair Coin Flipping and Stirling's Approximation

Suppose we have an unbiased coin, and our experiment consists of flipping the coin 4 times. The sample space Ω consists of the sixteen possible sequences of H's and T's.

For a fair coin, the probabilities are assigned uniformly; the probability of each sample point is $\frac{1}{16}$.

Consider event A_2: the event that there are exactly two heads. The probability of any particular outcome with two heads (such as $HTHT$) is $\frac{1}{16}$. So the key is to count $|A_2|$. How many such outcomes are there? There are $\binom{4}{2} = 6$ ways of choosing the positions of the heads, and these choices completely specify the sequence. So $\Pr[A_2] = \frac{6}{16} = \frac{3}{8}$.

More generally, if we flip the coin n times, we get a sample space Ω of cardinality 2^n. The sample points are all possible sequences of n H's and T's.

Now consider the event A_r that we get exactly r H's when we flip the coin n times. This event consists of exactly $\binom{n}{r}$ sample points. Each has probability $\frac{1}{2^n}$. So the probability of this event, $P[A_r] = \frac{\binom{n}{r}}{2^n}$.

It is interesting to observe that as n gets larger, the denominator above gets larger exponentially in n. But the number of different A_r is just $n+1$ — and so grows only linearly with n. So how does the probability distribute itself across the different A_r? That is what the fair coin tossing examples were showing you experimentally in the first probability lecture note.

But to get an analytic grasp on what is going on, we're going to have to find a way to translate from the world of factorials (that we see in the numerator $\binom{n}{r}$) and the world of exponentials (that we see in the denominator).

The key tool for doing this is the famous Stirling's Approximation:

$$n! \approx \sqrt{2\pi n}(\frac{n}{e})^n. \tag{1}$$

In practice, this is a very good approximation. (Plot it to see this!) It is also very surprising at first glance. To understand something as simple as factorial — which just involves the multiplication of wholesome integers in order — we need to invoke three irrational numbers $\sqrt{2}, \pi, e$ that seem to come from nowhere. Why should multiplying numbers in order have anything to do with the length of a diagonal of a square, the ratio of the circumference of a circle to its diameter, and the base of natural logarithms?!? This kind of linkage across very different areas is a part of the deep beauty of mathematics.

We can use Stirling's approximation to get a better intuitive handle on $\binom{n}{r}$. To help us simplify, let's define $q = \frac{r}{n}$.

$$\begin{aligned}
\binom{n}{r} &= \frac{n!}{r!(n-r)!} && (2)\\
&\approx \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{\sqrt{2\pi r}(\frac{r}{e})^r\sqrt{2\pi(n-r)}(\frac{n-r}{e})^{n-r}} && (3)\\
&= \frac{\sqrt{n}n^n}{\sqrt{2\pi}\sqrt{r(n-r)}r^r(n-r)^{n-r}} && (4)\\
&= \frac{n^n}{\sqrt{2\pi}\sqrt{nq(1-q)}(qn)^{qn}((1-q)n)^{(1-q)n}} && (5)\\
&= \frac{1}{\sqrt{2\pi}\sqrt{nq(1-q)}q^{qn}(1-q)^{(1-q)n}} && (6)\\
&= \frac{1}{\sqrt{2\pi}\sqrt{nq(1-q)}}\left(\frac{1}{q^q(1-q)^{(1-q)}}\right)^n && (7)
\end{aligned}$$

Notice that last term is actually something exponential in n. For example, plug in $q = \frac{1}{2}$ and you will get $\frac{1}{(q)^q(1-q)^{(1-q)}} = 2$.

This gives us the curious observation that about $\frac{\sqrt{2}}{\sqrt{\pi n}}$ of all possible n-length binary strings have exactly the same number of zeros and ones. Or in the language of probability, that is the approximate chance of getting exactly the same number of heads and tails when tossing a large (even) number of coins.

Could this $\sqrt{n}$ have something to do with what we had observed earlier experimentally in a sequence of coin tosses? We'll see later in the course.

Card Shuffling

The random experiment consists of shuffling a deck of cards. Ω is equal to the set of the 52! permutations of the deck. The probability space is uniform. Note that we're really talking about an idealized mathematical model of shuffling here; in real life, there will always be a bit of bias in our shuffling. However, the mathematical model is close enough to be useful.

Poker Hands

Here's another experiment: shuffling a deck of cards and dealing a poker hand. In this case, S is the set of 52 cards and our sample space $\Omega = \{\text{all possible poker hands}\}$, which corresponds to choosing $k = 5$ objects without replacement from a set of size $n = 52$ where order does not matter. Hence, as we saw in the previous Note, $|\Omega| = \binom{52}{5} = \frac{52\times 51\times 50\times 49\times 48}{5\times 4\times 3\times 2\times 1} = 2,598,960$. Since the deck is assumed to be randomly shuffled, the probability of each outcome is equally likely and we are therefore dealing with a uniform probability space.

Let A be the event that the poker hand is a flush. [For those who are not (yet) addicted to gambling, a *flush* is a hand in which all cards have the same suit, say Hearts.] Since the probability space is uniform, computing $\Pr[A]$ reduces to simply computing $|A|$, or the number of poker hands which are flushes. There are 13 cards in each suit, so the number of flushes in each suit is $\binom{13}{5}$. The total number of flushes is therefore $4\cdot\binom{13}{5}$. Then we have

$$\Pr[\text{hand is a flush}] = \frac{4\cdot\binom{13}{5}}{\binom{52}{5}} = \frac{4\cdot 13!\cdot 5!\cdot 47!}{5!\cdot 8!\cdot 52!} = \frac{4\cdot 13\cdot 12\cdot 11\cdot 10\cdot 9}{52\cdot 51\cdot 50\cdot 49\cdot 48} \approx 0.002.$$

As an exercise, you should compare to what the Stirling's approximation would yield for the above exact calculation.

Balls and Bins

In this experiment, we will throw 20 (labeled) balls into 10 (labeled) bins. Assume that each ball is equally likely to land in any bin, regardless of what happens to the other balls.

If you wish to understand this situation in terms of sampling a sequence of k elements from a set S of cardinality n: here the set S consists of the 10 bins, and we are sampling with replacement $k = 20$ times. The order of sampling matters, since the balls are labeled.

The sample space Ω is equal to $\{(b_1, b_2, \dots, b_{20}) : 1 \le b_i \le 10\}$, where the component b_i denotes the bin in which ball i lands. The cardinality of the sample space, $|\Omega|$, is equal to 10^{20}- each element b_i in the sequence has 10 possible choices, and there are 20 elements in the sequence. More generally, if we throw m balls into n bins, we have a sample space of size n^m. The probability space is uniform; as we said earlier, each ball is equally likely to land in any bin.

Let A be the event that bin 1 is empty. Since the probability space is uniform, we simply need to count how many outcomes have this property. This is exactly the number of ways all 20 balls can fall into the remaining nine bins, which is 9^{20}. Hence, $\Pr[A] = \frac{9^{20}}{10^{20}} = (\frac{9}{10})^{20} \approx 0.12$.

Let B be the event that bin 1 contains at least one ball. This event is the *complement* $\bar{A}$ of A, i.e., it consists of precisely those sample points which are not in A. So $Pr[B] = 1 - \Pr[A] \approx .88$. More generally, if we throw m balls into n bins, we have:

$$\Pr[\text{bin 1 is empty}] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m.$$

As we shall see, balls and bins is another probability space that shows up very often in EECS: for example, we can think of it as modeling a load balancing scheme, in which each job is sent to a random processor.

It is also a more general model for problems we have previously considered. For example, flipping a fair coin 3 times is a special case in which the number of balls (m) is 3 and the number of bins (n) is 2. Rolling two dice (an example in the previous lecture notes) is a special case in which $m = 2$ and $n = 6$.

Birthday Paradox

The "birthday paradox" is a remarkable phenomenon that examines the chances that two people in a group have the same birthday. It is a "paradox" not because of a logical contradiction, but because it goes against intuition. For ease of calculation, we take the number of days in a year to be 365. Then $U = \{1, \dots, 365\}$, and the random experiment consists of drawing a sample of n elements from U, where the elements are the birth dates of n people in a group. Then $|\Omega| = 365^n$. This is because each sample point is a sequence of possible birthdays for n people; so there are n points in the sequence and each point has 365 possible values.

Let A be the event that at least two people have the same birthday. If we want to determine $\Pr[A]$, it might be simpler to instead compute the probability of the complement of A, $\Pr[\bar{A}]$. $\bar{A}$ is the event that no two people have the same birthday. Since $\Pr[A] = 1 - \Pr[\bar{A}]$, we can then easily compute $\Pr[A]$.

We are again working in a uniform probability space, so we just need to determine $|\bar{A}|$. Equivalently, we are computing the number of ways there are for no two people to have the same birthday. There are 365 choices for the first person, 364 for the second, ..., $365 - n + 1$ choices for the n^{th} person, for a total of

$365 \times 364 \times \cdots \times (365-n+1)$. Note that this is simply an application of the first rule of counting; we are sampling without replacement and the order matters.

Thus we have $\Pr[\bar{A}] = \frac{|\bar{A}|}{|\Omega|} = \frac{365\times364\times\cdots\times(365-n+1)}{365^n}$. Then $\Pr[A] = 1 - \frac{365\times364\times\cdots\times(365-n+1)}{365^n}$. This allows us to compute $\Pr[A]$ as a function of the number of people, n. Of course, as n increases $\Pr[A]$ increases. In fact, with $n = 23$ people you should be willing to bet that at least two people do have the same birthday, since then $\Pr[A]$ is larger than 50%! For $n = 60$ people, $\Pr[A]$ is over 99%.

This is a somewhat surprising aspect of the nature of random fluctuations. We will study the engineering impacts of this later when we think about load balancing and hash tables.

Unfair coins and how to combine experiments into a single probability space

Suppose that instead of a fair coin, we have a coin of bias p, and our experiment consists of flipping the coin 4 times. The sample space Ω still consists of the sixteen possible sequences of H's and T's.

However, the probability space depends on p. If $p = \frac{1}{2}$ the probabilities are assigned uniformly; the probability of each sample point is $\frac{1}{16}$. What if the coin comes up heads with probability $\frac{2}{3}$ and tails with probability $\frac{1}{3}$ (i.e. the bias is $p = \frac{2}{3}$)? Then the probabilities of different outcomes are different.

For example, $\Pr[HHHH] = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{16}{81}$, while $\Pr[TTHH] = \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{81}$. [*Note: We have cheerfully* multiplied *probabilities here; But why is this allowed? It is* not *always OK!*]

The multiplication here can be justified in terms of uniform probabilities by thinking of the underlying experiment as rolling a fair three-sided die instead. We say "H" when the die comes up 1 or 2 and say "T" when the die comes up 3. In that case, there are a grand total of $81 = 3^4$ possible die-roll sequences. And the 16 above comes from the $2 \cdot 2 \cdot 2 \cdot 2$ different outcomes that all get labeled HHHH. The same argument works for justifying the calculation of TTHH's probability. What we are doing here is arguing what the $\frac{2}{3}$ probability of heads means.

In general, when we take make a new unified probability space out of two completely independent and separate experiments (these are experiments that cannot influence each other in any way), then the new outcomes are pairs of outcomes. So the new sample space $\Omega = \Omega_1 \times \Omega_2$ using set theoretic notation for the set of all pairs where the first element is from Ω_1 and the second is from Ω_2. The rule for evaluating the probability of the individual outcomes $\omega = (\omega_1, \omega_2)$ is just $\Pr[\omega] = \Pr_1[\omega_1] \cdot \Pr_2[\omega_2]$. This is generalized in the natural way to lists longer than two.

For the unfair coins here, we are looking at 4-tuples.

What type of events can we consider in this setting? Let event S be the event that all four coin tosses are the same. Then $S = \{HHHH, TTTT\}$. $HHHH$ has probability $(\frac{2}{3})^4$ and $TTTT$ has probability $(\frac{1}{3})^4$. Thus, $\Pr[S] = \Pr[HHHH] + \Pr[TTTT] = (\frac{2}{3})^4 + (\frac{1}{3})^4 = \frac{17}{81}$.

Next, consider the event A_2: the event that there are exactly two heads. We had seen this one earlier in the fair coin setting. This time, the probability of any particular outcome with two heads (such as $HTHT$) is $(\frac{2}{3})^2(\frac{1}{3})^2$. Notice that the order of the heads doesn't matter in doing this calculation. This is because real multiplication commutes. Once again, the key is to count $|A_2|$. There are $\binom{4}{2} = 6$ ways of choosing the positions of the heads, and so $\Pr[A_2] = 6(\frac{2}{3})^2(\frac{1}{3})^2 = \frac{24}{81} = \frac{8}{27}$.

More generally, if we flip the biased coin n times, we get a sample space Ω of cardinality 2^n. This is the same as the fair coin toss case. If the coin has bias p, and if we consider any sequence of n coin flips with exactly r H's, then the probability of this sequence is $p^r(1-p)^{n-r}$.

As before, we can consider the event A_r that we get exactly r H's when we flip the coin n times. This event consists of exactly $\binom{n}{r}$ sample points and so the probability of this event, $P[A_r] = \binom{n}{r} p^r (1-p)^{n-r}$.

We can use our earlier Stirling's approximation-based calculation to shed some light on what this is like. As before, let $q = \frac{r}{n}$.

$$P[A_r] = \binom{n}{r} p^r (1-p)^{n-r} \tag{8}$$

$$\approx \frac{1}{\sqrt{2\pi}\sqrt{nq(1-q)}} \left(\frac{1}{q^q(1-q)^{(1-q)}}\right)^n (p^q(1-p)^{1-q})^n \tag{9}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{nq(1-q)}} \left((\frac{p}{q})^q (\frac{1-p}{1-q})^{(1-q)} \right)^n \tag{10}$$

It is very interesting to observe what happens when $p = q$ above. In that case, Stirling's approximation says that the probability of getting this outcome is like $\frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}}$.

Biased coin-tossing sequences show up in many contexts: for example, they might model the behavior of n trials of a faulty system, which fails each time with probability p.

Conditional Probability Examples

With counting available to us, we can consider some more simple examples that deal with conditional probability.

Card Dealing

What is the probability that, when dealing 2 cards and the first card is known to be an ace, the second card is also an ace?

Let B be the event that the first card is an ace, and let A be the event that the second card is an ace. Note that $P[A] = P[B] = \frac{1}{13}$.

To compute $\Pr[A|B]$, we need to figure out $\Pr[A \cap B]$. This is the probability that both cards are aces. Note that there are $52 \cdot 51$ sample points in the sample space, since each sample point is a sequence of two cards. A sample point is in $A \cap B$ if both cards are aces. This can happen in $4 \cdot 3 = 12$ ways.

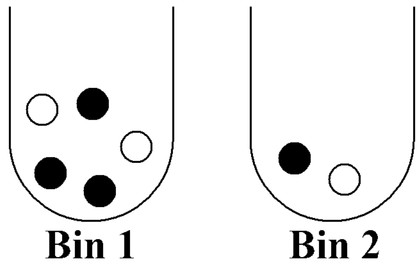
Since each sample point is equally likely, $\Pr[A \cap B] = \frac{12}{52 \cdot 51}$. The probability of event B, drawing an ace in the first trial, is $\frac{4}{52}$. Therefore,

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{3}{51}.$$

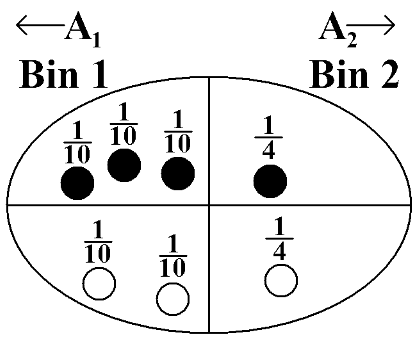
Note that this says that if the first card is an ace, it makes it less likely that the second card is also an ace.

Balls and Bins Revisited

Imagine we have two bins containing black and white balls, and further suppose that we wanted to know what is the chance that we picked Bin 1 given that we picked a white ball, i.e., $\Pr[\text{Bin } 1|\bigcirc]$. Assume that we are unbiased when choosing a bin so that each bin is chosen with probability $\frac{1}{2}$.



A wrong approach is to say that the answer is clearly $\frac{2}{3}$, since we know there are a total of three white balls, two of which are in bin 1. However, this picture is misleading because the bins have equal "weight". Instead, what we should do is appropriately scale each sample point as the following picture shows:



This images shows that the sample space Ω is equal to the union of the events contained in bin 1(A_1) and bin 2(A_2), so $\Omega = A_1 \cup A_2$. This is NOT the same as viewing this as two independent experiments. We get either something from bin 1 or something from bin 2. We don't get one from each.

We can use the definition of conditional probability to see that

$\Pr[\text{Bin } 1|\bigcirc] = \frac{\frac{1}{10}+\frac{1}{10}}{\frac{1}{10}+\frac{1}{10}+\frac{1}{4}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}$

Let us try to achieve this probability using Bayes' rule. To apply Bayes' rule, we need to compute $\Pr[\bigcirc|\text{Bin } 1]$, $\Pr[\text{Bin } 1]$ and $\Pr[\bigcirc]$. $\Pr[\bigcirc|\text{Bin } 1]$ is the chance that we pick a white ball given that we picked bin 1, which is $\frac{2}{5}$. $\Pr[\text{ Bin } 1]$ is $\frac{1}{2}$ as given in the description of the problem. Finally, $\Pr[\bigcirc]$ can be computed using the Total Probability rule:

$\Pr[\bigcirc] = \Pr[\bigcirc|\text{Bin } 1] \times \Pr[\text{Bin } 1] + \Pr[\bigcirc|\text{Bin } 2] \times \Pr[\text{Bin } 2] = \frac{2}{5} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{9}{20}$.

Observe that we can apply the Total Probability rule here because $\Pr[\text{ Bin } 1]$ is the complement of $\Pr[\text{ Bin } 2]$. Finally, if we plug the above values into Bayes' rule we obtain the probability that we picked bin 1 given that we picked a white ball:

$\Pr[\text{Bin } 1|\bigcirc] = \frac{\frac{2}{5} \times \frac{1}{2}}{\frac{9}{20}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}$.

All we have done above is combined Bayes' rule and the Total Probability rule. We could have equivalently applied Bayes' rule directly.[1]

[1]Notice here that we were able to do the calculation without being very precise about what the underlying probability space is. This is often the case when dealing with conditional probability, but it is good to be able to precisely define what outcomes are if you need to. Here, the most natural choice of the underyling sample space is to consider the composition of three independent and distinct experiments. Tossing a fair coin. Picking something from Bin 1. And picking something from Bin 2.

If we further view each of the balls as being distinct, this is a uniform probability space in which each of the 20 possible outcomes

Summary

The examples above illustrate the importance of doing probability calculations systematically, rather than just "intuitively." Recall the key steps in all our calculations:

- What is the sample space (i.e., the experiment and its set of possible outcomes)?
- What is the probability of each outcome (sample point)?
- What is the event we are interested in (i.e., which subset of the sample space)?
- Finally, compute the probability of the event by adding up the probabilities of the sample points inside it.

Whenever you meet a probability problem, you should always go back to these basics to avoid potential pitfalls. Even experienced researchers make mistakes when they forget to do this — witness many erroneous "proofs", submitted by mathematicians to newspapers at the time, of the fact that the switching strategy in the Monty Hall problem does not improve the odds.

has probability $\frac{1}{2} \cdot \frac{1}{5} \cdot \frac{1}{2} = \frac{1}{20}$. If we instead view balls only in terms of their color, then there are only $2 \cdot 2 \cdot 2 = 8$ possibilities and they are not all equally likely.