3.1.5 Special Distributions

As it turns out, there are some specific distributions that are used over and over in practice, thus they have been given special names. There is a random experiment behind each of these distributions. Since these random experiments model a lot of real life phenomenon, these special distributions are used frequently in different applications. That's why they have been given a name and we devote a section to study them. We will provide PMFs for all of these special random variables, but rather than trying to memorize the PMF, you should understand the random experiment behind each of them. If you understand the random experiments, you can simply derive the PMFs when you need them. Although it might seem that there are a lot of formulas in this section, there are in fact very few new concepts. Do not get intimidated by the large number of formulas, look at each distribution as a practice problem on discrete random variables.

Bernoulli Distribution

What is the simplest discrete random variable (i.e., simplest PMF) that you can imagine? My answer to this question is a PMF that is nonzero at only one point. For example, if you define

$$P_X(x) = \begin{cases} 1 & \text{for } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

then X is a discrete random variable that can only take one value, i.e., $X = 1$ with a probability of one. But this is not a very interesting distribution because it is not actually random. Then, you might ask what is the next simplest discrete distribution. And my answer to that is the **Bernoulli** distribution. A Bernoulli random variable is a random variable that can only take two possible values, usually 0 and 1. This random variable models random experiments that have two possible outcomes, sometimes referred to as "success" and "failure." Here are some examples:

- You take a pass-fail exam. You either pass (resulting in $X = 1$) or fail (resulting in $X = 0$).
- You toss a coin. The outcome is ether heads or tails.
- A child is born. The gender is either male or female.

Formally, the Bernoulli distribution is defined as follows:

Definition 3.4

A random variable X is said to be a *Bernoulli* random variable with *parameter* p, shown as $X \sim Bernoulli(p)$, if its PMF is given by

$$P_X(x) = \begin{cases} p & \text{for } x = 1 \\ 1-p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Figure 3.2 shows the PMF of a $Bernoulli(p)$ random variable.

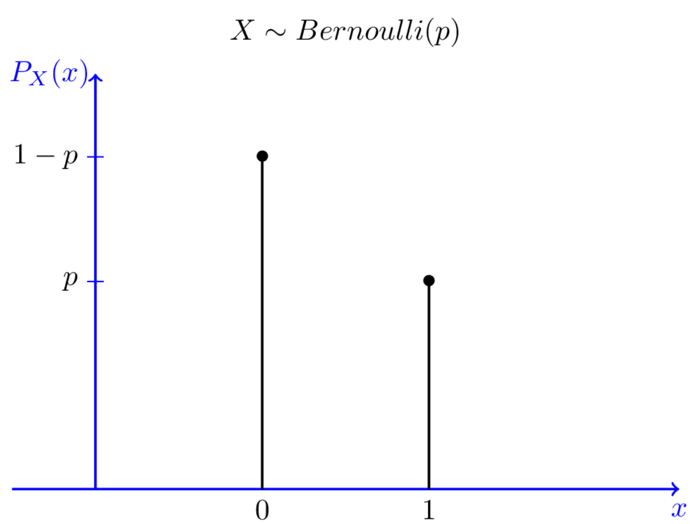

Fig.3.2 - PMF of a $Bernoulli(p)$ random variable.

A Bernoulli random variable is associated with a certain event A. If event A occurs (for example, if you pass the test), then $X = 1$; otherwise $X = 0$. For this reason the Bernoulli random variable, is also called the **indicator** random variable. In particular, the indicator random variable I_A for an event A is defined by

$$I_A = \begin{cases} 1 & \text{if the event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

The indicator random variable for an event A has Bernoulli distribution with parameter $p = P(A)$, so we can write

$$I_A \sim Bernoulli\big(P(A)\big).$$

Geometric Distribution

The random experiment behind the geometric distribution is as follows. Suppose that I have a coin with $P(H) = p$. I toss the coin until I observe the first heads. We define X as the total number of coin tosses in this experiment. Then X is said to have geometric distribution with parameter p. In other words, you can think of this experiment as repeating independent Bernoulli trials until observing the first success. This is exactly the same distribution that we saw in Example 3.4. The range of X here is $R_X = \{1, 2, 3, \ldots\}$. In Example 3.4, we obtained

$$P_X(k) = P(X = k) = (1-p)^{k-1}p, \text{ for } k = 1, 2, 3, \ldots$$

We usually define $q = 1 - p$, so we can write $P_X(k) = pq^{k-1}, \text{ for } k = 1, 2, 3, \ldots$. To say that a random variable has geometric distribution with parameter p, we write $X \sim Geometric(p)$. More formally, we have the following definition:

Definition 3.5

A random variable X is said to be a *geometric* random variable with *parameter* p, shown as $X \sim Geometric(p)$, if its PMF is given by

$$P_X(k) = \begin{cases} p(1-p)^{k-1} & \text{for } k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Figure 3.3 shows the PMF of a $Geometric(0.3)$ random variable.

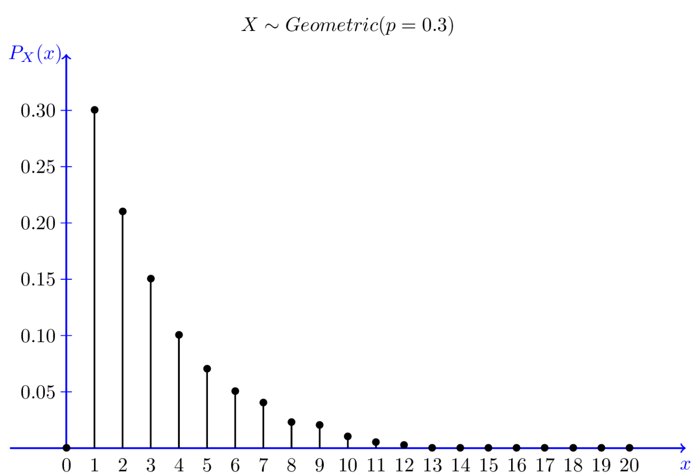

Fig.3.3 - PMF of a $Geometric(0.3)$ random variable.

We should note that some books define geometric random variables slightly differently. They define the geometric random variable X as the total number of failures before observing the first success. By this definition the range of X is $R_X = \{0, 1, 2, \dots\}$ and the PMF is given by

$$P_X(k) = \begin{cases} p(1-p)^k & \text{for } k = 0, 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

In this book, whenever we write $X \sim Geometric(p)$, we always mean X as the total number of trials as defined in Definition 3.5. Note that as long as you are consistent in your analysis, it does not matter which definition you use. That is why we emphasize that you should understand how to derive PMFs for these random variables rather than memorizing them.

Binomial Distribution

The random experiment behind the binomial distribution is as follows. Suppose that I have a coin with $P(H) = p$. I toss the coin n times and define X to be the total number of heads that I observe. Then X is binomial with parameter n and p, and we write $X \sim Binomial(n, p)$. The range of X in this case is $R_X = \{0, 1, 2, \dots, n\}$. As we have seen in Section 2.1.3, the PMF of X in this case is given by binomial formula

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, 2, \dots, n.$$

We have the following definition:

Definition 3.6

A random variable X is said to be a *binomial* random variable with parameters n and p, shown as $X \sim Binomial(n, p)$, if its PMF is given by

$$P_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Figures 3.4 and 3.5 show the $Binomial(n, p)$ PMF for $n = 10$, $p = 0.3$ and $n = 20$, $p = 0.6$ respectively.

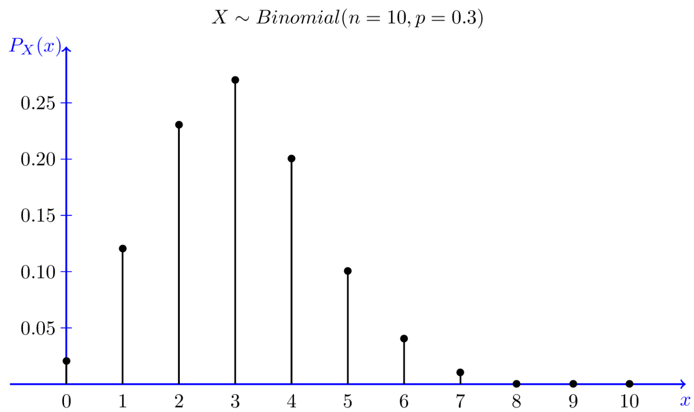

Fig.3.4 - PMF of a $Binomial(10, 0.3)$ random variable.

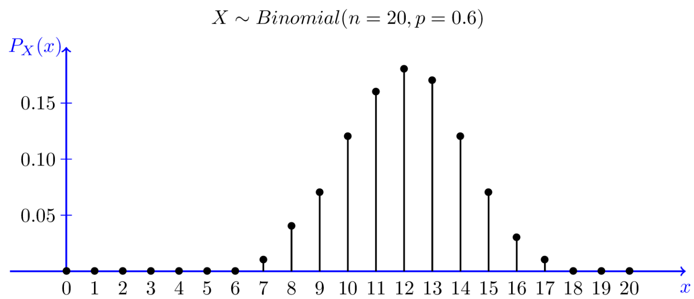

Fig.3.5 - PMF of a $Binomial(20, 0.6)$ random variable.

Binomial random variable as a sum of Bernoulli random variables
Here is a useful way of thinking about a binomial random variable. Note that a $Binomial(n, p)$ random variable can be obtained by n independent coin tosses. If we think of each coin toss as a $Bernoulli(p)$ random variable, the $Binomial(n, p)$ random variable is a sum of n independent $Bernoulli(p)$ random variables. This is stated more precisely in the following lemma.

Lemma 3.1
If $X_1, X_2, \ldots, X_n$ are independent $Bernoulli(p)$ random variables, then the random variable X defined by $X = X_1 + X_2 + \ldots + X_n$ has a $Binomial(n, p)$ distribution.

To generate a random variable $X \sim Binomial(n, p)$, we can toss a coin n times and count the number of heads. Counting the number of heads is exactly the same as finding $X_1 + X_2 + \ldots + X_n$, where each X_i is equal to one if the corresponding coin toss results in heads and zero otherwise. This interpretation of binomial random variables is sometimes very helpful. Let's look at an example.

Example 3.7

Let $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ be two independent random variables. Define a new random variable as $Z = X + Y$. Find the PMF of Z.

Solution

Since $X \sim Binomial(n,p)$, we can think of X as the number of heads in n independent coin tosses, i.e., we can write

$$X = X_1 + X_2 + \ldots + X_n,$$

where the X_i's are independent $Bernoulli(p)$ random variables. Similarly, since $Y \sim Binomial(m,p)$, we can think of Y as the number of heads in m independent coin tosses, i.e., we can write

$$Y = Y_1 + Y_2 + \ldots + Y_m,$$

where the Y_j's are independent $Bernoulli(p)$ random variables. Thus, the random variable $Z = X + Y$ will be the total number of heads in $n + m$ independent coin tosses:

$$Z = X + Y = X_1 + X_2 + \ldots + X_n + Y_1 + Y_2 + \ldots + Y_m,$$

where the X_i's and Y_j's are independent $Bernoulli(p)$ random variables. Thus, by Lemma 3.1, Z is a binomial random variable with parameters $m + n$ and p, i.e., $Binomial(m + n, p)$. Therefore, the PMF of Z is

$$P_Z(k) = \begin{cases} \binom{m+n}{k} p^k (1-p)^{m+n-k} & \text{for } k = 0, 1, 2, 3, \ldots, m+n \\ 0 & \text{otherwise} \end{cases}$$

The above solution is elegant and simple, but we may also want to directly obtain the PMF of Z using probability rules. Here is another method to solve Example 3.7. First, we note that $R_Z = \{0, 1, 2, \ldots, m+n\}$. For $k \in R_Z$, we can write

$$P_Z(k) = P(Z = k) = P(X + Y = k).$$

We will find $P(X + Y = k)$ by using conditioning and the law of total probability. In particular, we can write

$$\begin{aligned} P_Z(k) &= P(X + Y = k) \\ &= \sum_{i=0}^{n} P(X + Y = k | X = i) P(X = i) && \text{(law of total probability)} \\ &= \sum_{i=0}^{n} P(Y = k - i | X = i) P(X = i) \\ &= \sum_{i=0}^{n} P(Y = k - i) P(X = i) \end{aligned}$$

$$
\begin{aligned}
&= \textstyle\sum_{i=0}^{n} \binom{m}{k-i} p^{k-i}(1-p)^{m-k+i}\binom{n}{i}p^i(1-p)^{n-i} \\
&= \textstyle\sum_{i=0}^{n} \binom{m}{k-i}\binom{n}{i} p^{k}(1-p)^{m+n-k} \\
&= p^k(1-p)^{m+n-k}\textstyle\sum_{i=0}^{n} \binom{m}{k-i}\binom{n}{i} \\
&= \binom{m+n}{k} p^k(1-p)^{m+n-k} \qquad\qquad \text{(by Example 2.8 (part 3)).}
\end{aligned}
$$

Thus, we have proved $Z \sim Binomial(m+n, p)$ by directly finding the PMF of Z.

Negative Binomial (Pascal) Distribution

The negative binomial or Pascal distribution is a generalization of the geometric distribution. It relates to the random experiment of repeated independent trials until observing m successes. Again, different authors define the Pascal distribution slightly differently, and as we mentioned before if you understand one of them you can easily derive the other ones. Here is how we define the Pascal distribution in this book. Suppose that I have a coin with $P(H) = p$. I toss the coin until I observe m heads, where $m \in \mathbb{N}$. We define X as the total number of coin tosses in this experiment. Then X is said to have Pascal distribution with parameter m and p. We write $X \sim Pascal(m, p)$. Note that $Pascal(1, p) = Geometric(p)$. Note that by our definition the range of X is given by R_X=$\{m$, $m+1$, $m+2$, $m+3$, $\cdots\}$.

Let us derive the PMF of a $Pascal(m, p)$ random variable X. Suppose that I toss the coin until I observe m heads, and X is defined as the total number of coin tosses in this experiment. To find the probability of the event $A = \{X = k\}$, we argue as follows. By definition, event A can be written as $A = B \cap C$, where

- B is the event that we observe $m-1$ heads (successes) in the first $k-1$ trials, and
- C is the event that we observe a heads in the kth trial.

Note that B and C are independent events because they are related to different independent trials (coin tosses). Thus we can write

$$P(A) = P(B \cap C) = P(B)P(C).$$

Now, we have $P(C) = p$. Note also that $P(B)$ is the probability that I observe observe $m-1$ heads in the $k-1$ coin tosses. This probability is given by the binomial formula, in particular

$$P(B) = \binom{k-1}{m-1} p^{m-1}(1-p)^{\big((k-1)-(m-1)\big)} = \binom{k-1}{m-1} p^{m-1}(1-p)^{k-m}.$$

Thus, we obtain

$$P(A) = P(B \cap C) = P(B)P(C) = \binom{k-1}{m-1} p^m (1-p)^{k-m}.$$

To summarize, we have the following definition for the Pascal random variable

Definition 3.7
A random variable X is said to be a *Pascal* random variable with parameters m and p, shown as $X \sim Pascal(m, p)$, if its PMF is given by

$$P_X(k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m} & \text{for } k = m, m+1, m+2, m+3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Figure 3.6 shows the PMF of a $Pascal(m, p)$ random variable with $m = 3$ and $p = 0.5$.

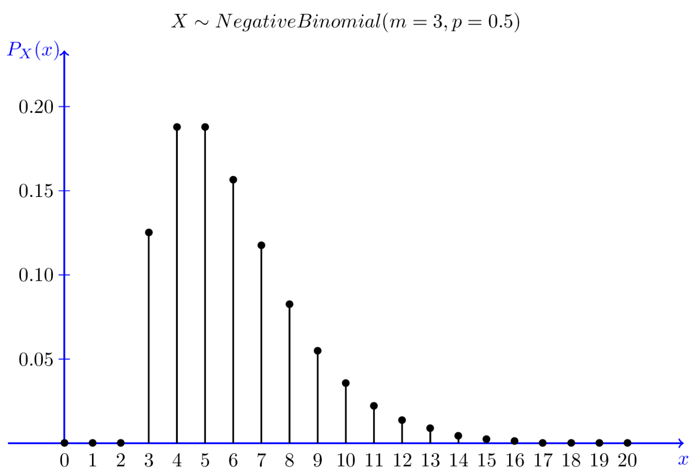

Fig.3.6 - PMF of a $Pascal(3, 0.5)$ (negative binomial) random variable.

Hypergeometric Distribution
Here is the random experiment behind the hypergeometric distribution. You have a bag that contains b blue marbles and r red marbles. You choose $k \leq b + r$ marbles at random (without replacement). Let X be the number of blue marbles in your sample. By this definition, we have $X \leq \min(k, b)$. Also, the number of red marbles in your sample must be less than or equal to r, so we conclude $X \geq \max(0, k - r)$. Therefore,

the range of X is given by
$R_X = \{\max(0, k-r), \max(0, k-r)+1, \max(0, k-r)+2, \ldots, \min(k, b)\}$.

To find $P_X(x)$, note that the total number of ways to choose k marbles from $b+r$ marbles is $\binom{b+r}{k}$. The total number of ways to choose x blue marbles and $k-x$ red marbles is $\binom{b}{x}\binom{r}{k-x}$. Thus, we have

$$P_X(x) = \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}}, \qquad \text{for } x \in R_X.$$

The following definition summarizes the discussion above.

Definition 3.8
A random variable X is said to be a *Hypergeometric* random variable with parameters b, r and k, shown as $X \sim Hypergeometric(b, r, k)$, if its range is $R_X = \{\max(0, k-r), \max(0, k-r)+1, \max(0, k-r)+2, \ldots, \min(k, b)\}$, and its PMF is given by

$$P_X(x) = \begin{cases} \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}} & \text{for } x \in R_X \\ 0 & \text{otherwise} \end{cases}$$

Again, there is no point to memorizing the PMF. All you need to know is how to solve problems that can be formulated as a hypergeometric random variable.

Poisson Distribution
The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable. Here is an example of a scenario where a Poisson random variable might be used. Suppose that we are counting the number of customers who visit a certain store from $1pm$ to $2pm$. Based on data from previous days, we know that on average $\lambda = 15$ customers visit the store. Of course, there will be more customers some days and fewer on others. Here, we may model the random variable X showing the number customers as a Poisson random variable with parameter $\lambda = 15$. Let us introduce the Poisson PMF first, and then we will talk about more examples and interpretations of this distribution.

Definition 3.9
A random variable X is said to be a *Poisson* random variable with parameter λ, shown as $X \sim Poisson(\lambda)$, if its range is $R_X = \{0, 1, 2, 3, \dots\}$, and its PMF is given by

$$P_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!} & \text{for } k \in R_X \\ 0 & \text{otherwise} \end{cases}$$

Before going any further, let's check that this is a valid PMF. First, we note that $P_X(k) \geq 0$ for all k. Next, we need to check $\sum_{k \in R_X} P_X(k) = 1$. To do that, let us first remember the Taylor series for e^x, $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Now we can write

$$\begin{aligned} \sum_{k \in R_X} P_X(k) &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} e^{\lambda} \\ &= 1. \end{aligned}$$

Figures 3.7, 3.8, and 3.9 show the $Poisson(\lambda)$ PMF for $\lambda = 1$, $\lambda = 5$, and $\lambda = 10$ respectively.

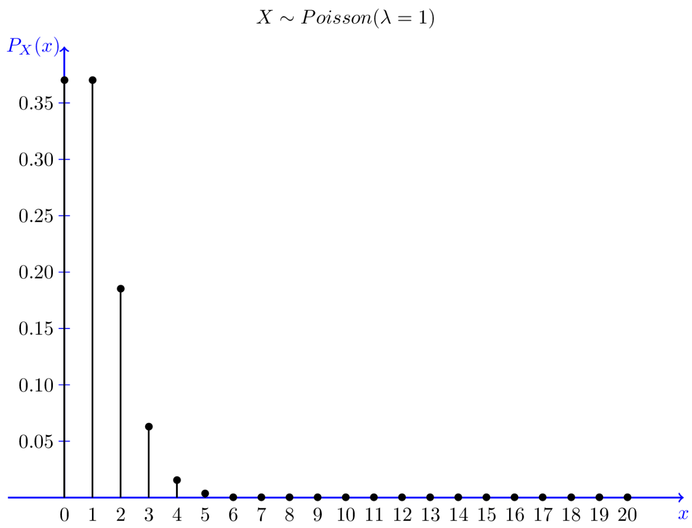

Fig.3.7 - PMF of a $Poisson(1)$ random variable.

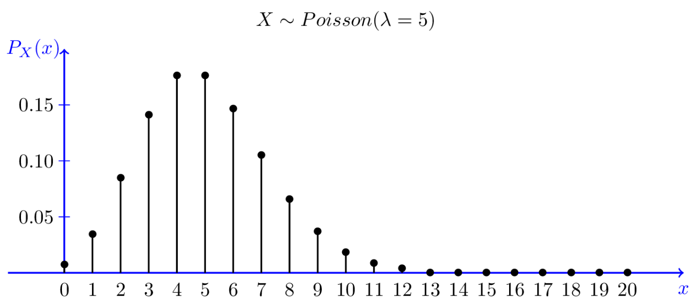

Fig.3.8 - PMF of a $Poisson(5)$ random variable.

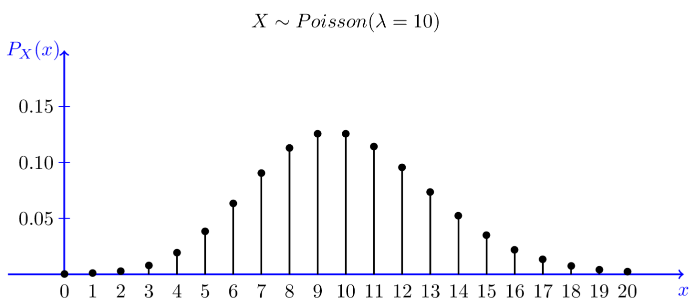

Fig.3.9 - PMF of a $Poisson(10)$ random variable.

Now let's look at an example.

Example 3.8

The number of emails that I get in a weekday can be modeled by a Poisson distribution with an average of 0.2 emails per minute.

1. What is the probability that I get no emails in an interval of length 5 minutes?
2. What is the probability that I get more than 3 emails in an interval of length 10 minutes?

Solution

1. Let X be the number of emails that I get in the 5-minute interval. Then, by the assumption X is a Poisson random variable with parameter $\lambda = 5(0.2) = 1$,

$$P(X = 0) = P_X(0) = \frac{e^{-\lambda}\lambda^0}{0!} = \frac{e^{-1} \cdot 1}{1} = \frac{1}{e} \approx 0.3679$$

2. Let Y be the number of emails that I get in the 10-minute interval. Then by the assumption Y is a Poisson random variable with parameter $\lambda = 10(0.2) = 2$,

$$\begin{aligned}
P(Y > 3) &= 1 - P(Y \leq 3) \\
&= 1 - \left(P_Y(0) + P_Y(1) + P_Y(2) + P_Y(3)\right) \\
&= 1 - e^{-\lambda} - \frac{e^{-\lambda}\lambda}{1!} - \frac{e^{-\lambda}\lambda^2}{2!} - \frac{e^{-\lambda}\lambda^3}{3!} \\
&= 1 - e^{-2} - \frac{2e^{-2}}{1} - \frac{4e^{-2}}{2} - \frac{8e^{-2}}{6} \\
&= 1 - e^{-2}\left(1 + 2 + 2 + \frac{8}{6}\right) \\
&= 1 - \frac{19}{3e^2} \approx 0.1429
\end{aligned}$$

Poisson as an approximation for binomial

The Poisson distribution can be viewed as the limit of binomial distribution. Suppose $X \sim Binomial(n, p)$ where n is very large and p is very small. In particular, assume that $\lambda = np$ is a positive constant. We show that the PMF of X can be approximated by the PMF of a $Poisson(\lambda)$ random variable. The importance of this is that Poisson PMF is much easier to compute than the binomial. Let us state this as a theorem.

Theorem 3.1

Let $X \sim Binomial(n, p = \frac{\lambda}{n})$, where $\lambda > 0$ is fixed. Then for any $k \in \{0, 1, 2, \dots\}$, we have

$$\lim_{n \to \infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Proof
We have

$$\begin{aligned}
\lim_{n\to\infty} P_X(k) &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lambda^k \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{1}{n^k}\right) \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!}. \lim_{n\to\infty} \left(\left[\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\right] \left[\left(1-\frac{\lambda}{n}\right)^n\right] \left[\left(1-\frac{\lambda}{n}\right)^{-k}\right] \right).
\end{aligned}$$

Note that for a fixed k, we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} = 1 \\
&\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1 \\
&\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}.
\end{aligned}$$

Thus, we conclude

$$\lim_{n \to \infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$
