Poisson distributions

Like the binomial distribution, the Poisson distribution models a discrete random variable, and it's particularly useful for finding the probability that a specific number of events will occur in a given period of time.

The Poisson process

A **Poisson process** calculates the number of times an event occurs in a period of time, or in a particular area, or over some distance, or within any other kind of measurement, and the process has particular characteristics:

1. The experiment counts the number of occurrences of an event over some other measurement,

2. The mean is the same for each interval,

3. The count of events in each interval is independent of the other intervals, and

4. The intervals don't overlap.

5. The probability of the event occurring is proportional to the period of time.

The Poisson process is useful in modeling many real-life events, such as radioactive decay, the number of visitors to a website, or even the number of trees in an acre of forest.

Let's say we want to use a Poisson process to model the number of cars that pass through an intersection each hour. In this case, because we're counting the number of occurrences of an event over time, we've met the first condition.

In order to meet the second condition, we need to be able to assume that the average number of cars that passes through the intersection each hour is the same. In other words, if the average number of cars passing through the intersection daily between $9:00$ a.m. and $10:00$ a.m. is 15, then the average for each other hour of the day needs to be 15 as well.

To meet the third condition, we need to be able to say that the number of cars that pass through the intersection this hour is not affected by the number of cars that came through during a previous hour, and that the number of cars passing through this hour will not affect the number of cars that pass through in the coming hours. In other words, the car count for each hour is independent of the count for any other hour.

To meet the fourth condition, we need our intervals to be non-overlapping. So if we take data from $9:00$ a.m. to $10:00$ a.m. and from $10:00$ a.m. to $11:00$ a.m., then we can't include a data set for $9:30$ a.m. to $10:30$ a.m., because that data would overlap with the data from the other intervals.

To meet the fifth condition, we need to the probability of the number of cars to scale based on the size of the interval. For instance, if the probability that C cars pass through the intersection in one hour is $P(C)$, then the probability that $2C$ cars pass through the intersection in two hours is $2P(C)$.

Assuming we meet all of those conditions, then we'll be able to use a Poisson process. In this example, the discrete random variable would be the actual number of cars passing through the intersection.

Probability formula for Poisson

The probability of exactly x occurrences of the event, when the mean number of occurrences in the interval is λ, is

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

So if we believe the mean number of cars passing through the intersection in any particular hour is $\lambda = 15$, and if we want to know the probability that $x = 13$ cars will pass through it in the next hour, that probability will be

$$P(13) = \frac{15^{13} e^{-15}}{13!}$$

$$P(13) \approx 0.0956$$

$$P(13) \approx 9.56\,\%$$

So there's an approximately $9.56\,\%$ chance that 13 cars will pass through the intersection in the next hour.

The Poisson formula can also be used to calculate cumulative probabilities. For instance, if we want to know the probability that at most 7 cars pass through the intersection in an hour, we calculate that as

$$P(x \le 7) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$$

$$+P(x=4)+P(x=5)+P(x=6)+P(x=7)$$

$$P(x \le 7)=\frac{15^0e^{-15}}{0!}+\frac{15^1e^{-15}}{1!}+\frac{15^2e^{-15}}{2!}+\frac{15^3e^{-15}}{3!}$$

$$+\frac{15^4e^{-15}}{4!}+\frac{15^5e^{-15}}{5!}+\frac{15^6e^{-15}}{6!}+\frac{15^7e^{-15}}{7!}$$

$$P(x \le 7)=\frac{1}{e^{15}}+\frac{15}{e^{15}}+\frac{15^2}{2e^{15}}+\frac{15^3}{6e^{15}}$$

$$+\frac{15^4}{24e^{15}}+\frac{15^5}{120e^{15}}+\frac{15^6}{720e^{15}}+\frac{15^7}{5{,}040e^{15}}$$

$$P(x \le 7)=\frac{1}{e^{15}}+\frac{15}{e^{15}}+\frac{15^2}{2e^{15}}+\frac{15^3}{6e^{15}}+\frac{15^4}{24e^{15}}+\frac{15^5}{120e^{15}}+\frac{15^6}{720e^{15}}+\frac{15^7}{5{,}040e^{15}}$$

$$P(x \le 7)=\frac{1}{e^{15}}\left(1+15+\frac{15^2}{2}+\frac{15^3}{6}+\frac{15^4}{24}+\frac{15^5}{120}+\frac{15^6}{720}+\frac{15^7}{5{,}040}\right)$$

$$P(x \le 7)\approx 0.018$$

So there's an approximately $1.8\,\%$ chance that 7 or fewer cars pass through the intersection in an hour.

Poisson distribution to approximate the binomial distribution

We already know how to calculate binomial probabilities, but using the Poisson distribution instead can actually save us some time. The Poisson distribution very closely approximates the binomial distribution when the number of binomial trials n is at least 20 and when the probability of success p in the binomial trial is at most 0.05.

So if those two conditions are met, then using the Poisson probability formula will give us a great approximation of the binomial probability. And the math for the Poisson probability is easier than the math for the binomial probability, so we'll save some time.

The probability of x successes in n attempts, given by the Poisson formula is

$$P(x) = \frac{(np)^x e^{-np}}{x!}$$

Notice that we've just substituted $\lambda = np$ into the original Poisson formula. Let's do an example with the Poisson formula for a binomial random variable.

Example

There are 30 students in a Kindergarten class and each one of them has a 4% chance of forgetting their lunch on any given day. What is the probability that exactly 5 of them will forget their lunch today.

This is a binomial experiment with $n = 30$, $p = 0.04$, and $x = 5$. Because we have at least 20 "attempts," and because the probability of a "success" is less than 5%, we can use the Poisson formula to estimate this binomial probability.

$$P(x) = \frac{(np)^x e^{-np}}{x!}$$

$$P(5) = \frac{(30 \cdot 0.04)^5 e^{-30 \cdot 0.04}}{5!}$$

$$P(5) = \frac{1.2^5 e^{-1.2}}{120}$$

$$P(5) \approx 0.006246$$

So the chance that exactly 5 of the Kindergarteners forget their lunch today is approximately $0.62\,\%$.