

to the mean:

$$\text{coef. of var.} = \frac{\sqrt{\text{Var}(X)}}{EX} \quad (3.41)$$

This is a scale-free measure (e.g. inches divided by inches), and serves as a good way to judge whether a variance is large or not.

### 3.6 Indicator Random Variables, and Their Means and Variances

**Definition 5** *A random variable that has the value 1 or 0, according to whether a specified event occurs or not is called an **indicator random variable** for that event.*

You'll often see later in this book that the notion of an indicator random variable is a very handy device in certain derivations. But for now, let's establish its properties in terms of mean and variance.

**Handy facts:** Suppose  $X$  is an indicator random variable for the event  $A$ . Let  $p$  denote  $P(A)$ .  
Then

$$E(X) = p \quad (3.42)$$

$$\text{Var}(X) = p(1 - p) \quad (3.43)$$

This two facts are easily derived. In the first case we have, using our properties for expected value,

$$EX = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = P(X = 1) = P(A) = p \quad (3.44)$$

The derivation for  $\text{Var}(X)$  is similar (use (3.29)).

### 3.7 A Combinatorial Example

A committee of four people is drawn at random from a set of six men and three women. Suppose we are concerned that there may be quite a gender imbalance in the membership of the committee. Toward that end, let  $M$  and  $W$  denote the numbers of men and women in our committee, and let  $D = M - W$ . Let's find  $E(D)$ , in two different ways.

D can take on the values 4-0, 3-1, 2-2 and 1-3, i.e. 4, 2, 0 and -2. So,

$$ED = -2 \cdot P(D = -2) + 0 \cdot P(D = 0) + 2 \cdot P(D = 2) + 4 \cdot P(D = 4) \quad (3.45)$$

Now, using reasoning along the lines in Section 2.12, we have

$$P(D = -2) = P(M = 1 \text{ and } W = 3) = \frac{\binom{6}{1} \binom{3}{3}}{\binom{9}{4}} \quad (3.46)$$

After similar calculations for the other probabilities in (3.45), we find the  $ED = \frac{4}{3}$ .

Note what this means: If we were to perform this experiment many times, i.e. choose committees again and again, on average we would have a little more than one more man than women on the committee.

Now let's use our "mailing tubes" to derive ED a different way:

$$ED = E(M - W) \quad (3.47)$$

$$= E[M - (4 - M)] \quad (3.48)$$

$$= E(2M - 4) \quad (3.49)$$

$$= 2EM - 4 \text{ (from (3.14))} \quad (3.50)$$

Now, let's find EM by using indicator random variables. Let  $G_i$  denote the indicator random variable for the event that the  $i^{\text{th}}$  person we pick is male,  $i = 1, 2, 3, 4$ . Then

$$M = G_1 + G_2 + G_3 + G_4 \quad (3.51)$$

so

$$EM = E(G_1 + G_2 + G_3 + G_4) \quad (3.52)$$

$$= EG_1 + EG_2 + EG_3 + EG_4 \text{ [ from (3.13)]} \quad (3.53)$$

$$= P(G_1 = 1) + P(G_2 = 1) + P(G_3 = 1) + P(G_4 = 1) \text{ [ from (3.42)]} \quad (3.54)$$

Note carefully that the second equality here, which uses (3.13), is true in spite of the fact that the  $G_i$  are not independent. Equation (3.13) does not require independence.

Another key point is that, due to symmetry,  $P(G_i = 1)$  is the same for all  $i$ . same expected value. (Note that we did not write a *conditional* probability here.) To see this, suppose the six men that are available

for the committee are named Alex, Bo, Carlo, David and Eduardo. When we select our first person, any of these men has the same chance of being chosen ( $1/9$ ). *But that is also true for the second pick.* Think of a notebook, with a column named “second pick.” In some lines, that column will say Alex, in some it will say Bo, and so on, and in some lines there will be women’s names. But in that column, Bo will appear the same fraction of the time as Alex, due to symmetry, and that will be the same fraction is for, say, Alice, again  $1/9$ .

Now,

$$P(G_1 = 1) = \frac{6}{9} = \frac{2}{3} \quad (3.55)$$

Thus

$$ED = 2 \cdot \left(4 \cdot \frac{2}{3}\right) - 4 = \frac{4}{3} \quad (3.56)$$

### 3.8 A Useful Fact

For a random variable  $X$ , consider the function

$$g(c) = E[(X - c)^2] \quad (3.57)$$

Remember, the quantity  $E[(X - c)^2]$  is a number, so  $g(c)$  really is a function, mapping a real number  $c$  to some real output.

We can ask the question, What value of  $c$  minimizes  $g(c)$ ? To answer that question, write:

$$g(c) = E[(X - c)^2] = E(X^2 - 2cX + c^2) = E(X^2) - 2cEX + c^2 \quad (3.58)$$

where we have used the various properties of expected value derived in recent sections.

Now differentiate with respect to  $c$ , and set the result to 0. Remembering that  $E(X^2)$  and  $EX$  are constants, we have

$$0 = -2EX + 2c \quad (3.59)$$

so the minimizing  $c$  is  $c = EX$ !

In other words, the minimum value of  $E[(X - c)^2]$  occurs at  $c = EX$ .