Variance

We are interested in calculating how much a random variable deviates from its mean. This measure is called variance. Formally, for a random variable $X$ we are interested in $E[X - E[X]]$. By the linearity of expectation we have

$$E[X - E[X]] = E[X] - E[E[X]] = E[X] - E[X] = 0$$

Note that we have used the fact that $E[X]$ is a constant and hence $E[E[X]] = E[X]$. This is not very informative. While calculating the deviations from the mean we do not want the positive and the negative deviations to cancel out each other. This suggests that we should take the absolute value of $X - E[X]$. But working with absolute values is messy. It turns out that squaring of $X - E[X]$ is more useful. This leads to the following definition.

**Definition.** The variance of a random variable $X$ is defined as

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

The standard deviation of a random variable $X$ is

$$\sigma[X] = \sqrt{\text{Var}[X]}$$

The standard deviation undoes the squaring in the variance. In doing the calculations it does not matter whether we use variance or the standard deviation as we can easily compute one from the other.

We show as follows that the two forms of variance in the definition are equivalent.

$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2]$$
$$= E[X^2] - 2E[XE[X]] + E[X]^2$$
$$= E[X^2] - 2E[X]^2 + E[X]^2$$
$$= E[X^2] - E[X]^2$$

In step 2 we used the linearity of expectation and the fact that $E[X]$ is a constant.

**Example 1.** Consider three random variables $X, Y, Z$. Their probability mass distribution is as follows.

$$\Pr[X = x] = \begin{cases} \frac{1}{2}, & x = -2 \\ \frac{1}{2}, & x = 2 \end{cases}$$
Pr\[Y = y\] = \[
\begin{cases}
0.001, & y = -10 \\
0.998, & y = 0 \\
0.001, & y = 10
\end{cases}
\]

Pr\[Z = z\] = \[
\begin{cases}
\frac{1}{3}, & z = -10 \\
\frac{1}{3}, & z = 0 \\
\frac{1}{3}, & z = 10
\end{cases}
\]

Which of the above random variables is more “spread out”?

**Solution.** It is easy to see that $E[X] = E[Y] = E[Z] = 0$.

\[
\begin{align*}
\text{Var}[X] & = E[X^2] - (E[X])^2 \\
& = 0.5 \cdot (-2)^2 + 0.5 \cdot (2)^2 \\
& = 4
\end{align*}
\]

\[
\begin{align*}
\text{Var}[Y] & = E[Y^2] - (E[Y])^2 \\
& = 0.001 \cdot (-10)^2 + 0.998 \cdot 0^2 + 0.001 \cdot (10)^2 \\
& = 0.2
\end{align*}
\]

\[
\begin{align*}
\text{Var}[Z] & = E[Z^2] - (E[Z])^2 \\
& = (1/3) \cdot (-5)^2 + (1/3) \cdot 0^2 + (1/3) \cdot (5)^2 \\
& = 16.67
\end{align*}
\]

Thus $Z$ is the most spread out and $Y$ is the most concentrated.

**Example 2.** In the experiment where we roll one die let $X$ be the random variable denoting the number that appears on the top face. What is $\text{Var}[X]$?

**Solution.** From the definition of variance, we have

\[
\begin{align*}
\text{Var}[X] & = E[X^2] - (E[X])^2 \\
& = \frac{1}{6} \left(1 + 4 + 9 + 16 + 25 + 36\right) + \left(\frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6)\right)^2 \\
& = \frac{91}{6} - \frac{49}{4} \\
& = \frac{35}{12}
\end{align*}
\]

Before we proceed let’s prove two useful results.

**Example 3.** In the hat-check problem that we did in one of the earlier lectures, what is the variance of the random variable $X$ that denotes the number of people who get their own hat back?
Solution. We can express $X$ as

$$X = X_1 + X_2 + \cdots + X_n$$

where $X_i$ is the random variable that denotes that is 1 if the $i$th person receives his/her own hat back and 0 otherwise. We already know from an earlier lecture that $E[X] = 1$. $E[X^2]$ can be calculated as follows.

$$E[X^2] = \sum_{i=1}^{n} E[X_i^2] + 2 \sum_{i<j} E[X_i \cdot X_j]$$

$$= \sum_{i=1}^{n} E[X_i^2] + 2 \sum_{i<j} 1 \cdot \Pr[X_i = 1 \cap X_j = 1]$$

$$= \sum_{i=1}^{n} \frac{1}{n} + 2 \left( \frac{n(n-1)}{2} \right) \left( \frac{1}{n(n-1)} \right)$$

$$= n \cdot \frac{1}{n} + 1$$

$$= 2$$

$\text{Var}[X]$ is given by

$$\text{Var}[X] = E[X^2] - E[X]^2 = 2 - 1 = 1$$

Note that like the expectation, the variance is independent of $n$. This means that it is not likely for many people to get their own hat back even if $n$ is large.

**Probability Distributions**

Tossing a coin is an experiment with exactly two outcomes: heads (“success”) with a probability of, say $p$, and tails (“failure”) with a probability of $1 - p$. Such an experiment is called a Bernoulli trial. Let $Y$ be a random variable that is 1 if the experiment succeeds and is 0 otherwise. $Y$ is called a Bernoulli or an indicator random variable. For such a variable we have

$$E[Y] = p \cdot 1 + (1 - p) \cdot 0 = p = \Pr[Y = 1]$$

Thus for a fair coin if we consider heads as “success” then the expected value of the corresponding indicator random variable is 1/2.

A sequence of Bernoulli trials means that the trials are independent and each has a probability $p$ of success. We will study two important distributions that arise from Bernoulli trials: the geometric distribution and the binomial distribution.

**The Geometric Distribution**

Consider the following question. Suppose we have a biased coin with heads probability $p$ that we flip repeatedly until it lands on heads. What is the distribution of the number of flips? This is an example of a geometric distribution. It arises in situations where we
perform a sequence of independent trials until the first success where each trial succeeds with a probability \( p \).

Note that the sample space \( \Omega \) consists of all sequences that end in \( H \) and have exactly one \( H \). That is
\[
\Omega = \{ H, TH, TTH, TTTH, TTTTH, \ldots \}
\]
For any \( \omega \in \Omega \) of length \( i \), \( \Pr[\omega] = (1-p)^i p \).

**Definition.** A geometric random variable \( X \) with parameter \( p \) is given by the following distribution for \( i = 1, 2, \ldots \):
\[
\Pr[X = i] = (1-p)^{i-1}p
\]

We can verify that the geometric random variable admits a valid probability distribution as follows:
\[
\sum_{i=1}^{\infty}(1-p)^{i-1}p = p \sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{p}{1-p} \sum_{i=1}^{\infty}(1-p)^i = \frac{p}{1-p} \cdot \frac{1-p}{1-(1-p)} = 1
\]

Note that to obtain the second-last term we have used the fact that \( \sum_{i=1}^{\infty}c^i = \frac{c}{1-c}, |c| < 1. \)

Let’s now calculate the expectation of a geometric random variable, \( X \). We can do this in several ways. One way is to use the definition of expectation.

\[
E[X] = \sum_{i=0}^{\infty} i \Pr[X = i]
\]
\[
= \sum_{i=0}^{\infty} i(1-p)^{i-1}p
\]
\[
= \frac{p}{1-p} \sum_{i=0}^{\infty} i(1-p)^{i} = \left( \frac{p}{1-p} \right) \left( \frac{1-p}{(1-(1-p))^2} \right) \quad \left( \because \sum_{i=0}^{\infty} kx^k = \frac{x}{(1-x)^2}, \text{ for } |x| < 1. \right)
\]
\[
= \left( \frac{p}{1-p} \right) \left( \frac{1-p}{p^2} \right) = \frac{1}{p}
\]

Another way to compute the expectation is to note that \( X \) is a random variable that takes on non-negative values. From a theorem proved in last class we know that if \( X \) takes on only non-negative values then
\[
E[Y] = \sum_{i=1}^{\infty} \Pr[Y \geq i]
\]
Using this result we can calculate the expectation of the geometric random variable $X$. For the geometric random variable $X$ with parameter $p$, 

$$Pr[X \geq i] = \sum_{j=i}^{\infty} (1-p)^{j-1}p = (1-p)^{i-1}p \sum_{j=0}^{\infty} (1-p)^j = (1-p)^{i-1}p \times \frac{1}{1-(1-p)} = (1-p)^{i-1}$$

Therefore

$$E[X] = \sum_{i=1}^{\infty} Pr[X \geq i] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-p} \sum_{i=1}^{\infty} (1-p)^i = \frac{1}{1-p} \times \frac{1-p}{1-(1-p)} = \frac{1}{p}$$

**Binomial Distributions**

Consider an experiment in which we perform a sequence of $n$ coin flips in which the probability of obtaining heads is $p$. How many flips result in heads?

If $X$ denotes the number of heads that appear then

$$Pr[X = j] = \binom{n}{j} p^j (1-p)^{n-j}$$

**Definition.** A *binomial* random variable $X$ with parameters $n$ and $p$ is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$Pr[X = j] = \binom{n}{j} p^j (1-p)^{n-j}$$