Applications of the Poisson probability distribution

Jerzy Letkowski

Western New England University

Abstract

The Poisson distribution was introduced by Simone Denis Poisson in 1837. It has since been subject of numerous publications and practical applications. The purpose of this paper is to raise awareness of numerous application opportunities and to provide more complete case coverage of the Poisson distribution. First a formal definition and basic characteristics of a Poisson variable and its distribution are summarized. Next cases, representing time and space oriented Poisson situations, are presented. Probability assessment solutions, using functions built in spreadsheet programs, are presented. Finally, pedagogical issues, regarding challenges in teaching Statistics are discussed.

Keywords: statistics, probability, distribution, random variable, Poisson.

POISSON VARIABLE AND DISTRIBUTION

The Poisson distribution is a probability distribution of a discrete random variable that stands for the number (count) of statistically independent events, occurring within a unit of time or space (Wikipedia-Poisson, 2012), (Doane, Seward, 2010, p.232), (Sharpie, De Veaux, Velleman, 2010, p. 654), (Jaggia, S., Kelly, A., 2012, p. 157) , (Donnelly, 2012, p. 215), (Anderson, Sweeney, Williams, 2012, p. 236). Given the expected value, μ, of the variable, X, the probability function is defined as:

$$f(n) = P(X = n) = \frac{e^{-u}\mu^n}{n!}, \quad F(n) = \sum_{k=0}^{k=n} f(k), \quad n = 0,1,2,...$$

A more generalized attempt to define this distribution is shown in (Levine, Stephan , Krehbiel, Berenson, 2011) where the unit of time or space is referred to as an area of opportunity (p. 177). It is an interesting generalization of this important context of the Poisson distribution.

Wolfram MathWorld (2012) arrives at this distribution by taking the limit of the Binomial (Bernoulli) distribution when the number, N, of trials becomes very large ($N \to \infty$). A Binomial random variable represents the number of successes in a series of independent and probabilistically homogenous trials (Wikipedia-Binomial, 2012). A relation of the Poisson distribution to the Binomial distribution is also motioned in (Pelosi, Sandifer, 2003) and (De Veaux, Velleman, Bock 2006, p. 388).

Assessment of probabilities for Poisson variables is not complicated. There are many Web sites that provide calculators to this end ("Poisson Calculator ...", 2012). All major, contemporary statistical packages and spreadsheet programs are well equipped with functions for computing Poisson probabilities. For example, Google spreadsheet and Microsoft Excel provide the same syntax for the distribution functions:

$P(X = n)$ =Poisson(n, μ, *false*), $P(X \leq n)$ =Poisson(n, μ, *true*)

Notice that the last parameter of the Poisson function indicates whether or not the function returns the cumulative probability. An example for computing Poisson probabilities in a Google spreadsheet is shown in Figure 1 (Appendix). A similar function is available in the Open Office Spreadsheet program, =*Poisson*(x; μ; 0|1). Notice that this spreadsheet program uses a semi-colon to separate the function's arguments.

TIME ORIENTED POISSON VARIABLES

Feller (1966, p. 17) shows how the Poisson distribution can be derived from a series of Exponentially distributed random variables, $S_n = X_1 + X_2 + ... + X_n$. Considering a random variable, $N(t)$, such that $N(t) = \max\{n: S_n \leq t\}$ and given all variables X_k , k=1,2,...,n, have the same exponential distribution, $f(x) = \mu e^{-\mu x}$, $x \geq 0$, the $N(t)$ variable has this distribution:

$$f(t,n)) = P\{N(t) = n\} = \frac{e^{-\mu t}(\mu t)^n}{n!}, \quad F(t,n) = P\{N(t) \leq n\} = \sum_{k=0}^{k=n} f(t,k), \quad n = 0,1,2,...$$

Consequently, the Poisson random variable can be "stretched" over longer or shorter time intervals. Since μ is the expected (average) number of events per one unit of time or space, μt

will be such a number per t units. One has to make sure that process $N(t)$ is stationary within time interval $(0,t)$.

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 soldiers of the Prussian army killed accidentally by horse kick per year (von Bortkewitsch, 1898, p. 25).
- The number of mutations on a given strand of DNA per time unit (Wikipedia-Poisson, 2012).
- The number of bankruptcies that are filed in a month (Jaggia, Kelly, 2012 p.158).
- The number of arrivals at a car wash in one hour (Anderson et al., 2012, p. 236).
- The number of network failures per day (Levine, 2010, p. 197).
- The number of file server virus infection at a data center during a 24-hour period . The number of Airbus 330 aircraft engine shutdowns per 100,000 flight hours. The number of asthma patient arrivals in a given hour at a walk-in clinic (Doane, Seward, 2010, p. 232).
- The number of hungry persons entering McDonald's restaurant. The number of work-related accidents over a given production time, The number of birth, deaths, marriages, divorces, suicides, and homicides over a given period of time (Weiers, 2008, p. 187).
- The number of customers who call to complain about a service problem per month (Donnelly, Jr., 2012, p. 215) .
- The number of visitors to a Web site per minute (Sharpie, De Veaux, Velleman, 2010, p. 654).
- The number of calls to consumer hot line in a 5-minute period (Pelosi, Sandifer, 2003, p. D1).
- The number of telephone calls per minute in a small business. The number of arrivals at a turnpike tollbooth par minute between 3 A.M. and 4 A.M. in January on the Kansas Turnpike (Black, 2012, p. 161).

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%

$P\{N(60) \leq 10\}$ =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%

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

Notice that in real world situations, due to the aspect of "seasonality", the arrival rate may remain constant only within limited time intervals ("seasons"). In such situations, variable *t* must not go beyond the upper limits of such seasonal intervals. For example, a morning rate may stay constant between 5:00 and 11:00 a.m., a noon rate— between 1:00 a.m. and 2:00 p.m., etc. Figure 2 (Appendix) shows implementation of the above formulas in a Google spreadsheet.

Time based Poisson variables are more popular. One can find numerous examples and cases, involving such random variables in most to the contemporary statistical textbooks or papers.

SPACE ORIENTED POISSON VARIABLES

Space oriented Poisson variables are less common or popular. Nonetheless, one can find some interesting examples of such variables in most recent Statistics books. For example:

- The number of typographical errors found in a manuscript or the total number of home runs hit in Major League Baseball games (Donnelly, Jr., 2012, p. 215).
- The number of surface defects on a new refrigerator or the number of fleas on the body of a dog (Levine et al., 2011, p.).
- The number of blemishes per sheet of white bond paper (Doane, Seward, 2010, p. 232) .
- The number of repairs needed in 10 miles of highway or the number of leaks per 100 miles of pipeline (Anderson et al., 2012, p. 237).
- The number of defects in a 50-yard roll of fabric or the number of bacteria in a specified culture (Jaggia, Kelly, 2012, p.158).
- The number of defects that occur in a computer monitor of a certain size (Sharpie et al., 2010, p. 654).
- The number of a certain type of insect that can be found in a 1-square-foot area of farmland (Pelosi, Sandifer, 2003, p. D1).
- The number of hazardous waste sites per county in the United States or the number of sewing flaws per pair of jeans during production (Black, 2012, p. 161).
- The number of eagles nesting in a region or the number of hits by V-1 buzz bombs in WWII London (Triola, 2007, p. 252-253) .

Among the above cases, of a particular interest is the last one. During the World War II, Germans fired thousands of so called V-1 buzz bombs over the English channel toward London. About 800 bombs managed to hit some targets in London. The rest either did not make it over the water or were destroyed by RAF. A particular case, presented by Triola (2007, p. 253), focuses on South London. This region was subdivided into 576 regions, each of 0.25 km^2. The entire region was hit by 535 V-1 bombs.

By analogy to the number of patients arriving at an emergency room per one minute, the random variable for the V-1 bomb case, represents the number of bomb hits per region. Since the each area is of the same size, the average number of hits par one area can be expressed as: μ = 537/575 = 0.9288. Figure 3 shows a Google spreadsheet based model for this case. The probabilities and expected number of areas hit the x V1 bombs are calculated in the following way:

P(X=n) =Poisson(n, 0.9288, False)

E(AreaCount | n) =576*Poisson(n, 0.9288, False)

Detail spreadsheet solution is shown in (Google-V1 Bombs, 2012).

Feller (1961, p. 145) shows remarkably close fit of the actual numbers of bomb hits as compared to the expected numbers provided by the Poisson distribution. Figure 3 (Appendix) shows the numbers along with a goodness-of-fit test outcomes.

APPROXIMATION OF THE BINOMIAL DISTRIBUTION

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. According to Triola (2007, p. 254) the Poisson distribution provides a good approximation of the Binomial distribution, if $n \geq 100$, and $np \leq 10$. 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 De Veaux et al. (2006 p. 387-388) 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. Detail calculations are shown in (De Veaux et al., 2006 p. 388-389) and in a spreadsheet at Google-Woburn (2012). Figures 4 and 5 (Appendix) show the output. With the Woburn population of 35,000, the probability of at least 8 leukemia cases, as computed in at Google-Woburn (2012), using the Binomial and Poisson functions, is:

$P(X \geq 8)$ =1 - BINOMDIST(7,35000, 0.00011,True) = #NUM

$P(X \geq 8)$ =1 - POISSON(7,35000*0.00011,True) = 4.25%

So the Binomial function in a Google spreadsheet fails to compute this probability, reporting that "argument 35000" is too large, as shown in Figure 4 (Appendix). On the other hand, the Poisson function provides a satisfactory approximation. Interestingly, in an Excel spreadsheet both the Binomial and Poisson function provide the same [correct] outcome, as shown in Figure 5 (Appendix).

The assessed probability of 4.25% is small and can be attributed to sources other than chance. Some may consider a too-close-to-call value. It is below but quite close to the common significance level of 5%. As De Veaux et al. (2006 p. 388) concludes: "That's small but not terribly unusual."

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. For example, a presentation of the WWII V1 London bombing case may be accompanied by a short youtube.com movie, showing how V1 Buzz bombs were being launched to hit London targets (V- 1 Flying Bomb, 2011). Such an exposure to background information makes "technical" cases more interesting and instructive.

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., (2012), Essentials of Modern Business Statistics with Microsoft® Excel. Mason, OH: South-Western, Cengage Learning.

Black, K. (2012) Business Statistics For Contemporary Decision Making. New York, NY: John Wiley and Sons, Inc.

Bortkewitsch, L. (1898). Das Gesetz der Kleinen Zahlen. Leipzig, Germay: Teubner.

De Veaux, R. D., Velleman, P. F., Bock D.E. (2006) Intro to Stats. Boston, MA: Addison Wesley, Pearson Education, Inc.

Doane, D., Seward, L. (2010) Applied Statistics in Business and Economics, 3rd Edition, Mcgraw-Hill, 2010.

Donnelly, Jr., R. A. (2012) Business Statistics. Upper Saddle River, NJ: Pearson Education, Inc.

Feller, W. (1961) An Introduction to Probability Theory and Its Applications, Vol. 1. New York, NY: John Wiley & Sons Inc.

Feller, W. (1966) An Introduction to Probability Theory and Its Applications, Vol. 2. New York, NY: John Wiley & Sons Inc.

Google-Emergency Room (2012) Exploring Poisson Distribution (a Google spreadsheet application for case Emergency Room). Retrieved from: https://docs.google.com/spreadsheet/ccc?key=0AsmhQG4y08HcdFBwc0ttVTNuSDI5a2ZTbnBoNDk4Snc

Google-Poisson (2012) Exploring Poisson Distribution (a Google spreadsheet application). Retrieved from: https://docs.google.com/spreadsheet/ccc?key=0AsmhQG4y08HcdGwzT0FmZDFVMC1GbDlLVGd0Z0k3Y2c

Google-V1 Bombs (2012) Exploring Poisson Distribution (a Google spreadsheet application for case WWII South London Bombing by V1 Buzz Bombs). Retrieved from: https://docs.google.com/spreadsheet/ccc?key=0AsmhQG4y08HcdHBMNTBZcS1YZ0JQd3pHVmlaM2pQOFE

Google-Woburn (2012) Exploring Binomial and Poisson Distribution (a Google spreadsheet application for case Woburn, MA Leukemia). Retrieved from: https://docs.google.com/spreadsheet/ccc?key=0AsmhQG4y08HcdGZzTzhORklGdEZwczVGbDRfdXdZMXc

Hu, H. (2008) Poisson distribution and application. Retrieved from: http://sces.phys.utk.edu/~moreo/mm08/Haohu.pdf.

Jaggia, S., Kelly, A. (2012) Business Statistics - Communicating with Numbers. New York, NY: McGraw-Hill Irvin.

Levine, D. M., Stephan, D.F., Krehbiel, T.C., Berenson, M.L. (2011) Statistics for Managers Using Microsoft® Excel, Sixth Edition. Boston, MA: Prentice Hall.

Pelosi, M. K., Sandifer, T.M. (2003) Elementary Statistics. New York, NY: John Wiley and Sons, Inc.

Poisson Calculator: Online Statistical Table (2012) Retrieved from: http://stattrek.com/tables/poisson.aspx

Sharpie, N. R., De Veaux, R. D., Velleman, P. F. (2010) Business Statistics. Boston, MA: Addison Wesley.

Spiegel, M. R., (1975) Probability and Statistics, Schaum's Outline Series in Mathematics. New York, NY: McGraw-Hill Book Company.

Triola, M. F. (2007) Elementary Statistics Using Excel®. Boston, MA: Addison Wesley, Pearson Education, Inc.

V- 1 Flying Bomb (2011) Fieseler Fi 103 (Vergeltungswaffe). Retrieved from: http://www.youtube.com/watch?NR=1&v=QY308O42Ur4&feature=endscreen

Weiers, R. M. (2008) Introduction to Business Statistics. Mason, OH: South-Western, Cengage Learning.

Wikipedia-Poisson (2012) Poisson Distribution. Retrieved from: http://en.wikipedia.org/wiki/Poisson_distribution

Wikipedia-Binomial (2012) Binomial Distribution. Retrieved from: http://en.wikipedia.org/wiki/Binomial_distribution

Wolfram MathWord (2012) Poisson Distribution. Retrieved from: http://mathworld.wolfram.com/PoissonProcess.html.

Appendix

fx =POISSON(C13,C6,true)-POISSON(C12-1,C6,true)

	A	B	C	D	E	F	G	H
1								
2		Exploring Poisson Distribution						
3								
4		X			a Poisson random variable			
5								
6		μ	10		the expected number of events per one unit of time or space			
7								
8		n	15		the number of the events to occur within one unit of time or space			
9		P(X = 15)	3.47%		the probability that exactly 15 events will occur			
10		P(X ≤ 15)	95.13%		the probability that up to (no more than) 15 events will occur			
11								
12		n1	8		at least 8	events will occur		
13		n2	12		no more 12			
14		P(8 ≤ X ≤ 12)	57.13%		the probability that between 8 and 12 events will occur			

Figure 1 Doing the Poisson distribution in a Google spreadsheet (Google-Poisson, 2012).

fx =1-POISSON(C8-1,C6*C7,true)

	A	B	C	D	E	F	G	H
1								
2		Exploring Poisson Distribution - Case: Emergency Room						
3								
4		N(t)			the number of patients arriving in t minutes			
5								
6		μ	0.2		the expected number of events per one minute			
7		t	60		the length of the time interval			
8		n	10		the number of patients to arrive in 60 minutes			
9		P{N(60) = 10}	10.48%		the probability that exactly 10 patients will arrive in 60 minutes			
10		P{N(60) ≤ 10}	34.72%		the probability that up to (no more than) 10 patients will arrive in 60 minutes			
11		P{N(60) < 10}	24.24%		the probability that fewer than 10 patients will arrive in 60 minutes			
12		P{N(60) > 10}	65.28%		the probability that more than 10 patients will arrive in 60 minutes			
13		P{N(60) ≥ 10}	75.76%		the probability that at least 10 patients will arrive in 60 minutes			
14								
15		n1	8		at least 8	patients will arrive in 60 minutes		
16		n2	12		no more than 12			
17		P(8 ≤ X ≤ 12)	48.65%		the probability that between 8 and 12 patients will arrive in 60 minutes			

Figure 2 The Poisson distribution. An emergency room situation (Google-Emergency Room, 2012) .

fx =POISSON(C9,C7,false)

Exploring Poisson Distribution - V1 Buzz Bombing of WWII London

X		a Poisson random variable: the number of V1 bombs hitting a region.
bomCnt	537	the total number of V1 bombs that hit South Loandon
areaCnt	576	the number of areas (subdivisions) of South London
μ	0.9323	the expected number of V1 bomb hits per area
n	2	the number of the bombs to hit a randomly selected region
P(X = 2)	17.11%	the probability that exactly 2 events will occur
E(areaCnt | n=2)	98.54	the expected number of areas hit 2 times

Observed vs. Expected Area Count

X	f(x)	Expected Number of Areas Hit by x Bombs	Observed Number of Areas Hit by x Bombs	Squared Difference	χ2
0	0.3937	226.7	229	5.2900	0.0233
1	0.3670	211.4	211	0.1600	0.0008
2	0.1711	98.5	93	30.2500	0.3071
3	0.0532	30.6	35	19.3600	0.6327
4	0.0124	7.1	7	0.0100	0.0014
	1-F(x)				
5+	0.0027	1.6	1	0.36	0.225
		Total	576	55.43	1.1903

χ2 Test	
df	4
α	5%
Critical χ2	9.4877
Observed χ2	1.1903

Comment: Since the observed χ2 is way below the critical one there is no reason to believe that the observed data do not support the Poisson distribution. This is actually an example of almost perfect fit.

Figure 3 The Poisson distribution. A V1 buzz WWII London bombing situation (Google-V1 Bombs, 2012).

fx =1-POISSON(C8-1,C14,True)

Binomial Distribution

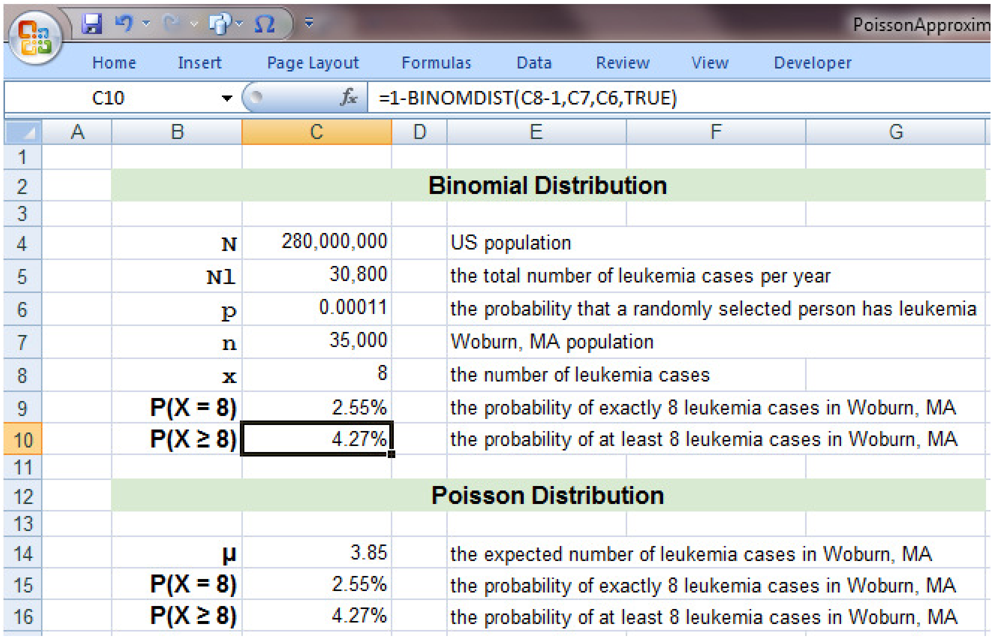
N	280,000,000	US population
N1	30,800	the total number of leukemia cases per year
p	0.00011	the probability that a randomly selected person has leukemia
n	35,000	Woburn, MA population
x	8	the number of leukemia cases
P(X = 8)	#NUM!	error: Argument too large: 35000 kemia cases in Woburn, MA
P(X ≥ 8)	#NUM!	

Poisson Distribution

μ	3.85	the expected number of leukemia cases in Woburn, MA
P(X = 8)	2.55%	the probability of exactly 8 leukemia cases in Woburn, MA
P(X ≥ 8)	4.27%	the probability of at least 8 leukemia cases in Woburn, MA

Figure 4 The Binomial and Poisson distributions. A Woburn, MA leukemia situation (Google-Woburn, 2012).

Applications of the Poisson probability



	A	B	C	D	E
1					
2		**Binomial Distribution**			
3					
4		N	280,000,000		US population
5		Nl	30,800		the total number of leukemia cases per year
6		p	0.00011		the probability that a randomly selected person has leukemia
7		n	35,000		Woburn, MA population
8		x	8		the number of leukemia cases
9		P(X = 8)	2.55%		the probability of exactly 8 leukemia cases in Woburn, MA
10		P(X ≥ 8)	4.27%		the probability of at least 8 leukemia cases in Woburn, MA
11					
12		**Poisson Distribution**			
13					
14		μ	3.85		the expected number of leukemia cases in Woburn, MA
15		P(X = 8)	2.55%		the probability of exactly 8 leukemia cases in Woburn, MA
16		P(X ≥ 8)	4.27%		the probability of at least 8 leukemia cases in Woburn, MA

Figure 5 The Binomial and Poisson distributions. A Woburn, MA leukemia situation (an Excel workbook).

Applications of the Poisson probability