

### Lemma 5.2

If  $X$  and  $Y$  are independent, then  $E[XY] = EXEY$ . Using LOTUS, we have

$$\begin{aligned} E[XY] &= \sum_{x \in R_x} \sum_{y \in R_y} xy P_{XY}(x, y) \\ &= \sum_{x \in R_x} \sum_{y \in R_y} xy P_X(x) P_Y(y) \\ &= \left( \sum_{x \in R_x} x P_X(x) \right) \left( \sum_{y \in R_y} y P_Y(y) \right) \\ &= EXEY. \end{aligned}$$

Note that the converse is **not** true. That is, if the only thing that we know about  $X$  and  $Y$  is that  $E[XY] = EXEY$ , then  $X$  and  $Y$  may or may not be independent. Using essentially the same proof as above, we can show if  $X$  and  $Y$  are independent, then  $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$  for any functions  $g : \mathbb{R} \mapsto \mathbb{R}$  and  $h : \mathbb{R} \mapsto \mathbb{R}$ .

If  $X$  and  $Y$  are independent random variables, then

1.  $E[X|Y] = EX$ ;
2.  $E[g(X)|Y] = E[g(X)]$ ;
3.  $E[XY] = EXEY$ ;
4.  $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ .

## Conditional Variance:

Similar to the conditional expectation, we can define the conditional variance of  $X$ ,  $\text{Var}(X|Y = y)$ , which is the variance of  $X$  in the conditional space where we know  $Y = y$ . If we let  $\mu_{X|Y}(y) = E[X|Y = y]$ , then

$$\begin{aligned} \text{Var}(X|Y = y) &= E[(X - \mu_{X|Y}(y))^2 | Y = y] \\ &= \sum_{x_i \in R_X} (x_i - \mu_{X|Y}(y))^2 P_{X|Y}(x_i) \\ &= E[X^2 | Y = y] - \mu_{X|Y}(y)^2. \end{aligned}$$

Note that  $\text{Var}(X|Y = y)$  is a function of  $y$ . Similar to our discussion on  $E[X|Y = y]$  and  $E[X|Y]$ , we define  $\text{Var}(X|Y)$  as a function of the random variable  $Y$ . That is,  $\text{Var}(X|Y)$  is a random variable whose value equals  $\text{Var}(X|Y = y)$  whenever  $Y = y$ . Let us look at an example.

---

### Example 5.13

Let  $X$ ,  $Y$ , and  $Z = E[X|Y]$  be as in Example 5.11. Let also  $V = \text{Var}(X|Y)$ .

- Find the PMF of  $V$ .
- Find  $EV$ .
- Check that  $\text{Var}(X) = E(V) + \text{Var}(Z)$ .

#### Solution

In Example 5.11, we found out that  $X, Y \sim \text{Bernoulli}(\frac{2}{5})$ . We also obtained

$$\begin{aligned} X|Y=0 &\sim \text{Bernoulli}\left(\frac{2}{3}\right), \\ P(X=0|Y=1) &= 1, \\ \text{Var}(Z) &= \frac{8}{75}. \end{aligned}$$

- To find the PMF of  $V$ , we note that  $V$  is a function of  $Y$ . Specifically,

$$V = \text{Var}(X|Y) = \begin{cases} \text{Var}(X|Y=0) & \text{if } Y=0 \\ \text{Var}(X|Y=1) & \text{if } Y=1 \end{cases}$$

Therefore,

$$V = \text{Var}(X|Y) = \begin{cases} \text{Var}(X|Y=0) & \text{with probability } \frac{3}{5} \\ \text{Var}(X|Y=1) & \text{with probability } \frac{2}{5} \end{cases}$$

Now, since  $X|Y=0 \sim \text{Bernoulli}\left(\frac{2}{3}\right)$ , we have

$$\text{Var}(X|Y=0) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9},$$

and since given  $Y=1$ ,  $X=0$ , we have

$$\text{Var}(X|Y=1) = 0.$$

Thus,

$$V = \text{Var}(X|Y) = \begin{cases} \frac{2}{9} & \text{with probability } \frac{3}{5} \\ 0 & \text{with probability } \frac{2}{5} \end{cases}$$

So we can write

$$P_V(v) = \begin{cases} \frac{3}{5} & \text{if } v = \frac{2}{9} \\ \frac{2}{5} & \text{if } v = 0 \\ 0 & \text{otherwise} \end{cases}$$

b. To find  $EV$ , we write

$$EV = \frac{2}{9} \cdot \frac{3}{5} + 0 \cdot \frac{2}{5} = \frac{2}{15}.$$

c. To check that  $\text{Var}(X) = E(V) + \text{Var}(Z)$ , we just note that

$$\text{Var}(X) = \frac{2}{5} \cdot \frac{3}{5} = \frac{6}{25},$$

$$EV = \frac{2}{15},$$

$$\text{Var}(Z) = \frac{8}{75}.$$

In the above example, we checked that  $\text{Var}(X) = E(V) + \text{Var}(Z)$ , which says

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E[X|Y]).$$

It turns out this is true in general and it is called *the law of total variance*, or *variance decomposition formula* [3]. Let us first prove the law of total variance, and then we explain it intuitively. Note that if  $V = \text{Var}(X|Y)$ , and  $Z = E[X|Y]$ , then

$$\begin{aligned} V &= E[X^2|Y] - (E[X|Y])^2 \\ &= E[X^2|Y] - Z^2. \end{aligned}$$

Thus,

$$\begin{aligned} EV &= E[E[X^2|Y]] - E[Z^2] \\ &= E[X^2] - E[Z^2] \quad (\text{law of iterated expectations (Equation 5.7)}) \end{aligned} \quad (5.8)$$

Next, we have

$$\begin{aligned} \text{Var}(Z) &= E[Z^2] - (EZ)^2 \\ &= E[Z^2] - (EX)^2 \quad (\text{law of iterated expectations}) \end{aligned} \quad (5.9)$$

Combining Equations 5.8 and 5.9, we obtain the law of total variance.

Law of Total Variance:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \quad (5.10)$$

There are several ways that we can look at the law of total variance to get some intuition. Let us first note that all the terms in Equation 5.10 are positive (since variance is always positive). Thus, we conclude

$$\text{Var}(X) \geq E(\text{Var}(X|Y)) \quad (5.11)$$

This states that when we condition on  $Y$ , the variance of  $X$  reduces on average. To describe this intuitively, we can say that variance of a random variable is a measure of our uncertainty about that random variable. For example, if  $\text{Var}(X) = 0$ , we do not have any uncertainty about  $X$ . Now, the above inequality simply states that if we obtain some extra information, i.e., we know the value of  $Y$ , our uncertainty about the value of the random variable  $X$  reduces on average. So, the above inequality makes sense. Now, how do we explain the whole law of total variance?

To describe the law of total variance intuitively, it is often useful to look at a population divided into several groups. In particular, suppose that we have this random experiment: We pick a person in the world at random and look at his/her height. Let's call the resulting value  $X$ . Define another random variable  $Y$  whose value depends on the country of the chosen person, where  $Y = 1, 2, 3, \dots, n$ , and  $n$  is the number of countries in the world. Then, let's look at the two terms in the law of total variance.

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E[X|Y]).$$

Note that  $\text{Var}(X|Y = i)$  is the variance of  $X$  in country  $i$ . Thus,  $E(\text{Var}(X|Y))$  is the average of variances in each country. On the other hand,  $E[X|Y = i]$  is the average height in country  $i$ . Thus,  $\text{Var}(E[X|Y])$  is the variance between countries. So, we can interpret the law of total variance in the following way. Variance of  $X$  can be decomposed into two parts: the first is the average of variances in each individual country, while the second is the variance between height averages in each country.

---

### Example 5.14

Let  $N$  be the number of customers that visit a certain store in a given day. Suppose that we know  $E[N]$  and  $\text{Var}(N)$ . Let  $X_i$  be the amount that the  $i$ th customer spends on

average. We assume  $X_i$ 's are independent of each other and also independent of  $N$ . We further assume they have the same mean and variance

$$\begin{aligned} EX_i &= EX, \\ \text{Var}(X_i) &= \text{Var}(X). \end{aligned}$$

Let  $Y$  be the store's total sales, i.e.,

$$Y = \sum_{i=1}^N X_i.$$

Find  $EY$  and  $\text{Var}(Y)$ .

### Solution

To find  $EY$ , we cannot directly use the linearity of expectation because  $N$  is random. But, conditioned on  $N = n$ , we can use linearity and find  $E[Y|N = n]$ ; so, we use the law of iterated expectations:

$$\begin{aligned} EY &= E[E[Y|N]] && \text{(law of iterated expectations)} \\ &= E \left[ E \left[ \sum_{i=1}^N X_i | N \right] \right] \\ &= E \left[ \sum_{i=1}^N E[X_i | N] \right] && \text{(linearity of expectation)} \\ &= E \left[ \sum_{i=1}^N E[X_i] \right] && (X_i \text{'s and } N \text{ are independent)} \\ &= E[NEX] && \text{(since } EX_i = EX \text{)} \\ &= E[X]E[N] && \text{(since } EX \text{ is not random).} \end{aligned}$$

To find  $\text{Var}(Y)$ , we use the law of total variance:

$$\begin{aligned} \text{Var}(Y) &= E(\text{Var}(Y|N)) + \text{Var}(E[Y|N]) \\ &= E(\text{Var}(Y|N)) + \text{Var}(NEX) && \text{(as above)} \\ &= E(\text{Var}(Y|N)) + (EX)^2 \text{Var}(N) && (5.12) \end{aligned}$$

To find  $E(\text{Var}(Y|N))$ , note that, given  $N = n$ ,  $Y$  is a sum  $n$  independent random variables. As we discussed before, for  $n$  independent random variables, the variance of the sum is equal to sum of the variances. This fact is officially proved in [Section 5.3](#) and also in Chapter 6, but we have occasionally used it as it simplifies the analysis. Thus, we can write

$$\begin{aligned}
\text{Var}(Y|N) &= \sum_{i=1}^N \text{Var}(X_i|N) \\
&= \sum_{i=1}^N \text{Var}(X_i) \quad (\text{since } X_i\text{'s are independent of } N) \\
&= N\text{Var}(X).
\end{aligned}$$

Thus, we have

$$E(\text{Var}(Y|N)) = EN\text{Var}(X) \quad (5.13)$$

Combining Equations 5.12 and 5.13, we obtain

$$\text{Var}(Y) = EN\text{Var}(X) + (EX)^2\text{Var}(N).$$