1 Application of a Geometric Random Variable: Coupon Collecting

We begin these notes with an application of a Geometric random variable.

Recall the coupon collecting problem from a previous note: we are trying to collect a set of n different baseball cards by buying boxes of cereal, each of which contains exactly one random card (uniformly distributed among the n cards). How many boxes do we need to buy until we have collected at least one copy of every card?

Let S_n denote the number of boxes we need to buy in order to collect all n cards. The distribution of S_n is difficult to compute directly (try it for $n = 3$!), and we did some calculations in a previous note to estimate the likely values of S_n. But if we are only interested in the expected value $\mathbb{E}[S_n]$, then we can evaluate it easily using linearity of expectation and what we have just learned about the geometric distribution.

We start by writing

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

for simpler random variables X_i, where X_i is the number of boxes we buy while trying to get the i-th new card (starting immediately after we have gotten the $(i-1)$st new card). With this definition, make sure you believe (1) before proceeding.

What does the distribution of X_i look like? Well, X_1 is trivial: no matter what happens, we always get a new card in the first box (since we have none to start with). So $\mathbb{P}[X_1 = 1] = 1$, and thus $\mathbb{E}[X_1] = 1$.

How about X_2? Each time we buy a box, we will get the same old card with probability $\frac{1}{n}$, and a new card with probability $\frac{n-1}{n}$. So we can think of buying boxes as flipping a biased coin with Heads probability $p = \frac{n-1}{n}$; then X_2 is just the number of tosses until the first Head appears. So X_2 has the geometric distribution with parameter $p = \frac{n-1}{n}$, and

$$\mathbb{E}[X_2] = \frac{n}{n-1}.$$

How about X_3? This is very similar to X_2 except that now we only get a new card with probability $\frac{n-2}{n}$ (since there are now two old ones). So X_3 has the geometric distribution with parameter $p = \frac{n-2}{n}$, and

$$\mathbb{E}[X_3] = \frac{n}{n-2}.$$

Arguing in the same way, we see that, for $i = 1, 2, \ldots, n$, X_i has the geometric distribution with parameter $p = \frac{n-i+1}{n}$, and hence that

$$\mathbb{E}[X_i] = \frac{n}{n-i+1}.$$

Finally, applying linearity of expectation to (1), we get

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{2} + \frac{n}{1} = n \sum_{i=1}^{n} \frac{1}{i}. \tag{2}$$

This is an exact expression for $\mathbb{E}[S_n]$. We can obtain a tidier form by noting that the sum in (2) actually has a very good approximation,[1] namely:

$$\sum_{i=1}^{n} \frac{1}{i} \approx \ln n + \gamma_E,$$

where $\gamma_E = 0.5772\ldots$ is known as *Euler's constant*.

Thus, the expected number of cereal boxes needed to collect n cards is about $n(\ln n + \gamma)$. This is an excellent approximation to the exact formula from (2) even for quite small values of n. So for example, for $n = 100$, we expect to buy about 518 boxes.

2 Expectation of a Function of a Random Variable

In the previous note, we saw that for any constant c, we have

$$\mathbb{E}[cX] = c\,\mathbb{E}[X].$$

In this note, we will consider more general functions of a random variable. For example, if we know the distribution of a random variable X, how can we compute the expectation of X^2?

Formally, let X be a random variable on a sample space Ω with probability distribution $\mathbb{P}_X$. Then for a function $f(\cdot)$ on the range of X, $f(X)$ is the random variable Y on the same sample space Ω, where $Y(\omega) = f(X(\omega))$ for $\omega \in \Omega$.

In terms of sample spaces, the event "$Y = y$" is equivalent to the event "$X \in f^{-1}(y)$". Here, we take $f^{-1}(y)$ to be the set $\{x \mid f(x) = y\}$. (When $f(\cdot)$ is one-to-one, it is a bit simpler, i.e., "$X = x$" is the same event as "$Y = f(x)$".)

With this view, the distribution of $Y = f(X)$, $\mathbb{P}_Y$, can be derived from the distribution of X as follows:

$$\mathbb{P}_Y[Y = y] = \sum_{x : f(x) = y} \mathbb{P}_X[X = x]. \tag{3}$$

From this, we observe that

$$\mathbb{E}[Y] = \sum_y y\ \mathbb{P}_Y[Y = y] = \sum_y \sum_{x : f(x) = y} y\ \mathbb{P}_X[X = x] = \sum_x f(x)\,\mathbb{P}_X[X = x].$$

The first equality is by the definition of expectation, the second is (3), and the last comes by observing that each value of x is included exactly once in the inner summation. Thus, we have the following identity:

$$\mathbb{E}[f(X)] = \sum_x f(x)\,\mathbb{P}_X[X = x].$$

In words: to compute the expectation of $f(X)$, we take the same weighted average as in the usual expectation of X, but replace the value x of X by $f(x)$. (This identity is sometimes called the Law of the Unconscious Statistician (LOTUS) as it is useful and used without much thought at times.)

An example that we will use extensively is the following. For any r.v. X, we have $\mathbb{E}\left[X^2\right] = \sum_x x^2\,\mathbb{P}_X[X = x]$. **Warning:** Note that $\mathbb{E}\left[X^2\right]$ is **not** the same as $(\mathbb{E}[X])^2$; and, more generally, $\mathbb{E}[f(X)]$ is **not** the same as $f(\mathbb{E}[X])$.

[1] This is another of the little tricks you might like to carry around in your toolbox.

Random Variables: Variance and Covariance

We have seen in the previous note that if we take a biased coin that comes up 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?

3 Random Walk

Let us consider a simpler setting that is equivalent to tossing a fair coin n times, but is easier to picture. 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

$$S_n = X_1 + X_2 + \cdots + X_n, \tag{4}$$

where $X_i = +1$ if the ith 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 squared distance after n steps is equal to n:

Proposition 20.1. *For the random variable S_n defined in (4), we have $\mathbb{E}\left[S_n^2\right] = n$.*

Proof. We use the expression from (4) and expand the square:

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[(X_1 + X_2 + \cdots + X_n)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + 2\sum_{i<j} \mathbb{E}[X_i X_j]. \tag{5}$$

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

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

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 $X_j = -1$. Since the only other possible value for X_iX_j is -1, we must also have $\mathbb{P}[X_iX_j = -1] = \frac{1}{2}$, and hence $\mathbb{E}[X_iX_j] = 0$. [**Note:** Later we will see a much simpler way to compute $\mathbb{E}[X_iX_j]$ using the fact that X_i, X_j are independent.]

Finally, plugging $\mathbb{E}\left[X_i^2\right] = 1$ and $\mathbb{E}[X_iX_j] = 0$, for $i \neq j$, into (5) gives $\mathbb{E}\left[S_n^2\right] = \sum_{i=1}^n 1 + 2\sum_{i<j} 0 = n$, as claimed. $\square$

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 $\mathbb{E}\left[S_n^2\right]$. For the moment, however, let us agree to view $\mathbb{E}\left[S_n^2\right]$ 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 $\mathbb{E}\left[(X-\mu)^2\right]$, the expected squared distance *from the mean*. In our symmetric random walk example, we have $\mathbb{E}[S_n] = \mu = 0$, so $\mathbb{E}\left[(S_n-\mu)^2\right]$ just reduces to $\mathbb{E}\left[S_n^2\right]$.

Definition 20.1 (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, Proposition 20.1 implies that $\mathrm{Var}(S_n) = n$, and the standard deviation is $\sigma(S_n) = \sqrt{n}$.

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

Theorem 20.1. *For a r.v. X with expectation $\mathbb{E}[X] = \mu$, we have* $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mu^2$.

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; note that, since $\mu = \mathbb{E}[X]$ is a constant, $\mathbb{E}[\mu X] = \mu\,\mathbb{E}[X] = \mu^2$ and $\mathbb{E}\left[\mu^2\right] = \mu^2$. $\square$

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

$$\operatorname{Var}(cX) = c^2 \operatorname{Var}(X). \tag{6}$$

Verifying this is straightforward and left as an **exercise**.

4 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 $\mathbb{E}[X^2]$, which is a routine calculation:

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

Thus, from Theorem 20.1,

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

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

$$\mathbb{E}[X] = \frac{n+1}{2}, \qquad \operatorname{Var}(X) = \frac{n^2-1}{12}, \qquad \sigma(X) = \sqrt{\frac{n^2-1}{12}}. \tag{7}$$

You should verify these as an **exercise**. [Recall that $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ and $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.]

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 $\mathbb{E}[X_n^2]$, 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

$$\mathbb{E}[X_n^2] = \sum_{i=1}^n \mathbb{E}[I_i^2] + 2\sum_{i<j} \mathbb{E}[I_i I_j]. \tag{8}$$

Since I_i is an indicator r.v., we have that $\mathbb{E}[I_i^2] = \mathbb{P}[I_i = 1] = \frac{1}{n}$. 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 (8) we get

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

Thus, $\operatorname{Var}(X_n) = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = 2 - 1 = 1$. 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.

4. **Variance of a Geometric random variable.**

Theorem 20.2. *For* $X \sim \text{Geometric}(p)$*, we have* $\text{Var}(X) = \frac{1-p}{p^2}$.

Proof. We use the formula

$$\text{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \mathbb{E}\left[X^2\right] - \frac{1}{p^2}. \tag{9}$$

Thus we just have to compute $\mathbb{E}\left[X^2\right]$, which we can do using a calculus trick as follows. Starting from

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

and differentiating with respect to p, we get

$$\sum_{i=1}^{\infty} i(1-p)^{i-1} = \frac{1}{p^2}. \tag{10}$$

Here we dropped the first term, for $i = 0$, since its derivative is zero, and multiplied both sides by -1. (Note incidentally that (10) gives an alternative proof of the fact that $\mathbb{E}[X] = \frac{1}{p}$; why?)

Now we multiply both sides of (10) by $(1-p)$ and differentiate again to get

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

Since the left-hand side here is exactly $\frac{1}{p}\mathbb{E}\left[X^2\right]$, we conclude that

$$\mathbb{E}\left[X^2\right] = \frac{2-p}{p^2}.$$

Finally, plugging this into (9) gives us

$$\text{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2},$$

as claimed. □

5. **Variance of a Poisson random variable.**

Theorem 20.3. *For a Poisson random variable* $X \sim \text{Poisson}(\lambda)$*, we have* $\text{Var}(X) = \lambda$.

Proof. We know from an earlier note that $\mathbb{E}[X] = \lambda$.

We can calculate $\mathbb{E}\left[X^2\right]$ as follows:

$$
\begin{aligned}
\mathbb{E}\left[X^2\right] &= \sum_{i=0}^{\infty} i^2 \times \mathbb{P}[X=i] \\
&= \sum_{i=1}^{\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} && \text{(the } i=0 \text{ term is equal to 0 so we omit it)} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} i \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \left[\sum_{i=2}^{\infty} \frac{\lambda^{i-1}}{(i-2)!} + \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \right] && \text{(replacing } i \text{ by } (i-1)+1\text{)} \\
&= \lambda e^{-\lambda} \left[\lambda e^{\lambda} + e^{\lambda} \right] && \text{(using } e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ twice, with } j=i-1 \text{ and } j=i-2\text{)} \\
&= \lambda^2 + \lambda.
\end{aligned}
$$

Therefore,

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

as desired. □

5 Sum of Independent Random Variables

One of the most important and useful facts about variance is that if a random variable X is the sum of *independent* random variables $X = X_1 + \cdots + X_n$, then its variance is the sum of the variances of the individual r.v.'s. In particular, if the individual r.v.'s X_i are identically distributed (i.e., they have the same distribution), then $\operatorname{Var}(X) = \sum_{i=1}^{n} \operatorname{Var}(X_i) = n \cdot \operatorname{Var}(X_1)$. This means that the standard deviation is $\sigma(X) = \sqrt{n} \cdot \sigma(X_1)$. Note that by contrast, the expected value is $\mathbb{E}[X] = n \cdot \mathbb{E}[X_1]$. Intuitively this means that whereas the average value of X grows proportionally to n, the spread of the distribution grows proportionally to $\sqrt{n}$, which is much smaller than n. In other words, the distribution of X tends to *concentrate* around its mean.

Let us now formalize these ideas. First, we have the following result which states that the expected value of the product of two *independent* random variables is equal to the product of their expected values.

Theorem 20.4. *For* independent *random variables X, Y, we have* $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Proof. We have

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_a \sum_b ab \times \mathbb{P}[X=a, Y=b] \\
&= \sum_a \sum_b ab \times \mathbb{P}[X=a] \times \mathbb{P}[Y=b] \\
&= \left(\sum_a a \times \mathbb{P}[X=a] \right) \times \left(\sum_b b \times \mathbb{P}[Y=b] \right) \\
&= \mathbb{E}[X] \times \mathbb{E}[Y],
\end{aligned}
$$

as required. In the second line here we made crucial use of independence. □

We now use the above theorem to conclude the nice property that the variance of the sum of independent random variables is equal to the sum of their variances.

Theorem 20.5. *For* independent *random variables X,Y, we have*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Proof. From the alternative formula for variance in Theorem 20.1 and linearity of expectation, we have

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 \\ &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\,\mathbb{E}[XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= (\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2) + (\mathbb{E}\left[Y^2\right] - \mathbb{E}[Y]^2) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]).\end{aligned}$$

Since X,Y are independent, Theorem 20.4 implies that the final term in this expression is zero. □

It is very important to remember that **neither** of the above two results is true in general when X,Y are not independent. As a simple example, note that even for a $\{0,1\}$-valued r.v. X with $\mathbb{P}[X=1] = p$, $\mathbb{E}\left[X^2\right] = p$ is not equal to $\mathbb{E}[X]^2 = p^2$ (because of course X and X are not independent!). This is in contrast to linearity of expectation, where we saw that the expectation of a sum of r.v.'s is the sum of the expectations of the individual r.v.'s, regardless of whether or not the r.v.'s are independent.

Example

Let us return to our motivating example of a sequence of n coin tosses. Let X_n denote the number of Heads in n tosses of a biased coin with Heads probability p (i.e., $X_n \sim \mathrm{Binomial}(n,p)$). As usual, we write $X_n = I_1 + I_2 + \cdots + I_n$, where $I_i = 1$ if the ith toss is H, and $I_i = 0$ otherwise.

We already know $\mathbb{E}[X_n] = \sum_{i=1}^n \mathbb{E}[I_i] = np$. We can compute $\mathrm{Var}(I_i) = \mathbb{E}\left[I_i^2\right] - \mathbb{E}[I_i]^2 = p - p^2 = p(1-p)$. Since the I_i's are independent, by Theorem 20.5 we get $\mathrm{Var}(X_n) = \sum_{i=1}^n \mathrm{Var}(I_i) = np(1-p)$.

As an example, for a fair coin ($p = \frac{1}{2}$) the expected number of Heads in n tosses is $\frac{n}{2}$, and the standard deviation is $\sqrt{\frac{n}{4}} = \frac{\sqrt{n}}{2}$. Note that since the maximum number of Heads is n, the standard deviation is much less than this maximum number for large n. This is in contrast to the previous example of the uniformly distributed random variable in (7), where the standard deviation $\sigma(X) = \sqrt{\frac{n^2-1}{12}} \approx \frac{n}{\sqrt{12}}$ (for large n) is of the same order as the largest value, n. In this sense, the spread of a binomially distributed r.v. is much smaller than that of a uniformly distributed r.v.

6 Exercises

1. For any random variable X and constant c, show that $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$.
2. For $X \sim \mathrm{Uniform}\{1,\ldots,n\}$, show that $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ are as shown in (7).