



[Unit 5: Averages, Law of Large
Numbers, and Central Limit](#)

[Course](#) > [Theorem](#)

> [5.1 Reading](#) > 5.7 Poisson

5.7 Poisson

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

The last famous discrete distribution that we'll discuss in this unit is the Poisson, which is an extremely popular distribution for modeling discrete data. We'll introduce its [PME](#), mean, and [variance](#), and then discuss its story in more detail.

Recall that we briefly introduced the Poisson in [Unit 3](#), and saw it in the Poisson process context in [Unit 4](#).

DEFINITION 5.7.1 (POISSON DISTRIBUTION).

An r.v. X has the *Poisson distribution* with parameter λ , where $\lambda > 0$, if the PMF of X is

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

We write this as $X \sim \text{Pois}(\lambda)$.

This is a valid PMF because of the Taylor series $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$.

Example 5.7.2 (Poisson expectation and variance).

Let $X \sim \text{Pois}(\lambda)$. Then the mean and variance are both equal to λ . For the mean, we have

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

First we dropped the $k = 0$ term because it was 0. Then we took a λ out of the sum so that what was left inside was just the Taylor series for e^{λ} .

To get the variance, we first find $E(X^2)$. By LOTUS,

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!}.$$

Using the same method we used to get the variance of a Geometric r.v., we can obtain

$$E(X^2) = \lambda(1 + \lambda),$$

so

$$\text{Var}(X) = E(X^2) - (EX)^2 = \lambda(1 + \lambda) - \lambda^2 = \lambda.$$

The figure below shows the PMF and CDF of the **Pois(2)** and **Pois(5)** distributions from $k = 0$ to $k = 10$. It appears that the mean of the **Pois(2)** is around 2 and the mean of the **Pois(5)** is around 5, consistent with our findings above. The PMF of the **Pois(2)** is highly skewed, but as λ grows larger, the skewness is reduced and the PMF becomes more bell-shaped.



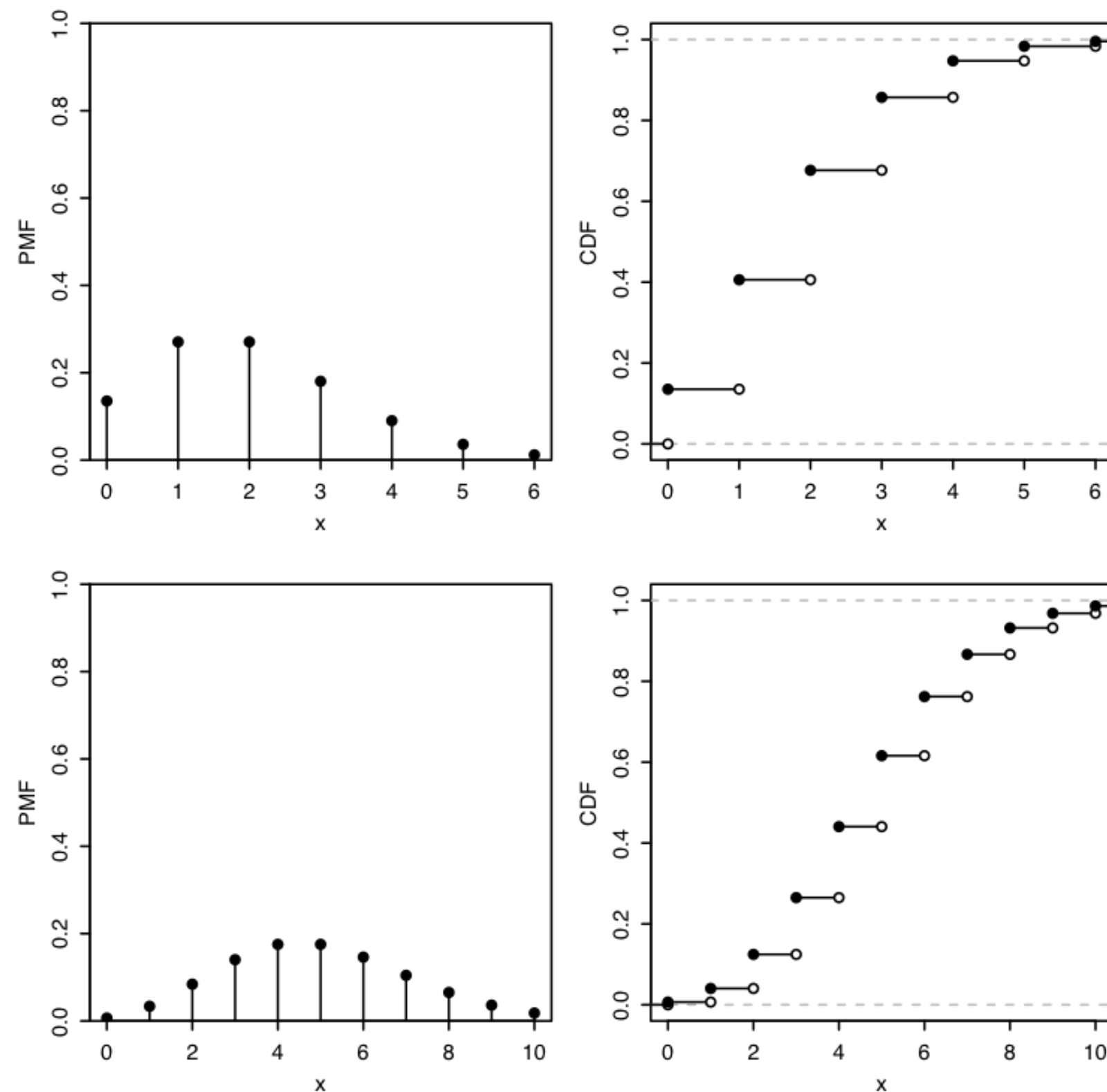


Figure 5.7.3: Top: **Pois(2)** PMF and CDF. Bottom: **Pois(5)** PMF and CDF.

[View Larger Image](#)

[Image Description](#)

The Poisson distribution is often used in situations where we are counting the number of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of success. For example, the following random variables could follow a distribution that is approximately Poisson.

- The number of emails you receive in an hour. There are a lot of people who could potentially email you in that hour, but it is unlikely that any specific person will actually email you in that hour. Alternatively, imagine subdividing the hour into milliseconds. There are 3.6×10^6 seconds in an hour, but in any specific millisecond it is unlikely that you will get an email.
- The number of chips in a chocolate chip cookie. Imagine subdividing the cookie into small cubes; the probability of getting a chocolate chip in a single cube is small, but the number of cubes is large.
- The number of earthquakes in a year in some region of the world. At any given time and location, the probability of an earthquake is small, but there are a large number of possible times and locations for earthquakes to occur over the course of the year.

The parameter λ is interpreted as the *rate* of occurrence of these rare events; in the examples above, λ could be **20** (emails per hour), **10** (chips per cookie), and **2** (earthquakes per year). In applications similar to the ones above, we can approximate the distribution of the number of events that occur by a Poisson distribution.

Poisson approximation.

Let A_1, A_2, \dots, A_n be events with $p_j = P(A_j)$, where n is large, the p_j are small, and the A_j are independent or weakly dependent. Let

$$X = \sum_{j=1}^n I(A_j)$$

count how many of the A_j occur. Then X is approximately $\text{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^n p_j$.

The Poisson paradigm is also called the *law of rare events*. The interpretation of "rare" is that the p_j are small, not that λ is small. For example, in the email example, the low probability of getting an email from a specific person in a particular hour is offset by the large number of people who could send you an email in that hour.

In the examples we gave above, the number of events that occur isn't *exactly* Poisson because a Poisson random variable has no upper bound, whereas how many of A_1, \dots, A_n occur is at most n , and there is a limit to how many chocolate chips can be crammed into a cookie. But the Poisson distribution often gives good *approximations*. Note that the conditions for the Poisson paradigm to hold are fairly flexible: the n trials can have different success probabilities, and the trials don't have to be independent, though they should not be very dependent. So there are a wide variety of situations that can be cast in terms of the Poisson paradigm. This makes the Poisson a popular model, or at least a starting point, for data whose values are nonnegative integers (called *count data* in statistics).

The Poisson approximation greatly simplifies obtaining a good approximate solution to the birthday problem, and makes it possible to obtain good approximations to various variations which would be very difficult to solve exactly.

Example 5.7.4 (Birthday problem continued).

If we have m people and make the usual assumptions about birthdays, then each pair of people has probability $p = 1/365$ of having the same birthday, and there are $\binom{m}{2}$ pairs. By the Poisson paradigm the distribution of the number X of birthday matches is approximately $\text{Pois}(\lambda)$, where $\lambda = \binom{m}{2} \frac{1}{365}$. Then the probability of at least one match is

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - e^{-\lambda}.$$

For $m = 23$, $\lambda = 253/365$ and $1 - e^{-\lambda} \approx 0.500002$, which agrees with the earlier result that we need 23 people to have a 50-50 chance of a matching birthday. Note that even though $m = 23$ is fairly small, the relevant quantity in this problem is actually $\binom{m}{2}$, which is the total number of "trials" for a successful birthday match, so the Poisson approximation still performs well.

Example 5.7.5 (Near-birthday problem).

What if we want to find the number of people required in order to have a 50-50 chance that two people would have birthdays within one day of each other (i.e., on the same day or one day apart)? Unlike the original birthday problem, this is difficult to obtain an exact answer for, but the Poisson paradigm still applies. The probability that any two people have birthdays within one day of each other is $3/365$ (choose a birthday for the first person, and then the second person needs to be born on that day, the day before, or the day after). Again there are $\binom{m}{2}$ possible pairs, so the number of within-one-day matches is approximately $\text{Pois}(\lambda)$ where $\lambda = \binom{m}{2} \frac{3}{365}$. Then a calculation similar to the one above tells us that we need $m = 14$ or more. This was a quick approximation, but it turns out that $m = 14$ is the exact answer!

Another useful property of the Poisson distribution is that the sum of independent Poissons is Poisson.

THEOREM 5.7.6 (SUM OF INDEPENDENT POISSONS).

If $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$, and X is independent of Y , then $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

Proof

To get the PMF of $X + Y$, condition on X and use the law of total probability:

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^k P(X + Y = k | X = j) P(X = j) \\ &= \sum_{j=0}^k P(Y = k - j) P(X = j) \\ &= \sum_{j=0}^k \frac{e^{-\lambda_2} \lambda_2^{k-j}}{(k-j)!} \frac{e^{-\lambda_1} \lambda_1^j}{j!} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{j=0}^k \binom{k}{j} \lambda_1^j \lambda_2^{k-j} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^k}{k!}. \end{aligned}$$

The last step used the binomial theorem. Since we've arrived at the $\text{Pois}(\lambda_1 + \lambda_2)$ PMF, we have $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

The story of the Poisson distribution provides intuition for this result. If there are two different types of events occurring at rates λ_1 and λ_2 , independently, then the overall event rate is $\lambda_1 + \lambda_2$.

[Learn About Verified Certificates](#)

