Some Important Distributions

Recall our basic probabilistic experiment of tossing a biased coin n times. This is a very simple model, yet surprisingly powerful. Many important probability distributions that are widely used to model real-world phenomena can be derived from looking at this basic coin tossing model.

The first example, which we have seen in Note 16, is the *binomial distribution Bin(n, p)*. This is the distribution of the number of Heads, S_n, in n tosses of a biased coin with probability p to be Heads. Recall that the distribution of S_n is $\Pr[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$. The expected value is $\mathrm{E}(S_n) = np$ and the variance is $\mathrm{Var}(S_n) = np(1-p)$. The binomial distribution frequently appears to model the number of successes in a repeated experiment.

Geometric Distribution

Consider tossing a biased coin with Heads probability p repeatedly. Let X denote the number of tosses until the first Head appears. Then X is a random variable that takes values in the set of positive integers $\{1, 2, 3, \ldots\}$. The event that $X = i$ is equal to the event of observing Tails for the first $i-1$ tosses and getting Heads in the i-th toss, which occurs with probability $(1-p)^{i-1}p$. Such a random variable is called a geometric random variable.

The geometric distribution frequently occurs in applications because we are often interested in how long we have to wait before a certain event happens: how many runs before the system fails, how many shots before one is on target, how many poll samples before we find a Democrat, how many retransmissions of a packet before successfully reaching the destination, etc.

Definition 19.1 (Geometric distribution). *A random variable X for which*

$$\Pr[X = i] = (1-p)^{i-1}p \qquad \text{for } i = 1, 2, 3, \ldots$$

is said to have the geometric distribution with parameter p. This is abbreviated as $X \sim Geom(p)$.

As a sanity check, we can verify that the total probability of X is equal to 1:

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

where in the second-to-last step we have used the formula for geometric series.

If we plot the distribution of X (i.e., the values $\Pr[X = i]$ against i) we get a curve that decreases monotonically by a factor of $1-p$ at each step, as shown in Figure 1.

Let us now compute the expectation $\mathrm{E}(X)$. Applying the definition of expected value directly gives us:

$$\mathrm{E}(X) = \sum_{i=1}^{\infty} i \times \Pr[X = i] = p \sum_{i=1}^{\infty} i(1-p)^{i-1}.$$
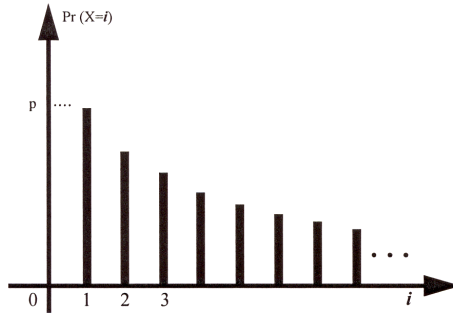
Figure 1: The Geometric distribution.

However, the final summation is difficult to evaluate and requires some calculus trick. Instead, we will use the following alternative formula for expectation.

Theorem 19.1. *Let X be a random variable that takes values in $\{0,1,2,\ldots\}$. Then*

$$E(X) = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

Proof. For notational convenience, let's write $p_i = \Pr[X = i]$, for $i = 0,1,2,\ldots$. From the definition of expectation, we have

$$\begin{aligned}
E(X) &= (0 \times p_0) + (1 \times p_1) + (2 \times p_2) + (3 \times p_3) + (4 \times p_4) + \cdots \\
&= p_1 + (p_2 + p_2) + (p_3 + p_3 + p_3) + (p_4 + p_4 + p_4 + p_4) + \cdots \\
&= (p_1 + p_2 + p_3 + p_4 + \cdots) + (p_2 + p_3 + p_4 + \cdots) + (p_3 + p_4 + \cdots) + (p_4 + \cdots) + \cdots \\
&= \Pr[X \geq 1] + \Pr[X \geq 2] + \Pr[X \geq 3] + \Pr[X \geq 4] + \cdots.
\end{aligned}$$

In the third line, we have regrouped the terms into convenient infinite sums, and each infinite sum is exactly the probability that $X \geq i$ for each i. You should check that you understand how the fourth line follows from the third.

Let us repeat the proof more formally, this time using more compact mathematical notation:

$$E(X) = \sum_{j=1}^{\infty} j \times \Pr[X = j] = \sum_{j=1}^{\infty} \sum_{i=1}^{j} \Pr[X = j] = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \Pr[X = j] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

$\square$

We can now use Theorem 19.1 to compute $E(X)$ more easily.

Theorem 19.2. *For a geometric random variable $X \sim Geom(p)$, we have $E(X) = \frac{1}{p}$.*

Proof. The key observation is that for a geometric random variable X,

$$\Pr[X \geq i] = (1-p)^{i-1} \quad \text{for } i = 1,2,\ldots. \tag{1}$$

We can obtain this simply by summing $\Pr[X = j]$ for $j \geq i$. Another way of seeing this is to note that the event "$X \geq i$" means at least i tosses are required. This is equivalent to saying that the first $i - 1$ tosses are all

Tails, and the probability of this event is precisely $(1-p)^{i-1}$. Now, plugging equation (1) into Theorem 19.1, we get

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

where we have used the formula for geometric series. $\qquad \square$

So, the expected number of tosses of a biased coin until the first Head appears is $\frac{1}{p}$. Intuitively, if in each coin toss we expect to get p Heads, then we need to toss the coin $\frac{1}{p}$ times to get 1 Head. So for a fair coin, the expected number of tosses is 2, but remember that the actual number of coin tosses that we need can be any positive integers.

Remark: Another way of deriving $E(X) = \frac{1}{p}$ is to use the interpretation of a geometric random variable X as the number of coin tosses until we get a Head. Consider what happens in the first coin toss: If the first toss comes up Heads, then $X = 1$. Otherwise, we have used one toss, and we repeat the coin tossing process again; the number of coin tosses after the first toss is again a geometric random variable with parameter p. Therefore, we can calculate:

$$E(X) = \underbrace{p \cdot 1}_{\text{first toss is H}} + \underbrace{(1-p) \cdot (1 + E(X))}_{\text{first toss is T, then toss again}}.$$

Solving for $E(X)$ yields $E(X) = \frac{1}{p}$, as claimed.

Let us now compute the variance of X.

Theorem 19.3. *For a geometric random variable $X \sim Geom(p)$, we have $\text{Var}(X) = \frac{1-p}{p^2}$.*

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

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

Now to show $E(X(X-1)) = \frac{2(1-p)}{p^2}$, 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):

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

Now using the geometric distribution of X and identity (2), we can calculate:

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

as desired. $\qquad \square$

Application: The Coupon Collector's Problem

Suppose we are trying to collect a set of n different baseball cards. We get the cards by buying boxes of cereal: each box contains exactly one card, and it is equally likely to be any of the n cards. How many boxes do we need to buy until we have collected at least one copy of every card?

Let X denote the number of boxes we need to buy in order to collect all n cards. The distribution of X is difficult to compute directly (try it for $n=3$). But if we are only interested in its expected value $E(X)$, then we can evaluate it easily using linearity of expectation and what we have just learned about the geometric distribution.

As usual, we start by writing

$$
X = X_1 + X_2 + \ldots + X_n \tag{3}
$$

for suitable simple random variables X_i. What should the X_i be? Naturally, X_i is the number of boxes we buy while trying to get the i-th new card (starting immediately after we've got the $(i-1)$-st new card). With this definition, make sure you believe equation (3) 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 $\Pr[X_1 = 1] = 1$, and thus $E(X_1) = 1$.

How about X_2? Each time we buy a box, we'll 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_1 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

$$
E(X_2) = \frac{n}{n-1}.
$$

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

$$
E(X_3) = \frac{n}{n-2}.
$$

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

$$
E(X_i) = \frac{n}{n-i+1}.
$$

Finally, applying linearity of expectation to equation (3), we get

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{2} + \frac{n}{1} = n\sum_{i=1}^{n} \frac{1}{i}. \tag{4}$$

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

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

where $\gamma = 0.5772\ldots$ 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 (4) even for quite small values of n. So for example, for $n = 100$, we expect to buy about 518 boxes.

Poisson Distribution

Consider the number of clicks of a Geiger counter, which measures radioactive emissions. The average number of such clicks per unit time, λ, is a measure of radioactivity, but the actual number of clicks fluctuates according to a certain distribution called the Poisson distribution. What is remarkable is that the average value, λ, completely determines the probability distribution on the number of clicks X.

Definition 19.2 (Poisson distribution). *A random variable X for which*

$$\Pr[X = i] = \frac{\lambda^i}{i!} e^{-\lambda} \qquad \text{for } i = 0, 1, 2, \ldots \tag{5}$$

is said to have the Poisson distribution with parameter λ. This is abbreviated as $X \sim Poiss(\lambda)$.

To make sure this is a valid definition, let us check that (5) is in fact a distribution, i.e., that the probabilities sum to 1. We have

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \times e^{\lambda} = 1.$$

In the second-to-last step, we used the Taylor series expansion $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$.

The Poisson distribution is also a very widely accepted model for so-called "rare events," such as misconnected phone calls, radioactive emissions, crossovers in chromosomes, the number of cases of disease, the number of births per hour, etc. This model is appropriate whenever the occurrences can be assumed to happen randomly with some constant density in a continuous region (of time or space), such that occurrences in disjoint subregions are independent. One can then show that the number of occurrences in a region of unit size should obey the Poisson distribution with parameter λ.

Example: Suppose when we write an article, we make an average of 1 typo per page. We can model this with a Poisson random variable X with $\lambda = 1$. So the probability that a page has 5 typos is

$$\Pr[X = 5] = \frac{1^5}{5!} e^{-1} = \frac{1}{120\,e} \approx \frac{1}{326}.$$

---

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

Now suppose the article has 200 pages. If we assume the number of typos in each page is independent, then the probability that there is at least one page with exactly 5 typos is

$$\Pr[\exists \text{ a page with exactly 5 typos}] = 1 - \Pr[\text{every page has} \neq 5 \text{ typos}]$$

$$= 1 - \prod_{k=1}^{200} \Pr[\text{page } k \text{ has} \neq 5 \text{ typos}]$$

$$= 1 - \prod_{k=1}^{200} (1 - \Pr[\text{page } k \text{ has exactly 5 typos}])$$

$$= 1 - \left(1 - \frac{1}{120\,e}\right)^{200},$$

where in the last step we have used our earlier calculation for $\Pr[X = 5]$. $\qquad\square$

Let us now calculate the expectation and variance of a Poisson random variable. As we noted before, the expected value is simply λ. Here we see that the variance is also equal to λ.

Theorem 19.4. *For a Poisson random variable $X \sim Poiss(\lambda)$, we have $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.*

Proof. We can calculate $\mathrm{E}(X)$ directly from the definition of expectation:

$$\mathrm{E}(X) = \sum_{i=0}^{\infty} i \times \Pr[X = i]$$

$$= \sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} \qquad \text{(the } i = 0 \text{ term is equal to 0 so we omit it)}$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda e^{-\lambda} e^{\lambda} \qquad \text{(since } e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i - 1)$$

$$= \lambda.$$

Similarly, we can calculate $\mathrm{E}(X(X-1))$ as follows:

$$\mathrm{E}(X(X-1)) = \sum_{i=0}^{\infty} i(i-1) \times \Pr[X = i]$$

$$= \sum_{i=2}^{\infty} i(i-1) \frac{\lambda^i}{i!} e^{-\lambda} \qquad \text{(the } i = 0 \text{ and } i = 1 \text{ terms are equal to 0 so we omit them)}$$

$$= \lambda^2 e^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!}$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} \qquad \text{(since } e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i - 2)$$

$$= \lambda^2.$$

Therefore,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - \mathrm{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

as desired. $\qquad\square$

A plot of the Poisson distribution reveals a curve that rises monotonically to a single peak and then decreases monotonically. The peak is as close as possible to the expected value, i.e., at $i = \lfloor \lambda \rfloor$. Figure 2 shows an example for $\lambda = 5$.
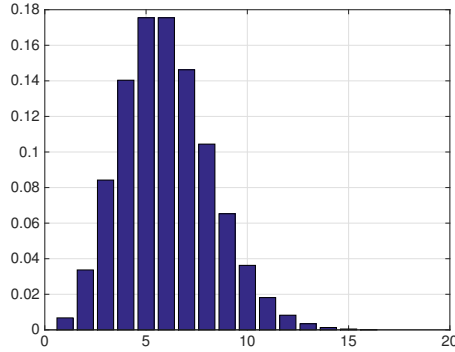

Figure 2: The Poisson distribution with $\lambda = 5$.

Poisson and Coin Flips

To see a concrete example of how Poisson distribution arises, suppose we want to model the number of cell phone users initiating calls in a network during a time period, of duration (say) 1 minute. There are many customers in the network, and all of them can potentially make a call during this time period. However, only a very small fraction of them actually will. Under this scenario, it seems reasonable to make two assumptions:

- The probability of having more than 1 customer initiating a call in any small time interval is negligible.

- The initiation of calls in disjoint time intervals are independent events.

Then if we divide the one-minute time period into n disjoint intervals, than the number of calls X in that time period can be modeled as a binomial random variable with parameter n and probability of success p, i.e., p is the probability of having a call initiated in a time interval of length $1/n$. But what should p be in terms of the relevant parameters of the problem? If calls are initiated at an average rate of λ calls per minute, then $\mathrm{E}(X) = \lambda$ and so $np = \lambda$, i.e., $p = \lambda/n$. So $X \sim Bin(n, \frac{\lambda}{n})$. As we shall see below, as we let n tend to infinity, this distribution tends to the Poisson distribution with parameter λ. We can also see why the Poisson distribution is a model for "rare events." We are thinking of it as a sequence of a large number, n, of coin flips, where we expect only a finite number λ of Heads.

Now we will prove that the Poisson distribution $Poiss(\lambda)$ is the limit of the binomial distribution $Bin(n, \frac{\lambda}{n})$, as n tends to infinity.

Theorem 19.5. *Let $X \sim Bin(n, \frac{\lambda}{n})$ where $\lambda > 0$ is a fixed constant. Then for every $i = 0, 1, 2, \ldots$,*

$$\Pr[X = i] \longrightarrow \frac{\lambda^i}{i!} e^{-\lambda} \quad as \ n \to \infty.$$

That is, the probability distribution of X converges to the Poisson distribution with parameter λ.

Proof. Fix $i \in \{0, 1, 2, \dots\}$, and assume $n \geq i$ (because we will let $n \to \infty$). Then, because X has binomial distribution with parameter n and $p = \frac{\lambda}{n}$,

$$\Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}.$$

Let us collect the factors into

$$\Pr[X = i] = \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}. \tag{6}$$

The first parenthesis above becomes, as $n \to \infty$,

$$\frac{n!}{(n-i)!} \cdot \frac{1}{n^i} = \frac{n \cdot (n-1) \cdots (n-i+1) \cdot (n-i)!}{(n-i)!} \cdot \frac{1}{n^i} = \frac{n}{n} \cdot \frac{(n-1)}{n} \cdots \frac{(n-i+1)}{n} \to 1.$$

From calculus, the second parenthesis in (6) becomes, as $n \to \infty$,

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}.$$

And since i is fixed, the third parenthesis in (6) becomes, as $n \to \infty$,

$$\left(1 - \frac{\lambda}{n}\right)^{-i} \to (1-0)^{-i} = 1.$$

Substituting these results back to (6) gives us

$$\Pr[X = i] \to \frac{\lambda^i}{i!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^i}{i!} e^{-\lambda},$$

as desired. $\qquad \square$