Chapter 6

Meeting Expectations

When a large collection of data is gathered, one is typically interested not necessarily in every individual data point, but rather in certain descriptive quantities such as the average or the median. The same is true for random variables. The PMF of discrete random variable X provides a complete characterization of the distribution of X by specifying the probability of every possible value of X. Still, it may be desirable to summarize the information contained in the PMF of a random variable. One way to accomplish this task and thereby obtain meaningful descriptive quantities is through the expectation operator.

6.1 Expected Values

The *expected value* $\mathrm{E}[X]$ of discrete random variable X is defined by

$$\mathrm{E}[X] = \sum_{x \in X(\Omega)} x p_X(x), \qquad (6.1)$$

whenever this sum converges absolutely. If this sum is not absolutely convergent, then X is said not to possess an expected value. As mentioned above, the expected value $\mathrm{E}[X]$ provides insightful information about the underlying random variable X without giving a comprehensive and overly detailed description. The expected value of a random variable, as defined in (6.1), is also called the *mean* of X. It is important to realize that $\mathrm{E}[X]$ is not a function of random variable X; rather, it is a function of the PMF of X.

Example 43. *A fair die is rolled once, with the number of dots appearing on the top face taken as the value of the corresponding random variable. The expected value of the roll can be computed as*

$$\sum_{k=1}^{6} \frac{k}{6} = \frac{42}{12} = 3.5.$$

In other words, the mean of this random variable is 3.5.

Example 44. *Assume that a fair coin is flipped repetitively until heads is observed. The value of random variable X is taken to be the total number of tosses performed during this experiment. The possible values for X are therefore given by $X(\Omega) = \{1, 2, \ldots\}$. Recall that, in this case, the PMF of X is equal to $p_X(k) = 2^{-k}$, where k is a positive integer. The expected value of this geometric random variable can be computed as*

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

The expected number of tosses until the coin produces heads is equal to two.

In general, determining the expectation of a random variable requires as input its PMF, a detailed characterization of the random variable, and returns a much simpler scalar attribute, its mean. Hence, computing the expected value of the random variable yields a concise summary of its overall behavior.

6.2   Functions and Expectations

The mean forms one instance where the distribution of a random variable is condensed into a scalar quantity. There are several additional examples. The notion of an *expectation* can be combined with traditional functions to create alternate descriptions and other meaningful quantities. Suppose that X is a discrete random variable. Let $g(\cdot)$ be a real-valued function on the range of X, and consider the expectation of $g(X)$. This expected value, $\mathrm{E}[g(X)]$, is a scalar quantity that can also provide partial information about the distribution of X.

One way to determine the expected value of $g(X)$ is to first note that $Y = g(X)$ is itself a random variable. Thus, we can find the derived distribution

of Y and then apply the definition of expected value provided in (6.1). Yet, there is a more direct way to compute this quantity; the *expectation* of $g(X)$ can be expressed as

$$\mathrm{E}\left[g(X)\right] = \sum_{x \in X(\Omega)} g(x) p_X(x). \tag{6.2}$$

It is worth re-emphasizing that there exist random variables and functions for which the above sum does not converge. In such cases, we simply say that the expected value of $g(X)$ does not exist. Also, notice that the mean $\mathrm{E}[X]$ is a special case of (6.2), where $g(X) = X$. Hence the definition of $\mathrm{E}[g(X)]$ given above subsumes our original description of $\mathrm{E}[X]$, which appeared in (6.1). We explore pertinent examples below.

Example 45. *The simplest possible scenario for* (6.2) *is the case where the function* $g(\cdot)$ *is a constant. The expectation of* $g(X) = c$ *becomes*

$$\mathrm{E}[c] = \sum_{x \in X(\Omega)} c p_X(x) = c \sum_{x \in X(\Omega)} p_X(x) = c.$$

The last inequality follows from the normalization axiom of probability laws. The expectation of a constant is always the constant itself.

Example 46. *Let S be a subset of the real numbers, and define the* indicator function *of S by*

$$\mathbf{1}_S(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S. \end{cases}$$

The expectation of $\mathbf{1}_S(X)$ is equal to

$$\mathrm{E}\left[\mathbf{1}_S(X)\right] = \sum_{x \in X(\Omega)} \mathbf{1}_S(x) p_X(x)$$
$$= \sum_{x \in S \cap X(\Omega)} p_X(x) = \Pr(X \in S).$$

That is, the expectation of the indicator function of S is simply the probability that X takes on a value in S. This alternate way of computing the probability of an event can sometimes be employed to solve otherwise difficult probability problems.

Let random variable Y be defined by applying real-valued function $g(\cdot)$ to X, with $Y = g(X)$. The mean of Y is equal to the expectation of $g(X)$, and we know from our ongoing discussion that this value can be obtained by applying two different formulas. To ensure consistency, we verify that these two approaches lead to a same answer.

First, we can apply (6.1) directly to Y, and obtain

$$\mathrm{E}[Y] = \sum_{y \in g(X(\Omega))} y p_Y(y),$$

where $p_Y(\cdot)$ is the PMF of Y provided by (5.3). Alternatively, using (6.2), we have

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p_X(x).$$

We prove that these two expressions describe a same answer as follows. Recall that the PMF of Y evaluated at y is obtained by summing the values of $p_X(\cdot)$ over all $x \in X(\Omega)$ such that $g(x) = y$. Mathematically, this can be expressed as $p_Y(y) = \sum_{\{x \in X(\Omega) | g(x) = y\}} p_X(x)$. Using this equality, we can write

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_{y \in g(X(\Omega))} y p_Y(y) = \sum_{y \in g(X(\Omega))} y \sum_{\{x \in X(\Omega) | g(x) = y\}} p_X(x) \\
&= \sum_{y \in g(X(\Omega))} \sum_{\{x \in X(\Omega) | g(x) = y\}} y p_X(x) \\
&= \sum_{y \in g(X(\Omega))} \sum_{\{x \in X(\Omega) | g(x) = y\}} g(x) p_X(x) \\
&= \sum_{x \in X(\Omega)} g(x) p_X(x) = \mathrm{E}[g(X)].
\end{aligned}$$

Note that first summing over all possible values of Y and then over the preimage of every $y \in Y(\Omega)$ is equivalent to summing over all $x \in X(\Omega)$. Hence, we have shown that computing the expectation of a function using the two methods outlined above leads to a same solution.

Example 47. *Brazos Extreme Events Radio creates the "Extreme Trio" contest. To participate, a person must fill out an application card. Three cards are drawn from the lot and each winner is awarded \$1,000. While a same participant can send multiple cards, he or she can only win one grand prize.*

At the time of the drawing, the radio station has accumulated a total of 100 cards. David, an over-enthusiastic listener, is accountable for half of these cards. We wish to compute the amount of money David expects to win under this promotion.

Let X be the number of cards drawn by the radio station written by David. The PMF of X is given by

$$p_X(k) = \frac{\binom{50}{k}\binom{50}{3-k}}{\binom{100}{3}} \quad k \in \{0, 1, 2, 3\}.$$

The money earned by David can be expressed as $g(k) = 1000 \min\{k, 1\}$. It follows that the expected amount of money he receives is equal to

$$\sum_{k=0}^{3} (1000 \min\{k, 1\}) \, p_X(k) = 1000 \cdot \frac{29}{33}.$$

Alternatively, we can define $Y = 1000 \min\{X, 1\}$. Clearly, Y can only take on one of two possible values, 0 or 1000. Evaluating the PMF of Y, we get $p_Y(0) = p_X(0) = 4/33$ and $p_Y(1000) = 1 - p_Y(0) = 29/33$. The expected value of Y is equal to

$$0 \cdot p_Y(0) + 1000 \cdot p_Y(1000) = 1000 \cdot \frac{29}{33}.$$

As anticipated, both methods lead to the same answer. The expected amount of money won by David is roughly \$878.79.

6.2.1 The Mean

As seen at the beginning of this chapter, the simplest non-trivial expectation is the *mean*. We provide two additional examples for the mean, and we explore a physical interpretation of its definition below.

Example 48. *Let X be a geometric random variable with parameter p and PMF*

$$p_X(k) = (1 - p)^{k-1}p, \quad k = 1, 2, \ldots$$

The mean of this random variable is

$$E[X] = \sum_{k=1}^{\infty} k(1 - p)^{k-1}p = p \sum_{k=1}^{\infty} k(1 - p)^{k-1} = \frac{1}{p}.$$

Example 49. *Let X be a binomial random variable with parameters n and p. The PMF of X is given by*

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

The mean of this binomial random variable can therefore be computed as

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{\ell=0}^{n-1} \frac{n!}{\ell!(n-1-\ell)!} p^{\ell+1} (1-p)^{n-\ell-1} \\
&= np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell} = np.
\end{aligned}
$$

Notice how we rearranged the sum into a familiar form to compute its value.

It can be insightful to relate the mean of a random variable to classical mechanics. Let X be a random variable and suppose that, for every $x \in X(\Omega)$, we place an infinitesimal particle of mass $p_X(x)$ at position x along a real line. The mean of random variable X as defined in (6.1) coincides with the center of mass of the system of particles.

Example 50. *Let X be a Bernoulli random variable such that*

$$p_X(x) = \begin{cases} 0.25, & \text{if } x = 0 \\ 0.75, & \text{if } x = 1. \end{cases}$$

The mean of X is given by

$$\mathrm{E}[X] = 0 \cdot 0.25 + 1 \cdot 0.75 = 0.75.$$

Consider a two-particle system with masses $m_1 = 0.25$ and $m_2 = 0.75$, respectively. In the coordinate system illustrated below, the particles are located at positions $x_1 = 0$ and $x_2 = 1$. From classical mechanics, we know that their center of mass can be expressed as

$$\frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} = 0.75.$$

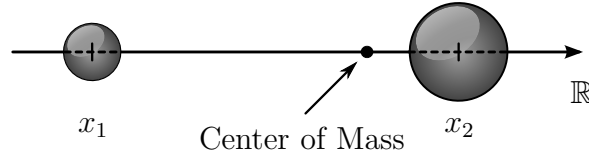*As anticipated, the center of mass corresponds to the mean of X.*

Figure 6.1: The center of mass on the figure is indicated by the tip of the arrow. In general, the mean of a discrete random variable corresponds to the center of mass of the associated particle system

6.2.2 The Variance

A second widespread descriptive quantity associated with random variable X is its *variance*, which we denote by $\text{Var}[X]$. It is defined by

$$\text{Var}[X] = \text{E}\left[(X - \text{E}[X])^2\right]. \tag{6.3}$$

Evidently, the variance is always nonnegative. It provides a measure of the dispersion of X around its mean. For discrete random variables, it can be computed explicitly as

$$\text{Var}[X] = \sum_{x \in X(\Omega)} (x - \text{E}[X])^2 \, p_X(x).$$

The square root of the variance is referred to as the *standard deviation* of X, and it is often denoted by σ.

Example 51. *Suppose X is a Bernoulli random variable with parameter p. We can compute the mean of X as*

$$\text{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Its variance is given by

$$\text{Var}[X] = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p).$$

Example 52. *Let X be a Poisson random variable with parameter λ. The mean of X is given by*

$$\text{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda}$$

$$= \lambda \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} e^{-\lambda} = \lambda.$$

The variance of X can be calculated as

$$\mathrm{Var}[X] = \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \sum_{k=0}^{\infty} \left(\lambda^2 + k(1 - 2\lambda) + k(k - 1) \right) \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \lambda - \lambda^2 + \sum_{k=2}^{\infty} \frac{\lambda^k}{(k - 2)!} e^{-\lambda} = \lambda.$$

Both the mean and the variance of a Poisson random variable are equal to its parameter λ.

6.2.3   Affine Functions

Proposition 3. *Suppose X is a random variable with finite mean. Let Y be the affine function of X defined by $Y = aX + b$, where a and b are fixed real numbers. The mean of random variable Y is equal to $\mathrm{E}[Y] = a\mathrm{E}[X] + b$.*

Proof. This can be computed using (6.2);

$$\mathrm{E}[Y] = \sum_{x \in X(\Omega)} (ax + b) p_X(x)$$

$$= a \sum_{x \in X(\Omega)} x p_X(x) + b \sum_{x \in X(\Omega)} p_X(x)$$

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

We can summarize this property with $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$. $\square$

It is not much harder to show that the expectation is a linear functional. Suppose that $g(\cdot)$ and $h(\cdot)$ are two real-valued functions such that $\mathrm{E}[g(X)]$ and $\mathrm{E}[h(X)]$ both exist. We can write the expectation of $ag(X) + h(X)$ as

$$\mathrm{E}[ag(X) + h(X)] = \sum_{x \in X(\Omega)} (ag(x) + h(x)) p_X(x)$$

$$= a \sum_{x \in X(\Omega)} g(x) p_X(x) + \sum_{x \in X(\Omega)} h(x) p_X(x)$$

$$= a\mathrm{E}[g(X)] + \mathrm{E}[h(X)].$$

This demonstrates that the expectation is both homogeneous and additive.

Proposition 4. *Assume that X is a random variable with finite mean and variance, and let Y be the affine function of X given by $Y = aX + b$, where a and b are constants. The variance of Y is equal to $\text{Var}[Y] = a^2\text{Var}[X]$.*

Proof. Consider (6.3) applied to $Y = aX + b$,

$$
\begin{aligned}
\text{Var}[Y] &= \sum_{x \in X(\Omega)} (ax + b - \text{E}[aX + b])^2 \, p_X(x) \\
&= \sum_{x \in X(\Omega)} (ax + b - a\text{E}[X] - b)^2 \, p_X(x) \\
&= a^2 \sum_{x \in X(\Omega)} (x - \text{E}[X])^2 \, p_X(x) = a^2\text{Var}[X].
\end{aligned}
$$

The variance of an affine function only depends on the distribution of its argument and parameter a. A translation of the argument by b does not affect the variance of $Y = aX + b$; in other words, it is shift invariant. $\qquad\square$

6.3 Moments

The *moments* of a random variable X are likewise important quantities used in providing partial information about the PMF of X. The nth moment of random variable X is defined by

$$
\text{E}[X^n] = \sum_{x \in X(\Omega)} x^n p_X(x). \tag{6.4}
$$

Incidentally, the mean of random variable X is its first moment.

Proposition 5. *The variance of random variable X can be expressed in terms of its first two moments, $\text{Var}[X] = \text{E}\left[X^2\right] - (\text{E}[X])^2$.*

Proof. Suppose that the variance of X exists and is finite. Starting from (6.3), we can expand the variance of X as follows,

$$
\begin{aligned}
\text{Var}[X] &= \sum_{x \in X(\Omega)} (x - \text{E}[X])^2 \, p_X(x) \\
&= \sum_{x \in X(\Omega)} \left(x^2 - 2x\text{E}[X] + (\text{E}[X])^2\right) p_X(x) \\
&= \sum_{x \in X(\Omega)} x^2 p_X(x) - 2\text{E}[X] \sum_{x \in X(\Omega)} x p_X(x) + (\text{E}[X])^2 \sum_{x \in X(\Omega)} p_X(x) \\
&= \text{E}\left[X^2\right] - (\text{E}[X])^2.
\end{aligned}
$$

This alternate formula for the variance is sometimes convenient for computational purposes. $\square$

We include below an example where the above formula for the variance is applied. This allows the straightforward application of standard sums from calculus.

Example 53. *Let X be a uniform random variable with PMF*

$$p_X(k) = \begin{cases} 1/n, & \text{if } k = 1, 2, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

The mean of this uniform random variable is equal to

$$\mathrm{E}[X] = \frac{n+1}{2}.$$

The variance of X can then be obtained as

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \sum_{k=1}^{n} \frac{k^2}{n} - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{n^2 - 1}{12}.$$

Closely related to the moments of a random variable are its *central moments*. The kth central moment of X is defined by $\mathrm{E}\left[(X - E[X])^k\right]$. The variance is an example of a central moment, as we can see from definition (6.3). The central moments are used to define the *skewness* of random variable, which is a measure of asymmetry; and its *kurtosis*, which assesses whether the variance is due to infrequent extreme deviations or more frequent, modest-size deviations. Although these quantities will not play a central role in our exposition of probability, they each reveal a different characteristic of a random variable and they are encountered frequently in statistics.

6.4   Ordinary Generating Functions

In the special yet important case where $X(\Omega)$ is a subset of the non-negative integers, it is occasionally useful to employ the *ordinary generating function*.

This function bears a close resemblance to the *z-transform* and is defined by

$$G_X(z) = \mathrm{E}\left[z^X\right] = \sum_{k=0}^{\infty} z^k p_X(k). \tag{6.5}$$

It is also called the *probability-generating function* because of the following property. The probability $\Pr(X = k) = p_X(k)$ can be recovered from the corresponding generating function $G_X(z)$ through Taylor series expansion. Within the radius of convergence of $G_X(z)$, we have

$$G_X(z) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{d^k G_X}{dz^k}(0)\right) z^k.$$

Comparing this equation to (6.5), we conclude that

$$p_X(k) = \frac{1}{k!}\frac{d^k G_X}{dz^k}(0).$$

We note that, for $|z| \le 1$, we have

$$\left|\sum_{k=0}^{\infty} z^k p_X(k)\right| \le \sum_{k=0}^{\infty} |z|^k p_X(k) \le \sum_{k=0}^{\infty} p_X(k) = 1$$

and hence the radius of convergence of any probability-generating function must include one.

The ordinary generating function plays an important role in dealing with sums of discrete random variables. As a preview of what lies ahead, we compute ordinary generating functions for Bernoulli and Binomial random variables below.

Example 54. *Let X be a Bernoulli random variable with parameter p. The ordinary generating function of X is given by*

$$G_X(z) = p_X(0) + p_X(1)z = 1 - p + pz.$$

Example 55. *Let S be a binomial random variable with parameters n and p. The ordinary generating function of S can be computed as*

$$G_S(z) = \sum_{k=0}^{n} z^k p_S(k) = \sum_{k=0}^{n} z^k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1-p)^{n-k} = (1 - p + pz)^n.$$

We know, from Section 5.2.2, that one way to create a binomial random variable is to sum n independent and identically distributed Bernoulli random variables, each with parameter p. Looking at the ordinary generating functions above, we notice that the $G_S(z)$ is the product of n copies of $G_X(z)$. Intuitively, it appears that the sum of independent discrete random variables leads to the product of their ordinary generating functions, a relation that we will revisit shortly.

The mean and second moment of X can be computed based on its ordinary generating function, In particular, we have

$$\mathrm{E}[X] = \lim_{z \uparrow 1} \frac{dG_X}{dz}(z).$$

Similarly, the second moment of X can be derived as

$$\mathrm{E}\left[X^2\right] = \lim_{z \uparrow 1} \left(\frac{d^2 G_X}{dz^2}(z) + \frac{dG_X}{dz}(z) \right).$$

This can be quite useful, as illustrated in the following example.

Example 56 (Poisson Random Variable). *Suppose that X has a Poisson distribution with parameter $\lambda > 0$. The function $G_X(s)$ can be computed using the distribution of X,*

$$G_X(z) = \sum_{k=0}^{\infty} z^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}.$$

The first two moments of X are given by

$$\mathrm{E}[X] = \lim_{z \uparrow 1} \frac{dG_X}{dz}(z) = \lim_{z \uparrow 1} \lambda e^{\lambda(z-1)} = \lambda$$

$$\mathrm{E}\left[X^2\right] = \lim_{z \uparrow 1} \left(\frac{d^2 G_X}{dz^2}(z) + \frac{dG_X}{dz}(z) \right) = \lim_{z \uparrow 1} \left(\lambda^2 + \lambda\right) e^{\lambda(z-1)} = \lambda^2 + \lambda.$$

This provides a very efficient way to compute the mean and variance of X, which are both equal to λ. It may be helpful to compare this derivation with Example 52.

Further Reading

1. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Section 4.3.

2. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Section 2.4.

3. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Sections 2.4, 3.1.

4. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Sections 4.8.