5.6 Variance

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

One important application of LOTUS is for finding the *variance* of a random variable. Like expected value, variance is a single-number summary of the distribution of a random variable. While the expected value tells us the center of mass of a distribution, the variance tells us how spread out the distribution is.

Definition 5.6.1 (Variance and standard deviation).

The *variance* of an r.v. X is

$$\text{Var}(X) = E(X - EX)^2.$$

The square root of the variance is called the *standard deviation* (*SD*):

$$\text{SD}(X) = \sqrt{\text{Var}(X)}.$$

Recall that when we write $E(X - EX)^2$, we mean the expectation of the random variable $(X - EX)^2$, *not* $(E(X - EX))^2$ (which is 0 by linearity).

The variance of X measures how far X is from its mean on average, but instead of simply taking the average difference between X and its mean EX, we take the average *squared* difference. To see why, note that the average deviation from the mean, $E(X - EX)$, always equals 0 by linearity; positive and negative deviations cancel each other out. By squaring the deviations, we ensure that both positive and negative deviations contribute to the overall variability. However, because variance is an average squared distance, it has the wrong units: if X is in dollars, $\text{Var}(X)$ is in squared dollars. To get back to our original units, we take the square root; this gives us the standard deviation.

One might wonder why variance isn't defined as $E|X - EX|$, which would achieve the goal of counting both positive and negative deviations while maintaining the same units as X. This measure of variability isn't nearly as popular as $E(X - EX)^2$, for a variety of reasons. The absolute value function isn't differentiable at 0, so it doesn't have as nice properties as the squaring function. Squared distances are also connected to geometry via the distance formula and the Pythagorean theorem, which turn out to have corresponding statistical interpretations.

An equivalent expression for variance is $\mathrm{Var}(X) = E(X^2) - (EX)^2$. This formula is often easier to work with when doing actual calculations. Since this is the variance formula we will use over and over again, we state it as its own theorem.

Theorem 5.6.2.

For any r.v. X,

$$\mathrm{Var}(X) = E(X^2) - (EX)^2.$$

Proof

Let $\mu = EX$. Expand $(X - \mu)^2$ and use linearity:

$$\mathrm{Var}(X) = E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu EX + \mu^2 = E(X^2) - \mu^2.$$

Variance has the following properties. The first two are easily verified from the definition, the third will be addressed in a later chapter, and the last one is proven just after stating it.

- $\mathrm{Var}(X + c) = \mathrm{Var}(X)$ for any constant c. Intuitively, if we shift a distribution to the left or right, that should affect the center of mass of the distribution but not its spread.
- $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$ for any constant c.
- If X and Y are independent, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. We prove this and discuss it more in Chapter 7. This is not true in general if X and Y are dependent. For example, in the extreme case where X always equals Y, we have

  $$\mathrm{Var}(X + Y) = \mathrm{Var}(2X) = 4\mathrm{Var}(X) > 2\mathrm{Var}(X) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

  if $\mathrm{Var}(X) > 0$ (which will be true unless X is a constant, as the next property shows).
- $\mathrm{Var}(X) \geq 0$, with equality if and only if $P(X = a) = 1$ for some constant a. In other words, the only random variables that have zero variance are constants (which can be thought of as degenerate r.v.s); all other r.v.s have positive variance.

To prove the last property, note that $\mathrm{Var}(X)$ is the expectation of the *nonnegative* r.v. $(X - EX)^2$, so $\mathrm{Var}(X) \geq 0$. If $P(X = a) = 1$ for some constant a, then $E(X) = a$ and $E(X^2) = a^2$, so $\mathrm{Var}(X) = 0$. Conversely, suppose that $\mathrm{Var}(X) = 0$. Then $E(X - EX)^2 = 0$, which shows that $(X - EX)^2 = 0$ has probability 1, which in turn shows that X equals its mean with probability 1.

Example 5.6.3 (Geometric and Negative Binomial variance).

In this example we'll use LOTUS to compute the variance of the Geometric distribution. Let $X \sim \text{Geom}(p)$. We already know $E(X) = q/p$. By LOTUS,

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 pq^k = \sum_{k=1}^{\infty} k^2 pq^k.$$

We'll find this using a similar tactic to how we found the expectation, starting from the geometric series

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

and taking derivatives. After differentiating once with respect to q, we have

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}.$$

We start the sum from $k = 1$ since the $k = 0$ term is 0 anyway. If we differentiate again, we'll get $k(k-1)$ instead of k^2 as we want, so let's replenish our supply of q's by multiplying both sides by q. This gives

$$\sum_{k=1}^{\infty} kq^k = \frac{q}{(1-q)^2}.$$

Now we are ready to take another derivative:

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3},$$

so

$$E(X^2) = \sum_{k=1}^{\infty} k^2 pq^k = pq\frac{1+q}{(1-q)^3} = \frac{q(1+q)}{p^2}.$$

Finally,

$$\text{Var}(X) = E(X^2) - (EX)^2 = \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}.$$

This is also the variance of the First Success distribution, since shifting by a constant does not affect the variance. Since an $\text{NBin}(r, p)$ r.v. can be represented as a sum of r i.i.d. $\text{Geom}(p)$ r.v.s by Theorem 5.3.10, and since variance is additive for independent random variables, it follows that the variance of the $\text{NBin}(r, p)$ distribution is $r \cdot \frac{q}{p^2}$. LOTUS is an all-purpose tool for computing $E(g(X))$ for any g, but as it usually leads to complicated sums, it should be used as a last resort. For variance calculations, our trusty indicator r.v.s can sometimes be used in place of LOTUS, as in the next example.

Example 5.6.4 (Binomial variance).

Let's find the variance of $X \sim \text{Bin}(n, p)$ using indicator r.v.s to avoid tedious sums. Represent $X = I_1 + I_2 + \cdots + I_n$, where I_j is the indicator of the jth trial being a success. Each I_j has variance

$$\text{Var}(I_j) = E(I_j^2) - (E(I_j))^2 = p - p^2 = p(1-p).$$

(Recall that $I_j^2 = I_j$, so $E(I_j^2) = E(I_j) = p$.) Since the I_j are independent, we can add their variances to get the variance of their sum:

$$\text{Var}(X) = \text{Var}(I_1) + \cdots + \text{Var}(I_n) = np(1-p).$$

Learn About Verified Certificates