B

Probability Concepts Used in Sampling

I recollect nothing that passed that day, except Johnson's quickness, who, when Dr. Beattie observed, as something remarkable which had happened to him, that he had chanced to see both No. 1, and No. 1000, of the hackney-coaches, the first and the last; "Why, Sir, (said Johnson,) there is an equal chance for one's seeing those two numbers as any other two." He was clearly right; yet the seeing of the two extremes, each of which is in some degree more conspicuous than the rest, could not but strike one in a stronger manner than the sight of any other two numbers."

—James Boswell, *The Life of Samuel Johnson*

The essence of probability sampling is that we can calculate the probability with which any subset of observations in the population will be selected as the sample. Most of the randomization theory results used in this book depend on probability concepts for their proof. In this appendix we present a brief review of some of the basic ideas used. The reader should consult a more comprehensive reference on probability, such as Ross (1998) or Durrett (1994), for more detail and for derivations and proofs.

Because all work in randomization theory concerns discrete random variables, only results for discrete random variables are given in this section. We use the results in Sections B.1–B.3 in Chapters 2–4, and the results in Section B.4 in Chapters 5 and 6.

B.1 Probability

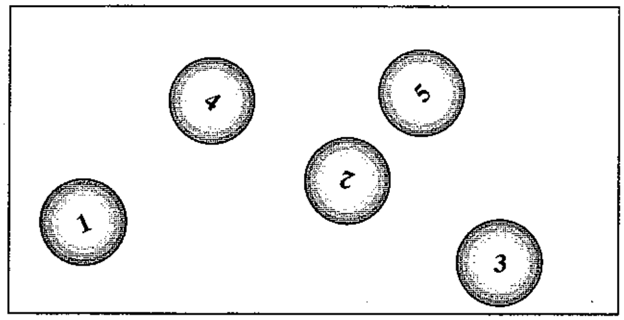
Consider performing an experiment in which you can write out all outcomes that could possibly happen, but you do not know exactly which one of those outcomes will occur. You might flip a coin, draw a card from a deck, or pick three names out of a hat containing 20 names. Probabilities are assigned to the different outcomes and to sets composed of outcomes (called **events**), in accordance with the likelihood that the events will occur. Let Ω be the **sample space**, the list of all possible outcomes. For flipping a coin, $\Omega = \{\text{heads, tails}\}$. Probabilities in finite sample spaces have three basic properties:

1 $P(\Omega) = 1$.

2 For any event A, $0 \le P(A) \le 1$.

3 If the events $A_1, \ldots, A_k$ are disjoint, then $P(\cup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i)$.

In sampling, we have a population of N units and use a probability sampling scheme to select n of those units. We can think of those N units as balls labeled 1 through N in a box, and we draw n balls from the box. For illustration, suppose $N = 5$ and $n = 2$. Then we draw two labeled balls out of the box:



If we take a simple random sample (SRS) of one ball, each ball has an equal probability $1/N$ of being chosen as the sample.

B.1.1 Simple Random Sampling with Replacement

In sampling with replacement, we put a ball back after it is chosen, so the same population is used on successive draws from the population. For the box with $N = 5$, there are 25 possible samples (a, b) in Ω, where a represents the first ball chosen and b represents the second ball chosen:

(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)
(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)
(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)
(1, 4)	(2, 4)	(3, 4)	(4, 4)	(5, 4)
(1, 5)	(2, 5)	(3, 5)	(4, 5)	(5, 5)

Since we are taking a random sample, each of the possible samples has the same probability, 1/25, of being the one chosen. When we take a sample, though, we usually do not care whether we chose unit 4 first and unit 5 second, or the other way around. Instead, we are interested in the probability that our sample consists of 4 and 5 in either order, which we write as $\mathcal{S} = \{4, 5\}$. By the third property in the definition of a probability,

$$P(\{4, 5\}) = P[(4, 5) \cup (5, 4)] = P[(4, 5)] + P[(5, 4)] = \frac{2}{25}.$$

Suppose we want to find P(unit 2 is in the sample). We can either count that nine of the outcomes above contain 2, so the probability is 9/25, or we can use the **addition formula:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{B.1}$$

Here, let A = {unit 2 is chosen on the first draw} and let B = {unit 2 is chosen on the second draw}.

$$\begin{aligned} P(\text{unit 2 is in the sample}) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{5} + \frac{1}{5} - \frac{1}{25} = \frac{9}{25}. \end{aligned}$$

Note that, for this example,

$$P(A \cap B) = P(A) \times P(B).$$

That occurs in this situation because events A and B are **independent**—that is, whatever happens on the first draw has no effect on the probabilities of what will happen on the second draw. Independence of the draws occurs in finite population sampling only when we sample with replacement.

B.1.2 Simple Random Sampling Without Replacement

Most of the time, we sample without replacement because it is more efficient—if Heather is already in the sample, why should we use resources by sampling her again? If we plan to take an SRS without replacement of our population with N balls, the ten possible samples (ignoring the ordering) are

{1, 2}	{1, 3}	{1, 4}	{1, 5}	{2, 3}
{2, 4}	{2, 5}	{3, 4}	{3, 5}	{4, 5}

Since there are ten possible samples and we are sampling with equal probabilities, the probability that a given sample will be chosen is $1/10$.

In general, there are

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{B.2}$$

possible samples of size n that can be drawn without replacement and with equal probabilities from a population of size N, where

$$k! = k(k-1)(k-2)\cdots 1 \quad \text{and} \quad 0! = 1.$$

For our example, there are

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(3 \cdot 2 \cdot 1)} = 10$$

possible samples of size 2, as we found when we listed them.

Note that in sampling without replacement, successive draws are *not* independent. For this example,

$$P(\text{2 chosen on first draw, 4 chosen on second draw}) = \frac{1}{20}.$$

But P(2 chosen on first draw) = $1/5$, and P(4 chosen on second draw) = $1/5$, so the product of the probabilities of the two events is not the probability of the intersection.

EXAMPLE B.1 Players of the Arizona State Lottery game "Fantasy 5" choose 5 numbers without replacement from the numbers 1 through 35. If the 5 numbers you choose match the 5 official winning numbers, you win \$50,000. What is the probability you will win \$50,000? You could select a total of

$$\binom{35}{5} = \frac{35!}{5!30!} = 324{,}632$$

possible sets of 5 numbers. But only

$$\binom{5}{5} = 1$$

of those sets will match the official winning numbers, so your probability of winning \$50,000 is 1/324,632.

Cash prizes are also given if you match 3 or 4 of the numbers. To match 4, you must select 4 numbers out of the set of 5 winning numbers and the remaining number out of the set of 30 nonwinning numbers, so the probability is

$$P(\text{match exactly 4 numbers}) = \frac{\binom{5}{4}\binom{30}{1}}{\binom{35}{5}} = \frac{150}{324{,}632}. \quad ■$$

EXERCISE B1 What is the probability you match exactly 3 of the numbers? Match at least 1 of the numbers? ■

EXERCISE B2 *Calculating the Sampling Distribution in Example 2.3*

A box has eight balls; three of the balls contain the number 7. You select an SRS without replacement of size 4. What is the probability that your sample contains no 7s? Exactly one 7? Exactly two 7s? ■

B.2 Random Variables and Expected Value

A **random variable** is a function that assigns a number to each outcome in the sample space. Which number the random variable will actually assume is determined only after we conduct the experiment and depends on a random process: Before we conduct the experiment, we only know probabilities with which the different outcomes can occur. The set of possible values of a random variable, along with the probability with which each value occurs, is called the **probability distribution** of the random variable. Random variables are denoted by capital letters in this book to distinguish them from the fixed values y_i. If X is a random variable, then $P(X = x)$ is the probability that the random variable takes on the value x. The quantity x is sometimes called a **realization** of the random variable X; x is one of the values that could occur if we performed the experiment.

EXAMPLE B.2 In the lottery game "Fantasy 5," let X be the amount of money you will win from your selection of numbers. You win \$50,000 if you match all 5 winning numbers, \$500 if you match 4, \$5 if you match 3, and nothing if you match fewer than 3. Then the probability distribution of X is given in the following table:

x	0	5	500	50,000
$P(X=x)$	$\frac{320{,}131}{324{,}632}$	$\frac{4350}{324{,}632}$	$\frac{150}{324{,}632}$	$\frac{1}{324{,}632}$

■

If you played "Fantasy 5" many, many times, what would you expect your average winnings per game to be? The answer is the **expected value** of X, defined by

$$E(X) = EX = \sum_x xP(X = x). \tag{B.3}$$

For "Fantasy 5,"

$$EX = \left(0 \times \frac{320{,}131}{324{,}632}\right) + \left(5 \times \frac{4350}{324{,}632}\right) + \left(500 \times \frac{150}{324{,}632}\right) + \left(50{,}000 \times \frac{1}{324{,}632}\right) = \frac{146{,}750}{324{,}632} = 0.45.$$

Think of a box containing 324,632 balls, in which 1 ball has the number 50,000 inside it, 150 balls have the number 500, 4350 balls have the number 5, and the remaining balls have the number 0. The expected value is simply the average of the numbers written inside all the balls in the box. One way to think about expected value is to imagine repeating the experiment over and over again and calculating the long-run average of the results. If you play "Fantasy 5" many, many times, you would expect to win about 45¢ per game, even though 45¢ is not one of the possible realizations of X.

Variance, covariance, correlation, and the coefficient of variation are defined directly in terms of the expected value:

$$V(X) = E[(X - EX)^2] = \text{Cov}(X, X). \tag{B.4}$$

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]. \tag{B.5}$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}. \tag{B.6}$$

$$\text{CV}(X) = \frac{\sqrt{V(X)}}{EX} \quad \text{for } EX \neq 0. \tag{B.7}$$

Expected value and variance have a number of properties that follow directly from the definitions above.

Properties of Expected Value

1 If g is a function, then $E[g(X)] = \sum_x g(x)P(X = x)$.

2 If a and b are constants, then $E(aX + b) = aE(X) + b$.

3 If X and Y are independent, then $E(XY) = (EX)(EY)$.

4 $\text{Cov}(X, Y) = E(XY) - (EX)(EY)$.

5 $\text{Cov}\left(\sum_{i=1}^{n} a_i X_i + b_i, \sum_{j=1}^{m} c_j Y_j + d_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i c_j \text{Cov}(X_i, Y_j)$.

6 $V(X) = E(X^2) - (EX)^2$.

7 $V(X + Y) = V(X) + V(Y) + 2\,\text{Cov}(X, Y)$.

8 $-1 \le \text{Corr}(X, Y) \le 1$.

EXERCISE B3 Prove properties 1 through 8 using the definitions in (B.3) through (B.7). ■

In sampling, we often use estimators that are ratios of two random variables. But $E(Y/X)$ usually does not equal EY/EX. To illustrate this, consider the following probability distribution for X and Y:

x	y	$\frac{y}{x}$	$P(X = x, Y = y)$
1	2	2	$\frac{1}{4}$
2	8	4	$\frac{1}{4}$
3	6	2	$\frac{1}{4}$
4	8	2	$\frac{1}{4}$

Then, $EY/EX = 6/2.5 = 2.4$, but $E(Y/X) = 2.5$. In this example, the values are close but are not equal.

The random variable we use most frequently in this book is

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ is in the sample.} \\ 0 & \text{if unit } i \text{ is not in the sample.} \end{cases} \tag{B.8}$$

This indicator variable tells us whether the ith unit is in the sample or not. In an SRS without replacement, n of the random variables $Z_1, Z_2, \ldots, Z_N$ will take on the value 1, and the remaining $N - n$ will be 0. For Z_i to equal 1, one of the units in the sample must be unit i, and the other $n - 1$ units must come from the remaining $N - 1$ units in the population, so

$$\begin{aligned} P(Z_i = 1) &= P(i\text{th unit is in the sample}) \\ &= \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} \\ &= \frac{n}{N}. \end{aligned} \tag{B.9}$$

Thus,

$$\begin{aligned} E(Z_i) &= 0 \times P(Z_i = 0) + 1 \times P(Z_i = 1) \\ &= P(Z_i = 1) = \frac{n}{N}. \end{aligned}$$

Similarly, for $i \neq j$,

$$\begin{aligned} P(Z_i Z_j = 1) &= P(Z_i = 1, Z_j = 1) \\ &= P(i\text{th unit is in the sample, and } j\text{th unit is in the sample}) \\ &= \frac{\binom{2}{2}\binom{N-2}{n-2}}{\binom{N}{n}} \\ &= \frac{n(n-1)}{N(N-1)}. \end{aligned}$$

Thus, for $i \neq j$,

$$\begin{aligned} E(Z_i Z_j) &= 0 \times P(Z_i Z_j = 0) + 1 \times P(Z_i Z_j = 1) \\ &= P(Z_i Z_j = 1) = \frac{n(n-1)}{N(N-1)}. \end{aligned}$$

EXERCISE B4 Show that

$$V(Z_i) = \text{Cov}(Z_i, Z_i) = \frac{n(N-n)}{N^2}$$

and that, for $i \neq j$,

$$\text{Cov}(Z_i, Z_j) = -\frac{n(N-n)}{N^2(N-1)}. \quad ■$$

The properties of expectation and covariance may be used to prove many results in finite population sampling. One result, used in Chapters 2 and 3, is given below.

Covariance of $\bar{x}$ and $\bar{y}$ from an SRS Let

$$\bar{x}_U = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \bar{y}_U = \frac{1}{N}\sum_{j=1}^{N} y_j,$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} Z_i x_i, \qquad \bar{y} = \frac{1}{n}\sum_{j=1}^{N} Z_j y_j,$$

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x}_U)(y_i - \bar{y}_U)}{(N-1)S_x S_y}.$$

Then,

$$\text{Cov}(\bar{x}, \bar{y}) = \left(1 - \frac{n}{N}\right)\frac{R S_x S_y}{n}. \tag{B.10}$$

We use properties 5 and 6 of expected value, along with some algebra, to show (B.10):

$$\text{Cov}(\bar{x}, \bar{y}) = \frac{1}{n^2}\,\text{Cov}\left(\sum_{i=1}^{N} Z_i x_i, \sum_{j=1}^{N} Z_j y_j\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i y_j \operatorname{Cov}(Z_i, Z_j)$$

$$= \frac{1}{n^2} \sum_{i=1}^{N} x_i y_i V(Z_i) + \frac{1}{n^2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} x_i y_j \operatorname{Cov}(Z_i, Z_j)$$

$$= \frac{1}{n} \frac{N-n}{N^2} \sum_{i=1}^{N} x_i y_i - \frac{1}{n} \frac{N-n}{N^2(N-1)} \sum_{i=1}^{N} \sum_{j \neq i}^{N} x_i y_j$$

$$= \frac{1}{n} \left[\frac{N-n}{N^2} + \frac{N-n}{N^2(N-1)} \right] \sum_{i=1}^{N} x_i y_i - \frac{1}{n} \frac{N-n}{N^2(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i y_j$$

$$= \frac{1}{n} \frac{N-n}{N(N-1)} \sum_{i=1}^{N} x_i y_i - \frac{1}{n} \frac{N-n}{N-1} \bar{x}_U \bar{y}_U$$

$$= \frac{1}{n} \frac{N-n}{N(N-1)} \sum_{i=1}^{N} (x_i - \bar{x}_U)(y_i - \bar{y}_U)$$

$$= \frac{1}{n} \left(1 - \frac{n}{N}\right) R S_x S_y.$$

EXERCISE B5 Show that

$$\operatorname{Corr}(\bar{x}, \bar{y}) = R. \quad \blacksquare \tag{B.11}$$

B.3 Conditional Probability

In an SRS without replacement, successive draws from the population are **dependent:** The unit we choose on the first draw changes the probabilities of selecting the other units on subsequent draws. For our box of five balls, each ball has probability $1/5$ of being chosen on the first draw. If we choose ball 2 on the first draw and sample without replacement, then

$$P(\text{ball 3 on second draw} \mid \text{ball 2 on first draw}) = \frac{1}{4}.$$

(Read as "the conditional probability that ball 3 is selected on the second draw *given that* ball 2 is selected on the first draw equals $1/4$.") Conditional probability allows us to adjust the probability of an event if we know that a related event occurred.

The **conditional probability** of A given B is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{B.12}$$

In sampling we usually use this definition the other way around:

$$P(A \cap B) = P(A \mid B) P(B). \tag{B.13}$$

If events A and B are independent—that is, knowing whether A occurred gives us absolutely no information about whether B occurred—then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

Suppose we have a population with eight households (HHs) and 15 persons living in the households, as follows:

Household	Persons
1	1, 2, 3
2	4
3	5
4	6, 7
5	8
6	9, 10
7	11, 12, 13, 14
8	15

In a one-stage cluster sample, as discussed in Chapter 5, we might take an SRS of two households, then interview each person in the selected households. Then,

$$\begin{aligned} P(\text{interview person 10}) &= P(\text{select HH 6})\ P(\text{interview person 10} \mid \text{select HH 6}) \\ &= \left(\frac{2}{8}\right)\left(\frac{2}{2}\right) = \frac{2}{8}. \end{aligned}$$

In fact, the probability that any individual in the population is interviewed is the same value, $2/8$, because the probability a person is selected is the same as the probability that the household is selected.

If we interview only one randomly selected person in each selected household, though, we are more likely to interview persons living alone than those living with others:

$$\begin{aligned} P(\text{interview person 4}) &= P(\text{select HH 2})\ P(\text{interview person 4} \mid \text{select HH 2}) \\ &= \left(\frac{2}{8}\right)\left(\frac{1}{1}\right) = \frac{2}{8}, \end{aligned}$$

but

$$\begin{aligned} P(\text{interview person 12}) &= P(\text{select HH 7})\ P(\text{interview person 12} \mid \text{select HH 7}) \\ &= \left(\frac{2}{8}\right)\left(\frac{1}{4}\right) = \frac{2}{32}. \end{aligned}$$

These calculations extend to multistage cluster sampling because of the general result

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_k) \\ &\quad = P(A_1 \mid A_2, \ldots, A_k)P(A_2 \mid A_3, \ldots, A_k) \cdots P(A_k). \end{aligned} \tag{B.14}$$

Suppose we take a three-stage cluster sample of grade school students. First, we take an SRS of schools, then sample classes within schools, then sample students within classes. Then, the event {Joe will be selected in the sample} is the same as {Joe's school is selected ∩ Joe's class is selected ∩ Joe is selected}, and we can find

Joe's probability of inclusion by

$$\begin{aligned} P(\text{Joe in sample}) = \; & P(\text{Joe's school is selected}) \\ & \times P(\text{Joe's class is selected} \mid \text{Joe's school is selected}) \\ & \times P(\text{Joe is selected} \mid \text{Joe's school and class are selected}). \end{aligned}$$

If we sample 10% of the schools, 20% of classes within selected schools, and 50% of students within selected classes, then

$$P(\text{Joe in sample}) = (0.10)(0.20)(0.50) = 0.01.$$

B.4 Conditional Expectation

Conditional expectation is used extensively in the theory of cluster sampling. Let X and Y be random variables. Then, using the definition of conditional probability,

$$P(Y = y \mid X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)}. \quad \textbf{(B.15)}$$

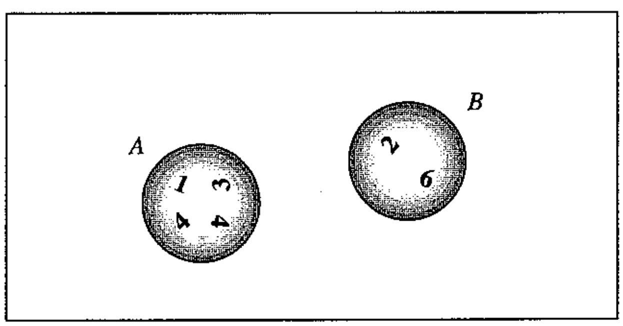
This gives the **conditional distribution** of Y given that $X = x$. The **conditional expectation** of Y given that $X = x$ simply follows the definition of expectation using the conditional distribution:

$$E(Y \mid X = x) = \sum_y y P(Y = y \mid X = x). \quad \textbf{(B.16)}$$

The **conditional variance** of Y given that $X = x$ is defined similarly:

$$V(Y \mid X = x) = \sum_y [y - E(Y \mid X = x)]^2 P(Y = y \mid X = x). \quad \textbf{(B.17)}$$

EXAMPLE B.3 Consider a box with two balls:



Choose one of the balls at random, then choose one of the numbers inside that ball. Let Y = the number that we choose and let

$$Z = \begin{cases} 1 & \text{if we choose ball A.} \\ 0 & \text{if we choose ball B.} \end{cases}$$

Then,

$$P(Y = 1 \mid Z = 1) = \frac{1}{4},$$

$$P(Y = 3 \mid Z = 1) = \frac{1}{4},$$

$$P(Y = 4 \mid Z = 1) = \frac{1}{2},$$

and

$$E(Y \mid Z = 1) = \left(1 \times \frac{1}{4}\right) + \left(3 \times \frac{1}{4}\right) + \left(4 \times \frac{1}{2}\right) = 3.$$

Similarly,

$$P(Y = 2 \mid Z = 0) = \frac{1}{2},$$

$$P(Y = 6 \mid Z = 0) = \frac{1}{2},$$

so

$$E(Y \mid Z = 0) = \left(2 \times \frac{1}{2}\right) + \left(6 \times \frac{1}{2}\right) = 4.$$

In short, if we know that ball A is picked, then the conditional expectation of Y is the average of the numbers in ball A since an SRS is taken; the conditional expectation of Y given that ball B is picked is the average of the numbers in ball B. ■

Note that $E(Y \mid X = x)$ is a function of x; call it $g(x)$. Define the conditional expectation of Y given X, $E(Y \mid X)$, to be $g(X)$, the same function but of the random variable instead. $E(Y \mid X)$ is a random variable and gives us the conditional expected value of Y for the general random variable X: For each possible value of x, the value $E(Y \mid X = x)$ occurs with probability $P(X = x)$.

EXAMPLE B.4 In Example B.3, we know the probability distribution of Z and can thus use the conditional expectations calculated to write the probability distribution of $E(Y \mid Z)$:

z	$E(Y \mid Z = z)$	Probability
0	4	$\frac{1}{2}$
1	3	$\frac{1}{2}$

■

In sampling, we need this general concept of conditional expectation largely so we can use the following properties of conditional expectation to find expected values and variances in cluster samples.

Properties of Conditional Expectation

1 $E(X \mid X) = X$.

2 $E[f(X)Y \mid X] = f(X)E(Y \mid X)$.

3 If X and Y are independent, then $E(Y \mid X) = E(Y)$.

4 $E(Y) = E[E(Y \mid X)]$.

5 $V(Y) = V[E(Y \mid X)] + E[V(Y \mid X)]$.

Conditional expectation can be confusing, so let's talk about what these properties mean. The interested reader should see Ross (1998) or Durrett (1994) for proofs of these properties.

1 $E(X \mid X) = X$. If we know what X is already, then we expect X to be X. The probability distribution of $E(X \mid X)$ is the same as the probability distribution of X.

2 $E[f(X)Y \mid X] = f(X)E(Y \mid X)$. If we know what X is, then we know X^2, or $\log X$, or any function $f(X)$ of X.

3 If X and Y are independent, then $E(Y \mid X) = E(Y)$. If X and Y are independent, then knowing X gives us no information about Y. Thus, the expected value of Y, the average of all the possible outcomes of Y in the experiment, is the same no matter what X is.

4 $E(Y) = E[E(Y \mid X)]$. This property, called **successive conditioning,** and property 5 are the ones we use the most in sampling to show that certain estimates in cluster sampling are unbiased and to calculate their variances. Successive conditioning simply says that if we average the conditional averages the result is the average of the response of interest. You use successive conditioning every time you take a weighted average of a quantity over subpopulations: If a population has 60 women and 40 men, and if the average height of the women is 64 inches and the average height of the men is 69 inches, then the average height for the class is

$$64 \times 0.6 + 69 \times 0.4 = 66 \text{ inches.}$$

In this example, 64 is the conditional expected value of height given that the person is a woman, and 66 is the expected value of height for all persons in the population.

5 $V(Y) = V[E(Y \mid X)] + E[V(Y \mid X)]$. This property gives an easy way of calculating variances in two-stage cluster samples. It says that the total variability has two parts: the variability that arises because (a) $E(Y \mid X = x)$ varies with different values of x and (b) not all y's associated with the same value of x have the same value.

EXAMPLE B.5 Here's how conditional expectation properties work in Example B.3. Successive conditioning implies that

$$\begin{aligned} EY &= E(Y \mid Z = 1)P(Z = 1) + E(Y \mid Z = 0)P(Z = 0) \\ &= \left(3 \times \frac{1}{2}\right) + \left(4 \times \frac{1}{2}\right) = 3.5. \end{aligned}$$

We can also find the distribution of $V(Y \mid Z)$, using property 6 of expected value:

$$\begin{aligned} V(Y \mid Z = 0) &= E(Y^2 \mid Z = 0) - [E(Y \mid Z = 0)]^2 \\ &= \left(4 \times \frac{1}{2}\right) + \left(36 \times \frac{1}{2}\right) - (4)^2 = 4. \end{aligned}$$

$$\begin{aligned} V(Y \mid Z = 1) &= E(Y^2 \mid Z = 1) - [E(Y \mid Z = 1)]^2 \\ &= \left(1 \times \frac{1}{4}\right) + \left(9 \times \frac{1}{4}\right) + \left(16 \times \frac{1}{2}\right) - (3)^2 = 1.5. \end{aligned}$$

z	$V(Y \mid Z = z)$	Probability
0	4	$\frac{1}{2}$
1	1.5	$\frac{1}{2}$

Thus,

$$V[E(Y \mid Z)] = \left(16 \times \frac{1}{2}\right) + \left(9 \times \frac{1}{2}\right) - \{E[E(Y \mid Z)]\}^2$$

$$= \left(16 \times \frac{1}{2}\right) + \left(9 \times \frac{1}{2}\right) - (3.5)^2 = 0.25,$$

$$E[V(Y \mid Z)] = \left(4 \times \frac{1}{2}\right) + \left(1.5 \times \frac{1}{2}\right) = 2.75;$$

so

$$V(Y) = 0.25 + 2.75 = 3.00. \quad \blacksquare$$

If we did not have the properties of conditional expectation, we would need to find the unconditional probability distribution of Y to calculate its expectation and variance—a relatively easy task for the small number of options in Example B.3 but cumbersome to do for general multistage cluster sampling.