LECTURE 5
EXPECTATION OF DISCRETE RANDOM VARIABLES

1 Expectation of a discrete random variable

In this section, we introduce the concept of *mathematical expectation* and discuss its properties.

Example. Consider the following game: you roll a dice once, and if the outcome is an even number you receive one dollar; however, if the outcome is an odd number, you have to pay one dollar. Let Y_1 denote the (random) payoff.

Table 1: Random payoffs (Y_1, Y_2)

ω	$P(\{\omega\})$	$Y_1(\omega)$	$Y_2(\omega)$
1	1/6	-1	-1
2	1/6	1	-1
3	1/6	-1	-1
4	1/6	1	-1
5	1/6	-1	-1
6	1/6	1	1

In this game, the support of the discrete random variable Y_1 is given by $S_{Y_1} = \{-1, 1\}$, and the PMF is given by:

$$\begin{aligned} p_{Y_1}(-1) &= P\left(Y_1 \in \{1,3,5\}\right) = 0.5, \\ p_{Y_1}(1) &= P(Y_1 \in \{2,4,6\}) = 0.5. \end{aligned}$$

The *average* payoff is

$$-1 \times 0.5 + 1 \times 0.5 = 0.$$

Now consider a game with slightly changed rules: the payoff, denoted Y_2, is -1 if $Y_2 \leq 5$ and 1 if $Y_2 = 6$. The support is the same as before, $S_{Y_2} = \{-1, 1\}$, however, the PMF is different:

$$\begin{aligned} p_{Y_2}(-1) &= 5/6, \\ p_{Y_2}(1) &= 1/6. \end{aligned}$$

Since the positive payoff now occurs with a smaller probability, it seems reasonable to give it a smaller weight when computing the average payoff. Consider a *weighted average* payoff with weights given by the probabilities:

$$-1 \times \frac{5}{6} + 1 \times \frac{1}{6} = -\frac{2}{3}.$$

In those examples, the average payoff is a *number* that represents the corresponding game. Since the average payoff is smaller in the second game than in the first game, a simple comparison of those two numbers demonstrates that the first game is preferable to the second.

The *weighted average* value of a discrete random variable, where the weights are given by the PMF, is called the *expected value* (or *mathematical expectation*, *expectation*, *mean value*, *mean*).

Definition 1. (Expectation) Let X be a discrete random variable with the support $S_X = \{x_1, x_2, \ldots\}$ and PMF p_X. Its expected value is defined as

$$EX = \sum_{x \in S_X} xP(X = x) = \sum_{x \in S_X} xp_X(x) = x_1 p_X(x_1) + x_2 p_X(x_2) + \ldots.$$

Remark.

1. Expectation is a number representing the distribution of a random variable. Thus, it is a property of the distribution.

2. While the notation used to denote expectations is EX or $E(X)$, it is wrong to think of it as a function of the random variable X. It is rather a function (property) of the distribution of X or p_X.

3. Expectation is just one of the characteristics of the distribution of X and, thus, does not provide a complete information about the distribution. It is possible to have many different distributions with the same expectation. For example, suppose that in the example above the payoff is -2 when the outcome is odd and 2 when the outcome is even. The expected value of this distribution is also zero, however, it is a different distribution. When given a choice between this game and the original one, most people would not be indifferent. A complete description for a discrete random variable is given by the PMF, and usually we lose some (or a lot of) information by focusing on one particular property of the distribution.

4. While expectation is just one of the properties of a distribution, it nevertheless plays a fundamental role in statistics, econometrics and economics. One of the reasons for that is the so called Law of Large Numbers (LLN). According to the LLN, if we repeat an experiment many times (independently), and then average the realizations of a random variable across the experiments, the average will be close to the expected value of the

random variable. (This concept will be discussed later in the course in details). Thus, in the above examples, the expected value is approximately equal to the average payoff if the game is played many times.

More generally, consider the expectation of a *function* of a random variable. Let X be a random variable, and suppose that $Y = u(X)$, where $u(\cdot)$ is some function. Since X is random, Y is random in general. Since Y is related to X through $u(\cdot)$, the expectation of Y must be determined by the distribution of X and the function $u(\cdot)$. The relationship is established in the following theorem.

Theorem 2. *Let X be a discrete random variable with the support $S_X = \{x_1, x_2, \ldots\}$ and PMF p_X. Define $Y = u(X)$, where $u : \mathbb{R} \to \mathbb{R}$ is some function. Then,*

$$EY = \sum_{x \in S_X} u(x)p_X(x) = u(x_1)p_X(x_1) + u(x_2)p_X(x_2) + \ldots.$$

Proof. By definition, $EY = \sum_{y \in S_Y} y p_Y(y)$, where S_Y is the support of Y and p_Y is its PMF. Given $y \in S_Y$, let $\mathcal{X}(y) \subset S_X$ be the set of all values $x \in S_X$ such that $u(x) = y$:

$$\mathcal{X}(y) = \{x \in S_X : u(x) = y\}.$$

Then,

$$\begin{aligned} p_Y(y) &= P(Y = y) \\ &= P_X(\mathcal{X}(y)) \\ &= \sum_{x \in \mathcal{X}(y)} p_X(x). \end{aligned} \tag{1}$$

We have:

$$\begin{aligned}
EY &= \sum_{y\in S_Y} y p_Y(y) \\
&= \sum_{y\in S_Y} y \left(\sum_{x\in\mathcal{X}(y)} p_X(x) \right) \\
&= \sum_{y\in S_Y} \left(\sum_{x\in\mathcal{X}(y)} y p_X(x) \right) \\
&= \sum_{y\in S_Y} \left(\sum_{x\in\mathcal{X}(y)} u(x) p_X(x) \right) \\
&= \sum_{y\in S_Y} \sum_{x\in\mathcal{X}(y)} u(x) p_X(x) \\
&= \sum_{x\in S_X} u(x) p_X(x).
\end{aligned}$$

The second equality holds by (1). The third equality holds by distributing y with the terms $p_X(x)$ in the inner sum. Note that the inner sum is computed over the the elements of $\mathcal{X}(y)$ while holding y constant. The fourth equality is by the definition of $\mathcal{X}(y)$, since $\mathcal{X}(y)$ consists of all points x such that $u(x) = y$. The last equality holds because summing over all x's in $\mathcal{X}(y)$ and then summing over all y's in S_Y is equivalent to summing over all $x \in S_X$. $\square$

Example. Suppose that $S_X = \{1, 2, 3, 4, 5, 6\}$, $p_X(x) = 1/6$ for all $x \in S_X$, and let

$$u(x) = \begin{cases} -1, & x \in \{1, 3, 5\}, \\ 1, & x \in \{2, 4, 6\}. \end{cases}$$

Suppose that $Y = u(X)$. In this case, $S_Y = \{-1, 1\}$, $\mathcal{X}(-1) = \{1, 3, 5\}$, and $\mathcal{X}(1) = \{2, 4, 6\}$.

$$\begin{aligned}
EY &= (-1) \times P(Y = -1) + 1 \times P(Y = 1) \\
&= (-1) \times P_X(\mathcal{X}(-1)) + 1 \times P_X(\mathcal{X}(1)) \\
&= (-1) \times (p_X(1) + p_X(3) + p_X(5)) + 1 \times (p_X(2) + p_X(4) + p_X(6)) \\
&= (-1) \times p_X(1) + 1 \times p_X(2) + (-1) \times p_X(3) + 1 \times p_X(4) + (-1) \times p_X(5) + 1 \times p_X(6) \\
&= \sum_{x\in\{1,2,3,4,5,6\}} u(x) p_X(x).
\end{aligned}$$

2 Properties of expectation

Here we will discuss the main properties of expectation.

First, note that a constant c can be viewed as a *degenerate* random variable: its support consists of one point $S_c = \{c\}$, and the PMF is given by $p_c(c) = 1$. Hence, by the definition of expectation:

$$Ec = c \times p_c(c) = c.$$

Thus, the expected value of a constant is the constant itself.

The following property is known as the *linearity* of expectation.

Theorem 3. *Let X be a discrete random variable with the support $S_X = \{x_1, x_2, \ldots\}$ and PMF p_X. Let $Y = a + bX$, where a and b are some constants. Then,*

$$EY = E(a + bX) = a + bEX.$$

Proof. By Theorem 2,

$$\begin{aligned}
E(a+bX) &= \sum_{x \in S_X} (a + bX) p_X(x) \\
&= (a + bx_1)p_X(x_1) + (a + bx_2)p_X(x_2) + \ldots \\
&= ap_X(x_1) + bx_1 p_X(x_1) + ap_X(x_2) + bx_2 p_X(x_2) + \ldots \\
&= a(p_X(x_1) + p_X(x_2) + \ldots) + b(x_1 p_X(x_1) + x_2 p_X(x_2) + \ldots) \\
&= a \sum_{x \in S_X} p_X(x) + b \sum_{x \in S_X} x p_X(x) \\
&= a \times 1 + bEX.
\end{aligned}$$

□

Note that the result follows because the expectation is defined as a sum, and sums satisfy linearity: let $\{c_i : i = 1, 2, \ldots\}$ be a sequence of numbers, then

$$\sum_i (a + bc_i) = \sum_i a + b \sum_i c_i.$$

More generally we have the following result:

$$E\left(a + b_1 u_1(X) + b_2 u_2(X)\right) = a + b_1 E u_1(X) + b_2 E u_2(X), \tag{2}$$

where a, b_1, b_2 are constants, $u_1(\cdot)$ and $u_2(\cdot)$ are two functions, and X is a random variable. The proof of the result is analogues to the proof of Theorem 3. The result can also be extended to more than two functions.

When the support of a random variable is finite, the expectation is always a number (finite). However, when the support is infinite, the expectation may be infinite since $\sum_{x \in S_X} x p_X(x)$ is a sum of infinitely many terms. Moreover, when the support of a random variable includes both negative and positive values and infinite, we can even have an *undefined* expectation as the formula for expectation can produce $\infty - \infty$. When expectation is infinite ($+\infty$ or $-\infty$) or undefined ($\infty - \infty$), we say that *expectation does not exist.* The following example demonstrates the case when $EX = \infty$.

Example. (St. Petersburg Paradox) The problem and its solution were described by Bernoulli in an article published in *Commentaries of the Imperial Academy of Science of Saint Petersburg* in 1738. Consider a game where a coin is tossed repeatedly until "tails" is drawn. Let X denote the number of tosses, and suppose that the payoff is given by 2^X, i.e. the payoff amount doubles with every toss: the payoff is \$2 when $X = 1$, \$4 when $X = 2$, and etc. In this game, the support of X (and of the random payoff) is infinite. Note also that the probability of having a payoff equal to 2^n dollars is given by $(1/2)^n$ and goes to zero exponentially fast as $n \to \infty$. Thus, the probability of having a very large payoff is minuscule. Nevertheless, since the amount of payoff increases at the same rate as the probability decreases, the expected payoff is infinite:

$$\begin{aligned} E\left(2^X\right) &= 2 \times \frac{1}{2} + 2^2 \times \frac{1}{2^2} + 2^3 \times \frac{1}{2^3} + \dots \\ &= 1 + 1 + 1 + \dots \\ &= \infty. \end{aligned}$$

If the decision whether to participate in this game or not was based on the expected value of the payoff, one would always choose to participate regardless of the participation cost. Suppose that the price of participation is c. One would choose to participate regardless of how large c is since $c < E2^X = \infty$ for any value of c.

The situation was considered a paradox because it seemed very unreasonable that any rational person would choose to pay an arbitrary large price to participate in the game: the probability of winning a very large prize is minuscule. For example, the probability of winning more than \$256 is

$$\begin{aligned} P(2^X > 256) &= P(2^X > 2^8) \\ &= P(X > 8) \\ &= 1 - \left(\frac{1}{2^1} + \frac{1}{2^2} + \dots + \frac{1}{2^8}\right) \\ &\approx 1 - 0.9961 \\ &= 0.0039. \end{aligned}$$

As a solution to the paradox, it was suggested that rational individuals evaluate random games not according to expected payoffs, but according to the expected utility of payoffs. Moreover, the utility function must have the property of diminishing marginal utility of money:

the utility gained from each additional dollar decreases with the wealth. Under diminishing marginal utility, $u(2^X)$ grows slower than 2^X, and therefore the expected value $Eu(2^X)$ can be finite. For example, with a log utility function, one can show that $E\log(2^X) = 2\log(2) \approx 1.39$.[1]

In this example, suppose that the cost of participation in the game is c. One decides to participate in the game if the expected utility from participation exceeds the utility he obtains if he decides not to participate. Let $u(\cdot)$ is a utility function, and suppose that w is the initial wealth of an individual. The individual will participate in this game if

$$Eu(w + 2^X - c) > u(w).$$

The following example suggest an important interpretation of expected value. Consider the following game. Let X be a random variable. You must pick a number $c \in \mathbb{R}$. If the realized value of X is x, you must pay $(x - c)^2$. How should one choose c? It seems logical to choose c that minimizes the expected loss from participating in this game. Thus, the optimal value c^* is the solution to $\min_{c\in\mathbb{R}} E(X - c)^2$. We will show below that $c^* = EX$.

Theorem 4. *Let X be a random variable, and suppose that $E(X - c)^2$ is finite.*[2] *Then,*

$$EX = \arg\min_{c\in\mathbb{R}} E(X - c)^2.$$

Proof. Note that since $(x-c)^2$ is a nonlinear function, we cannot use the linearity of expectation directly: $E(X - c)^2 \neq (EX - c)^2$. However, since the function is quadratic, we can expand it first and then use the linearity. Write

$$(X - c)^2 = X^2 - 2cX + c^2.$$

Hence,

$$\begin{aligned} E(X - c)^2 &= E\left(X^2 - 2cX + c^2\right) \\ &= EX^2 - E(2cX) + Ec^2 \\ &= EX^2 - 2c(EX) + c^2, \end{aligned}$$

where the result in the second line holds by (2), and the result in the last line holds by Theorem 3 and because c^2 is a constant. We can now find the first-order condition for our minimization

[1] We will discuss later in the course how to compute this.

[2] Note that E must be the last operation performed. To emphasize that, we could write the expression as $E((X - c)^2)$.

problem:

$$\begin{aligned}\frac{d}{dc}E(X-c)^2 &= \frac{d}{dc}(EX^2 - 2c(EX) + c^2)\\ &= -2EX + 2c.\end{aligned}$$

Solving the first-order condition for c, we obtain:

$$c^* = EX.$$

Note that

$$\frac{d^2}{dc^2}E(X-c)^2 = 2 > 0,$$

and therefore c^* is the minimum. □

We can view $E(X-c)^2$ as the average distance between a point c and the distribution of X, where we use the quadratic function $(x-c)^2$ as a measure of the distance. Since $c^* = EX$ minimizes the distance to the distribution, we can interpret EX as the *center of the distribution* of X.