Chapter 3. Discrete Random Variables

3.2: More on Expectation

Slides (Google Drive) Alex Tsun Video (YouTube)

3.2.1 Linearity of Expectation

Right now, the only way you've learned to compute expectation is by first computing the PMF of a random variable $p_X(k)$ and using the formula $\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k)$ which is just a weighted sum of the possible values of X. If you had two random variables X and Y, then to compute the expectation of their sum $Z = X + Y$, you could compute the PMF of Z and apply the same formula. But actually, if you knew both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$, you might be inclined to just say $\mathbb{E}[Z] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and we'll see that this is true! Linearity of expectation is one of the most fundamental and important concepts in probability theory, that you will use almost everywhere! We'll explain it in a simple example, prove it, and then use it to tackle hard problems.

Let's say that you and your friend sell fish for a living. Every day, you catch X fish, with $\mathbb{E}[X] = 3$ and your friend catches Y fish, with $\mathbb{E}[Y] = 7$. How many fish do the two of you bring in ($Z = X + Y$) on an average day? You might guess $3 + 7 = 10$. This is the formula you just guessed:

$$\mathbb{E}[Z] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 3 + 7 = 10$$

This property turns out to be true! Furthermore, let's say that you can sell each fish for \$5 at a store, but you need to pay \$20 in rent for the storefront. How much profit do you expect to make? The profit formula would be $5Z - 20$: \$5 times the number of total fish, minus \$20. You might guess $5 \cdot 10 - 20 = 30$ and you would be right once again! This is the formula you just guessed:

$$\mathbb{E}[5Z - 20] = 5\mathbb{E}[Z] - 20 = 5 \cdot 10 - 20 = 30$$

Theorem 3.2.1: Linearity of Expectation (LoE)

Let Ω be the sample space of an experiment, $X, Y : \Omega \to \mathbb{R}$ be (possibly "dependent") random variables both defined on Ω, and $a, b, c \in \mathbb{R}$ be scalars. Then,

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

and

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

Combining them gives,

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Proof of Linearity of Expectation. Note that X and Y are functions (since random variables are functions), so $X + Y$ is function that is the sum of the outputs of each of the functions. We have the following (in the first equation, $(X+Y)(\omega)$ is the function $(X+Y)$ applied to ω which is equal to $X(\omega)+Y(\omega)$, it is not a product):

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{\omega\in\Omega}(X+Y)(\omega)\cdot\mathbb{P}(\omega) && \text{[def of expectation for the rv } X+Y\text{]}\\
&= \sum_{\omega\in\Omega}(X(\omega)+Y(\omega))\cdot\mathbb{P}(\omega) && \text{[def of sum of functions]}\\
&= \sum_{\omega\in\Omega}X(\omega)\cdot\mathbb{P}(\omega)+\sum_{\omega\in\Omega}Y(\omega)\cdot\mathbb{P}(\omega) && \text{[property of summation]}\\
&= \mathbb{E}[X]+\mathbb{E}[Y] && \text{[def of expectation of } X \text{ and } Y\text{]}
\end{aligned}$$

For the second property, note that $aX+b$ is also a random variable and hence a function (e.g., if $f(x) = \sin(1/x)$, then $(2f-5)(x) = 2f(x)-5 = 2\sin(1/x)-5$.)

$$\begin{aligned}
\mathbb{E}[aX+b] &= \sum_{\omega\in\Omega}(aX+b)(\omega)\cdot\mathbb{P}(\omega) && \text{[def of expectation]}\\
&= \sum_{\omega\in\Omega}(aX(\omega)+b)\cdot\mathbb{P}(\omega) && \text{[def of the function } aX+b\text{]}\\
&= \sum_{\omega\in\Omega}aX(\omega)\cdot\mathbb{P}(\omega)+\sum_{\omega\in\Omega}b\cdot\mathbb{P}(\omega) && \text{[property of summation]}\\
&= a\sum_{\omega\in\Omega}X(\omega)\cdot\mathbb{P}(\omega)+b\sum_{\omega\in\Omega}\mathbb{P}(\omega) && \text{[property of summation]}\\
&= a\mathbb{E}[X]+b && \text{[def of } \mathbb{E}[X] \text{ and } \textstyle\sum_{\omega}\mathbb{P}(\omega)=1\text{]}
\end{aligned}$$

For the last property, we get to assume the first two that we proved already:

$$\begin{aligned}
\mathbb{E}[aX+bY+c] &= \mathbb{E}[aX]+\mathbb{E}[bY]+\mathbb{E}[c] && \text{[property 1]}\\
&= a\mathbb{E}[X]+b\mathbb{E}[Y]+c && \text{[property 2]}
\end{aligned}$$

□

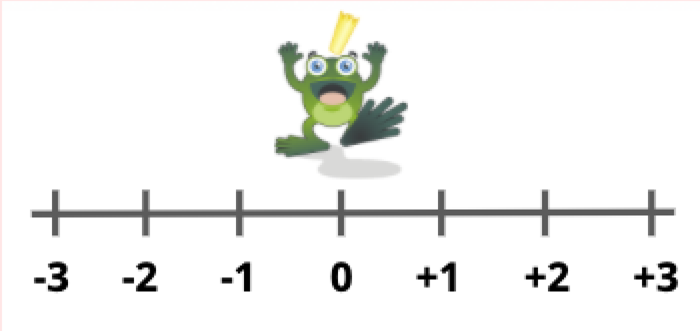
Again, you may think a result like this is "trivial" or "obvious", but we'll see the true power of linearity of expectation through examples. It is one of the most important ideas that you will continue to use (and probably take for granted), even when studying some of the most complex topics in probability theory.

Example(s)

A frog starts on a 1-dimensional number line at 0. At each time step, it moves

- left with probability p_L
- right with probability p_R
- stays with probability p_s

where $p_L+p_R+p_s=1$. Let X be the position of the frog after 2 (independent) time steps. What is $\mathbb{E}[X]$?



Brute Force Solution: When dealing with any random variable, the first thing you should do is identify its range. The frog must end up in one of these positions, since it can move at most 1 to the left and 1 to the right at each step:

$$\Omega_X = \{-2, -1, 0, +1, +2\}$$

So we need to compute 5 values: the probability of each of these. Let's start with the easier ones. The only way to end up at -2 is if the frog moves left at both steps, which happens with probability $p_L \cdot p_L = p_L^2 = \mathbb{P}(X = -2) = p_X(-2)$. The only reason we can multiply them is because of our independence assumption. Similarly, $p_X(2) = p_R \cdot p_R = p_R^2$.
To get to -1, there are two possibilities: first going left and staying ($p_L \cdot p_S$), or first staying and then going left ($p_S \cdot p_L$). Adding these disjoint cases gives $p_X(-1) = 2p_L p_S$. Again, we can only multiply due to independence. Similarly, $p_X(1) = 2p_R p_S$.
Finally, to compute $p_X(0)$, we have two options. One is considering all the possibilities (there are three: left right, right left, or stay stay) and adding them up, and you get $2p_L p_R + p_S^2$. Alternatively and equivalently, since you know the probabilities of 4 of the values $(p_X(-2), p_X(2), p_X(-1), p_X(1))$, the last one $p_X(0)$ must be 1 minus the other four since probabilities have to sum to 1! This is a often useful and clever trick - solving for all but one of the probabilities actually gives you the last one!
In summary, we would write the PMF as:

$$p_X(k) = \begin{cases} p_L^2 & k = -2 \quad \text{:Left left} \\ 2p_L p_S & k = -1 \quad \text{: Left and stay, or stay and left} \\ 2p_L p_R + p_S^2 & k = 0 \quad \text{: Right left, or left right, or stay stay} \\ 2p_R p_S & k = 1 \quad \text{: Right and stay, or stay and right} \\ p_R^2 & k = 2 \quad \text{: Right right} \end{cases}$$

Then to solve for the expectation we just multiply the value and probability mass function and take the sum and have the following:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k \in \Omega_X} k \cdot p_X(k) && \text{[def of expectation]} \\ &= (-2) \cdot p_L^2 + (-1) \cdot 2p_L p_S + (0) \cdot (2p_L p_R + p_S^2) + (1) \cdot 2p_R p_S + (2) \cdot p_R^2 && \text{[plug in our values]} \\ &= 2(p_R - p_L) && \text{[lots of messy algebra]} \end{aligned}$$

The last step of algebra is not important - once you get to more advanced mathematics (like this text), getting the second-to-last formula is sufficient. Everything else is algebra which you could do, or use a computer to do, and so we will omit the useless calculations.

This was quite tedious already; what if instead you were to find the expected location after 100 steps? Then, this method would be completely ridiculous: finding $\Omega_X = \{-100, -99, \dots, +99, +100\}$ and their 201 probabilities. Since you know the frog always moves with the same probabilities though, maybe we can do something more clever!

Linearity Solution:
Let X_1, X_2 be the distance the frog travels at time steps 1,2 respectively.
Important Observation: $X = X_1 + X_2$, since your location after 2 time steps is the sum of the displacement of the first time step and the second time step. Therefore, $\Omega_{X_1} = \Omega_{X_2} = \{-1, 0, +1\}$. They have the same simple PMF of:

$$p_{X_i}(k) = \begin{cases} p_L & k = -1 \\ p_S & k = 0 \\ p_R & k = 1 \end{cases}$$

So: $\mathbb{E}[X_i] = -1 \cdot p_L + 0 \cdot p_S + 1 \cdot p_R = p_R - p_L$, for both $i = 1$ and $i = 2$.
By linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 2(p_R - p_L)$$

Which method is easier? Maybe in this case it is debatable, but if we change the time steps from 2 to 100 or 1000, the brute force solution is entirely infeasible, and the linearity solution will basically be the same amount of work! You could say that $X_1, \dots, X_{100}$ is the displacement at each of 100 time steps, and hence by linearity:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{100} X_i\right] = \sum_{i=1}^{100} \mathbb{E}[X_i] = \sum_{i=1}^{100} (p_R - p_L) = 100(p_R - p_L)$$

Hopefully now you can come to appreciate more how powerful LoE truly is! We'll see more examples in the next section as well as at the end of this section.

3.2.2 Law of the Unconscious Statistician (LOTUS)

Recall the fish example at the beginning of this section regarding profit. Since the expected number of fish was $\mathbb{E}[Z] = 10$ and the profit was a function of the number of fish $g(Z) = 5Z - 20$, we were able to use linearity to say $\mathbb{E}[5Z - 20] = 5\mathbb{E}[Z] - 20$. But this formula only holds for nice linear functions (hence the name "**linearity** of expectation"). What if the profit function was instead something weird/non-linear like $h(Z) = Z^2$ or $h(Z) = \log(5^Z)$? It turns out we can't just say $\mathbb{E}[Z^2] = \mathbb{E}[Z]^2$ or $\mathbb{E}[\log(5^Z)] = \log(5^{\mathbb{E}[Z]})$ - this is actually almost never true! Let's see if we can't derive a nice formula for $\mathbb{E}[g(X)]$ for *any* function g, linear or not.

Consider we are flipping 2 coins again. Let X be the number of heads in two independent flips of a fair coin. Recall the range, PMF, and expectation (again, I'm using the dummy letter d to emphasize that p_X is the PMF for X, and the inner variable doesn't matter):

$$\Omega_X = \{0, 1, 2\}$$

,

$$p_X(d) = \begin{cases} \frac{1}{4} & d = 0 \\ \frac{1}{2} & d = 1 \\ \frac{1}{4} & d = 2 \end{cases}$$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

Let g be the cubing function; i.e., $g(t) = t^3$. Let $Y = g(X) = X^3$; what does this mean? It literally means the cubed number of heads! Let's try to compute $\mathbb{E}[Y] = \mathbb{E}[X^3]$, the expected cubed number of heads. We first find its range and PMF. Based on the range of X, we can calculate the range of Y to be:

$$\Omega_Y = \{0, 1, 8\}$$

since if we get 0 heads, the cubed number of heads is $0^3 = 0$; if we get 1 head, the cubed number of heads is $1^3 = 1$; and if we get 2 heads, the cubed number of heads is $2^3 = 8$.

Now to find the PMF of $Y = X^3$. (Again, below I use the notation p_Y to denote the probability mass function of $Y = X^3$; z is a dummy variable which could be any letter.)

$$p_Y(z) = \begin{cases} \frac{1}{4} & z = 0 \\ \frac{1}{2} & z = 1 \\ \frac{1}{4} & z = 8 \end{cases}$$

since there is a 1/4 chance of getting 0 cubed heads (the outcome TT), 1/2 chance of getting 1 cubed heads (the outcomes HT or TH), and a 1/4 chance of getting 8 cubed heads (the outcome HH).

$$\mathbb{E}[X^3] = \mathbb{E}[Y] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 8 \cdot \frac{1}{4} = 2.5$$

Is there an easier way to compute $\mathbb{E}[X^3] = \mathbb{E}[Y]$ without going through the trouble of writing out p_Y? Yes! Since we know X's PMF already, why should we have to find the PMF of $Y = g(X)$?

Note this formula below is the same formula as above, rewritten so you can observe something:

$$\mathbb{E}[X^3] = 0^3 \cdot \frac{1}{4} + 1^3 \cdot \frac{1}{2} + 2^3 \cdot \frac{1}{4} = 2.5$$

In fact:

$$\mathbb{E}[X^3] = \sum_{b \in \Omega_X} b^3 p_X(b)$$

That is, we can apply the function to each value in Ω_X, *and then* take the weighted average! We can generalize such that for any function $g : \Omega_X \to \mathbb{R}$, we have:

$$\mathbb{E}[g(X)] = \sum_{b \in \Omega_X} g(b) p_X(b)$$

Caveat: It is worth noting that $2.5 = \mathbb{E}\left[X^3\right] \neq (\mathbb{E}[X])^3 = 1$. You cannot just say $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$ as we just showed!

Theorem 3.2.2: Law of the Unconscious Statistician (LOTUS)

Let X be a discrete random variable with range Ω_X and $g : D \to \mathbb{R}$ be a function defined at least over $\Omega_X, (\Omega_X \subseteq D)$. Then

$$\mathbb{E}[g(X)] = \sum_{b \in \Omega_X} g(b) p_X(b)$$

Note that in general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$. For example, $\mathbb{E}\left[X^2\right] \neq (\mathbb{E}[X])^2$, and $\mathbb{E}[\log(X)] \neq \log(\mathbb{E}[X])$.

Before we formally prove this, it will help if we have some intuition for each step. As an example, let X have range $\Omega_X = \{-1, 0, 1\}$ and PMF

$$p_X(k) = \begin{cases} \frac{3}{12} & k = -1 \\ \frac{5}{12} & k = 0 \\ \frac{4}{12} & k = 1 \end{cases}$$

Notice that $Y = X^2$ has range $\Omega_Y = \{g(x) : x \in \Omega_X\} = \{(-1)^2, 0^2, 1^2\} = \{0, 1\}$ and the following PMF:

$$p_Y(k) = \begin{cases} \frac{3}{12} + \frac{4}{12} & k = 1 \\ \frac{5}{12} & k = 0 \end{cases}$$

Note that $p_Y(1) = \mathbb{P}(X = -1) + \mathbb{P}(X = 1)$ because $\{-1, 1\} = \{x : x^2 = 1\}$. The crux of the LOTUS proof depends on this fact. We just group things together and sum!

Proof of LOTUS. The proof isn't too complicated, but the notation is pretty tricky and may be an impediment to your understanding, so focus on understanding the setup in the next few lines.

Let $Y = g(X)$. Note that

$$p_Y(y) = \sum_{x \in \Omega_X : g(x) = y} p_X(x)$$

That is, the total probability that $Y = y$ is the sum of the probabilities over all $x \in \Omega_X$ where $g(x) = y$ (this is like saying $\mathbb{P}(Y = 1) = \mathbb{P}(X = -1) + \mathbb{P}(X = 1)$ because $\{x \in \Omega_X : x^2 = 1\} = \{-1, 1\}$.)

$$\begin{aligned}
\mathbb{E}[g(X)] &= \mathbb{E}[Y] && [Y = g(X)] \\
&= \sum_{y \in \Omega_Y} y p_Y(y) && [\text{def of expectation}] \\
&= \sum_{y \in \Omega_Y} y \sum_{x \in \Omega_X : g(x) = y} p_X(x) && [\text{above substitution}] \\
&= \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X : g(x) = y} y p_X(x) && [\text{move } y \text{ into a sum}] \\
&= \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X : g(x) = y} g(x) p_X(x) && [y = g(x) \text{ in the inner sum}] \\
&= \sum_{x \in \Omega_X} g(x) p_X(x) && [\text{the double sum is the same as summing over all } x]
\end{aligned}$$

□

3.2.3 Exercises

1. Let S be the sum of three rolls of a fair 6-sided die. What is $\mathbb{E}[S]$?

   Solution: Let X, Y, Z be the first, second, and third roll respectively. Then, $S = X + Y + Z$. We showed in the first exercise of 3.1 that $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[Z] = 3.5$, so by LoE,

   $$\mathbb{E}[S] = \mathbb{E}[X+Y+Z] = \mathbb{E}[X] + \mathbb{E}[Y] + \mathbb{E}[Z] = 3.5 + 3.5 + 3.5 = 10.5$$

   Alternatively, imagine if we didn't have this theorem. We would find the range of S, which is $\Omega_S = \{3, 4, \ldots, 18\}$ and find its PMF. What a nightmare!

2. Blind LOTUS Practice: This will all seem useless, but I promise we'll need this in the future. Let X have PMF

   $$p_X(k) = \begin{cases} \frac{3}{12} & k = 5 \\ \frac{5}{12} & k = 2 \\ \frac{4}{12} & k = 1 \end{cases}.$$

   (a) Compute $\mathbb{E}\left[X^2\right]$.

   (b) Compute $\mathbb{E}[\log(X)]$

   (c) Compute $\mathbb{E}\left[e^{\sin(X)}\right]$.

   Solution: LOTUS says that $\mathbb{E}[g(X)] = \sum_{k \in \Omega_X} g(k) p_X(k)$. That is,

   (a)
   $$\mathbb{E}\left[X^2\right] = \sum_{k \in \Omega_X} k^2 p_X(k) = 5^2 \cdot \frac{3}{12} + 2^2 \cdot \frac{5}{12} + 1^2 \cdot \frac{4}{12}$$

   (b)
   $$\mathbb{E}[\log X] = \sum_{k \in \Omega_X} \log(k) \cdot p_X(k) = \log(5) \cdot \frac{3}{12} + \log(2) \cdot \frac{5}{12} + \log(1) \cdot \frac{4}{12}$$

   (c)
   $$\mathbb{E}\left[e^{\sin(X)}\right] = \sum_{k \in \Omega_X} e^{\sin(k)} p_X(k) = e^{\sin(5)} \cdot \frac{3}{12} + e^{\sin(2)} \cdot \frac{5}{12} + e^{\sin(1)} \cdot \frac{4}{12}$$