Some Important Distributions

In this note we will introduce three important probability distributions that are widely used to model real-world phenomena. The first of these which we already learned about in the last Lecture Note 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. As we saw, $P[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$. $E[S_n] = np$, $Var[S_n] = np(1-p)$ and $\sigma(S_n) = \sqrt{np(1-p)}$.

Geometric Distribution

Question: A biased coin with Heads probability p is tossed repeatedly until the first Head appears. What is the distribution and the expected number of tosses?

As always, our first step in answering the question must be to define the sample space Ω. A moment's thought tells us that

$$\Omega = \{H, TH, TTH, TTTH, \ldots\},$$

i.e., Ω consists of all sequences over the alphabet $\{H,T\}$ that end with H and contain no other H's. This is our first example of an *infinite* sample space (though it is still discrete).

What is the probability of a sample point, say $\omega = TTH$? Since successive coin tosses are independent (this is implicit in the statement of the problem), we have

$$\Pr[TTH] = (1-p) \times (1-p) \times p = (1-p)^2 p.$$

And generally, for any sequence $\omega \in \Omega$ of length i, we have $\Pr[\omega] = (1-p)^{i-1}p$. To be sure everything is consistent, we should check that the probabilities of all the sample points add up to 1. Since there is exactly one sequence of each length $i \geq 1$ in Ω, we have

$$\sum_{\omega \in \Omega} \Pr[\omega] = \sum_{i=1}^{\infty} (1-p)^{i-1} p = p \sum_{i=0}^{\infty} (1-p)^i = p \times \frac{1}{1-(1-p)} = 1,$$

as expected. [In the second-last step here, we used the formula for summing a geometric series.]

Now let the random variable X denote the number of tosses in our sequence (i.e., $X(\omega)$ is the length of ω). Its distribution has a special name: it is called the *geometric distribution with parameter p* (where p is the probability that the coin comes up Heads on each toss).

Definition 15.1 (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 $X \sim Geom(p)$.

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. See Figure 1.

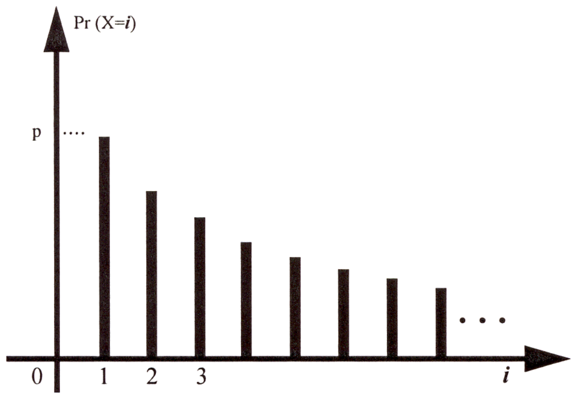

Figure 1: The Geometric distribution.

Our next goal is to compute $E(X)$. Despite the fact that X counts something, there's no obvious way to write it as a sum of simple r.v.'s as we did in many examples in an earlier lecture note. (Try it!) In a later lecture, we will give a slick way to do this calculation. For now, let's just dive in and try a direct computation of $E(X)$. Note that the distribution of X is quite simple:

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

So from the definition of expectation we have

$$E(X) = (1 \times p) + (2 \times (1-p)p) + (3 \times (1-p)^2 p) + \cdots \ = p\sum_{i=1}^{\infty} i(1-p)^{i-1}.$$

This series is a blend of an arithmetic series (the i part) and a geometric series (the $(1-p)^{i-1}$ part). There are several ways to sum it. Here is one way, using an auxiliary trick (given in the following Theorem) that is often very useful. [Ask your TA about other ways.]

Theorem 15.1: Let X be a random variable that takes on only non-negative integer values. Then

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

Proof: For notational convenience, let's write $p_i = \Pr[X = i]$, for $i = 0,1,2,\dots$. 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. You should check that you understand how the fourth line follows from the third.

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

$$\sum_{i=1}^{\infty} \Pr[X \geq i] = \sum_{i=1}^{\infty}\sum_{j \geq i} \Pr[X = j] = \sum_{j=1}^{\infty}\sum_{i \leq j} \Pr[X = j] = \sum_{j=1}^{\infty} j \Pr[X = j] = E[X]$$

□

Using Theorem 15.1, it is easy to compute $\mathrm{E}(X)$. The key observation is that, for our coin-tossing r.v. X,

$$\Pr[X \geq i] = (1-p)^{i-1}. \tag{1}$$

Why is this? Well, the event "$X \geq i$" means that at least i tosses are required. This is exactly 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 15.1, we get

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

So, the expected number of tosses of a biased coin until the first Head appears is $\frac{1}{p}$. For a fair coin, the expected number of tosses is 2.

For posterity, let's record two important facts we've learned about the geometric distribution:

Theorem 15.2: For a random variable X having the geometric distribution with parameter p,

1. $\mathrm{E}(X) = \frac{1}{p}$; and
2. $\Pr[X \geq i] = (1-p)^{i-1}$ for $i = 1, 2, \ldots$.

The next section discusses a rather more involved application, which is important in its own right.

The Coupon Collector's Problem

Question: 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?

The sample space here is similar in flavor to that for our previous coin-tossing example, though rather more complicated. It consists of all sequences ω over the alphabet $\{1, 2, \ldots, n\}$, such that

1. ω contains each symbol $1, 2, \ldots, n$ at least once; and
2. the final symbol in ω occurs only once.

[Check that you understand this!] For any such ω, the probability is just $\Pr[\omega] = \frac{1}{n^i}$, where i is the length of ω (why?). However, it is very hard to figure out how many sample points ω are of length i (try it for the case $n = 3$). So we will have a hard time figuring out the distribution of the random variable X, which is the length of the sequence (i.e., the number of boxes bought).

Fortunately, we can compute the expectation $\mathrm{E}(X)$ very easily, using (guess what?) linearity of expectation, plus the fact we have just learned about the expectation of the geometric distribution. As usual, we would like to write

$$X = X_1 + X_2 + \ldots + X_n \tag{2}$$

for suitable simple random variables X_i. But what should the X_i be? A natural thing to try is to make X_i equal to the number of boxes we buy while trying to get the ith new card (starting immediately after we've got the $(i-1)$st new card). With this definition, make sure you believe equation (2) 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 $\mathrm{E}(X_1) = 1$.

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

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

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

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

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

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

Finally, applying linearity of expectation to equation (2), we get

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{2} + \frac{n}{1} = n \sum_{i=1}^{n} \frac{1}{i}. \tag{3}$$

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 *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 (3) even for quite small values of n. So for example, for $n = 100$, we expect to buy about 518 boxes.

The 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 15.2 (Poisson distribution): A random variable X for which

$$\Pr[X = i] = \frac{\lambda^i}{i!} e^{-\lambda} \qquad \text{for } i = 0, 1, 2, \ldots \tag{4}$$

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

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

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

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.

To make sure this definition is valid, we had better check that (4) is in fact a distribution, i.e., that the probabilities sum to 1. We have

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

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

What is the expectation of a Poisson random variable X? This is a simple hands-on calculation, starting from the definition of expectation:

$$\begin{aligned} \mathrm{E}(X) &= \textstyle\sum_{i=0}^{\infty} i \times \Pr[X = i] \\ &= \textstyle\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} \mathrm{e}^{-\lambda} \\ &= \mathrm{e}^{-\lambda} \textstyle\sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} \\ &= \lambda \mathrm{e}^{-\lambda} \textstyle\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} \\ &= \lambda. \end{aligned}$$

So the expectation of a Poisson r.v. X with parameter λ is $\mathrm{E}(X) = \lambda$.

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 $i = \lfloor \lambda \rfloor$.

0.1 Poisson and Coin Flips

To see this in a concrete setting, 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 paying 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 binomially distributed with parameter n and probability of success p, 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 $E(X) = \lambda$ and so $np = \lambda$, i.e. $p = \lambda/n$. So $X \sim Bin(n, \lambda/n)$. 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 is the limit of the binomial distribution as $n \to \infty$. Let's look in more detail at the distribution of X as defined above. Recall that X is defined to be a special case of the binomial distribution, with $p = \frac{\lambda}{n}$ and $n \to \infty$. For convenience, we'll write $p_i = \Pr[X = i]$ for $i = 0, 1, 2, \ldots$.

We will first determine what the distribution is (i.e. the values of p_i) and then show that the expectation is λ.

Beginning with p_0, we have

$$p_0 = \left(1 - \frac{\lambda}{n}\right)^n \to \mathrm{e}^{-\lambda} \quad \text{as } n \to \infty.$$

What about the other p_i? Well, we know from the binomial distribution that $p_i = \binom{n}{i}(\frac{\lambda}{n})^i(1-\frac{\lambda}{n})^{n-i}$. Since we know how p_0 behaves, let's look at the ratio $\frac{p_1}{p_0}$:

$$\frac{p_1}{p_0} = \frac{n \times \frac{\lambda}{n} \times (1-\frac{\lambda}{n})^{n-1}}{(1-\frac{\lambda}{n})^n} = \frac{\lambda}{1-\frac{\lambda}{n}} = \frac{n\lambda}{n-\lambda} \to \lambda \quad \text{as } n \to \infty.$$

[Recall that we are assuming λ is a constant.] So, since $p_0 \to \mathrm{e}^{-\lambda}$, we see that $p_1 \to \lambda \mathrm{e}^{-\lambda}$ as $n \to \infty$.

Now let's look at the ratio $\frac{p_2}{p_1}$:

$$\frac{p_2}{p_1} = \frac{\binom{n}{2} \times (\frac{\lambda}{n})^2 \times (1-\frac{\lambda}{n})^{n-2}}{n \times (\frac{\lambda}{n}) \times (1-\frac{\lambda}{n})^{n-1}} = \frac{n-1}{2} \times \frac{\lambda}{n} \times \frac{1}{(1-\frac{\lambda}{n})} = \frac{n-1}{n-\lambda} \times \frac{\lambda}{2} \to \frac{\lambda}{2} \quad \text{as } n \to \infty.$$

So $p_2 \to \frac{\lambda^2}{2}\mathrm{e}^{-\lambda}$ as $n \to \infty$.

For each value of i, something very similar happens to the ratio $\frac{p_i}{p_{i-1}}$:

$$\frac{p_i}{p_{i-1}} = \frac{\binom{n}{i} \times (\frac{\lambda}{n})^i \times (1-\frac{\lambda}{n})^{n-i}}{\binom{n}{i-1}(\frac{\lambda}{n})^{i-1}(1-\frac{\lambda}{n})^{n-i+1}} = \frac{n-i+1}{i} \times \frac{\lambda}{n} \times \frac{n}{n-\lambda} = \frac{n-i+1}{n-\lambda} \times \frac{\lambda}{i} \to \frac{\lambda}{i} \quad \text{as } n \to \infty.$$

Putting this together, we see that, for each fixed value i,

$$p_i \to \frac{\lambda^i}{i!}\mathrm{e}^{-\lambda} \qquad \text{as } n \to \infty.$$

This is exactly according to our definition of the Poisson distribution, where $\Pr[X = i] = \frac{\lambda^i}{i!}\mathrm{e}^{-\lambda}$.

We have seen that the Poisson distribution arises as the limit of the number of balls in bin 1 when n balls are thrown into $\frac{n}{\lambda}$ bins. In other words, it is the limit of the binomial distribution with parameters n and $p = \frac{\lambda}{n}$ as $n \to \infty$, λ being a fixed constant.