DISCRETE RANDOM VARIABLES

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 4 and may not be concerned over whether the actual outcome was $(1,3)$, $(2,2)$ or $(3,1)$. Also, in flipping a coin, we may be interested in the total number of heads that occur and not care at all about the actual head–tail sequence that results. These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as random variables.

Example 0.1. Suppose that our experiment consists of tossing 3 fair coins. If we let X denote the number of heads that appear, then Y is a random variable taking on one of the values 0, 1, 2, and 3 with respective probabilities

$$\begin{aligned}
\mathbb{P}(X=0) &= \mathbb{P}(\{(T,T,H)\}) = 1/8,\\
\mathbb{P}(X=1) &= \mathbb{P}(\{(H,T,T),(T,H,T),(T,T,H)\}) = 3/8,\\
\mathbb{P}(X=2) &= \mathbb{P}(\{(H,H,T),(H,T,H),(T,,H,H)\}) = 3/8,\\
\mathbb{P}(X=3) &= \mathbb{P}(\{(H,H,H)\}) = 1/8.
\end{aligned}$$

Here X is an example of a discrete random variable that can take only four values: 0, 1, 2 or 3. It is a function from the sample space

$$S = \{(T,T,H),(H,T,T),(T,H,T),(T,T,H),(H,H,T),(H,T,H),(T,,H,H),(H,H,H)\}$$

to the set of integers: X assigns to each outcome the number of heads.

In general, a random variable that can take on at most a countable number of possible values is said to be **discrete**. For a discrete random variable X, we define the **probability mass function** p of X by

$$p(x) = \mathbb{P}(X = x).$$

If $x_1, x_2, x_3, ...$ are all the values that X can take (not necessarily integers) then

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = \mathbb{P}\left(\bigcup_{i=1}^{\infty}\{X = x_i\}\right) = 1.$$

In the above example, $p(0) = p(3) = 1/8$, $p(1) = p(2) = 3/8$ and $p(0)+p(1)+p(2)+p(3) = 1$, as expected.

1. Expectation

If X is a discrete random variable with possible values $x_1, x_2, ...$ then the expectation (or the expected value) of X, denoted by $\mathbb{E}[X]$ (or simply $\mathbb{E}X$), is defined by

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \cdot \mathbb{P}(X = x_i) = \sum_{i=1}^{\infty} x_i \cdot p(x_i).$$

X	0	1	2	3
$g(X)$	0	-1	0	3
Probability mass	1/8	3/8	3/8	1/8

TABLE 1. Probability mass function of X and $g(X)$

In words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.

For the random variable defined in Example 0.1, the expected value is

$$\mathbb{E}X = \frac{1}{8} \cdot 0 + \frac{3}{8} \cdot 1 + \frac{3}{8} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{3}{2}.$$

1.1. **Expectation of a function of a random variable.** Suppose that we are given a discrete random variable X along with its probability mass function and that we want to compute the expected value of some function of X, say, $g(X) = X^2 - 2X$. How can we accomplish this? One way is as follows: Since $g(X)$ is itself a discrete random variable, it has a probability mass function, which can be determined from the probability mass function of X. Once we have determined the probability mass function of $g(X)$, we can compute $E[g(X)]$ by using the definition of expected value.

Example 1.1. Let X be the random variable defined in Example 0.1 and $g(X) = X^2 - 2X$. By plugging all the values that X can take on, we get all the values that $g(X)$ can take, together with their probabilities, as shown in Table 1. We see that $g(X)$ can take three values -1, 0 and 3 with probabilities

$$\mathbb{P}(g(X) = -1) = 3/8, \quad \mathbb{P}(g(X) = 0) = 1/8 + 3/8 = 1/2, \quad \mathbb{P}(g(X) = 3) = 1/8.$$

Therefore by definition, the expected value of $g(X)$ is

$$\mathbb{E}g(X) = 3/8 \cdot (-1) + 1/2 \cdot 0 + 1/8 \cdot 3 = 0.$$

Although the preceding procedure will always enable us to compute the expected value of any function of X from a knowledge of the probability mass function of X, there is another way of thinking about $\mathbb{E}[g(X)]$: Since $g(X)$ will equal $g(x)$ whenever X is equal to x, it seems reasonable that $\mathbb{E}[g(X)]$ should just be a weighted average of the values $g(x)$, with $g(x)$ being weighted by the probability that X is equal to x. That is, the following result is quite intuitive

$$\mathbb{E}[g(X)] = \sum_{i=1}^{\infty} g(x_i) \cdot \mathbb{P}(X = x_i) = \sum_{i=1}^{\infty} g(x_i) \cdot p(x_i),$$

where $x_1, x_2, \ldots$ are the possible values that X can take.

The advantage of this formula is that we don't have to calculate the probability mass function of $g(X)$. For example, the expected value of $g(X)$ in the previous example can be calculated as

$$\begin{aligned}\mathbb{E}g(X) &= 1/8 \cdot g(0) + 3/8 \cdot g(1) + 3/8 \cdot g(2) + 1/8 \cdot g(3)\\ &= 1/8 \cdot 0 + 3/8 \cdot (-1) + 3/8 \cdot 0 + 1/8 \cdot 3\\ &= 0.\end{aligned}$$

As a consequence of the new formula for function $g(x) = ax + b$, we have

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_{i=1}^{\infty}(ax_i + b)p(x_i)\\ &= a\sum_{i=1}^{\infty} x_i p(x_i) + b\sum_{i=1}^{\infty} p(x_i)\\ &= a\mathbb{E}[X] + b,\end{aligned}$$

where the last equality holds because $\sum_{i=1}^{\infty} p(x_i) = 1$. This property is often referred to as the **linearity of expectation**.

1.2. **Expected value of sums of random variables.** Let X, Y be two random variables, i.e. two functions $X, Y : S \to \mathbb{R}$ from the sample space S to the set of real numbers. We then define the sum $X + Y$ as a new function $X + Y : S \to \mathbb{R}$ also from S to $\mathbb{R}$ in a usual way

$$(X + Y)(s) = X(s) + Y(s), \quad s \in S.$$

Example 1.2. Suppose that the experiment consists of flipping a coin 5 times, with the outcome being the resulting sequence of heads and tails. Suppose X is the number of heads in the first 3 flips and Y is the number of heads in the final 2 flips. Let $Z = X + Y$. Then, for instance, for the outcome $s = (H, T, H, T, H)$,

$$X(s) = 2, \quad Y(s) = 1, \quad Z(s) = X(s) + Y(s) = 3.$$

Since $X + Y$ is also a random variable, we are often interested in its expected value. The following important property of expectation allows us to calculate $\mathbb{E}[X + Y]$ easily.

Proposition 1.3. *The expected value of the sum of random variables is equal to the sum of their expected values:*

$$\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y.$$

To prove this formula, recall that if X takes possible values $x_1, x_2, \ldots$ then

$$\mathbb{E}X = \sum_{k=1}^{\infty} x_k \cdot \mathbb{P}(X = x_k). \tag{1.1}$$

For our purpose, it is more convenient to use a slightly different formula. To that end, let S_k be the event that $X = x_k$, that is

$$S_k = \{s \in S : X(s) = x_k\}.$$

In other words, S_k is the set of all outcomes for which the value of X is x_k. Then

$$\begin{aligned} x_k \cdot \mathbb{P}(X = x_k) &= x_k \cdot \sum_{s \in S_k} \mathbb{P}(\{s\}) \\ &= \sum_{s \in S_k} x_k \cdot \mathbb{P}(\{s\}) \\ &= \sum_{s \in S_k} X(s) \cdot \mathbb{P}(\{s\}) \\ &= \sum_{s \in S_k} X(s) \cdot p(s), \end{aligned}$$

where $p(s) = \mathbb{P}(\{s\})$ is the probability mass function of X.

Note that $S_1, S_2, \ldots$ are mutually exclusive and exhaustive, i.e. they are disjoint and their union is equal to the whole sample space. Therefore we can rewrite $\mathbb{E}X$ as follows:

$$\begin{aligned} \mathbb{E}X &= \sum_{k=1}^{\infty} x_k \cdot \mathbb{P}(X = x_k) \\ &= \sum_{k=1}^{\infty} \sum_{s \in S_k} X(s) \cdot p(s) \\ &= \sum_{s \in \bigcup_{k=1}^{\infty} S_k} X(s) \cdot p(s) \\ &= \sum_{s \in S} X(s) \cdot p(s). \end{aligned}$$

We have shown the following property

Proposition 1.4. *If X is a random variable with probability mass function p then*

$$\mathbb{E}X = \sum_{s \in S} X(s) \cdot p(s). \tag{1.2}$$

Note that in (1.1) the formula of $\mathbb{E}X$ does not *explicitly* depend on the outcomes $s \in S$. A drawback of (1.1) is that we have to list all the possible values x_k that X can take. In contrast, (1.2) depends explicitly on outcomes $s \in S$, but we do not have to use x_k. Using (1.2), we can easily prove Proposition 1.3 as follows:

$$\begin{aligned} \mathbb{E}[X + Y] &= \sum_{s \in S} [X(s) + Y(s)] \cdot p(s) \\ &= \sum_{s \in S} X(s) \cdot p(s) + \sum_{s \in S} Y(s) \cdot p(s) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

By applying this identity repeatedly, we see that for any finite number of random variables $X_1, \ldots, X_n$, we have

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n].$$

2. Variance

2.1. **Definition.** Given a random variable X, it would be extremely useful if we were able to summarize the essential properties of X by certain suitably defined measures. One such measure would be the expected value of X. However, although $\mathbb{E}[X]$ yields the weighted average of the possible values of X, it does not tell us anything about the variation, or spread, of these values. For instance, although random variables X and Y having probability mass functions determined by

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2 \tag{2.1}$$

$$\mathbb{P}(Y = -100) = \mathbb{P}(Y = -50) = \mathbb{P}(Y = 50) = \mathbb{P}(Y = 100) = 1/4 \tag{2.2}$$

all have the same expectation $\mathbb{E}X = \mathbb{E}Y = 0$, there is a much greater spread in the possible values of Y than in those of X.

Because we expect X to take on values around its mean $\mathbb{E}[X]$, it would appear that a reasonable way of measuring the possible variation of X would be to look at how far apart X would be from its mean, on the average. One possible way to measure this variation would be to consider the quantity $\mathbb{E}[|X - \mu|]$, where $\mu = \mathbb{E}[X]$. However, it turns out to be mathematically inconvenient to deal with the absolute value function $|X - \mu|$, so a more tractable quantity is usually considered, namely the expectation of the square of the difference between X and its mean. We thus have the following definition

$$Var(X) = \mathbb{E}(X - \mu)^2.$$

The main reason the square function is used instead of the absolute value function is that we can expand the square function easily. For example, we can expand $(X-\mu)^2$ to obtain an alternative formula for the variance

$$\begin{aligned} Var(X) &= \mathbb{E}(X - \mu)^2 \\ &= \mathbb{E}(X^2 - 2\mu X + \mu^2) \\ &= \mathbb{E}X^2 - 2\mu \cdot \mathbb{E}X + \mu^2 \qquad \text{(linearity of expectation)} \\ &= \mathbb{E}X^2 - 2\mu \cdot \mu + \mu^2 \\ &= \mathbb{E}X^2 - \mu^2. \end{aligned}$$

That is, $Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

Note that there is no easy way to expand $|X-\mu|$ as a sum of certain terms depending only on X or μ. In principle, we could use $(X - \mu)^4$ instead of $(X - \mu)^2$ (why not $(X - \mu)^3$?) to define the variance, but the higher the power, the more complicated the expansion.

The square root of the variance is called the *standard deviation*.

Example 2.1. Let us calculate the variance of X and Y defined by (2.1) and (2.2). Since $\mathbb{E}X = 0$, we have

$$\begin{aligned} Var(X) &= \mathbb{E}X^2 \\ &= 1/2 \cdot (-1)^2 + 1/2 \cdot (1)^2 = 1. \end{aligned}$$

Similarly,

$$\begin{aligned} Var(Y) &= \mathbb{E}Y^2 \\ &= 1/4 \cdot (-100)^2 + 1/4(-50)^2 + 1/4 \cdot (50)^2 + 1/4 \cdot (100)^2 \\ &= 6250. \end{aligned}$$

As expected, the variance of Y is much larger than the variance of X.

2.2. **Variance of sums of independent random variables.** Let X and Y be random variables with possible values $x_1, x_2, ...$ and $y_1, y_2, ...$, respectively. We say that X and Y are independent if event $X = x_i$ and event $Y = y_j$ are independent for all $i, j = 1, 2, ...$

For independent variables, we have the following very useful property.

Proposition 2.2. *] Variance of a sum of independent random variables is equal to the sum of their variances. That is, if X and Y are independent then*

$$Var(X + Y) = Var(X) + Var(Y).$$

To verify this property, note first that we can assume $\mathbb{E}X = \mathbb{E}Y = 0$ because otherwise we can define $X' = X - \mathbb{E}X$ and $Y' = Y - \mathbb{E}Y$ which have expected value zero and (please check)

$$Var(X) = Var(X'), \quad Var(Y) = Var(Y'), \quad Var(X + Y) = Var(X' + Y').$$

Therefore in the rest of this section we assume that $\mathbb{E}X = \mathbb{E}Y = 0$. Then

$$\begin{aligned} Var(X + Y) &= \mathbb{E}(X + Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) \\ &= \mathbb{E}X^2 + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) \\ &= Var(X) + Var(Y) + 2\mathbb{E}(XY). \end{aligned}$$

It remains to show that $\mathbb{E}(XY) = 0$. Let $x_1, x_2, ...$ and $y_1, y_2, ...$ be the possible values that X and Y can take, respectively. Then the possible values that XY as a random variable can take are $x_i y_j$ for all $i, j = 1, 2, ...$ Since X and Y are independent,

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{i,j=1}^{\infty} x_i y_j \cdot \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i,j=1}^{\infty} x_i y_j \cdot \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j) \\ &= \sum_{i,j=1}^{\infty} \Big(x_i \cdot \mathbb{P}(X = x_i)\Big) \cdot \Big(y_j \cdot \mathbb{P}(Y = y_j)\Big) \\ &= \sum_{i=1}^{\infty} x_i \cdot \mathbb{P}(X = x_i) \cdot \sum_{j=1}^{\infty} y_j \cdot \mathbb{P}(Y = y_j) \\ &= \mathbb{E}X \cdot \mathbb{E}Y \\ &= 0. \end{aligned}$$

We can easily extend this result for the sum of more than two random variables. We say that $X_1, ..., X_n$ are independent random variables if events $X_1 = x_1$, $X_2 = x_2$,..., $X_n = x_n$ are independent for any choice of $x_1, x_2, ..., x_n$. In that case,

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n).$$

3. Common random variables

3.1. **Bernoulli and binomial random variables.** Suppose that an experiment, whose outcome is either a success or a failure, is performed. Let X be a random variable with $X = 1$ when the outcome is a success and $X = 0$ when it is a failure. Then the probability mass function of X is given by

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p,$$

for some $p \in [0, 1]$. We say X is a **Bernoulli** random variable with success probability p. The mean and variance of X are

$$\begin{aligned} \mathbb{E}X &= p, \\ Var(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = p(1-p). \end{aligned}$$

Suppose now that n independent experiments with success probability p are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a **binomial** random variable with parameters (n, p), denoted by $X \sim \text{Binomial}(n, p)$.

Clearly, a binomial random variable with parameter (n, p) is a sum of n independent Bernoulli random variable with the same success probability p. When $n = 1$, it is just a Bernoulli random variable.

The probability mass function of $X \sim \text{Binomial}(n, p)$ is given by

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}, \quad k = 0, 1, ..., n.$$

To verify this formula, we first note that the probability of any particular sequence of n outcomes containing k successes and $n - k$ failures is, by the assumed independence of trials, $p^k(1-p)^{n-k}$. The formula then follows, since there are n choose k different sequences of the n outcomes leading to k successes and $n - k$ failures.

Example 3.1 (Wheel of fortune). A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears k times, $k = 1, 2, 3$, then the player wins k units; if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Calculate the expected return.

Solution. If we assume that the dice are fair and act independently of each other, then the number of times that the number bet appears is a binomial random variable with parameters $(3, 1/6)$. Hence, letting X denote the player's winnings in the game, we have

$$\begin{aligned}
\mathbb{P}(X=-1) &= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(1-\frac{1}{6}\right)^3 = \frac{125}{216}, \\
\mathbb{P}(X=1) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(1-\frac{1}{6}\right)^2 = \frac{75}{216}, \\
\mathbb{P}(X=2) &= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^1 = \frac{15}{216}, \\
\mathbb{P}(X=3) &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^0 = \frac{1}{216}.
\end{aligned}$$

The expected return is then

$$\mathbb{E}X = \frac{-1\cdot 125 + 1\cdot 75 + 2\cdot 15 + 3\cdot 1}{216} = \frac{-17}{216}.$$

Hence, in the long run, the player will lose 17 units per every 216 games he plays. □

The mean and variance of X are given by

$$\mathbb{E}X = np, \quad Var(X) = np(1-p).$$

This follows from Proposition 1.3, Proposition 2.2 and the fact that X is a sum of n *independent* Bernoulli random variables with mean p and variance $p(1-p)$.

We can also verify these formulas directly as follows. First, we calculate m-moment of X for any positive integer m,

$$\mathbb{E}X^m = \sum_{k=0}^{n} k^m \binom{n}{k} \cdot p^k(1-p)^{n-k} = \sum_{k=1}^{n} k^m \binom{n}{k} \cdot p^k(1-p)^{n-k}.$$

Since

$$k\cdot\binom{n}{k} = k\cdot\frac{n!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = n\cdot\frac{(n-1)!}{(k-1)!(n-k)!} = n\cdot\binom{n-1}{k-1},$$

it follows that

$$\begin{aligned}
\mathbb{E}X^m &= \sum_{k=1}^{n} k^{m-1} n\cdot\binom{n-1}{k-1}\cdot p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} k^{m-1}\cdot\binom{n-1}{k-1}\cdot p^{k-1}(1-p)^{n-k} \\
&= np\sum_{\ell=0}^{n-1}(\ell+1)^{m-1}\cdot\binom{n-1}{\ell}\cdot p^{\ell}(1-p)^{n-1-\ell} \quad \text{by letting } \ell = k-1 \\
&= np\cdot\mathbb{E}(Y+1)^{m-1},
\end{aligned}$$

where $Y \sim \text{Binomial}(n-1, p)$. Setting $m = 1$ yields

$$\mathbb{E}X = np\cdot\mathbb{E}(Y+1)^0 = np.$$

That is, the expected number of successes that occur in n independent trials when each is a success with probability p is equal to np. Setting $m = 2$ in the preceding equation and using linearity of expectation,

$$\mathbb{E}X^2 = np \cdot \mathbb{E}(Y+1) = np[\mathbb{E}Y + 1] = np[(n-1)p + 1].$$

Therefore

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np[(n-1)p+1] - (np)^2 = np(1-p).$$

This example shows the usefulness of Proposition 1.3 and Proposition 2.2 as they allow us avoid all this complicated calculation.

3.2. **Poisson random variable.** A random variable X that takes on one of the values 0,1,2,... is said to be a Poisson random variable with parameter $\lambda > 0$ if

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, ...$$

The Poisson random variable has a tremendous range of applications in diverse areas because it may be used as an approximation for a binomial random variable or model many random quantities of interest (see Section 4.7 of the textbook).

Expected value and variance of X are equal to the parameter λ:

$$\mathbb{E}X = Var(X) = \lambda.$$

We first verify the expectation:

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k-1}}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} \quad \text{by letting } \ell = k-1 \\
&= \lambda e^{-\lambda} \cdot e^{\lambda} \\
&= \lambda.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}X^2 &= \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{\ell=0}^{\infty} (\ell+1) \cdot \frac{\lambda^{\ell}}{\ell!} \quad \text{by letting } \ell = k-1 \\
&= \lambda e^{-\lambda} \cdot \left(\sum_{\ell=0}^{\infty} \ell \cdot \frac{\lambda^{\ell}}{\ell!} + \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} \right) \\
&= \lambda \left(\sum_{\ell=0}^{\infty} \ell \cdot e^{-\lambda} \cdot \frac{\lambda^{\ell}}{\ell!} + e^{-\lambda} \cdot \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} \right) \\
&= \lambda(\mathbb{E}X + 1) \\
&= \lambda(\lambda + 1).
\end{aligned}$$

Therefore

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.$$