

Geometric and Poisson 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 is the Binomial(n, p) distribution, introduced earlier. 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 $\mathbb{P}[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \dots, n\}$. The expected value is $\mathbb{E}[S_n] = np$ and the variance is $\text{Var}(S_n) = np(1-p)$. The binomial distribution is frequently used to model the number of successes in a repeated experiment.

1 Geometric Distribution

Consider repeatedly tossing a biased coin with Heads probability p . Let X denote the number of tosses *until the first Head appears*. Then X is a random variable that takes values in \mathbb{Z}^+ , the set of positive integers. 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. Note that X is unbounded: i.e., for any positive integer i (no matter how large), there is some positive probability that $X = i$.

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

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

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

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

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

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

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

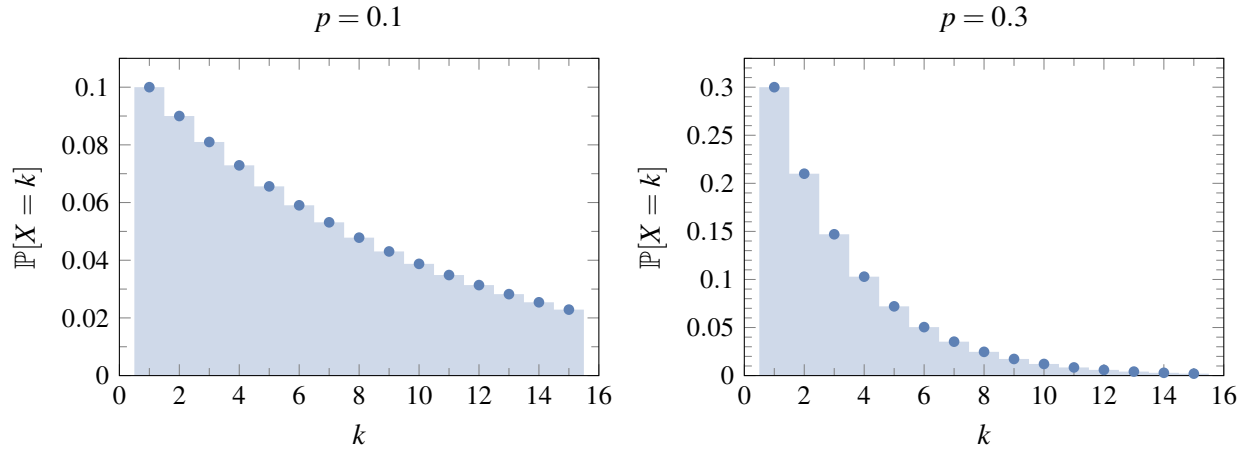


Figure 1: Illustration of the $\text{Geometric}(p)$ distribution for $p = 0.1$ and $p = 0.3$.

1.1 Mean and Variance of a Geometric Random Variable

Let us now compute the expectation $\mathbb{E}[X]$. Applying the definition of expected value directly gives us:

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \times \mathbb{P}[X = i] = p \sum_{i=1}^{\infty} i(1-p)^{i-1}.$$

However, the final summation is a little tricky to evaluate (though it can be done). Instead, we will use the following alternative formula for expectation that simplifies the calculation and is useful in its own right.

Theorem 19.1 (Tail Sum Formula). *Let X be a random variable that takes values in $\{0, 1, 2, \dots\}$. Then*

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i].$$

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

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

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:

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} j \times \mathbb{P}[X = j] = \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{P}[X = j] = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \mathbb{P}[X = j] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i],$$

where the third equality follows from interchanging the order of summations. □

We can now use Theorem 19.1 to compute $\mathbb{E}[X]$ more easily.

Theorem 19.2. For $X \sim \text{Geometric}(p)$, we have $\mathbb{E}[X] = \frac{1}{p}$.

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

$$\mathbb{P}[X \geq i] = (1-p)^{i-1} \quad \text{for } i = 1, 2, \dots \quad (1)$$

We can obtain this simply by summing $\mathbb{P}[X = j]$ 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

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

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

So, the expected number of tosses of a biased coin until the first Head appears is $\frac{1}{p}$. Thus for a fair coin, the expected number of tosses until the first Head is 2 (but of course the actual number of tosses can be any positive integer).

Let us now compute the variance of X .

Theorem 19.3. For $X \sim \text{Geometric}(p)$, we have $\text{Var}(X) = \frac{1-p}{p^2}$.

Proof. We use the formula

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

Thus we just have to compute $\mathbb{E}[X^2]$, 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

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

Here we dropped the first term, for $i=0$, since its derivative is zero, and multiplied both sides by -1 . (Note incidentally that (3) gives an alternative proof of the fact that $\mathbb{E}[X] = \frac{1}{p}$; why?)

Now we multiply both sides of (3) 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 $\frac{1}{p} \mathbb{E}[X^2]$, we conclude that

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

Finally, plugging this into (2) gives us

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

as claimed. \square

1.2 Memoryless Property

The geometric distribution has a special property called *memorylessness*¹. To get some intuition for this property, suppose we revisit our earlier motivational example of repeatedly tossing a biased coin with heads probability p , and let X denote the number of tosses until the first head appears. Thus $X \sim \text{Geometric}(p)$.

As we've seen, the probability that we need *more than* n tosses before getting the first head is $\mathbb{P}[X > n] = (1 - p)^n$, since the first n tosses must all have been tails.

What if we've already tossed the coin m times without getting a head? What is the probability that we need more than n *additional* tosses before getting our first head? We can calculate this probability as:

$$\begin{aligned}\mathbb{P}[X > n + m \mid X > m] &= \frac{\mathbb{P}[X > n + m]}{\mathbb{P}[X > m]} \\ &= \frac{(1 - p)^{n+m}}{(1 - p)^m} \\ &= (1 - p)^n \\ &= \mathbb{P}[X > n]\end{aligned}$$

This gives us the formal statement of the memoryless property of geometric distributions:

$$\mathbb{P}[X > n + m \mid X > m] = \mathbb{P}[X > n]..$$

The geometric distribution is memoryless in the sense that the length of time we have already been waiting does not affect the additional time we have to wait until we get our first head. (This, of course, defies the “common wisdom” that if we've tossed a coin ten times and it's come up tails every time, then surely it's more likely to come up heads on the 11th toss!)

1.3 Application: Coupon Collecting

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

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

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 equation (4) 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 $\mathbb{E}[X_1] = 1$.

¹This property is shared by the *exponential distribution*, the continuous analog of the geometric distribution, which we'll cover in a future note. Indeed, these are the *only* two distributions that are memoryless.

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 $i = 1, 2, \dots, n$, 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 equation (4), we get

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{2} + \frac{n}{1} = n \sum_{i=1}^n \frac{1}{i}. \quad (5)$$

This is an exact expression for $\mathbb{E}[S_n]$. We can obtain a tidier form by noting that the sum in (5) actually has a very good approximation,² namely:

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

where $\gamma_E = 0.5772\dots$ 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 (5) even for quite small values of n . So for example, for $n = 100$, we expect to buy about 518 boxes.

2 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*

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!} e^{-\lambda}, \quad \text{for } i = 0, 1, 2, \dots \quad (6)$$

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

To make sure this is a valid definition, let us check that (6) is in fact a distribution, i.e., that the probabilities sum to 1. We have

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

²This is another of the little tricks you might like to carry around in your toolbox.

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

The Poisson distribution is a very widely accepted model for so-called “rare events,” such as disconnected 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 λ 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

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

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

$$\begin{aligned} \mathbb{P}[\exists \text{ a page with exactly 5 typos}] &= 1 - \mathbb{P}[\text{every page has } \neq 5 \text{ typos}] \\ &= 1 - \prod_{k=1}^{200} \mathbb{P}[\text{page } k \text{ has } \neq 5 \text{ typos}] \\ &= 1 - \prod_{k=1}^{200} (1 - \mathbb{P}[\text{page } k \text{ has exactly 5 typos}]) \\ &= 1 - \left(1 - \frac{1}{120e}\right)^{200} \approx 0.46, \end{aligned}$$

where in the last step we have used our earlier calculation for $\mathbb{P}[X = 5]$. □

2.1 Mean and Variance of a Poisson Random Variable

Let us now calculate the expectation and variance of a Poisson random variable.

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

Proof. We can calculate $\mathbb{E}[X]$ directly from the definition of expectation:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=0}^{\infty} i \times \mathbb{P}[X = i] \\ &= \sum_{i=1}^{\infty} i \times \frac{\lambda^i}{i!} e^{-\lambda} && \text{(the } i = 0 \text{ term is equal to 0 so we omit it)} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} && \text{(since } e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i-1) \\ &= \lambda. \end{aligned}$$

Similarly, we can calculate $\mathbb{E}[X^2]$ as follows:

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

Therefore,

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

as desired. \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 illustrates the $\text{Poisson}(\lambda)$ distribution for $\lambda = 1, 2, 5, 20$.

2.2 Sum of Independent Poisson Random Variables

As illustrated in Figure 2, the $\text{Poisson}(\lambda)$ distribution becomes more symmetric and resembles a “bell curve” as λ increases. As we will learn later, this phenomenon is closely related to the following useful fact regarding a sum of independent Poisson random variables.

Theorem 19.5. *Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent Poisson random variables. Then, $X + Y \sim \text{Poisson}(\lambda + \mu)$.*

Proof. For all $k = 0, 1, 2, \dots$, we have

$$\begin{aligned}
\mathbb{P}[X + Y = k] &= \sum_{j=0}^k \mathbb{P}[X = j, Y = k - j] \\
&= \sum_{j=0}^k \mathbb{P}[X = j] \mathbb{P}[Y = k - j] \\
&= \sum_{j=0}^k \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{k-j}}{(k-j)!} e^{-\mu} \\
&= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{j=0}^k \frac{k!}{j!(k-j)!} \lambda^j \mu^{k-j} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!},
\end{aligned}$$

where the second equality follows from independence, and the last equality from the binomial theorem. \square

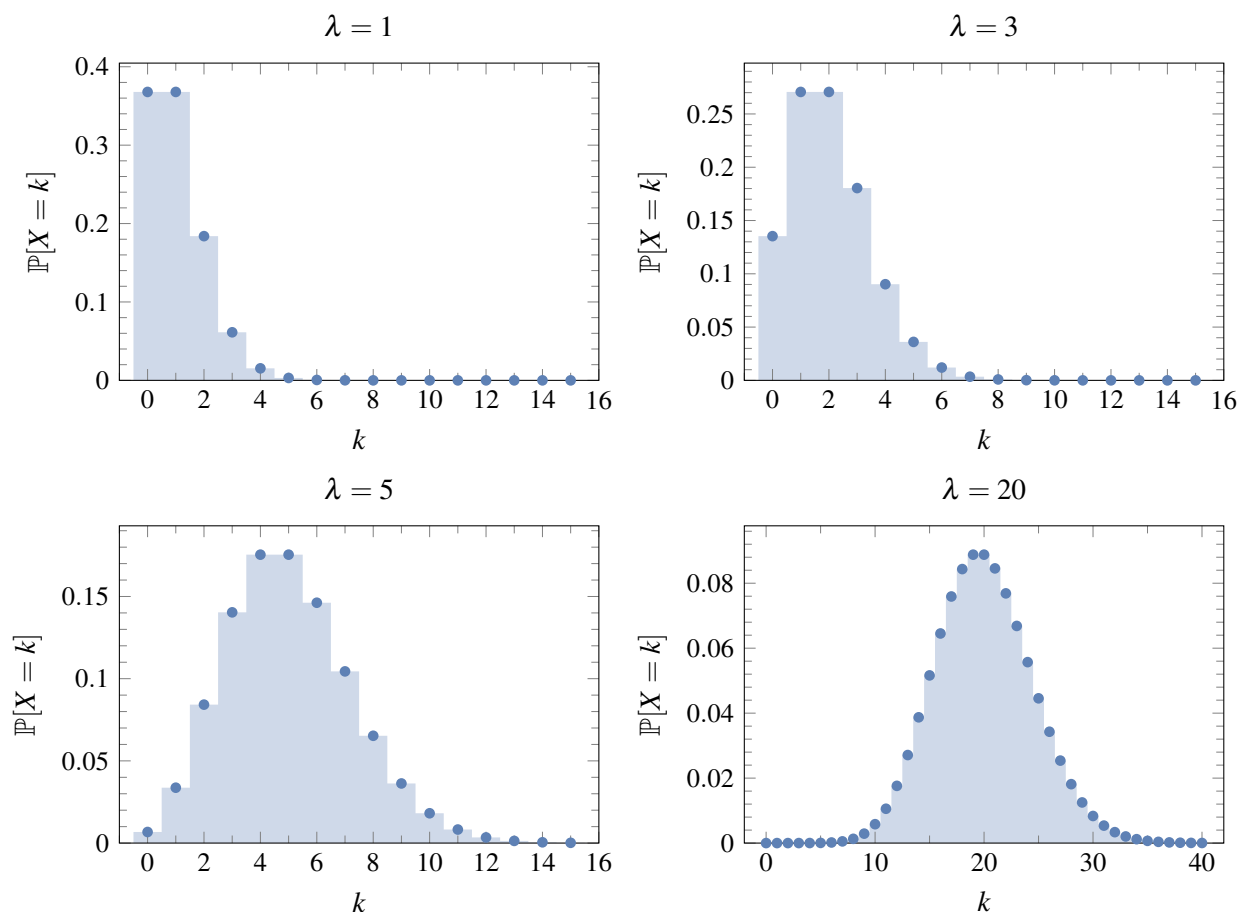


Figure 2: Illustration of the Poisson(λ) distribution for $\lambda = 1, 2, 5, 20$.

By induction, we conclude that if X_1, X_2, \dots, X_n are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively, then

$$X_1 + X_2 + \dots + X_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

2.3 Poisson as Limit of Binomial

To see a concrete example of how the 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 initiations of calls in disjoint time intervals are independent events.

Then, if we divide the one-minute time period into n disjoint intervals, the number of calls X in that time period can be modeled as a Binomial(n, p) random variable, where 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 $\mathbb{E}[X] = \lambda$ and so $np = \lambda$, i.e.,

$p = \lambda/n$. So $X \sim \text{Binomial}(n, \frac{\lambda}{n})$. As we shall see below, as we let n tend to infinity, this distribution tends to the Poisson distribution with parameter λ . This explains why the Poisson distribution is a model for “rare events”: it counts the number of heads in a very long sequence ($n \rightarrow \infty$) of coin flips, where the expected number of heads is a small finite number.

Now we will prove the claim above that $\text{Poisson}(\lambda)$ is the limit of $\text{Binomial}(n, \frac{\lambda}{n})$, as n tends to infinity.

Theorem 19.6. *Let $X \sim \text{Binomial}(n, \frac{\lambda}{n})$ where $\lambda > 0$ is a fixed constant. Then for every $i = 0, 1, 2, \dots$,*

$$\mathbb{P}[X = i] \longrightarrow \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{as } n \rightarrow \infty.$$

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 $n \rightarrow \infty$). Then, because X has binomial distribution with parameter n and $p = \frac{\lambda}{n}$,

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

Let us collect the factors into

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!} \left(\frac{n!}{(n-i)!} \cdot \frac{1}{n^i} \right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-i}. \quad (7)$$

For any fixed i , the first parenthesis above becomes, as $n \rightarrow \infty$,

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

Since λ is a fixed constant, the second parenthesis in (7) becomes, as $n \rightarrow \infty$,

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

And since i is fixed, the third parenthesis in (7) becomes, as $n \rightarrow \infty$,

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

Substituting these results back into (7) gives us

$$\mathbb{P}[X = i] \rightarrow \frac{\lambda^i}{i!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^i}{i!} e^{-\lambda},$$

as desired. □