

### 3.9: Random Variables, Expected Values, and Population Sets

When we sample a particular distribution, the value that we obtain depends on chance and on the nature of the distribution described by the function  $f(u)$ . The probability that any given trial will produce  $u$  in the interval  $a < u < b$  is equal to  $f(b) - f(a)$ . We often find situations in which a second function of  $u$ , call it  $g(u)$ , is also of interest. If we sample the distribution and obtain a value of the random variable,  $u_k$ , then the value of  $g$  associated with that trial is  $g(u_k)$ . The question arises: Given  $g(u)$  and the distribution function  $f(u)$ , what should we expect the value of  $g(u_k)$  to be? That is, if we get a value of  $u$  from the distribution and then find  $g(u)$ , what value should we expect to find for  $g(u)$ ? While this seems like a reasonable question, it is obvious that we can give a meaningful answer only when we can define more precisely just what we mean by “expect.”

To understand our definition of the **expected value** (sometimes called the **expectation value**) of  $g(u)$ , let us consider a game of chance. Suppose that we have a needle that rotates freely on a central axis. When spun, the needle describes a circular path, and its point eventually comes to rest at some point on this path. The location at which the needle stops is completely random. Imagine that we divide the circular path into six equal segments, which we number from one to six. When we spin the needle, it is equally likely to stop over any of these segments. Now, let us suppose that we conduct a lottery by selling six tickets, also numbered from one to six. We decide the winner of the lottery by spinning the needle. The holder of the ticket whose number matches the number on which the needle stops receives a payoff of \$6000. After the spin, one ticket is worth \$6000, and the other five are valueless. We ask: Before the spin, what is any one of the lottery tickets worth?

In this context, it is reasonable to define the expected value of a ticket as the amount that we should be willing to pay to buy a ticket. If we buy them all, we receive \$6000 when the winning ticket is selected. If we pay \$1000 per ticket to buy them all, we get our money back. If we buy all the tickets, the expected value of each ticket is \$1000. What if we buy only one ticket? Is it reasonable to continue to say that its expected value is \$1000? We argue that it is. One argument is that the expected value of a ticket should not depend on who owns the ticket; so, it should not depend on whether we buy one, two, or all of them. A more general argument supposes that repeated lotteries are held under the same rules. If we spend \$1000 to buy one ticket in each of a very large number of such lotteries, we expect that we will eventually “break even.” Since the needle comes to rest at each number with equal probability, we reason that

$$\begin{aligned}\text{Expected value of a ticket} \\ &= \$6000 \text{ (fraction of times our ticket would be selected)} \\ &= \$6000 (1/6) \\ &= \$1000\end{aligned}$$

Since we assume that the fraction of times our ticket would be selected in a long series of identical lotteries is the same thing as the probability that our ticket will be selected in any given drawing, we can also express the expected value as

$$\begin{aligned}\text{Expected value of a ticket} \\ &= \$6000 \text{ (probability that our ticket will be selected)} \\ &= \$6000 (1/6) \\ &= \$1000\end{aligned}$$

Clearly, the ticket is superfluous. The game depends on obtaining a value of a random variable from a distribution. The distribution is a spin of the needle. The random variable is the location at which the needle comes to rest. We can conduct essentially the same game by allowing any number of participants to bet that the needle will come to rest on any of the six equally probable segments of the circle. If an individual repeatedly bets on the same segment in many repetitions of this game, the total of his winnings eventually matches the total amount that he has wagered. (More precisely, the total of his winnings divided by the total amount he has wagered becomes arbitrarily close to one.)

Suppose now that we change the rules. Under the new rules, we designate segment 1 of the circle as the payoff segment. Participants pay a fixed sum to be eligible for the payoff for a particular game. Each game is decided by a spin of the needle. If the needle lands in segment 1, everyone who paid to participate in that game receives \$6000. Evidently, the new rules have no effect on the value of participation. Over the long haul, a participant in a large number of games wins \$6000 in one-sixth of these games. We take this to be equivalent to saying that he has a probability of one-sixth of winning \$6000 in a given game in which he participates. His expected payoff is

$$\begin{aligned}
 &\text{Expected value of game} \\
 &= \$6000 \text{ (probability of winning \$6000)} \\
 &= \$6000 (1/6) \\
 &= \$1000
 \end{aligned}$$

Let us change the game again. We sub-divide segment 2 into equal-size segments 2A and 2B. The probability that the needle lands in 2A or 2B is 1/12. In this new game, the payoff is \$6000 when the needle lands in either segment 1 or segment 2A. We can use any of the arguments that we have made previously to see that the expected payoff game is now  $\$6000 (1/4) = \$1500$ . However, the analysis that is most readily generalized recognizes that the payoff from this game is just the sum of the payout from the previous game plus the payout from a game in which the sole payout is \$6000 whenever the needle lands in segment 2A. For the new game, we have

$$\begin{aligned}
 &\text{Expected value of a game} \\
 &= \$6000 \times P(\text{segment 1}) + \$6000 \times P(\text{segment 2A}) \\
 &= \$6000 (1/6) + \$6000 (1/12) \\
 &= \$1500
 \end{aligned}$$

We can devise any number of new games by dividing the needle's circular path into  $\Omega$  non-overlapping segments. Each segment is a possible outcome. We number the possible outcomes 1, 2, ...,  $i$ , ...,  $\Omega$ , label these outcomes  $u_1, u_2, \dots, u_i, \dots, u_\Omega$ , and denote their probabilities as  $P(u_1), P(u_2), \dots, P(u_i), \dots, P(u_\Omega)$ . We say that the probability of outcome  $u_i$ ,  $P(u_i)$ , is the **expected frequency** of outcome  $u_i$ . We denote the respective payoffs as  $g(u_1), g(u_2), \dots, g(u_i), \dots, g(u_\Omega)$ . Straightforward generalization of our last analysis shows that the expected value for participation in any game of this type is

$$\sum_{i=1}^{\Omega} g(u_i) \times P(u_i)$$

Moreover, the spinner is representative of any distribution, so it is reasonable to generalize further. We can say that the expected value of the outcome of a single trial is always the probability-weighted sum, over all possible outcomes, of the value of each outcome. A common notation uses angular brackets to denote the expected value for a function of the random variable; the expected value of  $g(u)$  is  $\langle g(u) \rangle$ . For a discrete distribution with  $\Omega$  exhaustive mutually-exclusive outcomes  $u_i$ , probabilities  $P(u_i)$ , and outcome values (payoffs)  $g(u_i)$ , we define the **expected value** expected value of  $g(u)$  to be

$$\langle g(u) \rangle = \sum_{i=1}^{\Omega} g(u_i) \times P(u_i)$$

Now, let us examine the expected value of  $g(u)$  from a slightly different perspective. Let the number of times that each of the various outcomes is observed in a particular sample of  $N$  observations be  $N_1, N_2, \dots, N_3, \dots, N_\Omega$ . We have  $N = N_1 + N_2 + \dots + N_i + \dots + N_\Omega$ . The set  $\{N_1, N_2, \dots, N_i, \dots, N_\Omega\}$  specifies the way that the possible outcomes are populated in this particular series of  $N$  observations. We call  $\{N_1, N_2, \dots, N_i, \dots, N_\Omega\}$  a **population set**. If we make a second series of  $N$  observations, we obtain a second population set. We infer that the best forecast we can make for the number of occurrences of outcome  $u_i$  in any future series of  $N$  observations is  $N \times P(u_i)$ . We call  $N \times P(u_i)$  the **expected number** of observations of outcome  $u_i$  in a sample of size  $N$ .

In a particular series of  $N$  trials, the number of occurrences of outcome  $u_i$ , and hence of  $g(u_i)$ , is  $N_i$ . For the set of outcomes  $\{N_1, N_2, \dots, N_3, \dots, N_\Omega\}$ , the average value of  $g(u)$  is

$$\overline{g(u)} = \frac{1}{N} \sum_{i=1}^{\Omega} g(u_i) \times N_i$$

Collecting a second sample of  $N$  observations produces a second estimate of  $\overline{g(u)}$ . If  $N$  is small, successive estimates of  $\overline{g(u)}$  may differ significantly from one another. If we make a series of  $N$  observations multiple times, we obtain multiple population sets. In general, the population set from one series of  $N$  observations is different from the population set for a second series of  $N$  observations. If  $N \gg \Omega$ , collecting such samples of  $N$  a sufficiently large number of times must produce some population sets more than once, and among those that are observed more than once, one must occur more often than any other. We call it the **most probable** population set. Let the elements of the most probable population set be  $\{N_1, N_2, \dots, N_i, \dots, N_\Omega\}$ . We infer that the

most probable population set is the best forecast we can make about the outcomes of any future sample of  $N$  from this distribution. Moreover, we infer that the best estimate we can make of  $N_i$  is that it equals the expected number of observations of outcome  $u_i$ ; that is,

$$N_i \approx N \times P(u_i)$$

Now,  $N_i$  and  $N_i$  must be natural numbers, while  $N \times P(u_i)$  need only be real. In particular, we can have 0, but  $N_i$  must be 0 or 1 (or some higher integer). This is a situation of practical importance, because circumstances may limit the sample size to a number,  $N$ , that is much less than the number of possible outcomes,  $\Omega$ . (We encounter this situation in our discussion of statistical thermodynamics in Chapter 21. We find that the number of molecules in a system can be much smaller than the number of outcomes—observable energy levels—available to any given molecule.)

If many more than  $N$  outcomes have about the same probability, repeated collection of samples of  $N$  observations can produce a series of population sets (each population set different from all of the others) in each of which every element is either zero or one. When this occurs, it may be that no single population set is significantly more probable than any of many others. Nevertheless, every outcome occurs with a well-defined probability. We infer that the set  $\{N \times P(u_1), N \times P(u_2), \dots, N \times P(u_i), \dots, N \times P(u_\Omega)\}$  is always an adequate proxy for calculating the expected value for the most probable population set.

To illustrate this kind of distribution, suppose that there are 3000 possible outcomes, of which the first and last thousand have probabilities that are so low that they can be taken as zero, while the middle 1000 outcomes have approximately equal probabilities. Then  $P(u_i) \approx 0$  for  $1 < i < 1000$  and  $2001 < i < 3000$ , while  $P(u_i) \approx 10^{-3}$  for  $1001 < i < 2000$ . We are illustrating the situation in which the number of outcomes we can observe,  $N$ , is much less than the number of outcomes that have appreciable probability, which is 1000. So let us take the number of trials to be  $N = 4$ . If the value of  $g(u)$  for each of the 1000 middle outcomes is the same, say  $g(u_i) = 100$  for  $1001 < i < 2000$ , then our calculation of the expected value of  $g(u)$  will be

$$\langle g(u) \rangle = \frac{1}{4} \sum_{i=1}^{3000} g(u_i) \times N \times P(u_i) = \frac{1}{4} \sum_{i=1001}^{2000} 100 \times N_i = \frac{400}{4} = 100$$

regardless of which population set results from the four trials. That is, because all of the populations sets that have a significant chance to be observed have  $N_i = 1$  and  $g(u_i) = 100$  for exactly four values of  $i$  in the range  $1001 < i < 2000$ , all of the population sets that have a significant chance to be observed give rise to the same expected value.

Let us compute the arithmetic average,  $\overline{g(u)}$ , using the most probable population set for a sample of  $N$  trials. In this case, the number of observations of the outcome  $u_i$  is  $N_i = N \times P(u_i)$ .

$$\overline{g(u)} = \frac{1}{N} \sum_{i=1}^{\Omega} g(u_i) \times N_i = \frac{1}{N} \sum_{i=1}^{\Omega} g(u_i) \times N \times P(u_i) = \sum_{i=1}^{\Omega} g(u_i) \times P(u_i) = \langle g(u) \rangle$$

For a discrete distribution,  $\langle g(u) \rangle$  is the value of  $\overline{g(u)}$  that we calculate from the most probable population set,  $\{N_1, N_2, \dots, N_i, \dots, N_\Omega\}$ , or its proxy  $\{N \times P(u_1), N \times P(u_2), \dots, N \times P(u_i), \dots, N \times P(u_\Omega)\}$ .

We can extend the definition of the expected value,  $\langle g(u) \rangle$ , to cases in which the cumulative probability distribution function,  $f(u)$ , and the outcome-value function,  $g(u)$ , are continuous in the domain of the random variable,  $u_{\min} \leq u \leq u_{\max}$ . To do so, we divide this domain into a finite number,  $\Omega$ , of intervals,  $\Delta u_i$ . We let  $u_i$  be the lower limit of  $u$  in the interval  $\Delta u_i$ . Then the probability that a given trial yields a value of the random variable in the interval  $\Delta u_i$  is  $P(\Delta u_i) = f(u_i + \Delta u_i) - f(u_i)$ , and we can approximate the expected value of  $g(u)$  for the continuous distribution by the finite sum

$$\langle g(u) \rangle = \sum_{i=1}^{\Omega} g(u_i) \times P(\Delta u_i) = \sum_{i=1}^{\Omega} g(u_i) \times [f(u_i + \Delta u_i) - f(u_i)] = \sum_{i=1}^{\Omega} g(u_i) \times \left[ \frac{f(u_i + \Delta u_i) - f(u_i)}{\Delta u_i} \right] \times \Delta u_i$$

In the limit as  $\Omega$  becomes arbitrarily large and all of the intervals  $\Delta u_i$  become arbitrarily small, the expected value of  $g(u)$  for a continuous distribution becomes

$$\langle g(u) \rangle = \int_{-\infty}^{\infty} g(u) \left[ \frac{df(u)}{du} \right] du$$

This integral is the value of  $\langle g(u) \rangle$ , where  $df(u)/du$  is the probability density function for the distribution. If  $c$  is a constant, we have

$$\langle g(cu) \rangle = c \langle g(u) \rangle$$

If  $h(u)$  is a second function of the random variable, we have

$$\langle g(u) + h(u) \rangle = \langle g(u) \rangle + \langle h(u) \rangle$$

---

This page titled [3.9: Random Variables, Expected Values, and Population Sets](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Paul Ellgen](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.