Poisson Distribution

The Poisson distribution is another type of distribution involving discrete random variables.

The Poisson distribution arises when you count a number of events across time or over an area**.** You should think about the Poisson distribution for any situation that involves counting events**.**

Information about how the data was generated can help you decide whether the Poisson distribution fits. The Poisson distribution is based on four assumptions. We will use the term "interval" to refer to either a time interval or an area, depending on the context of the problem.

1. The probability of observing a single event over a small interval is approximately proportional to the size of that interval eg if there is one defect per cm^2 in a piece of metal, then there are two defects in 2 cm^2. If a business receives 10 phone calls per hour then it will receive 20 phone calls in two hours.
2. The probability of two events occurring in the same narrow interval is negligible.
3. The probability of an event within a certain interval does not change over different intervals.
4. The probability of an event in one interval is independent of the probability of an event in any other non-overlapping interval.

Some examples are:

- the number of Emergency Department visits by an infant during the first year of life,
- the number of pollen spores that impact on a slide in a pollen counting machine,
- the number of dingoes per km^2 in a desert region,
- The number of white blood cells found in a cm^3 of blood,
- The number of customers in a waiting line,
- The number of accidents at an intersection.

Poisson does not have a fixed number of trials. Instead, it uses the fixed interval of time or space in which the number of successes is recorded.

A More Detailed Example

The infection rate at a Neonatal Intensive Care Unit (NICU) is typically expressed as a number of infections per patient days. This is obviously counting a number of events across both time and patients. Does this data follow a Poisson distribution?

We need to assume that the probability of getting an infection over a short time period is proportional to the length of the time period. In other words, a patient who stays one hour in the NICU has twice the risk of a single infection as a patient who stays 30 minutes.

We also need to assume that for a small enough interval, the probability of getting two infections is negligible.

We need to assume that the probability of infection does not change over time or over infants. In other words, each infant is equally likely to get an infection over the same time interval and for a single infant, the probability of infection early in the NICU stay is the same as the probability of infection later in the NICU stay.

And we need to assume independence. Here independence means two things. The probability of seeing an infection in one child does not increase or decrease the probability of seeing an infection in another child. In other words, infections don't spread from one infant to another. We also need to that if an infant who gets an infection during one time interval, it doesn't change the probability that he or she will get another infection during a later time interval.

All of these assumptions are suspect, but especially the last two. If one infant gets an infection it increases the chance that other infants will get the same infection, the infection rate changes from early in the NICU stay to later in the stay, since older infants have better immune systems; and some infants are more infection prone than others.

Mathematical details

The Poisson distribution depends on a single parameter λ. The probability that the Poisson random variable equals k is

$$\frac{\lambda^x e^{-\lambda}}{x!}$$

for any value of k from 0 all the way up to infinity. Although there is no theoretical upper bound for the Poisson distribution, in practice these probabilities get small enough to be negligible when k is very large. Exactly how large k needs to be before the probabilities become negligible depends on the value of λ.

Here are some tables of probabilities for small values of λ.

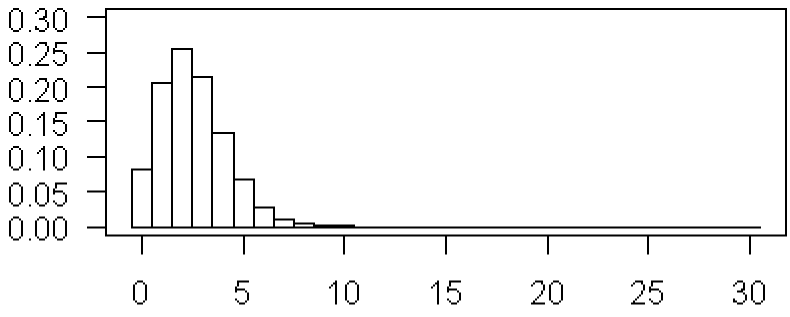
λ = 0.1 x =	0	1	2	3	4	5
k	0.905	0.090	0.005	0.000		

ie $\frac{0.1^0 \times e^{-0.1}}{0!}$ = 0.905, $\frac{0.1^1 \times e^{-0.1}}{1!}$ = 0.90, $\frac{0.1^2 \times e^{-0.1}}{2!}$ = 0.005 etc

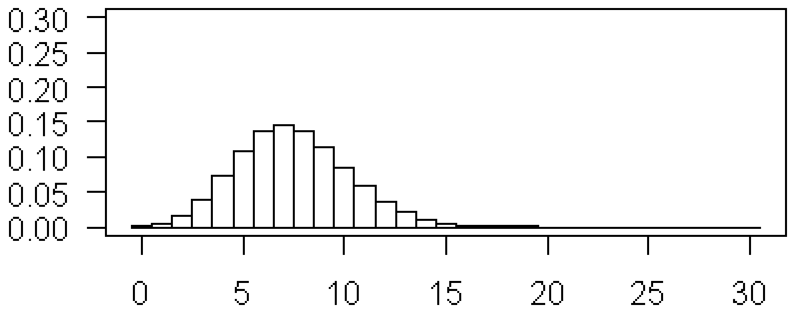
λ = 0.5 x =	0	1	2	3	4	5
k	0.607	0.303	0.076	0.013	0.002	0.000

λ = 1.5 x =	0	1	2	3	4	5	6
k	0.223	0.335	0.251	0.126	0.047	0.014	0.004

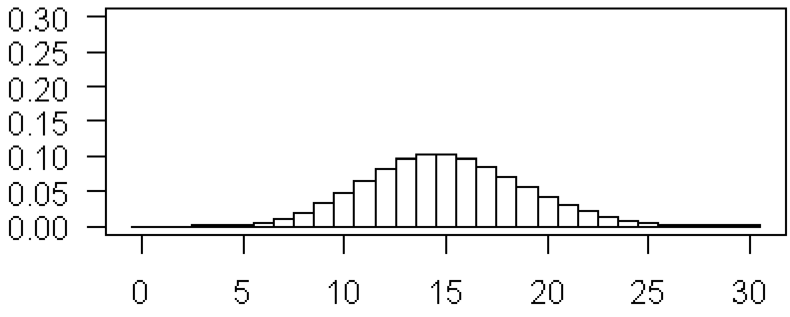
For larger values of λ it is easier to display the probabilities in a graph.



The plot shown above illustrates Poisson probabilities for λ = 2.5.



The above plot illustrates Poisson probabilities for λ = 7.5.



and this plot illustrates Poisson probabilities for λ = 15.

A histogram of the Poisson data should be skewed right, though the skewness becomes less pronounced as the mean increases.

The mean of the Poisson distribution is λ. For the Poisson distribution, the variance, λ, is the same as the mean, so the standard deviation is √λ.

This means that λ = μ and the formula for calculating probabilities becomes:

$$\frac{e^{-x}\mu^{x}}{x!}$$

Thus we only need to know the value of μ to calculate probabilities.