POISSON PROBABILITY DISTRIBUTION

Dr. Bibha Verma
PG Department of Zoology
Maharaja College Ara

PG Sem II, CC-6, Unit -V

INTRODUCTION

In probability theory and statistics, the Poisson distribution, named after French mathematician Siméon Denis Poisson, is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

The Poisson distribution played a key role in experiments that had a historic role in the development of molecular biology. In particular, the interpretation and design of experiments elucidating the actions of bacteriophages and their host bacteria during the infection process were based on the parameters of the Poisson distribution.

Example:

A certain fast-food restaurant gets an average of 3 visitors to the drive-through per minute. This is just an average, however. The actual amount can vary.

A Poisson distribution can be used to analyze the probability of various events regarding how many customers go through the drive-through. It can allow one to calculate the probability of a lull in activity (when there are 0 customers coming to the drive-through) as well as the probability of a flurry of activity (when there are 5 or more customers coming to the drive-through). This information can, in turn, help a manager plan for these events with staffing and scheduling.

ASSUMPTIONS AND VALIDITY

The Poisson distribution is an appropriate model if the following assumptions are true:

- k is the number of times an event occurs in an interval and k can take values 0, 1, 2, ...
- The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
- The average rate at which events occur is independent of any occurrences. For simplicity, this is usually assumed to be constant, but may in practice vary with time.

- Two events cannot occur at the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

If these conditions are true, then k is a Poisson random variable, and the distribution of k is a Poisson distribution.

The Poisson distribution is also the limit of a binomial distribution, for which the probability of success for each trial equals λ divided by the number of trials, as the number of trials approaches infinity (see Related distributions).

PROBABILITY MASS FUNCTION

The Poisson distribution is popular for modeling the number of times an event occurs in an interval of time or space.

A discrete random variable X is said to have a Poisson distribution with parameter λ > 0, if, for k = 0, 1, 2, ..., the probability mass function of X is given by

$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where

- e is Euler's number (e = 2.71828...)
- k! is the factorial of k.

The positive real number λ is equal to the expected value of X and also to its variance

$$\lambda = \mathrm{E}(X) = \mathrm{Var}(X).$$

The Poisson distribution can be applied to systems with a large number of possible events, each of which is rare. The number of such events that occur during a fixed time interval is, under the right circumstances, a random number with a Poisson distribution.

APPLICATIONS IN BIOSTATISTICS

Whether one observes patients arriving at an emergency room, cars driving up to a gas station, decaying radioactive atoms, bank customers coming to their bank, or shoppers being served at a cash register, the streams of such events typically follow the Poisson process. The underlying assumption is that the events are statistically independent and the rate, μ, of these events (the expected number of the events per time unit) is constant. The list of applications of the Poisson distribution is very long. To name just a few more:

- The number of mutations on a given strand of DNA per time unit
- The number of file server virus infection at a data center during a 24-hour period.

Consider a simple emergency room example where 2 patients arrive, on average, every 10 minutes (this is equivalent to 0.2 patients per one minute). Consecutive arrivals are statistically independent. This means that each given arrival has no impact on the probability of next arrivals. Letting N(t) represent the number of such arrival in t minutes, one can evaluate the following probabilities, using the Poisson distribution (as implemented in Excel or Google spreadsheet):

P{N(60) = 10} =Poisson(10,60*0.2,False) = 10.48%

P{N(60) ≤ 10} =Poisson(10,60*0.2,True) = 34.72%

P{N(60) < 10} =Poisson(9,60*0.2,True) = 24.24%

P{N(60) > 10} =1-Poisson(10,60*0.2,True) = 65.28%

P{N(60) ≥ 10} =1-Poisson(9,60*0.2,True) = 75.76%

As shown above, the Poisson distribution is a special case of the Binomial distribution. In some situations for former one can be used to approximate the latter one. It is particularly feasible if, for of a Binomial random variable, the number of trials, n, is extremely large and the probability of success, p, is very small.

The Poisson distribution provides a good approximation of the Binomial distribution, if n ≥ 100, and np ≤ 10.In such situations, events attributed to successes are called rare events. The Poisson distribution has been particularly useful in handling such events.

A leukemia case of Woburn, Massachusetts, falls into the category of rare-event situations shows the following case:

In early 1990s, a leukemia cluster was identified in the Massachusetts town of Woburn. Many more cases of leukemia, a malignant cancer that originates in a cell in the marrow of bone, appeared in this small town than would be predicted. Was it evidence of a problem in the town, or was it a chance? That question led to a famous trial in which the families of eight leukemia victims sued and became grist for a book and movie A Civil Action. Following an 80-day trial, the judge called for a retrial after dismissing the jury's contradictory and confusing findings. Shortly thereafter, the chemical companies and the families settled.

The issue of evidence versus chance is at a core of statistical studies. Using common data for the same period, one can estimate the probably of leukemia cases nationwide and then compare them with local results. The total US population of 280,000,000 and an annual average of leukemia cases of 30,8000 provide the Binomial probability, p, of success of about 30,8000/280,000,000 =0.00011.

CONCLUSIONS

The Poisson distribution has a strong theoretical background and very wide spectrum of practical applications. Bringing original and/or unusual cases, featuring Poisson processes, may provide opportunities for increasing students' attentiveness and interests in Statistics. One important lesson the author of this paper has learned is that presentation of statistical cases, including Poisson examples, should be accompanied and enriched by significant business, social or historical background description and discussion.

REFERENCES

https://en.wikipedia.org/wiki/Poisson_distribution
https://brilliant.org/wiki/poisson-distribution/
https://www.intmath.com/counting-probability/13-poisson-probability-distribution.php
https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459
https://www.itl.nist.gov/div898/handbook/eda/section3/eda366j.htm
https://www.zoology.ubc.ca/~bio300b/poissonnotes.html
http://www.aabri.com/SA12Manuscripts/SA12083.pdf
https://online.ucpress.edu/abt/article/79/1/10/18882/The-Poisson-Distribution-in-Biology-Instruction