Chapter 4
Discrete Probability Distribution

Objectives

After completing this chapter, you will be able to:

- The properties of a probability distribution
- To compute the expected value and variance of a probability distribution
- To calculate the covariance and understand its use in finance
- To compute probabilities from the binomial, Poisson, and hypergeometric distributions
- How the binomial, Poisson, and hypergeometric distributions can be used to solve business problems

INTRODUCTION

A random variable can be either discrete or continuous. In this module, we cover the first type, and the module Continuous probability distributions covers the second.

The idea of a random variable builds on the fundamental ideas of probability. Students need to understand that random variables are conceptually different from the mathematical variables that they have met before. A random variable is linked to observations in the real world, where uncertainty is involved.

An informal — but important — understanding of a random variable is that it is a variable whose numerical value is determined by the outcome of a random procedure. In this module, we also see the more formal understanding, which is that a random variable is a function from the event space of a random procedure to the set of real numbers.

Random variables are central to the use of probability in practice. They are used to model random phenomena, which mean that they are relevant to a wide range of human activity. In particular, they are used extensively in many branches of research, including agriculture, biology, ecology, economics, medicine, meteorology, physics, psychology and others. They provide a structure for making inferences about the world, when it is impossible to measure things comprehensively. They are used to model outcomes of processes that cannot be predicted in advance.

Random variables have distributions. In this module, we describe the essential properties of distributions of discrete random variables. Distributions can have many forms, but there are some special types of distributions that arise in many different practical contexts. In this module, we discuss two of these special distributions: discrete uniform distributions and geometric distributions.

This module also covers the mean of a discrete random variable, which is a measure of central location, and the variance and standard deviation, which are measures of spread.

Random variables

A random variable is a variable whose value is determined by the outcome of a random procedure. The concept of a random procedure was discussed in the module *Probability*. What makes the variable *random* is that — unlike the kind of variable we see in a quadratic equation — we cannot say what the observed value of the random variable is until we actually carry out the random procedure.

a. A **discrete random variable** takes values confined to a range of separate or `discrete' values. (More formally, a discrete random variable takes either a finite number of values or a countably infinite number of values.)
b. A **continuous random variable** can take any value in an interval.

This module concerns discrete random variables. The module Continuous probability distributions deal with continuous random variables. It is important to see that the definition of a random variable needs a specification of what is observed or recorded. In some situations, this is essentially implied; but it is always required implicitly or explicitly.

Discrete random variables: general ideas

We start with the classic example of rolling a fair six-sided die. This scenario has already been discussed extensively in the module Probability. You may wonder why introductory discussions of probability and random variables often start with examples involving dice, cards and coins. The reason is that these are concrete randomising devices with which we are familiar, and for which we are willing to entertain plausible assumptions about the idealised model.

As we have seen in the module Probability, for coin tossing in particular, this is not as straightforward as it seems. When considering the actual use of dice, cards and coins, the potential for other factors that will invalidate the model should always be kept in mind. Probability and statistics give us the framework to think about this rationally.

But for the initial example, we assume that the model is correct: the die is symmetric in every respect and is shaken well before it is rolled. The uppermost face is observed. The event space for this random procedure is $E = \{1, 2, 3, 4, 5, 6\}$, and the possible outcomes are equiprobable.

Define the discrete random variable X to be the number on the uppermost face when the die comes to rest. In this case, there is a distinct value of the random variable for each distinct possible outcome of the random procedure. X can take the values 1, 2, 3, 4, 5, 6, and no other values. It takes discrete values, integers in this case — the die cannot end up with an outcome of π or 1.01 — so X is a discrete random variable. Since X can only take six possible values, it is a simple random variable.

The probability that a discrete random variable X takes the value x is denoted Pr(X = x). We read this as 'the probability that X equals x', which means the probability that X takes the value x when we actually obtain an observation. For the die-rolling example, for each real number x. Often, for discrete random variables, it is sufficient to specify in some way the values with non-zero probability only; the values with zero probability are usually clear, or clearly implied.

$$\Pr(X = x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{otherwise,} \end{cases}$$

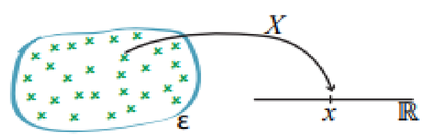
There are other simple random variables that can be defined for the random procedure of rolling a die. For example:

- Let Y be the number of even numbers appearing. Then Y takes value 1 if a 2, 4 or 6 is rolled, and Y takes value 0 otherwise.
- Let Z be the number of prime numbers appearing. Here Z takes value 1 if a 2, 3 or 5 is rolled, and takes value 0 otherwise.

These two examples are not terribly interesting; but they illustrate the important point that a single random procedure can accommodate several random variables.
Note that "Y = 1" and "Z = 0" are events, in that they define subsets of the event space E . The event "Y = 1" is {2, 4, 6}. This is a crucial insight; it makes it feasible to obtain the probability distribution of a random variable.

This leads to the definition of a random variable from a formal perspective. Students accustomed to formal mathematical treatments of topics sometimes find the description of a random variable given so far somewhat elusive. A random variable can be defined formally in a way that strongly relates to mathematical topics that students have covered elsewhere, specifically, functions.

A random variable is a numerical-valued function that maps the event space E to the set of real numbers. Students will not be familiar with a 'variable' that is a function. It is an important conceptual point, represented in the following diagram.



A random variable as a function
from the event space $\mathcal{E}$ to the set $\mathbb{R}$ of real numbers.

From this diagram, we see that the random variable must take exactly one value for each element of the event space E . So each possible outcome in the event space has a corresponding value for the random variable. As with functions generally, a number of possible outcomes in E may have the same value of the random variable, and in practice this occurs frequently.

Probability functions

To work out the probability that a discrete random variable X takes a particular value x, we need to identify the event (the set of possible outcomes) that corresponds to "X = x". In general, the function used to describe the probability distribution of a discrete random variable is called its probability function (abbreviated as pf). The probability function of X is the function pX : R → [0, 1] given by pX (x) = Pr(X = x).

In general, the probability function pX (x) may be specified in a variety of ways. One way is to specify a numerical value for each possible value of X; we have done that for the die-rolling example. In the die-rolling example, the random variable X can take exactly six values, and no others, and we assert that the probability that X takes any one of these values is the same, namely 1/6. As is the case generally for functions, the lower-case x here is merely the argument of the function. If we write Pr(X = y) it is essentially the same function, just as f (x) = 2x 2+3x−1 and f (y) = 2y 2+3y −1 are the same function. But it helps to associate the corresponding lower-case letter with the random variable we are considering.

Another way to specify the probability function is using a formula. We will see examples of this in the next section.

Less formally, the probability distribution may be represented using a graph, with a spike of height pX (x) at each possible value x of X. If there are too many possible values of X for this visual representation to work, we may choose to omit probabilities that are very close to zero; such values will typically be invisible on a graph anyway.

Discrete uniform distribution

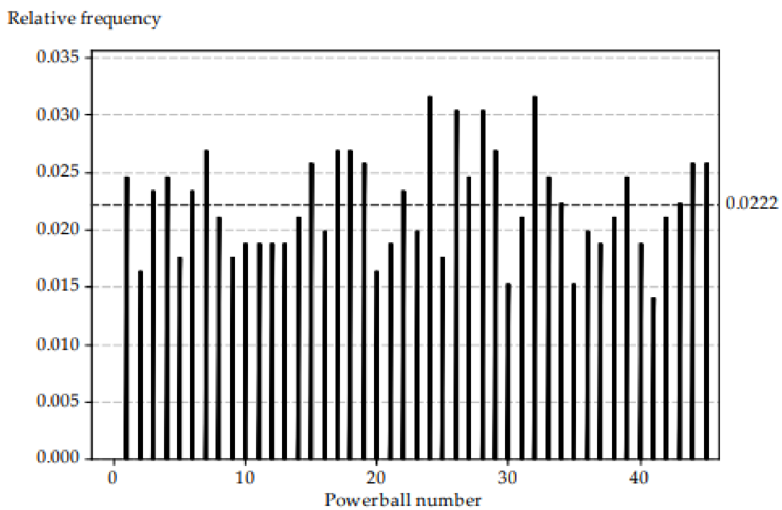
In the previous two sections, we considered the random variable defined as the number on the uppermost face when a fair die is rolled. This is an example of a random variable with a discrete uniform distribution. In general, for a positive integer m, let X be a discrete random variable with pf pX (x) given by

$$p_X(x) = \Pr(X = x) = \begin{cases} \frac{1}{m} & \text{if } x = 1, 2, \dots, m, \\ 0 & \text{otherwise.} \end{cases}$$

Then X has a discrete uniform distribution. This is a distribution that arises often in lotteries and games of chance. We have seen this distribution in the Powerball example considered in the module Probability. In commercial lotteries, such as Powerball, it is a regulatory requirement that each outcome is equally likely. There are 45 possible Powerball numbers (1, 2,..., 45). So if X is the Powerball drawn on a particular occasion, the pf pX (x) of X is given by

$$p_X(x) = \Pr(X = x) = \begin{cases} \frac{1}{45} & \text{if } x = 1, 2, \dots, 45, \\ 0 & \text{otherwise.} \end{cases}$$

If this model for the drawing of Powerball numbers is correct, we should expect that, over a large number of draws, the relative frequencies of the 45 possible numbers are all approximately equal to 1 45 ≈ 0.0222. The following graph shows the relative frequencies observed in the 853 draws from May 1996 to September 2012.



Relative frequencies of Powerball numbers over 853 draws.
The model probability of 0.0222 is shown as a reference line.

Geometric distribution

We have already met one special distribution that is given a name (the discrete uniform distribution). The distribution in the previous example also has a name; it is called a geometric distribution. Suppose that a sequence of independent 'trials' occur, and at each trial the probability of 'success' equals p. Define X to be the number of trials that occur before the first success is observed. Then X has a geometric distribution with parameter p. We introduce here a symbol used throughout the modules on probability and statistics. If X has a geometric distribution with parameter p, we write X d = G(p). The symbol d = stands for 'has the distribution', meaning the distribution indicated immediately to the right of the symbol.

Note the use of the generic terms 'trial' and 'success'. They are arbitrary, but they carry with them the idea of each observation involving a kind of test (i.e., trial) in which we ask the question: Which one of the two possibilities will be observed, a 'success' or a 'failure'? In this sense, the words 'success' and 'failure' are just labels to keep track of the two possibilities for each trial.

Note that X can take the value 0, if a success is observed at the very first trial. Or it can take the value 1, if a failure is observed at the first trial and then a success at the second trial. And so on. What is the largest value that X can take? There is no upper limit, in theory. As the values of X increase, the probabilities become smaller and smaller. Another important distribution arises in the context of a sequence of independent trials that each have the same probability of success p. This is the binomial distribution. There is an entire module devoted to it (Binomial distribution), so we do not consider it further here.

Mean of a discrete random variable

If you roll a fair die many times, what will be the average outcome? Imagine rolling it 6000 times. You would expect to roll about 1000 ones, 1000 twos, and so on: about

1000 occurrences of each possible outcome. What would be the average value of the outcomes obtained? Approximately, the average or mean would be

$$\frac{(1000 \times 1) + (1000 \times 2) + \cdots + (1000 \times 6)}{6000} = \frac{21\,000}{6000} = 3.5.$$

This can be thought of as the weighted average of the six possible values 1, 2,..., 6, with weights given by the relative frequencies. Note that 3.5 is not a value that we can actually observe. By analogy with data and relative frequencies, we can define the mean of a discrete random variable using probabilities from its distribution, as follows. The mean µX of a discrete random variable X with probability function *pX* (x) is given by

$$\mu_X = \sum x\, p_X(x),$$

where the sum is taken over all values x for which pX (x) > 0.

The mean can be regarded as a measure of 'central location' of a random variable. It is the weighted average of the values that X can take, with weights provided by the probability distribution.

The mean is also sometimes called the expected value or expectation of X and denoted by E(X). These are both somewhat curious terms to use; it is important to understand that they refer to the long-run average. The mean is the value that we expect the long-run average to approach. It is not the value of X that we expect to observe.

Consider a random variable U that has the discrete uniform distribution with possible values 1, 2,...,m. The mean is given by

$$\begin{aligned} \mu_U &= \sum_{x=1}^{m}\left(x \times \frac{1}{m}\right) \\ &= \frac{1}{m}\sum_{x=1}^{m} x \\ &= \frac{1}{m} \times \frac{m(m+1)}{2} \\ &= \frac{m+1}{2}. \end{aligned}$$

Variance of a discrete random variable

We have seen that the mean of a random variable X is a measure of the central location of the distribution of X. If we are summarising features of the distribution of X, it is clear that location is not the only relevant feature. The second most important feature is the spread of the distribution.

If values of X near its mean µX are very likely and values further away from µX have very small probability, then the distribution of X will be closely concentrated around µX . In this case, the spread of the distribution of X is small. On the other hand, if values of X some distance from its mean µX are likely, the spread of the distribution of X will be large.

These ideas lead to the most important measure of spread, the **variance**, and a closely related measure, the **standard deviation.**

Students have met the concepts of variance and standard deviation when summarising data. These were the sample variance and the sample standard deviation. The difference here is that we are referring to properties of the distribution of a random variable.

The variance of a random variable X is defined by

$$\operatorname{var}(X) = \mathrm{E}[(X-\mu)^2], \qquad \text{where } \mu = \mathrm{E}(X).$$

For a discrete random variable X, the variance of X is obtained as follows:

$$\operatorname{var}(X) = \sum (x-\mu)^2 p_X(x),$$

where the sum is taken over all values of x for which pX (x) > 0.

So the variance of X is the weighted average of the squared deviations from the mean µ, where the weights are given by the probability function pX (x) of X.

The standard deviation of X is defined to be the square root of the variance of X. That is

$$\operatorname{sd}(X) = \sigma_X = \sqrt{\operatorname{var}(X)}.$$

Because of this definition, the variance of X is often denoted by a^3x

In some ways, the standard deviation is the more tangible of the two measures, since it is in the same units as X. For example, if X is a random variable measuring lengths in metres, then the standard deviation is in metres (m), while the variance is in square metres (m^2).

Unlike the mean, there is no simple direct interpretation of the variance or standard deviation. The variance is analogous to the moment of inertia in physics, but that is not necessarily widely understood by students. What is important to understand is that, in relative terms:

- a small standard deviation (or variance) means that the distribution of the random variable is narrowly concentrated around the mean
- a large standard deviation (or variance) means that the distribution is spread out, with some chance of observing values at some distance from the mean.

Note that the variance cannot be negative, because it is an average of squared quantities. This is appropriate, as a negative spread for a distribution does not make sense. Hence, var(X) ≥ 0 and sd(X) ≥ 0 always.

Video Links:

Discrete Probability Distribution

- https://www.khanacademy.org/math/ap-statistics/random-variables-ap/discrete-random-variables/v/discrete-probability-distribution
- https://courses.lumenlearning.com/introstats1/chapter/probability-distribution-function-pdf-for-a-discrete-random-variable/
- https://study.com/academy/lesson/discrete-probability-distributions-equations-examples.html
- https://faculty.elgin.edu/dkernler/statistics/ch06/6-1.html

References

- https://amsi.org.au/ESA_Senior_Years/PDF/DiscreteProbability4c.pdf
- https://courses.lumenlearning.com/introstats1/chapter/probability-distribution-function-pdf-for-a-discrete-random-variable/
- https://www.coconino.edu/resources/files/pdfs/academics/sabbatical-reports/kate-kozak/chapter_5.pdf