Random Variables: Distributions, Independence, and Expectations

In the last note, we saw how useful it is to have a way of thinking about quantities that are inherently random. In this section, we will introduce the notion of a random variable, which allows us to more directly deal with questions like "what is the number of . . . ".

In our coin flipping examples, let us call the number of H's in n coin flips X. Of course, X is not a fixed number. It depends upon the actual sequence of coin flips that we toss. What we know is that X is an integer between 0 and n, and if the coin is fair. We would also like to show that if the coin is fair, then with very high probability X takes on a value very close to $\frac{n}{2}$. Before we formalize the notion of a random variable, let us consider another example.

Question: The homeworks of 20 students are collected in, randomly shuffled and returned to the students. How many students receive their own homework?

To answer this question, we first need to specify the probability space: plainly, it should consist of all 20! permutations of the homeworks, each with probability $\frac{1}{20!}$. [Note that this is the same as the probability space for card shuffling, except that the number of items being shuffled is now 20 rather than 52.] It helps to have a picture of a permutation. Think of 20 books lined up on a shelf, labeled from left to right with $1,2,\ldots,20$. A permutation π is just a reordering of the books, which we can describe just by listing their labels from left to right. Let's denote by π_i the label of the book that is in position i. We are interested in the number of books that are still in their original position, i.e., in the number of i's such that $\pi_i = i$. These are often known as fixed points of the permutation.

As in the coin flipping case above, our question does not have a simple numerical answer (such as 6), because the number depends on the particular permutation we choose (i.e., on the sample point). Let's call the number of fixed points X. To make life simpler, let's also shrink the class size down to 3 for a while. The following table gives a complete listing of the sample space (of size $3! = 6$), together with the corresponding value of X for each sample point. [We use our bookshelf convention for writing a permutation: thus, for example, the permutation 312 means that book 3 is on the left, book 1 in the center, and book 2 on the right. You should check you agree with this table.]

permutation π	value of X
123	3
132	1
213	1
231	0
312	0
321	1

Thus we see that X takes on values 0, 1 or 3, depending on the sample point. A quantity like this, which takes on some numerical value at each sample point, is called a *random variable* (or *r.v.*) on the sample space.

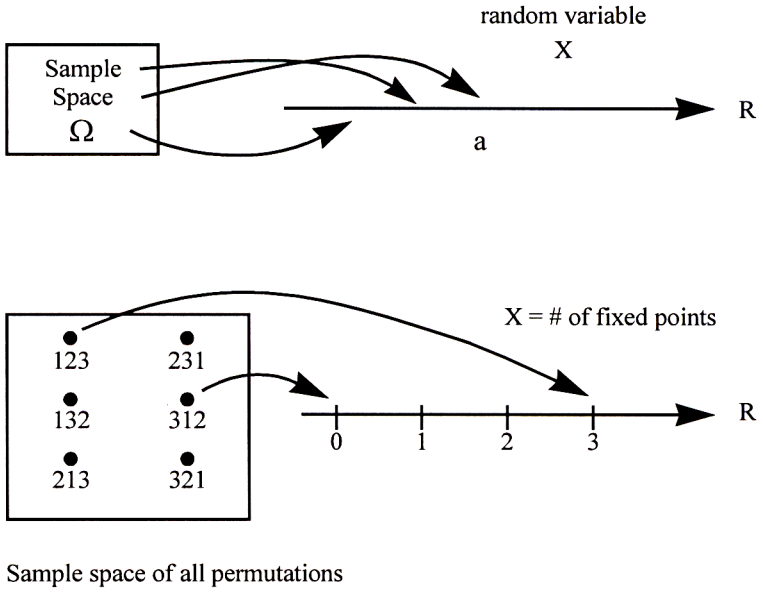

Figure 1: Visualization of how a random variable is defined on the sample space.

Random Variables

Let us formalize the concepts discussed above.

Definition 15.1 (random variable): A random variable X on a sample space Ω is a function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

Until further notice, we'll restrict out attention to random variables that are discrete, i.e., they take values in a range that is finite or countably infinite.

A random variable can be visualized in general by the picture in Figure 1[1]. Note that the term "random variable" is really something of a misnomer: it is a function so there is nothing random about it and it is definitely not a variable! What is random is which sample point of the experiment is realized and hence the value that the random variable maps the sample point to.

Distribution

When we introduced the basic probability space in Note 11, we defined two things: 1) the sample space Ω consisting of all the possible outcomes (sample points) of the experiment; 2) the probability of each of the sample points. Analogously, there are two things important about any random variable: 1) the set of values that it can take ; 2) the probabilities with which it takes on the values. Since a random variable is defined on a probability space, we can calculate these probabilities given the probabilities of the sample points. Let a be any number in the range of a random variable X. Then the set

$$\{\omega \in \Omega : X(\omega) = a\}$$

is an *event* in the sample space (why?). We usually abbreviate this event to simply "$X = a$". Since $X = a$ is an event, we can talk about its probability, $\Pr[X = a]$. The collection of these probabilities, for all possible

[1]This and other figures in this note are inspired by figures in Chapter 2 of "Introduction to Probability" by D. Bertsekas and J. Tsitsiklis.

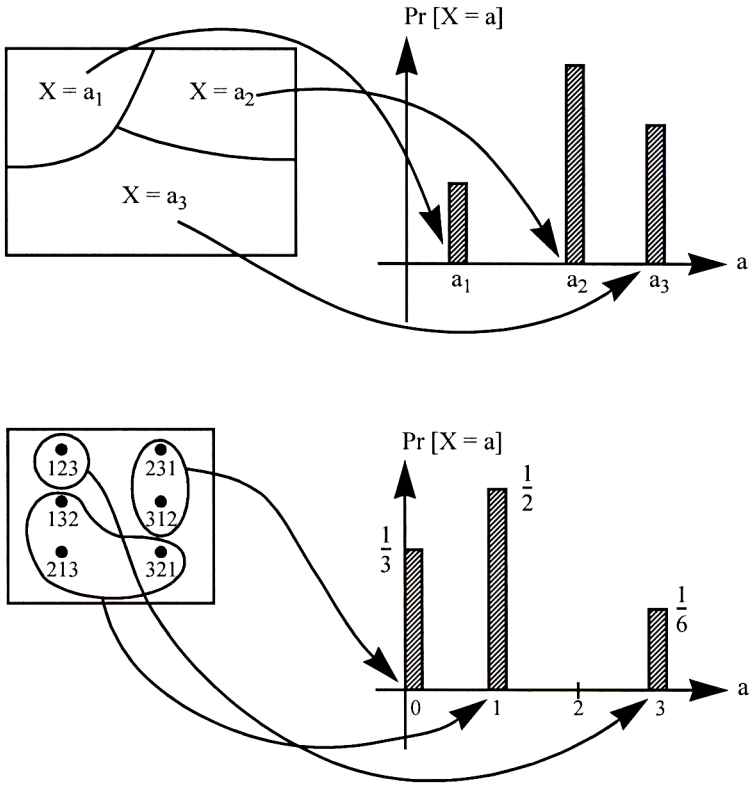

Figure 2: Visualization of how the distribution of a random variable is defined.

values of a, is known as the *distribution* of the r.v. X.

Definition 15.2 (distribution): The distribution of a discrete random variable X is the collection of values $\{(a, \Pr[X = a]) : a \in \mathscr{A}\}$, where $\mathscr{A}$ is the set of all possible values taken by X.

Thus the distribution of the random variable X in our permutation example above is

$$\Pr[X = 0] = \frac{1}{3}; \qquad \Pr[X = 1] = \frac{1}{2}; \qquad \Pr[X = 3] = \frac{1}{6};$$

and $\Pr[X = a] = 0$ for all other values of a.

The distribution of a random variable can be visualized as a bar diagram, shown in Figure 2. The x-axis represents the values that the random variable can take on. The height of the bar at a value a is the probability $\Pr[X = a]$. Each of these probabilities can be computed by looking at the probability of the corresponding event in the sample space.

Note that the collection of events $X = a$, $a \in \mathscr{A}$, satisfy two important properties:

- any two events $X = a_1$ and $X = a_2$ with $a_1 \neq a_2$ are disjoint.
- the union of all these events is equal to the entire sample space Ω.

The collection of events thus form a *partition* of the sample space (see Figure 2). Both properties follow directly from the fact that X is a function defined on Ω, i.e., X assigns a unique value to each and every possible sample point in Ω. As a consequence, the sum of the probabilities $\Pr[X = a]$ over all possible

values of a is exactly 1. So when we sum up the probabilities of the events $X = a$, we are really summing up the probabilities of all the sample points.

Another useful term is the *Probability Mass Function* (or PMF) of a random variable. This is usually denoted $P_X(i)$ and is simply defined to be $\Pr[X = i] = \Pr[\{\omega \in \Omega | X(\omega) = i\}]$. The probability mass function inherits certain properties from the probability space:

- $0 \leq P_X(i) \leq 1$ for all i.
- $\sum_i P_X(i) = 1$ where the sum is taken over all the i in the range of X.

If we have more than one random variable that we are interested in, and both of them share the same underlying probability space (i.e. they are random quantities that have to do with the same experiment — like the weight and sweetness of a randomly chosen apple), then we can also define a joint probability mass function (or joint PMF) as $P_{X,Y}(i,j) = \Pr[X = i \cap Y = j]$. Alternatively, you can think of the pair (X,Y) as being a 2-tuple-valued random variable.

Given a joint PMF, you can always recover the individual PMF (sometimes called the marginal PMF) by just summing. See if you can use the law of total probability to prove for yourself that:

$$P_X(i) = \sum_j P_{X,Y}(i,j).$$

This is a simple matter of manipulating sets and understanding that $P_X(i) = \sum_{\omega \in \Omega | X(\omega) = i} \Pr[\omega]$.

Example: The Binomial Distribution

Let's return to our example above, where we defined our random variable X to be the number of heads. More formally, consider the random experiment consisting of n independent tosses of a biased coin which lands on heads with probability p. Each sample point ω is a sequence of tosses. $X(\omega)$ is defined to be the number of heads in ω. For example, when $n = 3$, $X(THH) = 2$.

To compute the distribution of X, we first enumerate the possible values X can take on. They are simply $0, 1, \ldots, n$. Then we compute the probability of each event $X = i$ for $i = 0, \ldots, n$. The probability of the event $X = i$ is the sum of the probabilities of all the sample points with i Heads (if $n = 3$ and $i = 2$, there would be three such sample points $\{HHT, HTH, THH\}$). Any such sample point has a probability $p^i(1-p)^{n-i}$, since the coin flips are independent. There are exactly $\binom{n}{i}$ of these sample points. So

$$\Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0, 1, \ldots n \tag{1}$$

This distribution, called the *binomial* distribution, is one of the most important distributions in probability. A random variable with this distribution is called a *binomial* random variable (for brevity, we will say $X \sim Bin(n,p)$). An example of a binomial distribution is shown in Figure 3. Notice that, due to the properties of X mentioned above, it must be the case that $\sum_{i=0}^{n} \Pr[X = i] = 1$, which implies that $\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1$. This provides a probabilistic proof of the Binomial Theorem!

Although we define the binomial distribution in terms of an experiment involving tossing coins, this distribution is useful for modeling many real-world problems. Consider for example the error correction problem studied in Note 8. Recall that we wanted to encode n packets into $n+k$ packets such that the recipient can reconstruct the original n packets from any n packets received. But in practice, the number of packet

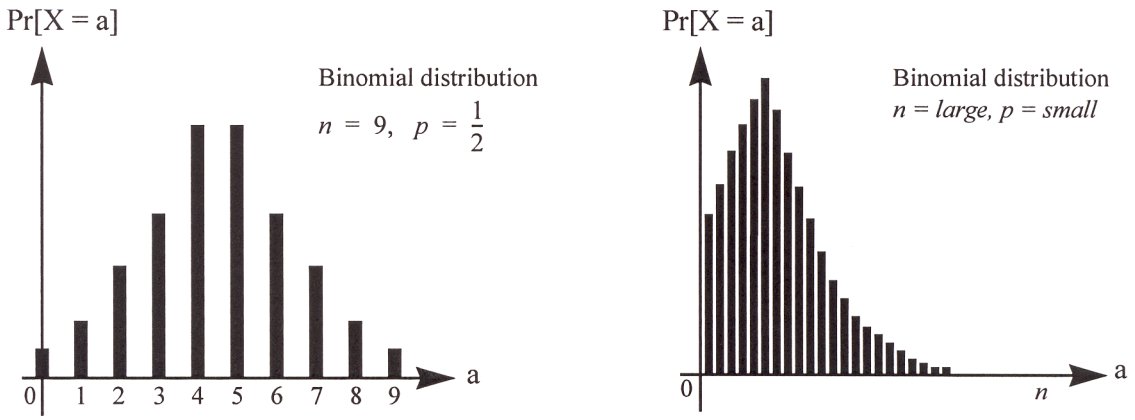

Figure 3: The binomial distributions for two choices of (n, p).

losses is random, so how do we choose k, the amount of redundancy? If we model each packet getting lost with probability p and the losses are independent, then if we transmit $n+k$ packets, the number of packets received is a random variable X and $X \sim Bin(n+k, 1-p)$. (We are tossing a coin $n+k$ times, and each coin turns out to be a Head (packet received) with probability $1-p$). So the probability of successfully decoding the original data is:

$$\Pr[X \geq n] = \sum_{i=n}^{n+k} \binom{n+k}{i} (1-p)^i p^{n+k-i}.$$

We can choose k such that this probability is no less than, say, 0.99.

Independent Random Variables

Independence for random variables is defined[2] in an analogous fashion to independence for events:

Definition 15.3 (independent r.v.'s): Random variables X and Y on the same probability space are said to be *independent* if the events $X = a$ and $Y = b$ are independent for all values a, b. Equivalently, the joint distribution of independent r.v.'s decomposes as

$$\Pr[X = a, Y = b] = \Pr[X = a] \Pr[Y = b] \quad \forall a, b.$$

To use PMF language, independence means $P_{X,Y}(i, j) = P_X(i) P_Y(j)$.

Mutual independence of more than two r.v.'s is defined similarly. The great thing about mutual independence of random variables as compared to events is that you don't have to check all subsets of random variables separately. If you show that the joint PMF can be written as a product of the individual PMFs, you're done. (Exercise: see if you can prove this for yourself.)

A very important example of independent r.v.'s is indicator r.v.'s for independent events. Thus, for example, if $\{X_i\}$ are indicator r.v.'s for the ith toss of a coin being Heads, then the X_i are mutually independent r.v.'s.

[2]Strictly speaking, independence of two random variables is defined as all the events that are expressible only in terms of one random variable are independent of all the events expressible only in terms of the other random variable. This captures the intuitive sense that these two random quantities must vary indepently of each other in that knowing anything about one tells you nothing about the other.

Expectation

The distribution of a r.v. (or collection or r.v.s) contains *all* the probabilistic information about the r.v. (or collection of r.v.s) In most applications, however, the complete distribution of a r.v. is very hard to calculate. For example, consider the homework example with 20 students. In principle, we'd have to enumerate $20! \approx 2.4 \times 10^{18}$ sample points, compute the value of X at each one, and count the number of points at which X takes on each of its possible values! (though in practice we could streamline this calculation a bit). Moreover, even when we can compute the complete distribution of a r.v., it is often not very helpful from an intuitive perspective.

For these reasons, we seek to *compress* the distribution into a more compact, convenient form that is also easier to compute. The most widely used such form is the *expectation* (or *mean* or *average*) of the r.v.

Definition 15.4 (expectation): The expectation of a discrete random variable X is defined as

$$\mathrm{E}(X) = \sum_{a \in \mathscr{A}} a \times \Pr[X = a],$$

where the sum is over all possible values taken by the r.v.

For our homework example, the expectation is

$$\mathrm{E}(X) = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{2}\right) + \left(3 \times \frac{1}{6}\right) = 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

i.e., the expected number of fixed points in a permutation of three items is exactly 1.

The expectation can be seen in some sense as a "typical" value of the r.v. (though note that it may not actually be a value that the r.v. ever takes on). The question of how typical the expectation is for a given r.v. is a very important one that we shall return to in a later lecture.

Here is a physical interpretation of the expectation of a random variable: imagine carving out a wooden cutout figure of the probability distribution as in the figure. Then the expected value of the distribution is the balance point (directly below the center of gravity) of this object.

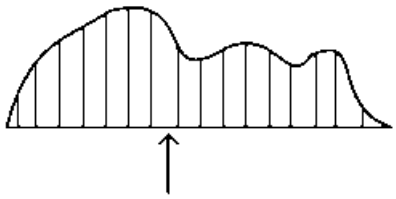

Figure 4: Physical interpretation of expected value as the balance point.

Examples

1. **Single die.** Throw one fair die. Let X be the number that comes up. Then X takes on values $1, 2, \ldots, 6$ each with probability $\frac{1}{6}$, so

   $$\mathrm{E}(X) = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}.$$

   Note that X never actually takes on its expected value $\frac{7}{2}$.

2. **Two dice.** Throw two fair dice. Let X be the sum of their scores. Then the distribution of X is

a	2	3	4	5	6	7	8	9	10	11	12
$\Pr[X=a]$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

   The expectation is therefore

$$\mathrm{E}(X) = \left(2 \times \frac{1}{36}\right) + \left(3 \times \frac{1}{18}\right) + \left(4 \times \frac{1}{12}\right) + \cdots + \left(12 \times \frac{1}{36}\right) = 7.$$

3. **Roulette.** A roulette wheel is spun. You bet \$1 on Black. If a black number comes up, you receive your stake plus \$1; otherwise you lose your stake. Let X be your net winnings in one game. Then X can take on the values $+1$ and -1, and $\Pr[X=1] = \frac{18}{38}$, $\Pr[X=-1] = \frac{20}{38}$. [Recall that a roulette wheel has 38 slots: the numbers $1, 2, \ldots, 36$, half of which are red and half black, plus 0 and 00, which are green.] Thus

$$\mathrm{E}(X) = \left(1 \times \frac{18}{38}\right) + \left(-1 \times \frac{20}{38}\right) = -\frac{1}{19};$$

   i.e., you expect to lose about a nickel per game. Notice how the zeros tip the balance in favor of the casino!

Linearity of expectation

So far, we've computed expectations by brute force: i.e., we have written down the whole distribution and then added up the contributions for all possible values of the r.v. The real power of expectations is that in many real-life examples they can be computed much more easily using a simple shortcut. The shortcut is the following:

Theorem 15.1: For any two random variables X and Y on the same probability space, we have

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y).$$

Also, for any constant c, we have

$$\mathrm{E}(cX) = c\mathrm{E}(X).$$

Proof: To make the proof easier, we will first rewrite the definition of expectation in a more convenient form. Recall from Definition 15.4 that

$$\mathrm{E}(X) = \sum_{a \in \mathscr{A}} a \times \Pr[X=a].$$

Let's look at one term $a \times \Pr[X=a]$ in the above sum. Notice that $\Pr[X=a]$, by definition, is the sum of $\Pr[\omega]$ over those sample points ω for which $X(\omega) = a$. And we know that every sample point $\omega \in \Omega$ is in exactly one of these events $X=a$. This means we can write out the above definition in a more long-winded form as

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega) \times \Pr[\omega]. \tag{2}$$

This equivalent definition of expectation will make the present proof much easier (though it is usually less convenient for actual calculations).

Now let's write out $\mathrm{E}(X+Y)$ using equation (2):

$$\begin{aligned}\mathrm{E}(X+Y) &= \textstyle\sum_{\omega\in\Omega}(X+Y)(\omega)\times\Pr[\omega]\\ &= \textstyle\sum_{\omega\in\Omega}(X(\omega)+Y(\omega))\times\Pr[\omega]\\ &= \textstyle\sum_{\omega\in\Omega}\Big(X(\omega)\times\Pr[\omega]\Big)+\sum_{\omega\in\Omega}\Big(Y(\omega)\times\Pr[\omega]\Big)\\ &= \mathrm{E}(X)+\mathrm{E}(Y).\end{aligned}$$

In the last step, we used equation (2) twice.

This completes the proof of the first equality. The proof of the second equality is much simpler and is left as an exercise. □

Theorem 15.1 is very powerful: it says that the expectation of a sum of r.v.'s is the sum of their expectations, with no assumptions about the r.v.'s. We can use Theorem 15.1 to conclude things like $\mathrm{E}(3X-5Y) = 3\mathrm{E}(X)-5\mathrm{E}(Y)$. This property is known as linearity of expectation. *Important caveat: Theorem 15.1 does* **not** *say that* $\mathrm{E}(XY)=\mathrm{E}(X)\mathrm{E}(Y)$*, or that* $\mathrm{E}(\frac{1}{X})=\frac{1}{\mathrm{E}(X)}$ *etc. These claims are not true in general. It is only* sums *and* differences *and* constant multiples *of random variables that behave so nicely. We talk about products at the end of this note.*

Examples

Now let's see some examples of Theorem 15.1 in action.

4. **Two dice again.** Here's a much less painful way of computing $\mathrm{E}(X)$, where X is the sum of the scores of the two dice. Note that $X = X_1+X_2$, where X_i is the score on die i. We know from example 1 above that $\mathrm{E}(X_1)=\mathrm{E}(X_2)=\frac{7}{2}$. So by Theorem 15.1 we have $\mathrm{E}(X)=\mathrm{E}(X_1)+\mathrm{E}(X_2)=7$.

5. **More roulette.** Suppose we play the above roulette game not once, but 1000 times. Let X be our expected net winnings. Then $X = X_1+X_2+\ldots+X_{1000}$, where X_i is our net winnings in the ith play. We know from earlier that $\mathrm{E}(X_i)=-\frac{1}{19}$ for each i. Therefore, by Theorem 15.1, $\mathrm{E}(X) = \mathrm{E}(X_1)+\mathrm{E}(X_2)+\cdots+\mathrm{E}(X_{1000}) = 1000\times(-\frac{1}{19}) = -\frac{1000}{19}\approx -53$. So if you play 1000 games, you expect to lose about \$53.

6. **Homeworks.** Let's go back and answer our original question about the class of 20 students. Recall that the r.v. X is the number of students who receive their own homework after shuffling (or equivalently, the number of fixed points). To take advantage of Theorem 15.1, we need to write X as the *sum* of simpler r.v.'s. But since X *counts* the number of times something happens, we can write it as a sum using the following trick:

$$X = X_1+X_2+\ldots+X_{20}, \qquad \text{where } X_i = \begin{cases}1 & \text{if student } i \text{ gets her own hw;}\\ 0 & \text{otherwise.}\end{cases} \tag{3}$$

[You should think about this equation for a moment. Remember that all the X's are random variables. What does an equation involving random variables mean? What we mean is that, *at every sample point* ω, we have $X(\omega)=X_1(\omega)+X_2(\omega)+\ldots+X_{20}(\omega)$. Why is this true?]

A 0/1-valued random variable such as X_i is called an indicator random variable of the corresponding event (in this case, the event that student i gets her own hw). For indicator r.v.'s, the expectation is particularly easy to calculate. Namely,

$$\mathrm{E}(X_i) = (0\times\Pr[X_i=0])+(1\times\Pr[X_i=1]) = \Pr[X_i=1].$$

But in our case, we have

$$\Pr[X_i = 1] = \Pr[\text{student } i \text{ gets her own hw}] = \frac{1}{20}.$$

Now we can apply Theorem 15.1 to (3), to get

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_{20}) = 20 \times \frac{1}{20} = 1.$$

So we see that the expected number of students who get their own homeworks in a class of size 20 is 1. But this is exactly the same answer as we got for a class of size 3! And indeed, we can easily see from the above calculation that we would get $\mathrm{E}(X) = 1$ for *any* class size n: this is because we can write $X = X_1 + X_2 + \ldots + X_n$, and $\mathrm{E}(X_i) = \frac{1}{n}$ for each i.

So the expected number of fixed points in a random permutation of n items is always 1, regardless of n. Amazing, but true.

7. **Coin tosses.** Toss a fair coin 100 times. Let the r.v. X be the number of Heads. As in the previous example, to take advantage of Theorem 15.1 we write

$$X = X_1 + X_2 + \ldots + X_{100},$$

where X_i is the indicator r.v. of the event that the ith toss is Heads. Since the coin is fair, we have

$$\mathrm{E}(X_i) = \Pr[X_i = 1] = \Pr[i\text{th toss is Heads}] = \tfrac{1}{2}.$$

Using Theorem 15.1, we therefore get

$$\mathrm{E}(X) = \sum_{i=1}^{100} \frac{1}{2} = 100 \times \tfrac{1}{2} = 50.$$

More generally, the expected number of Heads in n tosses of a fair coin is $\frac{n}{2}$. And in n tosses of a biased coin with Heads probability p, it is np (why?). So the expectation of a r.v. $X \sim Bin(n,p)$ is np. Note that it would have been harder to reach the same conclusion by computing this directly from definition of expectation.

8. **Balls and bins.** Throw m balls into n bins. Let the r.v. X be the number of balls that land in the first bin. Then X behaves exactly like the number of Heads in n tosses of a biased coin, with Heads probability $\frac{1}{n}$ (why?). So from example 7 we get $\mathrm{E}(X) = \frac{m}{n}$.

   In the special case $m = n$, the expected number of balls in any bin is 1. If we wanted to compute this directly from the distribution of X, we'd get into a messy calculation involving binomial coefficients.

   Here's another example on the same sample space. Let the r.v. Y be the number of empty bins. The distribution of Y is horrible to contemplate: to get a feel for this, you might like to write it down for $m = n = 3$ (3 balls, 3 bins). However, computing the expectation $\mathrm{E}(Y)$ is a piece of cake using Theorem 15.1. As usual, let's write

$$Y = Y_1 + Y_2 + \cdots + Y_n, \tag{4}$$

   where Y_i is the indicator r.v. of the event "bin i is empty". Again as usual, the expectation of Y_i is easy:

$$\mathrm{E}(Y_i) = \Pr[Y_i = 1] = \Pr[\text{bin } i \text{ is empty}] = \left(1 - \frac{1}{n}\right)^m;$$

recall that we computed this probability (quite easily) in an earlier lecture. Applying Theorem 15.1 to (4) we therefore have

$$\mathrm{E}(Y) = \sum_{i=1}^{n} \mathrm{E}(Y_i) = n\left(1 - \frac{1}{n}\right)^m,$$

a very simple formula, very easily derived.

Let's see how it behaves in the special case $m = n$ (same number of balls as bins). In this case we get $\mathrm{E}(Y) = n(1-\frac{1}{n})^n$. Now the quantity $(1-\frac{1}{n})^n$ can be approximated (for large enough values of n) by the number $\frac{1}{\mathrm{e}}$.[3] So we see that

$$\mathrm{E}(Y) \to \frac{n}{\mathrm{e}} \approx 0.368n \quad \text{as } n \to \infty.$$

The bottom line is that, if we throw (say) 1000 balls into 1000 bins, the expected number of empty bins is about 368.

Expectations and Independence

Linearity of expectations does not require independence. Where independence helps is when we are working with products of random variables.

Theorem 15.2: For *independent* random variables X, Y, we have $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

Proof: We have

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_a \sum_b ab \times \Pr[X = a, Y = b] \\
&= \sum_a \sum_b ab \times \Pr[X = a] \times \Pr[Y = b] \\
&= \left(\sum_a a \times \Pr[X = a]\right) \times \left(\sum_b b \times \Pr[Y = b]\right) \\
&= \mathrm{E}(X) \times \mathrm{E}(Y),
\end{aligned}$$

as required. In the second line here we made crucial use of independence. □

This property is a consequence of independence, but it is not sufficient to prove independence. For example, consider X to be a fair coin toss that we consider as taking values $+1$ and -1 equally likely. Suppose Y is an independent fair coin toss that takes values $+1$ and $+2$. The random variables X and Y are independent by construction. Let's consider a new random variable $Z = XY$. Is Z independent of Y ? Obviously not. Z takes on four possible values $-2, -1, +1, +2$ and the magnitude of Z reveals exactly what Y is. But let's consider $\mathrm{E}(YZ)$. You can calculate for yourself that this is equal to zero which is equal to $\mathrm{E}(Y)\mathrm{E}(Z)$. So be careful. Meanwhile, $\mathrm{E}(XZ)$ is definitely not equal to $\mathrm{E}(X)\mathrm{E}(Z)$ since $XZ = Y$ is always positive by construction while the expectations of both X and Z are zero individually.

This property of products and independence is what lets us prove the laws of large numbers that we saw empirically at the beginning of our discussion of probability.

[3] More generally, it is a standard fact that for any constant c,

$$(1 + \tfrac{c}{n})^n \to \mathrm{e}^c \quad \text{as } n \to \infty.$$

We just used this fact in the special case $c = -1$. The approximation is actually very good even for quite small values of n — try it yourself! E.g., for $n = 20$ we already get $(1-\frac{1}{n})^n$=0.358, which is very close to $\frac{1}{\mathrm{e}} = 0.367\ldots$. The approximation gets better and better for larger n.