Random Variables: Variance

We have seen in the previous note that if we take a biased coin that shows heads with probability p and toss it n times, then the expected number of heads is np. What this means is that if we repeat the experiment multiple times, where in each experiment we toss the coin n times, then on average we get np heads. But in any single experiment, the number of heads observed can be any value between 0 and n. What can we say about how far off we are from the expected value? That is, what is the typical deviation of the number of heads from np?

1 Random Walk

Let us consider a simpler setting that is equivalent to tossing a fair coin n times, but is more amenable to analysis. Suppose we have a particle that starts at position 0 and performs a random walk in one dimension. At each time step, the particle moves either one step to the right or one step to the left with equal probability (this kind of random walk is called *symmetric*), and the move at each time step is independent of all other moves. We think of these random moves as taking place according to whether a fair coin comes up heads or tails. The expected position of the particle after n moves is back at 0, but how far from 0 should we typically expect the particle to end up?

Denoting a right-move by $+1$ and a left-move by -1, we can describe the probability space here as the set of all sequences of length n over the alphabet $\{\pm 1\}$, each having equal probability $\frac{1}{2^n}$. Let the r.v. S_n denote the position of the particle (relative to our starting point 0) after n moves. Thus, we can write

$$S_n = X_1 + X_2 + \cdots + X_n, \tag{1}$$

where $X_i = +1$ if the i-th move is to the right and $X_i = -1$ if the move is to the left.

The expectation of S_n can be easily computed as follows. Since $\mathbb{E}[X_i] = (\frac{1}{2} \times 1) + (\frac{1}{2} \times (-1)) = 0$, applying linearity of expectation immediately gives $\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = 0$. But of course this is not very informative, and is due to the fact that positive and negative deviations from 0 cancel out.

What we are really asking is: What is the expected value of $|S_n|$, the *distance* of the particle from 0? Rather than consider the r.v. $|S_n|$, which is a little difficult to work with due to the absolute value operator, we will instead look at the r.v. S_n^2. Notice that this also has the effect of making all deviations from 0 positive, so it should also give a good measure of the distance from 0. However, because it is the *squared* distance, we will need to take a square root at the end.

We will now show that the expected square distance after n steps is equal to n:

Claim 16.1. *For the random variable S_n defined in* (1)*, we have $\mathbb{E}\left[S_n^2\right] = n$.*

Proof. We use the expression (1) and expand the square:

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[(X_1 + X_2 + \cdots + X_n)^2\right] = \mathbb{E}\left[\sum_{i=1}^n X_i^2 + 2\sum_{i<j} X_i X_j\right] = \sum_{i=1}^n \mathbb{E}\left[X_i^2\right] + 2\sum_{i<j} \mathbb{E}[X_i X_j]. \tag{2}$$

In the last equality we have used linearity of expectation. To proceed, we need to compute $\mathbb{E}[X_i^2]$ and $\mathbb{E}[X_iX_j]$ for $i \neq j$. Since X_i can take on only values ± 1, clearly $X_i^2 = 1$ always, so $\mathbb{E}[X_i^2] = 1$. To compute $\mathbb{E}[X_iX_j]$ for $i \neq j$, note $X_iX_j = +1$ when $X_i = X_j = +1$ or $X_i = X_j = -1$, and otherwise $X_iX_j = -1$. Therefore,

$$\begin{aligned}
\mathbb{P}[X_iX_j = 1] &= \mathbb{P}[(X_i = X_j = +1) \vee (X_i = X_j = -1)] \\
&= \mathbb{P}[X_i = X_j = +1] + \mathbb{P}[X_i = X_j = -1] \\
&= \mathbb{P}[X_i = +1] \times \mathbb{P}[X_j = +1] + \mathbb{P}[X_i = -1] \times \mathbb{P}[X_j = -1] \\
&= \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2},
\end{aligned}$$

where the second equality follows from the fact that the events $X_i = X_j = +1$ and $X_i = X_j = -1$ are mutually exclusive, while the third equality follows from the independence of the events $X_i = +1$ and $X_j = +1$, and likewise for the events $X_i = -1$ and $X_j = -1$. In a similar vein, one obtains $\mathbb{P}[X_iX_j = -1] = \frac{1}{2}$, and hence $\mathbb{E}[X_iX_j] = 0$.

Finally, plugging $\mathbb{E}[X_i^2] = 1$ and $\mathbb{E}[X_iX_j] = 0$, for $i \neq j$, into (2) gives $\mathbb{E}[S_n^2] = \sum_{i=1}^n 1 + 2\sum_{i<j} 0 = n$, as desired. □

So, for the symmetric random walk example, we see that <u>the expected squared distance from 0 is n</u>. One interpretation of this is that we might expect to be a distance of about $\sqrt{n}$ away from 0 after n steps. However, we have to be careful here: we **cannot** simply argue that $\mathbb{E}[|S_n|] = \sqrt{\mathbb{E}[S_n^2]} = \sqrt{n}$. (Why not?) We will see later in the course how to make precise deductions about $|S_n|$ from knowledge of $\mathbb{E}[S_n^2]$. For the moment, however, let us agree to view $\mathbb{E}[S_n^2]$ as an intuitive measure of "spread" of the r.v. S_n.

For a more general r.v. X with expectation $\mathbb{E}[X] = \mu$, what we are really interested in is $\mathbb{E}[(X-\mu)^2]$, the expected squared distance *from the mean*. In our symmetric random walk example, we had $\mu = 0$, so $\mathbb{E}[(X-\mu)^2]$ just reduced to $\mathbb{E}[X^2]$.

Definition 16.1 (Variance)**.** *For a r.v. X with expectation $\mathbb{E}[X] = \mu$, the <u>variance</u> of X is defined to be*

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right].$$

The square root $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ is called the <u>standard deviation</u> of X.

The point of taking the square root of variance is to put the standard deviation "on the same scale" as the r.v. itself. Since the variance and standard deviation differ just by a square, it really doesn't matter which one we choose to work with as we can always compute one from the other. We shall usually use the variance. For the random walk example above, we have that $\mathrm{Var}(S_n) = n$, and the standard deviation $\sigma(S_n)$ of X is $\sqrt{n}$.

The following observation provides a slightly different way to compute the variance, which sometimes turns out to be simpler.

Theorem 16.1. *For a r.v. X with expectation $\mathbb{E}[X] = \mu$, we have* $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mu^2$.

Proof. From the definition of variance, we have

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] = \mathbb{E}\left[X^2\right] - 2\mu\,\mathbb{E}[X] + \mu^2 = \mathbb{E}\left[X^2\right] - \mu^2.$$

In the third equality, we used linearity of expectation. We also used the fact that since $\mu = \mathbb{E}[X]$ is a constant, $\mathbb{E}[\mu X] = \mu\,\mathbb{E}[X] = \mu^2$ and $\mathbb{E}[\mu^2] = \mu^2$. □

Another important property that will come in handy is the following: For any random variable X and constant c, we have

$$\mathrm{Var}(cX) = c^2\mathrm{Var}(X).$$

The proof is simple and left as an exercise.

2 Variance Computation

Let us see some examples of variance calculations.

1. **Fair die.** Let X be the score on the roll of a single fair die. Recall from the previous note that $\mathbb{E}[X] = \frac{7}{2}$. So we just need to compute $\mathbb{E}\left[X^2\right]$, which is a routine calculation:

$$\mathbb{E}\left[X^2\right] = \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) = \frac{91}{6}.$$

Thus, from Theorem 16.1,

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

2. **Uniform distribution.** More generally, if X is a uniform random variable on the set $\{1,\ldots,n\}$, so X takes on values $1,\ldots,n$ with equal probability $\frac{1}{n}$, then the mean, variance and standard deviation of X are given by:

$$\mathbb{E}[X] = \frac{n+1}{2}, \qquad \mathrm{Var}(X) = \frac{n^2-1}{12}, \qquad \sigma(X) = \sqrt{\frac{n^2-1}{12}}. \tag{3}$$

You should verify these as an exercise.

3. **Fixed points of permutations.** Let X_n be the number of fixed points in a random permutation of n items (i.e., in the homework permutation example, X_n is the number of students in a class of size n who receive their own homework after shuffling). We saw in the previous note that $\mathbb{E}[X_n] = 1$, regardless of n. To compute $\mathbb{E}\left[X_n^2\right]$, write $X_n = I_1 + I_2 + \cdots + I_n$, where $I_i = 1$ if i is a fixed point, and $I_i = 0$ otherwise. Then as usual we have

$$\mathbb{E}\left[X_n^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[I_i^2\right] + 2\sum_{i<j} \mathbb{E}[I_i I_j]. \tag{4}$$

Since I_i is an indicator r.v., we have that $\mathbb{E}\left[I_i^2\right] = \mathbb{P}[I_i = 1] = \frac{1}{n}$. For $i<j$, since both I_i and I_j are indicators, we can compute $\mathbb{E}[I_i I_j]$ as follows:

$$\mathbb{E}[I_i I_j] = \mathbb{P}[I_i I_j = 1] = \mathbb{P}[I_i = 1 \wedge I_j = 1] = \mathbb{P}[\text{both } i \text{ and } j \text{ are fixed points}] = \frac{1}{n(n-1)}.$$

Make sure that you understand the last step here. Plugging this into equation (4) we get

$$\mathbb{E}\left[X_n^2\right] = \sum_{i=1}^{n} \frac{1}{n} + 2\sum_{i<j} \frac{1}{n(n-1)} = \left(n \times \frac{1}{n}\right) + \left[2\binom{n}{2} \times \frac{1}{n(n-1)}\right] = 1+1 = 2.$$

Thus, $\mathrm{Var}(X_n) = \mathbb{E}\left[X_n^2\right] - (\mathbb{E}[X_n])^2 = 2-1 = 1$. That is, the variance and the mean are both equal to 1. Like the mean, the variance is also independent of n. Intuitively at least, this means that it is unlikely that there will be more than a small number of fixed points even when the number of items, n, is very large.

2.1 Variance of a Geometric Random Variable [PROOF OPTIONAL]

Let us now compute the variance of X.

Theorem 16.2. *For* $X \sim \text{Geometric}(p)$*, we have* $\text{Var}(X) = \frac{1-p}{p^2}$.

Proof. We will show that $\mathbb{E}[X(X-1)] = \frac{2(1-p)}{p^2}$. Since we already know $\mathbb{E}[X] = \frac{1}{p}$, this will imply the desired result:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2(1-p)+p-1}{p^2} = \frac{1-p}{p^2}.\end{aligned}$$

Now to show $\mathbb{E}[X(X-1)] = \frac{2(1-p)}{p^2}$, we begin with the following identity of geometric series:

$$\sum_{i=0}^{\infty}(1-p)^i = \frac{1}{p}.$$

Differentiating the identity above with respect to p yields (the $i=0$ term is equal to 0 so we omit it):

$$-\sum_{i=1}^{\infty} i(1-p)^{i-1} = -\frac{1}{p^2}.$$

Differentiating both sides with respect to p again gives us (the $i=1$ term is equal to 0 so we omit it):

$$\sum_{i=2}^{\infty} i(i-1)(1-p)^{i-2} = \frac{2}{p^3}. \tag{5}$$

Now using the geometric distribution of X and identity (5), we can calculate:

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{i=1}^{\infty} i(i-1) \times \mathbb{P}[X=i] \\ &= \sum_{i=2}^{\infty} i(i-1)(1-p)^{i-1}p && \text{(the } i=1 \text{ term is equal to 0 so we omit it)} \\ &= p(1-p)\sum_{i=2}^{\infty} i(i-1)(1-p)^{i-2} \\ &= p(1-p) \times \frac{2}{p^3} && \text{(using identity (5))} \\ &= \frac{2(1-p)}{p^2},\end{aligned}$$

as desired. □

2.2 Variance of a Poisson Random Variable [PROOF OPTIONAL]

Theorem 16.3. *For* $X \sim \text{Poisson}(\lambda)$*, we have* $\text{Var}(X) = \lambda$.

Proof. Similarly, we can calculate $\mathbb{E}[X(X-1)]$ as follows:

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{i=0}^{\infty} i(i-1) \times \mathbb{P}[X=i] \\
&= \sum_{i=2}^{\infty} i(i-1)\,\frac{\lambda^i}{i!}\mathrm{e}^{-\lambda} && \text{(the } i=0 \text{ and } i=1 \text{ terms are equal to 0 so we omit them)} \\
&= \lambda^2 \mathrm{e}^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \\
&= \lambda^2 \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} && \text{(since } \mathrm{e}^{\lambda} = \textstyle\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j=i-2\text{)} \\
&= \lambda^2.
\end{aligned}$$

Therefore,

$$\mathrm{Var}(X) \;=\; \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \;=\; \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \;=\; \lambda^2 + \lambda - \lambda^2 \;=\; \lambda,$$

as desired. □