## 3.2.2 Expectation

If you have a collection of numbers $a_1, a_2, \ldots, a_N$, their average is a single number that describes the whole collection. Now, consider a random variable $X$. We would like to define its average, or as it is called in probability, its **expected value** or **mean**. The expected value is defined as the weighted average of the values in the range.

*Expected value* (= mean=average):

**Definition 3.11**

Let $X$ be a discrete random variable with range $R_X = \{x_1, x_2, x_3, \ldots\}$ (finite or countably infinite). The *expected* value of $X$, denoted by $EX$ is defined as

$$EX = \sum_{x_k \in R_X} x_k \mathrm{P}(X = x_k) = \sum_{x_k \in R_X} x_k P_X(x_k).$$

To understand the concept behind $EX$, consider a discrete random variable with range $R_X = \{x_1, x_2, x_3, \ldots\}$. This random variable is a result of random experiment. Suppose that we repeat this experiment a very large number of times $N$, and that the trials are independent. Let $N_1$ be the number of times we observe $x_1$, $N_2$ be the number of times we observe $x_2$, ...., $N_k$ be the number of times we observe $x_k$, and so on. Since $P(X = x_k) = P_X(x_k)$, we expect that

$$P_X(x_1) \approx \frac{N_1}{N},$$

$$P_X(x_2) \approx \frac{N_2}{N},$$

$$. \quad . \quad .$$

$$P_X(x_k) \approx \frac{N_k}{N},$$

$$. \quad . \quad .$$

In other words, we have $N_k \approx N P_X(x_k)$. Now, if we take the average of the observed values of $X$, we obtain

$$\begin{aligned} \text{Average } &= \frac{N_1 x_1 + N_2 x_2 + N_3 x_3 + \ldots}{N} \\ &\approx \frac{x_1 N P_X(x_1) + x_2 N P_X(x_2) + x_3 N P_X(x_3) + \ldots}{N} \end{aligned}$$

$$
\begin{aligned}
&= x_1 P_X(x_1) + x_2 P_X(x_2) + x_3 P_X(x_3) + \ldots \\
&= EX.
\end{aligned}
$$

Thus, the intuition behind $EX$ is that if you repeat the random experiment independently $N$ times and take the average of the observed data, the average gets closer and closer to $EX$ as $N$ gets larger and larger. We sometimes denote $EX$ by $\mu_X$.

Different notations for expected value of $X$: $EX = E[X] = E(X) = \mu_X$.

Let's compute the expected values of some well-known distributions.

---

**Example 3.11**
Let $X \sim Bernoulli(p)$. Find $EX$.

**Solution**

For the Bernoulli distribution, the range of $X$ is $R_X = \{0, 1\}$, and $P_X(1) = p$ and $P_X(0) = 1 - p$. Thus,

$$
\begin{aligned}
EX &= 0 \cdot P_X(0) + 1 \cdot P_X(1) \\
&= 0 \cdot (1 - p) + 1 \cdot p \\
&= p.
\end{aligned}
$$

---

For a Bernoulli random variable, finding the expectation $EX$ was easy. However, for some random variables, to find the expectation sum, you might need a little algebra. Let's look at another example.

---

**Example 3.12**
Let $X \sim Geometric(p)$. Find $EX$.

**Solution**

For the geometric distribution, the range is $R_X = \{1, 2, 3, \ldots\}$ and the PMF is given by

$$P_X(k) = q^{k-1}p, \qquad \text{for } k = 1, 2, \ldots$$

where, $0 < p < 1$ and $q = p - 1$. Thus, we can write

$$\begin{aligned} EX &= \textstyle\sum_{x_k \in R_X} x_k P_X(x_k) \\ &= \textstyle\sum_{k=1}^{\infty} kq^{k-1}p \\ &= p \textstyle\sum_{k=1}^{\infty} kq^{k-1}. \end{aligned}$$

Now, we already know the geometric sum formula

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \qquad \text{for } |x| < 1.$$

But we need to find a sum $\sum_{k=1}^{\infty} kq^{k-1}$. Luckily, we can convert the geometric sum to the form we want by taking derivative with respect to $x$, i.e.,

$$\frac{d}{dx} \sum_{k=0}^{\infty} x^k = \frac{d}{dx} \frac{1}{1-x}, \qquad \text{for } |x| < 1.$$

Thus, we have

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}, \qquad \text{for } |x| < 1.$$

To finish finding the expectation, we can write

$$\begin{aligned} EX &= p \textstyle\sum_{k=1}^{\infty} kq^{k-1} \\ &= p \tfrac{1}{(1-q)^2} \\ &= p \tfrac{1}{p^2} \\ &= \tfrac{1}{p}. \end{aligned}$$

So, for $X \sim Geometric(p)$, $EX = \frac{1}{p}$. Note that this makes sense intuitively. The random experiment behind the geometric distribution was that we tossed a coin until we observed the first heads, where $P(H) = p$. Here, we found out that on average you need to toss the coin $\frac{1}{p}$ times in this experiment. In particular, if $p$ is small (heads are unlikely), then $\frac{1}{p}$ is large, so you need to toss the coin a large number of times before you observe a heads. Conversely, for large $p$ a few coin tosses usually suffices.

---

**Example 3.13**

Let $X \sim Poisson(\lambda)$. Find $EX$.

**Solution**

Before doing the math, we suggest that you try to guess what the expected value would be. It might be a good idea to think about the examples where the Poisson distribution is used. For the Poisson distribution, the range is $R_X = \{0, 1, 2, \cdots\}$ and the PMF is given by

$$P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad \text{for } k = 0, 1, 2, \ldots$$

Thus, we can write

$$\begin{aligned} EX &= \textstyle\sum_{x_k \in R_X} x_k P_X(x_k) \\ &= \textstyle\sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda} \textstyle\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \textstyle\sum_{j=0}^{\infty} \frac{\lambda^{(j+1)}}{j!} && (\text{ by letting } j = k-1) \\ &= \lambda e^{-\lambda} \textstyle\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} && (\text{ Taylor series for } e^{\lambda}) \\ &= \lambda. \end{aligned}$$

So the expected value is $\lambda$. Remember, when we first talked about the Poisson distribution, we introduced its parameter $\lambda$ as the average number of events. So it is not surprising that the expected value is $EX = \lambda$.

---

Before looking at more examples, we would like to talk about an important property of expectation, which is *linearity*. Note that if $X$ is a random variable, any function of $X$ is also a random variable, so we can talk about its expected value. For example, if $Y = aX + b$, we can talk about $EY = E[aX + b]$. Or if you define $Y = X_1 + X_2 + \cdots + X_n$, where $X_i$'s are random variables, we can talk about $EY = E[X_1 + X_2 + \cdots + X_n]$. The following theorem states that expectation is linear, which makes it easier to calculate the expected value of linear functions of random variables.

Expectation is linear:

**Theorem 3.2**

We have

- $E[aX+b] = aEX+b$, for all $a, b \in \mathbb{R}$;
- $E[X_1+X_2+\cdots+X_n] = EX_1+EX_2+\cdots+EX_n$, for any set of random variables $X_1, X_2, \cdots, X_n$.

We will prove this theorem later on in Chapter 5, but here we would like to emphasize its importance with an example.

---

**Example 3.14**

Let $X \sim Binomial(n, p)$. Find $EX$.

**Solution**

We provide two ways to solve this problem. One way is as before: we do the math and calculate $EX = \sum_{x_k \in R_X} x_k P_X(x_k)$ which will be a little tedious. A much faster way would be to use linearity of expectation. In particular, remember that if $X_1, X_2, \ldots, X_n$ are independent $Bernoulli(p)$ random variables, then the random variable $X$ defined by $X = X_1 + X_2 + \ldots + X_n$ has a $Binomial(n, p)$ distribution. Thus, we can write

$$\begin{aligned}
EX &= E[X_1+X_2+\cdots+X_n] \\
&= EX_1+EX_2+\cdots+EX_n && \text{by linearity of expectation} \\
&= p+p+\cdots+p \\
&= np.
\end{aligned}$$

We will provide the direct calculation of $EX = \sum_{x_k \in R_X} x_k P_X(x_k)$ in the Solved Problems section and as you will see it needs a lot more algebra than above. The bottom line is that linearity of expectation can sometimes make our calculations much easier. Let's look at another example.

---

**Example 3.15**

Let $X \sim Pascal(m, p)$. Find $EX$. (*Hint:* Try to write $X = X_1 + X_2 + \cdots + X_m$, such that you already know $EX_i$.)

**Solution**

We claim that if the $X_i$'s are independent and $X_i \sim Geometric(p)$, for $i = 1, 2, \cdots, m$, then the random variable $X$ defined by $X = X_1 + X_2 + \cdots + X_m$ has $Pascal(m, p)$. To see this, you can look at Problem 5 in Section 3.1.6 and the discussion there. Now, since we already know $EX_i = \frac{1}{p}$, we conclude

$$\begin{aligned} EX &= E[X_1 + X_2 + \cdots + X_m] \\ &= EX_1 + EX_2 + \cdots + EX_m \qquad \text{by linearity of expectation} \\ &= \frac{1}{p} + \frac{1}{p} + \cdots + \frac{1}{p} \\ &= \frac{m}{p}. \end{aligned}$$

Again, you can try to find $EX$ directly and as you will see, you need much more algebra compared to using the linearity of expectation.

---