5.2 Linearity of expectation

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

The most important property of expectation is *linearity*: the expected value of a sum of r.v.s is the sum of the individual expected values.

> **Theorem 5.2.1 (Linearity of expectation).**
>
> For any r.v.s X, Y and any constant c,
>
> $$E(X+Y) = E(X) + E(Y),$$
> $$E(cX) = cE(X).$$

We will now show that linearity is true for discrete r.v.s X and Y. Before doing that, let's recall some basic facts about averages. If we have a list of numbers, say $(1,1,1,1,1,3,3,5)$, we can calculate their mean by adding all the values and dividing by the length of the list, so that each element of the list gets a weight of $\frac{1}{8}$:

$$\frac{1}{8}(1+1+1+1+1+3+3+5) = 2.$$

But another way to calculate the mean is to group together all the 1's, all the 3's, and all the 5's, and then take a weighted average, giving appropriate weights to 1's, 3's, and 5's:

$$\frac{5}{8}\cdot 1 + \frac{2}{8}\cdot 3 + \frac{1}{8}\cdot 5 = 2.$$

This insight - that averages can be calculated in two ways, *ungrouped* or *grouped*---is all that is needed to prove linearity (in the discrete case)! Recall that X is a function which assigns a real number to every outcome s in the sample space. The r.v. X may assign the same value to multiple sample outcomes. When this happens, our definition of expectation groups all these outcomes together into a *super-pebble* whose weight, $P(X = x)$, is the total weight of the constituent pebbles. This grouping process is illustrated in Figure 5.2.2 for a hypothetical r.v. taking values in $\{0, 1, 2\}$. So our definition of expectation corresponds to the grouped way of taking averages.

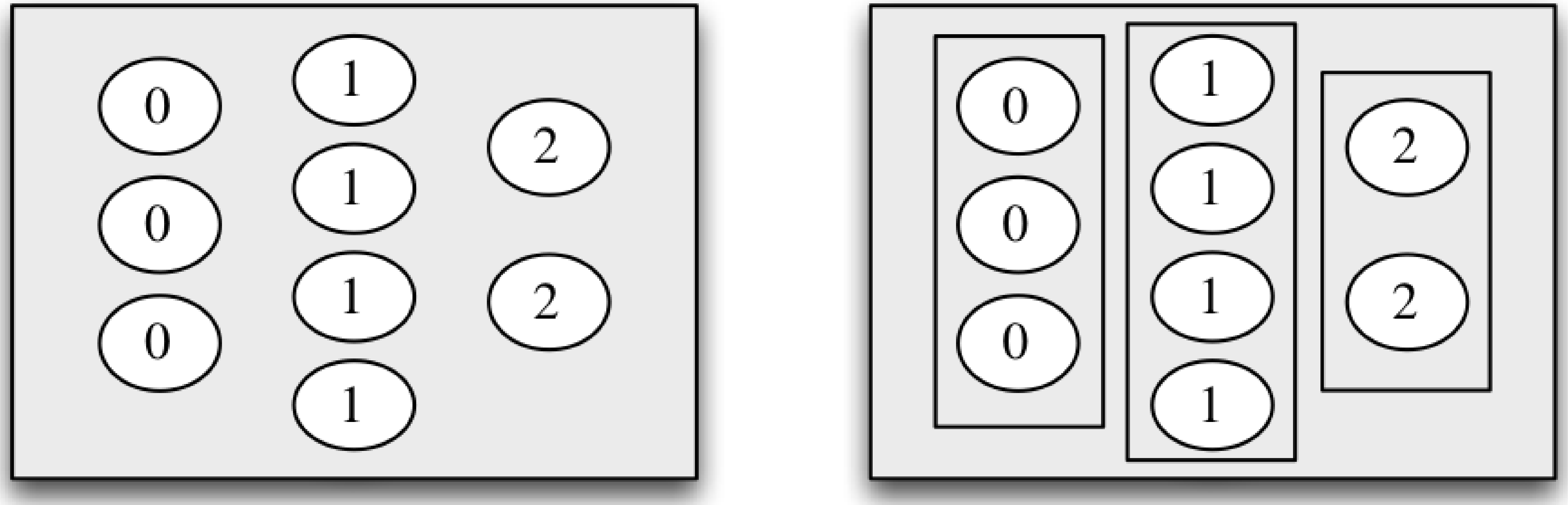

Figure 5.2.2: Left: X assigns a number to each pebble in the sample space. Right: Grouping the pebbles by the value that X assigns to them, the 9 pebbles become 3 super-pebbles. The weight of a super-pebble is the sum of the weights of the constituent pebbles.

View Larger Image
Image Description

The advantage of this definition is that it allows us to work with the distribution of X directly, without returning to the sample space. The disadvantage comes when we have to prove theorems like this one, for if we have another r.v. Y on the same sample space, the super-pebbles created by Y are different from those created from X, with different weights $P(Y = y)$; this makes it difficult to combine $\sum_x xP(X = x)$ and $\sum_y yP(Y = y)$.

Fortunately, we know there's another equally valid way to calculate an average: we can take a weighted average of the values of individual pebbles. In other words, if $X(s)$ is the value that X assigns to pebble s, we can take the weighted average

$$E(X) = \sum_s X(s)P(\{s\}),$$

where $P(\{s\})$ is the weight of pebble s.

This corresponds to the ungrouped way of taking averages. The advantage of this definition is that it breaks down the sample space into the smallest possible units, so we are now using the *same* weights $P(\{s\})$ for every random variable defined on this sample space. If Y is another random variable, then

$$E(Y) = \sum_s Y(s)P(\{s\}).$$

Now we *can* combine $\sum_s X(s)P(\{s\})$ and $\sum_s Y(s)P(\{s\})$, which gives $E(X+Y)$ as desired:

$$E(X)+E(Y)=\sum_s X(s)P(\{s\})+\sum_s Y(s)P(\{s\})=\sum_s (X+Y)(s)P(\{s\})=E(X+Y).$$

Linearity is an extremely handy tool for calculating expected values, often allowing us to bypass the definition of expected value altogether. Let's use linearity to find the expectations of the Binomial and Hypergeometric distributions.

Example 5.2.3 (Binomial expectation).

For $X \sim \text{Bin}(n,p)$, let's find $E(X)$. By definition of expectation,

$$E(X)=\sum_{k=0}^{n} kP(X=k)=\sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}.$$

This sum can be done with some work, but linearity of expectation provides a *much* shorter path to the same result. Let's write X as the sum of n independent $\text{Bern}(p)$ r.v.s:

$$X = I_1 + \cdots + I_n,$$

where each I_j has expectation $E(I_j)=1p+0q=p$. By linearity,

$$E(X)=E(I_1)+\cdots+E(I_n)=np.$$

Example 5.2.4 (Hypergeometric expectation).

Let $X \sim \text{HGeom}(w,b,n)$, interpreted as the number of white balls in a sample of size n drawn without replacement from an urn with w white and b black balls. As in the Binomial case, we can write X as a sum of Bernoulli random variables,

$$X = I_1 + \cdots + I_n,$$

where I_j equals 1 if the jth ball in the sample is white and 0 otherwise. By symmetry, $I_j \sim \text{Bern}(p)$ with $p = w/(w+b)$, since unconditionally the jth ball drawn is equally likely to be any of the balls.

Unlike in the Binomial case, the I_j are not independent, since the sampling is without replacement: given that a ball in the sample is white, there is a lower chance that another ball in the sample is white. However, linearity still holds for dependent r.v.s! Thus,

$$E(X)=nw/(w+b).$$

Learn About Verified Certificates