# Poisson and Binomial Distribution

Alessio Palmisano
Institute of Archaeology, University College London
a.palmisano@ucl.ac.uk
Word Count: 2,036

## Abstract
Probability distributions are important tools for assessing the probability of the outcomes that occur. In particular, archaeologists have used the Binomial and the Poisson distributions in order to model the likelihood of discrete archaeological phenomena. This section introduces these two methods and shows practical examples where they can be applied to archaeological contexts. The two distributions are, therefore, comparatively described and evaluated on their similarities and differences. The Binomial distribution models the probability of "successes" and "failures" in a fixed number of trials. Instead, the Poisson distribution counts the occurrences occurring in a given unit of time or space with no fixed cutoff.

**Keywords:** Probability distributions, spatial statistics, archaeology, probability models, discrete distributions

## Main Text

The Binomial and Poisson distributions are two statistical methods suitable for calculating discrete archaeological phenomena, which means that they have been used for assessing the probability of the outcomes that occur. This section introduces these two distributions and comparatively shows their differences and similarities.

The Binomial distribution is a means to model the likelihood of discrete multiple phenomena in which only two possible outcomes are possible such as the presence or the absence of material evidence in a given archaeological context (Banning 2000, 124-125). The two possible outcomes of this distribution are generally designated as "success" ($p$) and "failure" ($q$). The binomial distribution describes the behavior of a random count variable X only if all the following conditions are met: 1) the number of trials ($n$) is fixed; 2) each trial has two possible outcomes as "success" ($p$) or "failure" ($q$); 3) the probability of success $p$ is the same for each trial; 4) each trial is independent, meaning that the outcome of one trial does not affect that of any other. The Binomial distribution is applied to cases such as coin flipping, where there is a fixed number of trials, the outcome is either head or tail, the probability of success (e.g. getting a head) is $p = 0.5$ for each trial, and the outcome of one flip does not affect the outcome of other flips. For the case in which a coin is flipped ten times and counting the number of heads, the parameters of the binomial distribution will be the number of trials $n = 10$ and the probability of success $p = 0.5$, and the distribution is conventionally abbreviated as B (10, 0.5). This means that there is the fifty percent of probability of getting a head when flipping a coin ten times. Given that there are only two possible outcomes, the probability of success ($p$) in flipping a con for each trial is $P(p) = 0.5$, and the probability of failure ($q$) is $P(q) = (1-p)$, which is $1 – 0.5 = 0.5$. Instead, the number of sixes by rolling a die ten times has a Binomial distribution B (10, 1/6). So, the probability of success $p$ for

each trial is 0.167, while the probability of failure *q* is 1 – 0.167 = 0.83. In Archaeology there are several different applications for the Binomial Distribution. For instance, paleo-botanists use Binomial distribution to measure the presence or absence of particular plant taxa in a fixed numbers of samples (Popper 1988, 63). This distribution has also been used to assess at the site of Casas Grande the likelihood of a co-occurrence of particular decorative motif with male or female effigies (VanPool and VanPool 2006).

For instance, the Binomial distribution can be applied to a scenario where the probability of finding an arrowhead in an excavation unit of one squared meter is 0.3, and the excavated units are 10. In this case the Binomial distribution is symbolized as B (10, 0.3). Therefore, it is possible to calculate the probability of finding one or more arrowheads in a given number of trials. Probabilities of a random variable X can be found by applying the following formula (Van Pool and Leonard 2011, 80-81):

$$\mathrm{P(x)} = \frac{n!}{x!(n-x)!}\, p^x (1-p)^{n-x} \qquad (1)$$

where P(x) is the probability of success out of *n* trials, *x* is the specified number of successes (finding an arrowhead), *n* is the fixed number of trials, *p* is the probability of success on any given trial, and (1-*p)* is the probability of failure (*q*). The expression $\frac{n!}{x!(n-x)!}$ is known as binomial coefficient and the notation *n!* stands for the factorial of a number, which can be calculated by multiplying a given number by all integers between itself and 1 (e.g. 3! = 3 × 2 × 1).

Using the formula (1) it is possible to get the probability of finding 4 arrowheads (x) after excavating 10 excavation units:

$$\mathrm{P(4)} = \frac{10\times9\times8\times7\times6\times5\times4\times3\times2\times1}{4\times3\times2\times1(10-4)!}\ 0.3^4(1-0.3)^{10-4} = 210\ \times\ 0.3^4\times\ 0.7^6 = 0.20$$

Figure 1a-b shows the different probabilities density function of finding 0 to 10 arrowheads and cumulative distribution for the Binomial Distribution (10, 0.3).

The probability of finding from 0 to 4 arrowheads is then the sum of these probabilities. The probabilities are 0.03, 0.12, 0.23, 0.26, and 0.20. Hence, the probability of finding 4 or fewer arrowheads is 0.84. The cumulative probability of finding more than 4 arrowheads after excavating 10 excavation units is 1 – 0.84 = 0.16.

Continuing excavating over and over again it is possible to calculate the long-term average of the expected successes (number of arrowheads found) over the entire population of trials (excavation units). The mean of a binomial random variable X is calculated by the following formula:

$$\mu_x = np \qquad (2)$$

where *n* is the number of trials and *p* the probability of success. Hence, the mean of the binomial distribution after digging 20 excavation units is *np* = 20 × 0.3 = 6.

The variance of the random variable X is $\sigma^2 = np\,(1-p)$ and the standard deviation of X is just the square root of the variance, which is $\sigma = \sqrt{np\,(1-p)}$. The variance in the binomial distribution B(20, 0.3) is 6(1-0.3) = 6 × 0.7 = 4.84, and the standard deviation is the square root, which is 2.2.

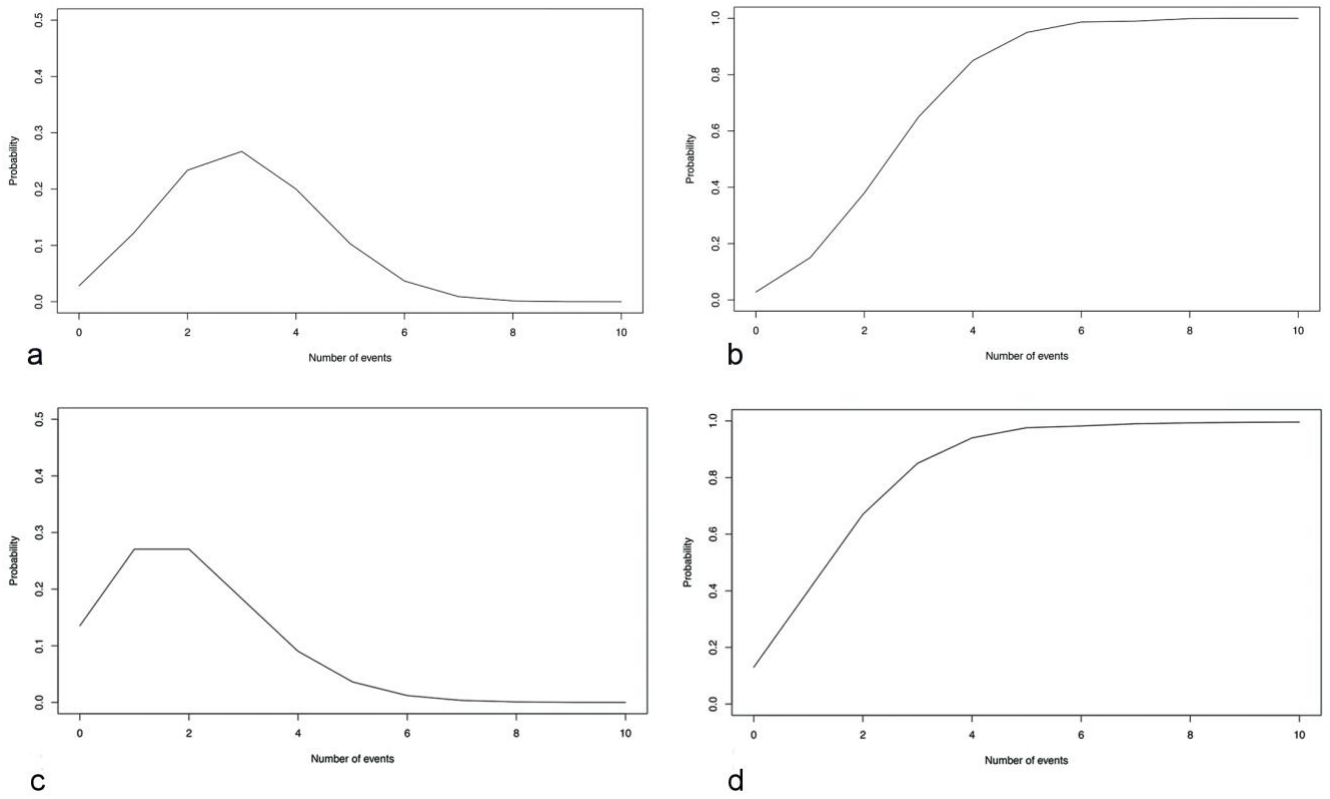


**Figure 1.** Probability (a) and cumulative distribution function (b) for binomial distribution B (10, 0.3), and Poisson distribution with $\lambda = 2$ (c, d).

The Poisson distribution is useful for modeling and finding the probability for the number of events occurring in a specific interval or time or within a given area of space. A random variable X, the number of events in a fixed unit of time or space, has a Poisson distribution when the following assumptions are true: 1) the events occur independently of each other; 2) the average frequency that an event occurs in a fixed time or space does not change through time; 3) Two events cannot happen simultaneously; 4) The probability of the occurrence of an event in a fixed interval is proportional to the length of the interval. Unlike the Binomial distribution, the Poisson distribution does not model the probability of the frequency of "successes" ($p$) or "failures|" ($q$) occurring in a fixed number ($n$) of trials. Instead, it provides the expected number of events in a fixed unit of time or space. This distribution is usually used to test if specific events were generated by a random process. In addition, because the Poisson distribution is characterized by no fixed cutoff, the random variable X can be a positive integer from zero to infinity. The Poisson distribution closely approximates the Binomial distribution when $n$ is large, $p$ is very small (one rule of thumb is that it should be no greater than 0.1), and $\lambda = \mu_X$. The Poisson probability of any given number of the variable X can be calculated with the following formula:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (3)$$

Where x is the number of the events whose probability is calculated, $\lambda$ is the average number of events occurring per unit of time or space, and *e* is a constant that is approximately equal to 2.71828, and x! is the factorial number of x.

The Poisson distribution can be useful to model events such as the number of patients arriving at an emergency room between 10 and12 am, the number of goals scored in a World Cup's football match, the number of students enrolling at the University in a year, the number of phone calls received by the Police's phone centre between 11-12 pm, and so on. In Archaeology the Poisson distribution has been generally used to test if past events have randomly occurred per unit of time and space. For instance, the rate of disintegrations of radioactive $^{14}C$ per minute is described by a Poisson distribution (Banning 2000, 125; Buck. Cavanagh, and Litton 1996, 105). This distribution has also been used by to assess if the spatial configuration of archaeological artifacts or sites in a given study area is the result of past random processes (Bevan et al. 2013; Crema, Bevan, and Lake 2010; Palmisano 2013).

Using the formula (3) it is possible to calculate the probability of the number of events (e.g. artifacts or sites) found within a given spatial unit (e.g. a square). For instance, in a scenario where 20 events are found in 10 squares, and $\lambda$ is 20/10 = 2, the probability of an empty square is given by the following formula:

$$P(0) = \frac{2^0(2.718..)^{-2}}{0!} = (2.718..)^{-2} = 0.135$$

and the probability of a square containing one event is:

$$P(1) = \frac{2^1(2.718..)^{-2}}{1!} = 2 \times 0.135 = 0.27$$

Probabilities for larger values of X are calculated in the same way to get the distribution shown in Figure 1c.

Although the graph does not show the probabilities for values higher than 10, this is just to save space. In fact, there is a low probability of having a square containing 10 or more events. In the Poisson distribution it is theoretically possible having hundreds of events within on square, even though that probability would be extremely low. The Figure 1d shows the cumulative probability from zero to X for any specific value of X. This is useful to get the probability of finding more or less X events within a square. In the present example the probability of finding 1 or less events in a square is 0.135 + 0.27 = 0.405. Once all probabilities are found, then it is possible to calculate the expected number of squares that can contain X events by multiplying the relevant probability by the total number of squares (Lloyd 2011, 248-249). In this case the expected numbers of squares respectively containing 0 and 1 events is 1.35 and 2.7. Instead, the expected number of squares containing 4 events will be 0.09 x 10 = 0.9. Comparing the expected frequencies with the observed data allows one to assess if the events are distributed randomly in a given study area. For instance, the expected number of squares containing 10 events (0.0003) is close to 0 according to the Poisson distribution. Nevertheless, if in the observed data there is one or more squares containing 10 events, it means that the number of squares containing X events is more than expected and the events are clustered in one or more parts of the surveyed area. Poisson models are, therefore, useful to assess how well a null hypothesis of complete spatial randomness explains the observed distribution of artifacts or sites in a given study area.

**SEE ALSO:** saseas0322, saseas0553, saseas0478, saseas0545

## References

Banning, E. B. 2000. *The Archaeologist's laboratory: The Analysis of Archaeological Data*. New York: Kluver Academic Publishers.

Bevan, A, Crema, E.R., Li, Xiuzhen, and Palmisano, A. 2013. "Intensities, interactions and uncertainties: some new approaches to archaeological distributions". In *Computational Approaches to Archaeological Space,* edited by A. Bevan, and M. Lake, M., 27-52. Walnut Creek: Left Coast Press.

Buck, C. E, Cavanagh, W. G., and Litton, C. D. 1996. *Bayesian Approach to Interpreting Archaeological Data: Statistics in Practice.* New York: John Wiley & Sons.

Crema, E. R., Bevan, A. and Lake, M. 2010. "A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record ". *Journal of Archaeological Science,* 37: 1118-1130.

Lloyd, C. 2011. *Local Models For Spatial Analysis*. London: CRC Press.

Palmisano, A. 2013. "Zooming Patterns Among the Scales: a Statistics Technique to Detect Spatial Patterns Among Settlements." In *CAA 2012: Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, England,* edited by G. Earl, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska, and D. Wheatley, 348-356. Amsterdam: Amsterdam University Press.

Popper, V. S., 1988. "Selecting quantitative measurements in paleoethnobotany." In *Current Paleoethnobotany, Analytical Methods and Cultural Interpretations of Archaeological Plant Remains,* edited by C. Hastorf, and V. Popper, 53-71. Chicago: University of Chicago Press.

Van Pool, T. L., and Leonard, R. D. 2011.*Quantitative Analysis in Archaeology*. Chichester: Wiley-Blackwell.

VanPool , C.S. and VanPool, T. L. 2006. "Gender in middle range societies: A case study in Casas Grandes iconography." *American Antiquity*, 71: 53 – 75.

## Further Readings

Banning, E. B. 2002. *Archaeological survey*. New York: Springer Science & Business Media, 49-54.

Barceló, J. A., Achino, K. F., Bogdanovic, I., Capuzzo, G., Del Castillo, F., de Almeida, V. M., and Negre, J. 2015. "Measuring, Counting and Explaining: An Introduction to Mathematics in Archaeology". In *Mathematics and Archaeology*, edited by J. A. Barceló, and I. Bogdanovic, 3-64. Boca Raton, Florida: CRC Press.

Dalgaard, P. 2008. *Introductory statistics with R*. New York: Springer Science & Business Media.

Goreaud, F., and Pélissier, R., 2000. "Spatial structure analysis of heterogeneous point patterns: examples of application to forest stands." *ADS in ADE-4 Topic Documentation* 8: 1-49.

Kintigh, Keith W. 1988. "The effectiveness of subsurface testing: a simulation approach." *American Antiquity* 53(4): 686-707.

Orton, C. 1982. *Mathematics in archaeology*. Cambridge: Cambridge University Press.

Orton, C. 2007. "Horse kicks, flying bombs and potsherds: statistical theory contributes to archaeological survey." *Archaeology International* 10: 24-27.

Perry, G. L., Miller, B. P., and Enright, N. J. 2006. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant ecology* 187(1): 59-82.

Ripley, Brian D. 1976. "The second-order analysis of stationary point processes." *Journal of applied probability*: 255-266.

Thomas, D. H. 1972. "A computer simulation model of Great Basin Shoshonean subsistence and settlement patterns." In *Models in archaeology* , edited by D. L. Clarke, 671-704. New York: Routledge.