Expectation of a random variable

Since random variables give us a way to talk quantitatively about uncertain quantities, they should also give us a way to make predictions about future outcomes. After seeing a lot of trials we could find their average. Could we compute what we expect this average to be before seeing any data? This will lead us to our most basic way to describe a random variable: the *expectation* (or *expected value*).

First, an example: suppose you are playing a game where you roll a (fair) die and get paid in dollars the amount shown on the die. If we let X denote the number of dollars we win on a particular roll of the die, then X is a random variable with pmf given by

$$p_X(k) = \begin{cases} \frac{1}{6} & \text{for } k = 1, \ldots, 6 \\ 0 & \text{otherwise} \end{cases}$$

How much would you pay to play this game (i.e., how much money can you expect to make "per roll")?

If we roll the die N times, the total amount of money you make is

$$1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + 4 \cdot n_4 + 5 \cdot n_5 + 6 \cdot n_6,$$

where n_k is the number of times the die landed on k. We could then compute the *average amount earned per roll* as

$$M = \frac{1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + 4 \cdot n_4 + 5 \cdot n_5 + 6 \cdot n_6}{N}.$$

As N gets large, we expect that

$$\frac{n_k}{N} \approx \mathrm{P}\left(X = k\right) = p_X(k) = \frac{1}{6}$$

for all k. (After all, what do probabilities mean if not this?) This motivates the definition of the *expected payout* as

$$\begin{aligned}\mathrm{E}[X] &= 1 \cdot p_X(1) + 2 \cdot p_X(2) + 3 \cdot p_X(3) + 4 \cdot p_X(4) + 5 \cdot p_X(5) + 6 \cdot p_X(6) \\ &= \frac{1+2+3+4+5+6}{6} = 3.5.\end{aligned}$$

Definition: The *expectation* of a random variable X with pmf $p_X(k)$ is

$$\mathrm{E}[X] = \sum_k k\, p_X(k).$$

This is really just a "weighted average" of all the values X can take, where the weights are given by the probabilities that X takes each of those values. Think of it as a center of mass for the PMF.

Important Points:

- The expectation of X **is not random**, it is a completely deterministic function of the pmf of X.
- The expectation of X is not necessarily the same thing as the "expected outcome". In the example above, the expected payout is \$3.50, but on no actual roll of the die will you ever win \$3.50.

Exercise:
Suppose that X has the pmf

$$p_X(k) = \begin{cases} k/10 & \text{for } k = 1, 2, 3, 4 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\mathrm{E}[X]$.

Example. Suppose that X is a Poisson random variable with parameter λ, so

$$p_X(k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \ldots.$$

Then

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \\ &= \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\ &= \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\ &= \lambda\mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda\mathrm{e}^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} \\ &= \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} \\ &= \lambda. \end{aligned}$$

Exercise:
Google is interested in hiring you to do some consulting work for them. They have two projects they would like your help with, but they will only pay you if they are satisfied with your work.

- Project 1 pays \$1000 and you believe that the probability you will complete the project to Google's satisfaction is 0.8.
- Project 2 pays \$2000 and you believe that the probability you will complete the project to Google's satisfaction is 0.5.

If Google is happy with your work on whichever project you choose to do first, they will give you the chance to do the second project, but if they don't like your work they will send you on your way. Which project should you take first to maximize your expected earnings?

Note:
It is possible that the pmf of a random variable is well-defined, but the expectation is not well-defined. For example, say X has pmf

$$p_X(k) = \frac{6}{\pi^2}\frac{1}{k^2}, \quad k = 1, 2, \ldots$$

This is a proper pmf, since it is a fact that $\sum_{k\geq 1} 1/k^2 = \pi^2/6$. But

$$\mathrm{E}[X] = \frac{6}{\pi^2}\sum_{k=1}^{\infty} k\,\frac{1}{k^2} = \frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty.$$

This is still a valid PMF; we just can't talk about it's expectation.

Expectations of functions of a random variable

It is straightforward to define the expectation of a random variable $g(X)$:

$$\mathrm{E}[g(X)] = \sum_k g(k) p_X(k).$$

Exercise:
Suppose that X is a discrete random variable with pmf

$$p_X(k) = \begin{cases} 1/9, & \text{when } k \text{ is an integer with } -4 \leq k \leq 4 \\ 0, & \text{otherwise} \end{cases}.$$

1. Let $Y = g(X)$, where $g(k) = |k|$. Compute $E[Y]$.

2. In this case, does $E[g(X)] = g(E[X])$?
   (That is, does $\mathrm{E}[|X|] = |\,\mathrm{E}[X]|$?)

It is important to note that in general:

$$\mathbf{E}[\boldsymbol{g(X)}] \neq \boldsymbol{g}(\mathbf{E}[\boldsymbol{X}]).$$

I cannot stress this enough. Making these two things equal is one of the most common mistakes that probability students make.

Exercise:
If the weather is good, which happens with probability 0.7, I walk the half mile from Tech Square to Van Leer at a speed of $V = 3$ miles per hour; otherwise, I take the Tech Trolley. Assume that when I take the trolley my average speed is $V = 15$ miles per hour. What is the expected value of the time T it takes me to get to class?

Is this the same as if you calculated your expected velocity and used that to calculate your travel time?

Variance

While the expectation tells us something about the average outcome, we are also interested in quantifying how likely X is to be close to that expectation. Due to this, another important quantity associated with a random variable is its **variance**:

$$\mathrm{var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

When we compute the variance, we are calculating the expected value of $(X - E[X])^2$ to tell us roughly how X varies from its expectation $E[X]$ on average. There is nothing particularly sacred about measuring how X varies from $E[X]$ via $(X - E[X])^2$. We could also measure this via something like $|X - E[X]|$ (which is called the "absolute deviation") or even something like $(|X - E[X]|)^3$. However, the particular choice of $(X - E[X])^2$ has a very special role in probability theory, as we will see throughout the rest of the semester.

Notice that since $(X - \mathrm{E}[X])^2 \geq 0$, the variance is always non-negative: $\mathrm{var}(X) \geq 0$.

Related to the variance is the **standard deviation**:

$$\sigma_X = \sqrt{\mathrm{var}(X)}$$

The variance and the standard deviation are measures of the *dispersion* of X around its mean. We will use both, but σ_X is often easier to interpret since it has the same units as X. (For example, if X is in "feet", the $\mathrm{var}(X)$ is in "feet^2" while σ_X is also in "feet".)

Example. Let X be a Bernoulli random variable, with

$$p_X(k) = \begin{cases} 1-p & k=0 \\ p & k=1 \end{cases}.$$

Then

$$\mathrm{E}[X] = p, \qquad \mathrm{var}(X) = \mathrm{E}[(X-p)^2] = p(1-p)$$

Exercise:
Suppose X has pmf

$$p_X(k) = \begin{cases} \frac{1}{3} & \text{for } k = 1,2,3 \\ 0 & \text{otherwise,} \end{cases}$$

and Y has pmf

$$p_Y(k) = \begin{cases} \frac{1}{3} & \text{for } k = 0,2,4 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\mathrm{var}(X)$ and $\mathrm{var}(Y)$.

Exercise:
Suppose X has pmf

$$p_X(k) = \begin{cases} \frac{1}{9} & \text{for } k = -4, \dots, 4 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\text{var}(X)$.

Exercise:
Suppose X has pmf

$$p_X(k) = \begin{cases} \frac{1}{2N+1} & \text{for } k = -N, \dots, N \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\text{var}(X)$.
It may be helpful to recall that $\sum_{k=1}^{N} k^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}$.

Properties of mean and variance

Below, X is a random variable, and $a, b \in \mathbb{R}$ are constants.

1. $\mathrm{E}[X + b] = \mathrm{E}[X] + b$

2. $\mathrm{E}[aX] = a\,\mathrm{E}[X]$

3. We can collect the two results above into one statement:

$$\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b.$$

   So if $g(X)$ has the form $g(x) = ax + b$, the we actually do have $\mathrm{E}[g(X)] = g(\mathrm{E}[X])$ — but again, this is **not true** for general $g(x)$.

4. $\mathrm{var}(X) = \mathrm{E}[X^2] - (E[X])^2$.
   It is easy to see why this is true:

$$\begin{aligned}\mathrm{var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2] &= \mathrm{E}[X^2] - 2\,\mathrm{E}[X\,\mathrm{E}[X]] + \mathrm{E}[(\mathrm{E}[X])^2] \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2,\end{aligned}$$

   where we have used the fact that since $\mathrm{E}[X]$ is not random at all, $\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]$, etc.

5. $\mathrm{var}(X + b) = \mathrm{var}(X)$.
   (You can prove that at home.)

6. $\mathrm{var}(aX) = a^2\,\mathrm{var}\,X$