Lesson 2.2: Mathematical Expectation of Discrete Random Variables

Learning Outcomes

At the end of the lesson, the students must be able to

1. Articulate the meaning of mathematical expectation,
2. Explain the properties of expected value,
3. Evaluate the mean and variance of a discrete random variable,
4. Determine the expected value of functions of a discrete random variable, and
5. Derive the mean and variance of a random variable using the moment-generating function.

1 What is mathematical expectation?

If you have a collection of numbers $a_1, a_2, ..., a_n$, 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.

Definition:
Let Y be a discrete random variable with pmf $p_Y(y)$ and domain D. The expected value of Y is given by

$$E(Y) = \sum_{y \in D} y \times p_Y(y) = \sum_{y \in D} y \times P(Y = y)$$

The above definition simply means that the expected value $E(Y)$ is a weighted average of the possible values of Y. Each value $y \in D$ is weighted by its corresponding probability $P(Y = y)$.

Example 1:

What is the average number of dots if we toss a die?

Solution:
Let Y be the number of dots on the upturned face if we toss a die. Recall that the pmf of Y is given by

Y	1	2	3	4	5	6
$P(Y=y)$	1/6	1/6	1/6	1/6	1/6	1/6

So based on the definition, the expected value of Y is

$$\begin{aligned} E(Y) &= \sum_{y=1}^{6} y \times P(Y=y) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \\ &= 3.5 \end{aligned}$$

What does this value mean? Hmm... if we toss a fair, six-sided die once, should we expect the toss to be 3.5? No, of course not! All the expected value tells us is what we would expect the average of a large number of tosses to be in the long run. If we toss a die a thousand times, say, and calculate the average of the tosses, will the average of the 1000 tosses be exactly 3.5? No, probably not! But, we can certainly expect it to be close to 3.5. It is important to keep in mind that the expected value of Y is a theoretical average, not an actual value of Y!

Remarks:

The expected value, or mean, of Y can be interpreted in different ways:

1. $E(Y)$ is the "center of gravity" on the pmf of Y. It's located where the pmf would balance.
2. $E(Y)$is a "long run average." In other words, if we observed the value of Y over and over again, then the average value would be close to $E(Y)$.
3. We also use μ_Y to denote the expected value of Y. That is, $E(Y) = \mu_Y$.

Definition:
Let Y be a discrete random variable with pmf $p_Y(y)$ and domain D. Suppose that g is a real-valued function. Then, $g(Y)$ is a random variable and

$$E[g(Y)] = \sum_{y \in D} g(y) \times p_Y(y)$$

Example 2:

Patient responses to a generic drug to control pain are scored on a 5-point scale (1 = lowest pain level; 5 = highest pain level). In a certain population of patients, the pmf of the response Y is given by

Y	1	2	3	4	5
$P(Y=y)$	0.38	0.27	0.18	0.11	0.06

Calculate:

1. $E(Y)$
2. $E(Y^2)$
3. $E(2Y^3+1)$

Solution:

1.
$$\begin{aligned} E(Y) &= \sum_{y \in D} y \times p_Y(y) \\ &= 1 \times 0.38 + 2 \times 0.27 + 3 \times 0.18 + 4 \times 0.11 + 5 \times 0.06 \\ &= 2.2 \end{aligned}$$

2.
$$\begin{aligned} E(Y^2) &= \sum_{y \in D} y^2 \times p_Y(y) \\ &= 1^2 \times 0.38 + 2^2 \times 0.27 + 3^2 \times 0.18 + 4^2 \times 0.11 + 5^2 \times 0.06 \\ &= 6.34 \end{aligned}$$

3.
$$\begin{aligned} E(2Y^3+1) &= \sum_{y \in D} (2y^3+1) \times p_Y(y) \\ &= (2 \times 1^3 + 1) \times 0.38 + (2 \times 2^3 + 1) \times 0.27 + (2 \times 3^3 + 1) \times 0.18 + (2 \times 4^3 + 1) \times 0.11 \\ &\quad + (2 \times 5^3 + 1) \times 0.06 \\ &= 44.88 \end{aligned}$$

Properties of expected value

In general, the expectation operator $E()$ possesses certain properties. These are as follows.

1. The expected value of a constant k is equal to k. That is, $E(k) = k$.
2. Constant multiplier can be moved outside the expectation. That is, $E(kY) = kE(Y)$.
3. The expectation operator $E()$ is additive. That is, $E(Y_1 + Y_2) = E(Y_1) + E(Y_2)$.

2 The variance and the standard deviation

Suppose Y is a discrete random variable with mean $E(Y) = \mu_Y$. The variance of Y is

$$V(Y) = \sigma_Y^2 = E[(Y - \mu_Y)^2] = \sum_{y \in D} (y - \mu_Y)^2 \times P(Y = y)$$

The above definition tells us that the variance is a weighted average of the possible values of $g(Y) = (Y - \mu_Y)^2$ weighted by the corresponding probabilities $p_Y(y) = P(Y = y)$.

The variance comes in squared units of the data set. That is, if the original data are measured in centimeters, the variance are expressed in squared centimeters. Therefore, it is difficult to imagine using the variance as a measure of "distance" of each value from the mean of the data set. It is in this context that we get the positive square root of the variance and call it the **standard deviation**. The standard deviation is given by

$$\sigma_Y = \sqrt{V(Y)} = \sqrt{\sigma_Y^2}$$

Example 3:

Calculate the variance and standard deviation of the random variable Y in Example 2.

Solution:
From Example 2, we have $\mu_Y = E(Y) = 2.2$, hence

$$\begin{aligned}
\sigma_Y^2 &= E[(Y - \mu_Y)^2] \\
&= \sum_{y \in D} (y - \mu_Y)^2 \times P(Y = y) \\
&= (1 - 2.2)^2 \times 0.38 + (2 - 2.2)^2 \times 0.27 + (3 - 2.2)^2 \times 0.18 + (4 - 2.2)^2 \times 0.11 + (5 - 2.2)^2 \times 0.06 \\
&= 1.5
\end{aligned}$$

Thus, the standard deviation is $\sigma_Y = \sqrt{1.5} \approx 1.225$. This means that patients' pain tolerance scores deviates from the mean (2.2), on average, by 1.225 points.

Remark:

We can apply algebra to the definitional formula of the variance to derive its computing formula. It is given as

$$V(Y) = E(Y^2) - [E(Y)]^2 = E(Y^2) - \mu_Y^2$$

3 Moment generating function

Definition:
Let X be a discrete random variable with pmf $p_X(x)$ and domain D. The *moment generating function* (mgf) for X, denoted by $m_X(t)$, is given by

$$m_X(t) = E[e^{tX}] = \sum_{\forall x \in D} e^{tx} P(X = x)$$

Remarks:

1. The moment generating function for some random variables may not exist.
2. The moment generating function (mgf) can be used to generate moments. In fact, from the theory of Laplace transforms, it follows that if the mgf exists, it characterizes an infinite set of moments.

Theorem: Let X denote a random variable with domain D and mgf $m_X(t)$. Then,

$$E[Y^k] = \frac{d^k}{dt^k} m_X(t) \Big|_{t=0}$$

From the above theorem, we have

$$E(X) = \frac{d}{dt} m_X(t) \Big|_{t=0}$$
$$E(X^2) = \frac{d^2}{dt^2} m_X(t) \Big|_{t=0}$$

Example 4:

Suppose a random variable X has the following pmf:

$$P(X = x) = \begin{cases} \frac{1}{10}(5 - x), \text{ for } x = 1, 2, 3, 4 \\ 0, \text{ otherwise} \end{cases}$$

(a) Derive the mgf of X.

(b) Use the answer in (a) to calculate the mean and standard deviation of X.

Solution: Left as a classroom exercise!

Example 5:

Consider the random experiment of rolling a pair of dice. Let Y be the absolute difference in the number of dots on each die. Derive and/or calculate the following:

(a) Derive the mgf of Y.

(b) Use the mgf to calculate the mean and standard deviation of Y.

Solution: Left as a classroom exercise!