3.2.4 Variance

Consider two random variables X and Y with the following PMFs.

$$P_X(x) = \begin{cases} 0.5 & \text{for } x = -100 \\ 0.5 & \text{for } x = 100 \\ 0 & \text{otherwise} \end{cases} \qquad (3.3)$$

$$P_Y(y) = \begin{cases} 1 & \text{for } y = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3.4)$$

Note that $EX = EY = 0$. Although both random variables have the same mean value, their distribution is completely different. Y is always equal to its mean of 0, while X is either 100 or -100, quite far from its mean value. The **variance** is a measure of how spread out the distribution of a random variable is. Here, the variance of Y is quite small since its distribution is concentrated at a single value, while the variance of X will be larger since its distribution is more spread out.

The **variance** of a random variable X, with mean $EX = \mu_X$, is defined as

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

By definition, the variance of X is the average value of $(X - \mu_X)^2$. Since $(X - \mu_X)^2 \geq 0$, the variance is always larger than or equal to zero. A large value of the variance means that $(X - \mu_X)^2$ is often large, so X often takes values far from its mean. This means that the distribution is very spread out. On the other hand, a low variance means that the distribution is concentrated around its average.

Note that if we did not square the difference between X and its mean, the result would be 0. That is

$$E[X - \mu_X] = EX - E[\mu_X] = \mu_X - \mu_X = 0.$$

X is sometimes below its average and sometimes above its average. Thus, $X - \mu_X$ is sometimes negative and sometimes positive, but on average it is zero.

To compute $Var(X) = E\big[(X-\mu_X)^2\big]$, note that we need to find the expected value of $g(X) = (X-\mu_X)^2$, so we can use LOTUS. In particular, we can write

$$\text{Var}(X) = E\big[(X-\mu_X)^2\big] = \sum_{x_k \in R_X} (x_k - \mu_X)^2 P_X(x_k).$$

For example, for X and Y defined in Equations 3.3 and 3.4, we have

$$\text{Var}(X) = (-100-0)^2(0.5) + (100-0)^2(0.5) = 10,000$$

$$\text{Var}(Y) = (0-0)^2(1) = 0.$$

As we expect, X has a very large variance while Var$(Y) = 0$.

Note that Var(X) has a different unit than X. For example, if X is measured in $meters$ then Var(X) is in $meters^2$. To solve this issue, we define another measure, called the **standard deviation**, usually shown as σ_X, which is simply the square root of variance.

> The **standard deviation** of a random variable X is defined as
>
> $$\text{SD}(X) = \sigma_X = \sqrt{\text{Var}(X)}.$$

The standard deviation of X has the same unit as X. For X and Y defined in Equations 3.3 and 3.4, we have

$$\begin{aligned} \sigma_X &= \sqrt{10,000} = 100 \\ \sigma_Y &= \sqrt{0} = 0. \end{aligned}$$

Here is a useful formula for computing the variance.

> Computational formula for the variance:
>
> $$\text{Var}(X) = E\big[X^2\big] - \big[EX\big]^2 \tag{3.5}$$

To prove it note that

$$\begin{aligned}\mathrm{Var}(X) &= E\big[(X-\mu_X)^2\big] \\ &= E\big[X^2 - 2\mu_X X + \mu_X^2\big] \\ &= E\big[X^2\big] - 2E\big[\mu_X X\big] + E\big[\mu_X^2\big] \qquad \text{by linearity of expectation.}\end{aligned}$$

Note that for a given random variable X, μ_X is just a constant real number. Thus, $E\big[\mu_X X\big] = \mu_X E[X] = \mu_X^2$, and $E[\mu_X^2] = \mu_X^2$, so we have

$$\begin{aligned}\mathrm{Var}(X) &= E\big[X^2\big] - 2\mu_X^2 + \mu_X^2 \\ &= E\big[X^2\big] - \mu_X^2.\end{aligned}$$

quation 3.5 is usually easier to work with compared to $\mathrm{Var}(X) = E\big[(X-\mu_X)^2\big]$. To use this equation, we can find $E[X^2] = EX^2$ using LOTUS

$$EX^2 = \sum_{x_k \in R_X} x_k^2 P_X(x_k),$$

and then subtract μ_X^2 to obtain the variance.

Example 3.19

I roll a fair die and let X be the resulting number. Find EX, Var(X), and σ_X.

Solution

We have $R_X = \{1,2,3,4,5,6\}$ and $P_X(k) = \frac{1}{6}$ for $k = 1,2,\ldots,6$. Thus, we have

$$EX = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{7}{2};$$

$$EX^2 = 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6} = \frac{91}{6}.$$

Thus

$$\mathrm{Var}(X) = E\big[X^2\big] - \big(EX\big)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \approx 2.92,$$

$$\sigma_X = \sqrt{\mathrm{Var}(X)} \approx \sqrt{2.92} \approx 1.71$$

Note that variance is not a linear operator. In particular, we have the following theorem.

Theorem 3.3
For a random variable X and real numbers a and b,

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X) \tag{3.6}$$

Proof
If $Y = aX + b$, $EY = aEX + b$. Thus,

$$\begin{aligned}\mathrm{Var}(Y) &= E[(Y-EY)^2] \\ &= E[(aX+b-aEX-b)^2] \\ &= E[a^2(X-\mu_X)^2] \\ &= a^2E[(X-\mu_X)^2] \\ &= a^2\mathrm{Var}(X)\end{aligned}$$

From Equation 3.6, we conclude that, for standard deviation, $\mathrm{SD}(aX+b) = |a|\mathrm{SD}(X)$. We mentioned that variance is NOT a linear operation. But there is a very important case, in which variance behaves like a linear operation and that is when we look at sum of independent random variables.

Theorem 3.4
If $X_1, X_2, \cdots, X_n$ are independent random variables and $X = X_1 + X_2 + \cdots + X_n$, then

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \tag{3.7}$$

We will prove this theorem in Chapter 6, but for now we can look at an example to see how we can use it.

Example 3.20
If $X \sim Binomial(n, p)$ find Var(X).

Solution

We know that we can write a $Binomial(n, p)$ random variable as the sum of n **independent** $Bernoulli(p)$ random variables, i.e., $X = X_1 + X_2 + \cdots + X_n$. Thus, we conclude

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n).$$

If $X_i \sim Bernoulli(p)$, then its variance is

$$\mathrm{Var}(X_i) = E[X_i^2] - (EX_i)^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p(1-p).$$

Thus,

$$\begin{aligned}\mathrm{Var}(X) &= p(1-p) + p(1-p) + \cdots + p(1-p) \\ &= np(1-p).\end{aligned}$$