2.8 Variance and Standard Deviation

In Section 2.5, we describe an average as a typical value of a random variable. It is one number that summarizes an entire probability model. After finding an average, someone who wants to look further into the probability model might ask, "How typical is the average?" or, "What are the chances of observing an event far from the average?" In the example of the midterm exam, after you find out your score is 7 points above average, you are likely to ask, "How good is that? Is it near the top of the class or somewhere near the middle?" A measure of dispersion is an answer to these questions wrapped up in a single number. If this measure is small, observations are likely to be near the average. A high measure of dispersion suggests that it is not unusual to observe events that are far from the average.

The most important measures of dispersion are the standard deviation and its close relative, the variance. The variance of random variable X describes the difference between X and its expected value. This difference is the derived random variable, $Y = X - \mu_X$. Theorem 2.11 states that $\mu_Y = 0$, regardless of the probability model of X. Therefore μ_Y provides no information about the dispersion of X around μ_X. A useful measure of the likely difference between X and its expected value is the expected absolute value of the difference, $E[|Y|]$. However, this parameter is not easy to work with mathematically in many situations, and it is not used frequently.

Instead we focus on $E[Y^2] = E[(X - \mu_X)^2]$, which is referred to as $\text{Var}[X]$, the variance of X. The square root of the variance is σ_X, the standard deviation of X.

Definition 2.16 ***Variance***

*The **variance** of random variable X is*

$$\text{Var}[X] = E\left[(X - \mu_X)^2\right].$$

Definition 2.17 ***Standard Deviation***

*The **standard deviation** of random variable X is*

$$\sigma_X = \sqrt{\text{Var}\,[X]}.$$

It is useful to take the square root of $\text{Var}[X]$ because σ_X has the same units (for example, exam points) as X. The units of the variance are squares of the units of the random variable (exam points squared). Thus σ_X can be compared directly with the expected value. Informally we think of outcomes within $\pm\sigma_X$ of μ_X as being in the center of the distribution. Thus if the standard deviation of exam scores is 12 points, the student with a score of +7 with respect to the mean can think of herself in the middle of the class. If the standard deviation is 3 points, she is likely to be near the top. Informally, we think of sample values within σ_X of the expected value, $x \in [\mu_X - \sigma_X, \mu_X + \sigma_X]$, as "typical" values of X and other values as "unusual."

Because $(X - \mu_X)^2$ is a function of X, $\text{Var}[X]$ can be computed according to Theorem 2.10.

$$\text{Var}[X] = \sigma_X^2 = \sum_{x \in S_X} (x - \mu_X)^2 P_X(x). \tag{2.91}$$

By expanding the square in this formula, we arrive at the most useful approach to computing the variance.

Theorem 2.13

$$\text{Var}[X] = E\left[X^2\right] - \mu_X^2 = E\left[X^2\right] - (E[X])^2$$

Proof Expanding the square in (2.91), we have

$$\begin{aligned}\text{Var}[X] &= \sum_{x \in S_X} x^2 P_X(x) - \sum_{x \in S_X} 2\mu_X x P_X(x) + \sum_{x \in S_X} \mu_X^2 P_X(x) \\ &= E[X^2] - 2\mu_X \sum_{x \in S_X} x P_X(x) + \mu_X^2 \sum_{x \in S_X} P_X(x) \\ &= E[X^2] - 2\mu_X^2 + \mu_X^2\end{aligned}$$

We note that $E[X]$ and $E[X^2]$ are examples of *moments* of the random variable X. $\text{Var}[X]$ is a *central moment* of X.

Definition 2.18 ***Moments***

For random variable X:

(a) The nth **moment** *is $E[X^n]$.*

(b) The nth **central moment** *is $E[(X - \mu_X)^n]$.*

Thus, $E[X]$ is the *first moment* of random variable X. Similarly, $E[X^2]$ is the *second moment*. Theorem 2.13 says that the variance of X is the second moment of X minus the square of the first moment.

Like the PMF and the CDF of a random variable, the set of moments of X is a complete probability model. We learn in Section 6.3 that the model based on moments can be expressed as a *moment generating function*.

Example 2.34 In Example 2.6, we found that random variable R has PMF

$$P_R(r) = \begin{cases} 1/4 & r = 0, \\ 3/4 & r = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{2.92}$$

In Example 2.26, we calculated $E[R] = \mu_R = 3/2$. What is the variance of R?

In order of increasing simplicity, we present three ways to compute $\text{Var}[R]$.

- From Definition 2.16, define

$$W = (R - \mu_R)^2 = (R - 3/2)^2 \tag{2.93}$$

Theore

The PMF of W is

$$P_W(w) = \begin{cases} 1/4 & w = (0-3/2)^2 = 9/4, \\ 3/4 & w = (2-3/2)^2 = 1/4, \\ 0 & \text{otherwise.} \end{cases} \tag{2.94}$$

Then

$$\text{Var}[R] = E[W] = (1/4)(9/4) + (3/4)(1/4) = 3/4. \tag{2.95}$$

- Recall that Theorem 2.10 produces the same result without requiring the derivation of $P_W(w)$.

$$\text{Var}[R] = E\left[(R-\mu_R)^2\right] \tag{2.96}$$
$$= (0-3/2)^2 P_R(0) + (2-3/2)^2 P_R(2) = 3/4 \tag{2.97}$$

- To apply Theorem 2.13, we find that

$$E\left[R^2\right] = 0^2 P_R(0) + 2^2 P_R(2) = 3 \tag{2.98}$$

Thus Theorem 2.13 yields

$$\text{Var}[R] = E\left[R^2\right] - \mu_R^2 = 3 - (3/2)^2 = 3/4 \tag{2.99}$$

Note that $(X - \mu_X)^2 \geq 0$. Therefore, its expected value is also nonnegative. That is, for any random variable X

$$\text{Var}[X] \geq 0. \tag{2.100}$$

The following theorem is related to Theorem 2.12

Theorem 2.14

$$\text{Var}[aX + b] = a^2 \,\text{Var}[X]$$

Proof We let $Y = aX + b$ and apply Theorem 2.13. We first expand the second moment to obtain

$$E\left[Y^2\right] = E\left[a^2X^2 + 2abX + b^2\right] = a^2 E\left[X^2\right] + 2ab\mu_X + b^2. \tag{2.101}$$

Expanding the right side of Theorem 2.12 yields

$$\mu_Y^2 = a^2\mu_X^2 + 2ab\mu_x + b^2. \tag{2.102}$$

Because $\text{Var}[Y] = E[Y^2] - \mu_Y^2$, Equations (2.101) and (2.102) imply that

$$\text{Var}[Y] = a^2 E\left[X^2\right] - a^2\mu_X^2 = a^2(E\left[X^2\right] - \mu_X^2) = a^2 \,\text{Var}[X]. \tag{2.103}$$

If we let $a = 0$ in this theorem, we have $\text{Var}[b] = 0$ because there is no dispersion around the expected value of a constant. If we let $a = 1$, we have $\text{Var}[X + b] = \text{Var}[X]$ because shifting a random variable by a constant does not change the dispersion of outcomes around the expected value.

Example 2.35 A new fax machine automatically transmits an initial cover page that precedes the regular fax transmission of X information pages. Using this new machine, the number of pages in a fax is $Y = X + 1$. What are the expected value and variance of Y?

The expected number of transmitted pages is $E[Y] = E[X] + 1$. The variance of the number of pages sent is $\text{Var}[Y] = \text{Var}[X]$.

If we let $b = 0$ in Theorem 2.12, we have $\text{Var}[aX] = a^2 \text{Var}[X]$ and $\sigma_{aX} = a\sigma X$. Multiplying a random variable by a constant is equivalent to a scale change in the units of measurement of the random variable.

Example 2.36 In Example 2.30, the amplitude V in volts has PMF

$$P_V(v) = \begin{cases} 1/7 & v = -3, -2, \ldots, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.104}$$

A new voltmeter records the amplitude U in millivolts. What is the variance of U?

Note that $U = 1000V$. To use Theorem 2.14, we first find the variance of V. The expected value of the amplitude is

$$\mu_V = 1/7[-3 + (-2) + (-1) + 0 + 1 + 2 + 3] = 0 \text{ volts.} \tag{2.105}$$

The second moment is

$$E\left[V^2\right] = 1/7[(-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2] = 4 \text{ volts}^2 \tag{2.106}$$

Therefore the variance is $\text{Var}[V] = E[V^2] - \mu_V^2 = 4 \text{ volts}^2$. By Theorem 2.14,

$$\text{Var}[U] = 1000^2 \text{Var}[V] = 4{,}000{,}000 \text{ millivolts}^2. \tag{2.107}$$

The following theorem states the variances of the families of random variables defined in Section 2.3.

Theorem 2.15

(a) *If X is Bernoulli (p), then* $\text{Var}[X] = p(1 - p)$.

(b) *If X is geometric (p), then* $\text{Var}[X] = (1 - p)/p^2$.

(c) *If X is binomial (n, p), then* $\text{Var}[X] = np(1 - p)$.

(d) *If X is Pascal (k, p), then* $\text{Var}[X] = k(1 - p)/p^2$.

(e) *If X is Poisson (α), then* $\text{Var}[X] = \alpha$.

(f) *If X is discrete uniform (k, l), then* $\text{Var}[X] = (l - k)(l - k + 2)/12$.

Quiz 2.8 *In an experiment to monitor two calls, the PMF of N the number of voice calls, is*

$$P_N(n) = \begin{cases} 0.1 & n = 0, \\ 0.4 & n = 1, \\ 0.5 & n = 2, \\ 0 & \textit{otherwise}. \end{cases} \tag{2.108}$$

Find

(1) The expected value $E[N]$

(2) The second moment $E[N^2]$

(3) The variance $\text{Var}[N]$

(4) The standard deviation σ_N

2.9 Conditional Probability Mass Function

Recall from Section 1.5 that the conditional probability $P[A|B]$ is a number that expresses our new knowledge about the occurrence of event A, when we learn that another event B occurs. In this section, we consider event A to be the observation of a particular value of a random variable. That is, $A = \{X = x\}$. The conditioning event B contains information about X but not the precise value of X. For example, we might learn that $X \leq 33$ or that $|X| > 100$. In general, we learn of the occurrence of an event B that describes some property of X.

Example 2.37 Let N equal the number of bytes in a fax. A conditioning event might be the event I that the fax contains an image. A second kind of conditioning would be the event $\{N > 10{,}000\}$ which tells us that the fax required more than 10,000 bytes. Both events I and $\{N > 10{,}000\}$ give us information that the fax is likely to have many bytes.

The occurrence of the conditioning event B changes the probabilities of the event $\{X = x\}$. Given this information and a probability model for our experiment, we can use Definition 1.6 to find the conditional probabilities

$$P[A|B] = P[X = x|B] \tag{2.109}$$

for all real numbers x. This collection of probabilities is a function of x. It is the *conditional probability mass function* of random variable X, given that B occurred.

Definition 2.19 ***Conditional PMF***
Given the event B, with $P[B] > 0$, *the* ***conditional probability mass function*** *of X is*

$$P_{X|B}(x) = P[X = x|B].$$

Here we extend our notation convention for probability mass functions. The name of a PMF is the letter P with a subscript containing the name of the random variable. For a conditional PMF, the subscript contains the name of the random variable followed by a vertical bar followed by a statement of the conditioning event. The argument of the function is usually the lowercase letter corresponding to the variable name. The argument is a dummy variable. It could be any letter, so that $P_{X|B}(x)$ is the same function as $P_{X|B}(u)$. Sometimes we write the function with no specified argument at all, $P_{X|B}(\cdot)$.

In some applications, we begin with a set of conditional PMFs, $P_{X|B_i}(x), i = 1, 2, \ldots, m$, where $B_1, B_2, \ldots, B_m$ is an event space. We then use the law of total probability to find the PMF $P_X(x)$.

Theorem 2.16 *A random variable X resulting from an experiment with event space $B_1, \ldots, B_m$ has PMF*

$$P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i].$$

Proof The theorem follows directly from Theorem 1.10 with A denoting the event $\{X = x\}$.

Example 2.38 Let X denote the number of additional years that a randomly chosen 70 year old person will live. If the person has high blood pressure, denoted as event H, then X is a geometric ($p = 0.1$) random variable. Otherwise, if the person's blood pressure is regular, event R, then X has a geometric ($p = 0.05$) PMF with parameter. Find the conditional PMFs $P_{X|H}(x)$ and $P_{X|R}(x)$. If 40 percent of all seventy year olds have high blood pressure, what is the PMF of X?

The problem statement specifies the conditional PMFs in words. Mathematically, the two conditional PMFs are

$$P_{X|H}(x) = \begin{cases} 0.1(0.9)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise,} \end{cases} \tag{2.110}$$

$$P_{X|R}(x) = \begin{cases} 0.05(0.95)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.111}$$

Since H, R is an event space, we can use Theorem 2.16 to write

$$P_X(x) = P_{X|H}(x) P[H] + P_{X|R}(x) P[R] \tag{2.112}$$

$$= \begin{cases} (0.4)(0.1)(0.9)^{x-1} + (0.6)(0.05)(0.95)^{x-1} & x = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.113}$$

When a conditioning event $B \subset S_X$, the PMF $P_X(x)$ determines both the probability of B as well as the conditional PMF:

$$P_{X|B}(x) = \frac{P[X = x, B]}{P[B]}. \tag{2.114}$$

Now either the event $X = x$ is contained in the event B or it is not. If $x \in B$, then $\{X = x\} \cap B = \{X = x\}$ and $P[X = x, B] = P_X(x)$. Otherwise, if $x \notin B$, then $\{X = x\} \cap B = \phi$ and $P[X = x, B] = 0$. The next theorem uses Equation (2.114) to calculate the conditional PMF.

Theorem 2.17

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{P[B]} & x \in B, \\ 0 & \text{otherwise.} \end{cases}$$

The theorem states that when we learn that an outcome $x \in B$, the probabilities of all $x \notin B$ are zero in our conditional model and the probabilities of all $x \in B$ are proportionally higher than they were before we learned $x \in B$.

Example 2.39 In the probability model of Example 2.29, the length of a fax X has PMF

$$P_X(x) = \begin{cases} 0.15 & x = 1, 2, 3, 4, \\ 0.1 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.115}$$

Suppose the company has two fax machines, one for faxes shorter than five pages and the other for faxes that have five or more pages. What is the PMF of fax length in the second machine?

Relative to $P_X(x)$, we seek a conditional PMF. The condition is $x \in L$ where $L = \{5, 6, 7, 8\}$. From Theorem 2.17,

$$P_{X|L}(x) = \begin{cases} \dfrac{P_X(x)}{P[L]} & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.116}$$

From the definition of L, we have

$$P[L] = \sum_{x=5}^{8} P_X(x) = 0.4. \tag{2.117}$$

With $P_X(x) = 0.1$ for $x \in L$,

$$P_{X|L}(x) = \begin{cases} 0.1/0.4 = 0.25 & x = 5, 6, 7, 8, \\ 0 & \text{otherwise.} \end{cases} \tag{2.118}$$

Thus the lengths of long faxes are equally likely. Among the long faxes, each length has probability 0.25.

Sometimes instead of a letter such as B or L that denotes the subset of S_X that forms the condition, we write the condition itself in the PMF. In the preceding example we could use the notation $P_{X|X\geq 5}(x)$ for the conditional PMF.

Example 2.40 Suppose X, the time in integer minutes you must wait for a bus, has the uniform PMF

$$P_X(x) = \begin{cases} 1/20 & x = 1, 2, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{2.119}$$

Suppose the bus has not arrived by the eighth minute, what is the conditional PMF of your waiting time X?

Let A denote the event $X > 8$. Observing that $P[A] = 12/20$, we can write the conditional PMF of X as

$$P_{X|X>8}(x) = \begin{cases} \dfrac{1/20}{12/20} = \dfrac{1}{12} & x = 9, 10, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{2.120}$$

Note that $P_{X|B}(x)$ is a perfectly respectable PMF. Because the conditioning event B tells us that all possible outcomes are in B, we rewrite Theorem 2.1 using B in place of S.

Theorem 2.18

(a) For any $x \in B$, $P_{X|B}(x) \geq 0$.

(b) $\sum_{x \in B} P_{X|B}(x) = 1$.

(c) For any event $C \subset B$, $P[C|B]$, *the conditional probability that* X *is in the set* C, *is*

$$P[C|B] = \sum_{x \in C} P_{X|B}(x).$$

Therefore, we can compute averages of the conditional random variable $X|B$ and averages of functions of $X|B$ in the same way that we compute averages of X. The only difference is that we use the conditional PMF $P_{X|B}(\cdot)$ in place of $P_X(\cdot)$.

Definition 2.20 ***Conditional Expected Value***

The ***conditional expected value*** *of random variable* X *given condition* B *is*

$$E[X|B] = \mu_{X|B} = \sum_{x \in B} x P_{X|B}(x).$$

When we are given a family of conditional probability models $P_{X|B_i}(x)$ for an event space $B_1, \ldots, B_m$, we can compute the expected value $E[X]$ in terms of the conditional expected values $E[X|B_i]$.

Theorem 2.19 *For a random variable* X *resulting from an experiment with event space* $B_1, \ldots, B_m$,

$$E[X] = \sum_{i=1}^{m} E[X|B_i] P[B_i].$$

Proof Since $E[X] = \sum_x x P_X(x)$, we can use Theorem 2.16 to write

$$E[X] = \sum_x x \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i] \tag{2.121}$$

$$= \sum_{i=1}^{m} P[B_i] \sum_x x P_{X|B_i}(x) = \sum_{i=1}^{m} P[B_i] E[X|B_i]. \tag{2.122}$$

For a derived random variable $Y = g(X)$, we have the equivalent of Theorem 2.10.

rem 2.20 *The conditional expected value of $Y = g(X)$ given condition B is*

$$E[Y|B] = E[g(X)|B] = \sum_{x \in B} g(x) P_{X|B}(x).$$

It follows that the conditional variance and conditional standard deviation conform to Definitions 2.16 and 2.17 with $X|B$ replacing X.

Example 2.41 Find the conditional expected value, the conditional variance, and the conditional standard deviation for the long faxes defined in Example 2.39.

$$E[X|L] = \mu_{X|L} = \sum_{x=5}^{8} x P_{X|L}(x) = 0.25 \sum_{x=5}^{8} x = 6.5 \text{ pages} \tag{2.123}$$

$$E\left[X^2|L\right] = 0.25 \sum_{x=5}^{8} x^2 = 43.5 \text{ pages}^2 \tag{2.124}$$

$$\text{Var}[X|L] = E\left[X^2|L\right] - \mu_{X|L}^2 = 1.25 \text{ pages}^2 \tag{2.125}$$

$$\sigma_{X|L} = \sqrt{\text{Var}[X|L]} = 1.12 \text{ pages} \tag{2.126}$$

2.9 *On the Internet, data is transmitted in packets. In a simple model for World Wide Web traffic, the number of packets N needed to transmit a Web page depends on whether the page has graphic images. If the page has images (event I), then N is uniformly distributed between 1 and 50 packets. If the page is just text (event T), then N is uniform between 1 and 5 packets. Assuming a page has images with probability 1/4, find the*

(1) *conditional PMF* $P_{N|I}(n)$
(2) *conditional PMF* $P_{N|T}(n)$
(3) *PMF* $P_N(n)$
(4) *conditional PMF* $P_{N|N \le 10}(n)$
(5) *conditional expected value* $E[N|N \le 10]$
(6) *conditional variance* $\text{Var}[N|N \le 10]$

) MATLAB

For discrete random variables, this section will develop a variety of ways to use MATLAB. We start by calculating probabilities for any finite random variable with arbitrary PMF $P_X(x)$. We then compute PMFs and CDFs for the families of random variables introduced in Section 2.3. Based on the calculation of the CDF, we then develop a method for generating random sample values. Generating a random sample is a simple simulation of a experiment that produces the corresponding random variable. In subsequent chapters, we will see that MATLAB functions that generate random samples are building blocks for the simulation of