LINEARITY OF EXPECTATION

SIMON RUBINSTEIN-SALZEDO

1. A famous problem

One of the most famous problems in all of mathematics is the Buffon needle problem. It goes like this:

Question 1. *Suppose that the floor of a room is made up of parallel pieces of wood, each of width 1. We have a needle of length 1, and we drop it at random on the floor. What is the probability that the needle will hit one of the lines between two slices of wood?*

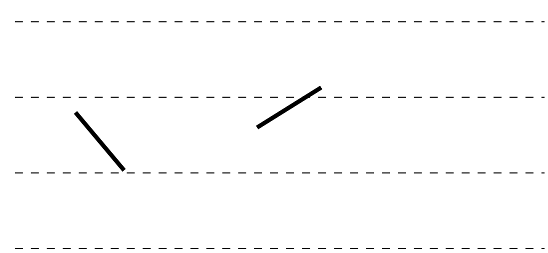


Figure 1. The needle on the left does not cross one of the lines on the floor, whereas the needle on the right does.

Usually, this problem is solved using calculus. However, it is also possible to prove it without calculus, and in fact using seemingly nothing at all except for a bit of creative thinking!

2. Random variables

We won't bother giving a formal definition of random variable, because it won't matter to us. But informally, a random variable is one whose value depends on chance. Think, for instance, of a roll of a die: it can be 1, 2, 3, 4, 5, or 6. Let X be the outcome of a die roll. Then the *event* that $X = 4$ happens with probability 1/6. We write $\mathbb{P}(X = 4) = \frac{1}{6}$.

In addition to probabilities, we may be interested in the *expected value*, also known as the *mean*. This is the average value that a random variable takes on. We write $\mathbb{E}(X)$ for the expected value of the random variable X.

Let us compute $\mathbb{E}(X)$ when X is a die roll. To do this, we look at all the possibilities (1,2,3,4,5,6), calculate the probabilities, multiply them, and add them all up: we have

$$\mathbb{E}(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5.$$

So, the expected value of a die roll is 3.5. Note that this does *not* mean that we expect the die to come up as 3.5; obviously we expect it to be an integer.

Date: April 3, 2019.

Instead, what we mean by the expected value is the expected average result if we roll the die many times. If we roll the die 10,000 times and average the results, we should get something quite close to 3.5. It probably won't be exact, but it will get closer as we roll more and more times and average them.

Now let's work out the case of a roll of *two* dice. There are 11 possible results that we can get: 2,3,4,5,6,7,8,9,10,11,12, and these occur with probability

$$\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}.$$

If we let X be the result of the sum of the two rolls, we have

$$\begin{aligned}\mathbb{E}(X) = \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 \\ + \frac{6}{36} \times 7 + \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 = 7.\end{aligned}$$

Observe that this is double the result of a single die roll. Coincidence? I think not!

3. Linearity of expectation

The most important property of expected value is called *linearity.*

Theorem 2. *Let X and Y be random variables, and let $c \in \mathbb{R}$. Then*

- $\mathbb{E}(cX) = c\mathbb{E}(X)$,
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

These look obvious, and indeed they aren't very hard to prove. But their consequences can be magical. The point is that linearity of expectation, in particular the second part $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y)$, is obvious if X and Y have nothing to do with one another; that is, if they are *independent.* But it's still true even if X and Y are closely related. And that's where the magic comes in.

Linearity of expectation makes many expected value calculations way easier. For example:

Question 3. *Suppose we flip a coin 10 times. What is the expected number of heads we get?*

First, the naïve solution: we can get heads 0,1,2,3,4,5,6,7,8,9,10 times, and the probability we get heads n times is $\frac{1}{2^{10}}\binom{10}{n}$. So, we (painfully) compute

$$\begin{aligned}\frac{1}{2^{10}} \times \left[\binom{10}{0} \times 0 + \binom{10}{1} \times 1 + \binom{10}{2} \times 2 + \binom{10}{3} \times 3 + \binom{10}{4} \times 4 + \binom{10}{5} \times 5 \right.\\ \left. + \binom{10}{6} \times 6 + \binom{10}{7} \times 7 + \binom{10}{8} \times 8 + \binom{10}{9} \times 9 + \binom{10}{10} \times 10\right] = 5.\end{aligned}$$

Now the smarter solution: define a random variable X_i for each flip, by setting

$$X_i = \begin{cases} 0 & \text{if the } i^{\text{th}} \text{ flip is tails,} \\ 1 & \text{if the } i^{\text{th}} \text{ flip is heads.} \end{cases}$$

Then $X = X_1 + X_2 + \cdots + X_{10}$ counts the total number of heads. We are interested in $\mathbb{E}(X)$, and by linearity of expectation, this is

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_{10}).$$

For each i, we have $\mathbb{E}(X_i) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$. Thus $\mathbb{E}(X) = 10 \times \frac{1}{2} = 5$. Wasn't that better?

In general, the idea behind linearity of expectation is to turn one hard problem into a whole bunch of easy ones. Working out the expected value of the result of a whole bunch of coin flips seems like a hard thing to do, whereas working out the expected value of a single coin flip is very easy.

4. Permutation statistics

Let's do some more tricks!

Definition 4. A *permutation* of $\{1, 2, \ldots, n\}$ is an arrangement or ordering of the numbers $1, 2, \ldots, n$.

For example, 521643 is a permutation of $\{1, 2, \ldots, n\}$.

Question 5. *How many permutations are there of* $\{1, 2, \ldots, n\}$*?*

It is sometimes convenient to think of a permutation as a function $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, where $f(1)$ is the first number, $f(2)$ is the second number, and so forth. For the permutation 521643, we have

$$f(1) = 5, \qquad f(2) = 2, \qquad f(3) = 1, \qquad f(4) = 6, \qquad f(5) = 4, \qquad f(6) = 3.$$

Observe that when we write permutations as functions in this way, then if $i \neq j$, then $f(i) \neq f(j)$.

Note that sometimes (for example, in the case of 2 above), we have $f(i) = i$.

Definition 6. Let f be a permutation of $\{1, 2, \ldots, n\}$. A *fixed point* of f is a number i so that $f(i) = i$.

What is the expected number of fixed points of a permutation? That is, choose a permutation at random, and count the number of fixed points. How many do we get, on average?

We could try to count the number of permutations with $0, 1, 2, \ldots, n$ fixed points. For specific values of n, we can do that, although we end up getting a horribly ugly sum involving lots of binomial coefficients. Fortunately, there's a better way, using our shiny new trick.

Let X be the number of fixed points of a random permutation f. We can break X up by looking at which numbers are fixed. As in the case of the coin flips, we write $X = X_1 + X_2 + \cdots + X_n$, where

$$X_i = \begin{cases} 0 & \text{if } i \text{ is not a fixed point of } f, \\ 1 & \text{if } i \text{ is a fixed point of } f. \end{cases}$$

Now, how do we calculate $\mathbb{E}(X_i)$? Well, $f(i)$ could be anything, and all values are equally likely. So $\mathbb{P}(f(i) = i)$ is $\frac{1}{n}$, and $\mathbb{P}(f(i) \neq i)$ is $\frac{n-1}{n}$. Hence $\mathbb{E}(X_i) = \frac{n-1}{n} \times 0 + \frac{1}{n} \times 1 = \frac{1}{n}$.

Now, we put it all together:

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = n \times \frac{1}{n} = 1.$$

On average, there's exactly one fixed point!

Note that the X_i's are *not* independent of one another: if you know that i is a fixed point, that makes it a bit more likely that any $j \neq i$ is also a fixed point, whereas if you know that

i is *not* a fixed point, then it is a bit *less* likely that any $j \neq i$ is a fixed point: it could happen that $f(i) = j$, in which case j certainly cannot be a fixed point!

There's a lot more we can do with linearity of expectation and permutation statistics. For a permutation f of $\{1, 2, \ldots, n\}$ call a pair (i, j) with $1 \leq i < j \leq n$ a *transposition* if $f(i) = j$ and $f(j) = i$.

Question 7. *What is the expected number of transpositions in a random permutation?*

And there's no reason to stop there.

Question 8. *On average, how many triples (i, j, k) are there with $f(i) = j$, $f(j) = k$, and $f(k) = i$? (You have to be a little careful with the ordering now; (i, j, k) and (i, k, j) are different.) What about more general k-cycles?*

5. Dirty socks

It is well-known that washing machines eat socks. However, it is still sometimes necessary to put your socks in the washing machine.

Question 9. *Suppose you have 10 pairs of socks, with each pair being a different color. (So, maybe you have a red pair, and a yellow pair, and so forth.) You put them all in the washing machine. The washing machine eats four socks at random. What is the expected value for the number of complete pairs that make it out alive?*

When I taught a probability class at Dartmouth College in 2013, I tried to convince my students that linearity of expectation is how you solve all problems in probability. Evidently, they didn't believe me. I put this problem on an exam, and most tried to bash it out with lots of cases (with most of them making mistakes and getting it wrong, as is very easy to do); I think only one student attempted to use linearity of expectation.

But we will do this problem correctly, with linearity of expectation. Let X be the number of complete pairs that survive the washing machine. Let us number the pairs $1, 2, \ldots, 10$. Let us write $X = X_1 + X_2 + \cdots + X_{10}$, where

$$X_i = \begin{cases} 0 & \text{if the washing machine eats at least one sock in pair } i, \\ 1 & \text{if both socks in pair } i \text{ survive the washing machine.} \end{cases}$$

Now we compute $\mathbb{E}(X_i)$: it is $0 \times \mathbb{P}(X_i = 0) + 1 \times \mathbb{P}(X_i = 1)$, so we just have to compute $\mathbb{P}(X_i = 1)$, which is the probability that both socks in pair i survive.

To compute $\mathbb{P}(X_i = 1)$, look first at the first sock in pair i. Its chance of survival is $\frac{16}{20}$. Now, what about the second sock? If the first sock was eaten, then we don't care about the fate of the second sock; the pair can't survive regardless. But if the first sock survives, then there are 19 socks left, and the washing machine still eats 4 of them, so the probability the second sock survives is $\frac{15}{19}$. Combining these, the probability that both socks survive is $\frac{16}{20} \times \frac{15}{19} = \frac{12}{19}$. This is also $\mathbb{E}(X_i)$.

Now, we put it all together.

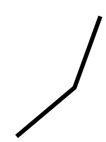
$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_{10}) = 10 \times \frac{12}{19} = \frac{120}{19} \approx 6.32.$$

6. Buffon's needle

Unlike the other problems we have looked at, Buffon's needle doesn't seem to be a good candidate for linearity of expectation. How are we supposed to break it up into little problems? And is it even asking for an expected value?

We can fix the last objection easily enough. Drop the needle at random, and let X be the number of lines the needle crosses. Since the needle is the same length as the width of the planks, this number is always going to be 0 or 1. (We don't worry about the case where the needle exactly touches two of the lines, since the probability of that happening is 0.)

Now, what happens if we take two needles, and call X_1 and X_2 the number of lines they each cross? Then $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 2\mathbb{E}(X)$. What happens if we glue the two needles together? Same thing: the expected number of crossings is still $2\mathbb{E}(X)$. And it doesn't matter if we glue them together in a funny way, like this:

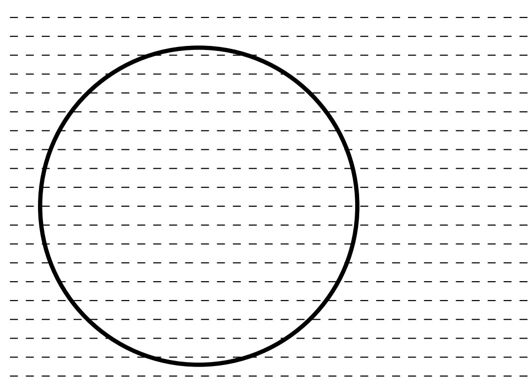


The expected number of intersections is still $2\mathbb{E}(X)$.

Okay, so now let's take LOTS of needles! In fact, let's take N of them, where N is really big. Extending our last observation, we learn that the expected number of crossings depends only on N, not how the needles are arranged. So, is there any shape that makes it really easy to count the number of intersections?

Yes: a circle! Well, we can't quite get a circle, because it's made out of straight lines. But we can get very close to a circle, close enough to pretend.

And how many crossings do we have with a circular needle with circumference N? (Assume that the width of each piece of wood is 1.)



Let $R = \frac{N}{2\pi}$ be the radius of the circle. Then it hits $2R$ lines on the left, and $2R$ lines on the right, or $4R$ total. So the expected number of hits is $4R = \frac{2N}{\pi}$. But recall that this number is just $N \times \mathbb{E}(X)$. So, solving for $\mathbb{E}(X)$, we have

$$\mathbb{E}(X) = \frac{2}{\pi}. \tag{1}$$

Note that this is a fun way to compute π: throw a needle many times and count the number of times it hits (out of the total number of tosses), and use (1) to figure out what π is, based on your simulation. This is an example of a *Monte Carlo method*: run an experiment many times to approximate some quantity, in this case π.

7. Further reading

The probabilistic method, one of Paul Erdős's greatest contributions to mathematics, is a powerful tool for showing the existence of combinatorial objects, such as graphs with certain properties. The idea is to pick a candidate object and show that, with some positive probability, this candidate object has the desired properties. Frequently, this is done using linearity of expectation. We will not discuss the probabilistic method further here, but I encourage you to learn about it on your own. The standard reference is the wonderful book *The Probabilistic Method* by Noga Alon and Joel Spencer. An easier introduction might be Alon and Krivelevich's article "Extremal and Probabilistic Combinatorics."[1]

8. Further problems

(1) $n \geq 3$ babies sit it a circle. Each baby pokes a neighbor, at random. What is the expected number of unpoked babies? (HMMT)

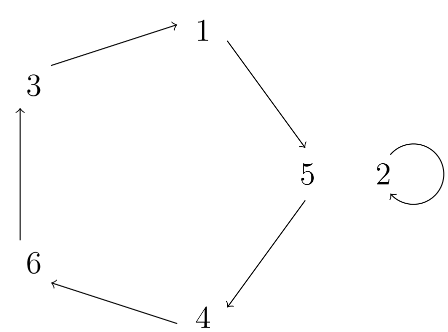
(2) m boys and n girls order themselves randomly in a line. What is the expected number of places where a boy and a girl are next to each other? (AHSME)

(3) n people play a game of laser tag. Each person tags one of the other people, at random. What is the expected number of untagged people?

(4) Let f be a permutation of $\{1, 2, \ldots, n\}$. We can draw a picture of f as follows, best illustrated by an example. Suppose $n = 6$, and

$$f(1) = 5, \qquad f(2) = 2, \qquad f(3) = 1, \qquad f(4) = 6, \qquad f(5) = 4, \qquad f(6) = 3.$$

Then we can draw an arrow from i to $f(i)$, for each i:



Each loop in this figure is called a *cycle*. So, here $1 \to 5 \to 4 \to 6 \to 3 \to 1$ is a cycle, and $2 \to 2$ is a cycle, for 2 cycles total. What is the expected number of cycles in a permutation of $\{1, 2, \ldots, n\}$?

(5) (Coupon collector problem.) You want to collect all 50 state quarters. So, you go to a bank and ask for a state quarter, at random. Each state is equally likely. How many trips do you have to make on average before you have at least one quarter from each state?

(6) In the Buffon needle problem, what is the expected number of crossings if the floor is made up of wooden squares? (You should now count both horizontal *and* vertical crossings.) What if the tiles are made of regular hexagons?

[1] `http://www.math.tau.ac.il/~nogaa/PDFS/epc7.pdf`

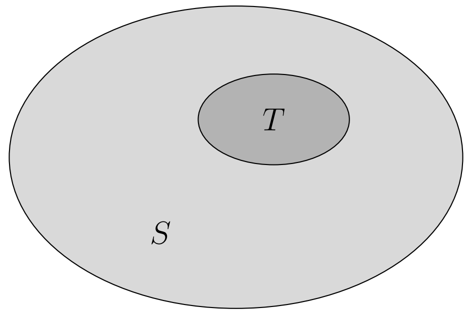


Figure 2. Two convex sets, one inside the other.

(7) (Hat check problem.) There are n people who go to a party, and each one of them wears a hat. Each person leaves eir[2] hat with the concierge when ey arrives at the party. At the end, the concierge gives each person a hat at random. What is the probability that everyone gets the wrong hat?

(8) (Erdős.) Let X be a set of $n \geq 1$ nonzero real numbers. Show that there is a subset $Y \subset X$ with $|Y| \geq |X|/3$ that is sum-free. (A set S is sum-free if there do not exist $x, y, z \in S$ with $x + y = z$.)

(9) (Sylvester's problem.) A set $S \subset \mathbb{R}^2$ is said to be *convex* if, for any two points $x, y \in S$, the segment connecting x to y is contained in S. A set $S \subset \mathbb{R}^2$ is said to be *bounded* if there is some number M so that the distance between any two points in S is less than M. (That is, S does not go off to infinity.) Now, suppose that S and T are two bounded convex regions in $\mathbb{R}^2$, with $T \subset S$. What is the probability that a random line that intersects S also intersects T? (See Figure 2.)

Euler Circle, Palo Alto, CA 94306
E-mail address: `simon@eulercircle.com`

[2]This is a Spivak pronoun, which is a type of third-person singular gender-neutral pronoun. They are formed (essentially) by chopping off the "th" from the front of the third-person plural pronouns "they," "them," "their," "theirs," etc. I am on a mission to make them into a standard part of the English language. Please do your part to help!