Stat 121 (Mathematical Statistics I)

Lesson 2.2 Mathematical Expectation of Discrete Random Variables

Learning Outcomes

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

1. Articulate the meaning of mathematical expectation,
2. Explain the properties of expected value,
3. Determine expected value of a discrete random variable, and
4. Determine the expected value of functions of a discrete random variable.

What is a 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 simply the **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

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

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

<u>Example 2.2.1</u>

What is the average toss of a fair six-sided die?

<u>Solution</u>

Let Y be random variable denoting the number of dots on the upturned face if we roll a die. I think everybody understands that the PMF of Y is given by the table.

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

Therefore,

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

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 fair, six-sided 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 X is a theoretical average, not an actual, realized one!

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 in *Example 2.2.1*. 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, $E(Y) = \mu_Y$.

Example 2.2.2

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

Solution

Based on the definition, the expected value of Y is

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

Definition

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

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

Example 2.2.3

In *Example 2.2.2*, We used the PMF below to describe the population of patients' responses to a generic drug to control pain:

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

Compute the following.

a. $E(Y^2)$

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

Solution

a. Let $g(Y) = Y^2$. By definition, we have

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

b. Let $g(Y) = 2Y^3 + 1$. By definition, we have

$$
\begin{aligned}
E[g(Y)] &= \sum_{\forall \in D} g(y)P(Y = y) \\
&= \sum_{\forall \in D} (2y^3 + 1) \times P(Y = y) \\
&= [2(1^3) + 1] \times 0.38 + [2(2^3) + 1] \times 0.27 + [2(3^3) + 1] \times 0.18 + [2(4^3) + 1] \times 0.11 + [2(5^3) + 1 \\
&= 44.88
\end{aligned}
$$

Properties of expected value

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

1. The expected value of a constant k is k. That is, $E(k) = k$.

Proof:

$$
\begin{aligned}
E(k) &= \sum_{\forall y} kP(Y = y) \\
&= k \sum_{\forall y} P(Y = y) \\
&= k(1) \\
&= k
\end{aligned}
$$

2. A constant multiplier can be moved outside the expectation. That is, $E(kY) = kE(Y)$.

Proof:

$$
\begin{aligned}
E(kY) &= \sum_{\forall y} kyP(Y = y) \\
&= k \sum_{\forall y} yP(Y = y) \\
&= kE(Y)
\end{aligned}
$$

3. The expectation operator $E()$ is additive. That is, $E[g_1(Y) + g_2(Y)] = E[g_1(Y)] + E[g_2(Y)]$.

Proof:

$$
\begin{aligned}
E[g_1(Y) + g_2(Y)] &= \sum_{\forall y \in D} [g_1(y) + g_2(y)]P(Y = y) \\
&= \sum_{\forall y} g_1(y)P(Y = y) + \sum_{\forall y} g_2(y)P(Y = y) \\
&= E[g_1(Y)] + E[g_2(Y)]
\end{aligned}
$$

This property can be extended to more than two terms. That is,

$$
E\Big[\sum_{i=1}^{n} g_i(Y)\Big] = \sum_{i=1}^{n} E[g_i(Y)]
$$

The proof is similar to the case of two terms and is left as an exercise.

The variance and the standard deviation

Another important measure to describe a set of data, in addition to the mean, is a measure of dispersion. It reflects how the values in a set are dispersed about the mean. That is, a measure of dispersion tells us the "distance" of each value from their mean. One such measure is the **variance**.

Definition

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

$$V(Y) = \sigma_Y^2 = E[(Y - \mu)^2] = \sum_{\forall y} (y - \mu)^2 P(Y = y)$$

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

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**. That is,

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

Example 2.2.4

In *Example 2.2.2*, we used the PMF below to describe the population of patients' responses to a generic drug to control pain:

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

Compute the variance and standard deviation of Y.

Solution

From *Example 2.2.2*, the mean of Y is equal to 2.2. Hence, the variance of Y is

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

Remarks

The variance of a discrete random variable Y has the following properties and interpretations:

1. The variance is nonnegative, that is, $\sigma_Y^2 \geq 0$. This is easy to understand since the variance is the expected value of squares.

2. Can $V(Y)$ ever be zero? It can, but only when all of the probability mass for Y resides at one point, namely $y = \mu$. A random variable Y with this property is called a **degenerate** random variable. Any constant k can be thought of as a degenerate random variable. Hence, $V(k) = 0$.

c. The larger (smaller) $V(Y)$ is, the more (less) spread in the possible values of Y about the mean $\mu = E(Y)$.

d. The variance $V(Y)$ is measured in the squared units of Y . The standard deviation is measured in the same units as Y. Because of this, the standard deviation is easier to interpret.

e. The variance can also be computed as $V(Y) = E(Y^2) - \mu^2$. The proof is given below.

Proof:

$$
\begin{aligned}
V(Y) &= E[(Y - \mu)^2] \\
&= E[Y^2 - 2Y\mu - \mu^2] \\
&= E(Y^2) - 2\mu E(Y) - \mu^2 \\
&= E(Y^2) - 2\mu^2 - \mu^2 \\
&= E(Y^2) - \mu^2
\end{aligned}
$$

Learning Tasks/Activities

1. Suppose the random variable Y has the following probability distribution.

Y	0	1	2	3	4
P(Y = y)	0.1	0.2	0.4	0.2	0.1

Compute the mean and variance of Y.

2. Let Y be a discrete random variable with mean μ_Y and variance σ_Y^2. Show using the definition of the variance that $V(aY + b) = a^2V(Y)$, where a and b are constants.

3. The maximum patent life for a new drug is 17 years. Subtracting the length of time required by the FDA for testing and approval of the drug provide the actual patent life for the drug—that is, the length of time that the company has to recover research and development costs and to make a profit. The distribution of the lengths of actual patent lives for new drugs is given below:

Y	3	4	5	6	7	8	9	10	11	12	13
P(Y = y)	0.03	0.05	0.07	0.10	0.14	0.20	0.18	0.12	0.07	0.03	0.01

a. Find the mean and standard deviation of Y.

b. What is the probability that the value of Y falls in the interval $\mu \pm 2\sigma_Y$?