The Application of Poisson Distribution Model

Yunjie Hou [1, *, †], Junxiang Zhang [2, †]

[1] College nine, University of California Santa Cruz, Santa Cruz, California, USA

[2] Dalian Leicester institution of Mathematics, Dalian University of Technology, Dalian, China & School of Mathematics, University of Leicester, Leicester, UK

* Corresponding author. Email: Yhou25@ucsc.edu

†These authors contributed equally.

Abstract. Poisson distribution, which provides the probability of events occurring in a settled time interval when events occur at a accurate average rate, is one of the most useful statistical distribution to solve many problems in the world today. In this paper, the definitions and properties of Poisson distribution are introduced. Besides, Poisson distribution is used to predict the household prices and infectious diseases. This prediction is supported by collecting the data and setting the Poisson distribution. Nonetheless, a direct experimental observation of the reviewing literature is lacking. The prediction of household's prices in the future and the key periods in the infectious disease prevention are discussed by observing the characteristics of distribution of probability of infectious disease in peak period which suit the Poisson distribution. Then Poisson model is set to discuss. Our work can facilitate the study for the application of Poisson distribution in household prices and infectious disease.

Keywords: Poisson distribution, Estimating households by household size, Infectious disease research.

1. Introduction

Poisson distribution is a tool that helps to predict the probability of certain events happening when the researchers know the frequency of the event has occurred. It gives us the probability of a given number of events happening in a fixed interval of time. In addition to its use for staffing and scheduling [1,2], the Poisson distribution also has applications in biology [3], finance [4], disaster readiness, and any other situation in which events are time independent. For instance, businesses can use the Poisson distribution to check how to take measures to improve their operational efficiency. An analysis applied with the Poisson Distribution may reveal how a business can organize personnel to better manage peak periods for customer service calls.

The Poisson distribution is suitable for describing the frequency of events occurred every unit time. As long as we have the average number of events every unit time, the other numbers which will occur are under prediction. Hence, people can make effective decisions to gain maximum profits or predict some information needed.

In this paper, firstly, a brief introduction of the Poisson distribution is given, including the definition, some basic properties and a classic example using the Poisson distribution, which is flipping a coin with an infinite number of times. Secondly, we give two applications of the Poisson distribution which will be studied, which is estimating households by household size and doing research on infectious disease. Finally, a conclusion of papers will be given.

2. Mainbody

2.1. Definitions And Properties of Poisson Distribution

2.1.1. Definitions

In statistics, the Poisson distribution is a probability distribution that reflects the frequency that a serious independent events occur in a fixed time interval. It was named after a French mathematician of 18th and 19th centuries [5]. The Poisson distribution is a discrete function, which means the independent variable can only take countable values in an array. In other word, the variable cannot take all the values in any continuous ranges. As for the Poisson distribution, Poisson random variable can only take natural numbers 0, 1, 2, 3, …. The Poisson distribution has only one parameter, which is commonly described as a lowercase Greek letter $\lambda > 0$, and used to express the frequency that a serious independent events occur in a fixed time interval. The probability mass function (P.M.F.) of Poisson distribution is

$$f(x) = P\{X = x\} = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0,1,2,\cdots \tag{1}$$

As expected, the P.M.F. of the Poisson distribution f(x) is non-negative. Then, it is easily to have that for all x in domain,

$$f(x) \geq 0 \tag{2}$$

The Poisson distribution is normalized so that the sum of probabilities equals 1, since

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1, \tag{3}$$

With the definition of expectation, it is easily that

$$E[X] = \lambda, E[X^2] = \lambda(\lambda+1). \tag{4}$$

$$Var[X] = E[X^2] - E^2[X] = \lambda. \tag{5}$$

Therefore, the expected value and the variance of a Poisson random variable are both equal to the parameter λ.

The Poisson random variable has many applications because it is used as an approximation for a binomial random variable with parameter (n, p) when n is large and p is small enough. To see this, suppose that X is a binomial random variable with parameters (n, p), and let $\lambda = np$.

After the substitution, it follows that

$$P\{X = i\} \approx \frac{e^{\lambda}\lambda^{i}}{i!} \tag{6}$$

In other word, if there are n independent trials, which have a probability p of success, when n is large enough and p is small enough, the number of an event occurring successfully is approximately a Poisson random variable with parameter [6].

2.1.2. Example and Properties

In Bernoulli trials of binomial distribution, if the number of trials n is large, the probability P of binomial distribution is small, and $\lambda = np$ is moderate, then the probability of the number of events can be approximated by Poisson distribution. In fact, the binomial distribution can be regarded as the counterpart of the Poisson distribution in discrete time.

Now we consider flipping a coin with an infinite number of times. The probability of getting the head of the coins in a coin flip is infinitely close to zero （$p = \frac{\lambda}{n}$）. More specifically, it's very hard to flip heads of the coins. But the expectation of times that the coin comes up in n trials (λ) is invariable.

An event is called the coin heads arrival in the Poisson process. The number of arrivals within the unit time is referred to as the arrival rate.

The properties that Poisson process satisfies are as follows:

In any intervals of unit time, the arrival rate is stable. For infinite number of coin flips, we essentially divide one unit of time into an infinite number of coin flips, with the same probability of heads $\frac{\lambda}{n}$. After an infinite number of flips, we expect to flip λ heads of a coin.

The results of future experiments have nothing to do with the results of past experiments. In the case of the infinite number of flips, no matter how many heads and tails you flip before, it doesn't affect how the coin turns up later.

With a very short period of time, the probability of 1 arrival is low, while the probability of no arrival is very high. For infinite number of coin flips, it is figured out that the probability $p = \frac{\lambda}{n}$ that the coin comes up approaches to 0.

2.2. Application in Estimating Households By Household Size

The United States created the American Community Survey (ACS) to provide data on the distribution of household size across census tracts in five-year data products. At small geographic scales, single-year data is not available. Furthermore, although the ACS provides useful historical data, infrastructure planners also require information on projected households by size.

The Poisson distribution describes the potential number of objects that may be found in a certain volume, or the number of events that will be observed in a given time period.

The Poisson distribution has a wide variety of applications in biology and engineering. There are four criteria for using Poisson distribution in a specific case. Firstly, natural numbers can describe events, and the number of occurring is unlimited. Households are described in natural numbers. In theory, the maximum size of a household is unlimited. While extremely large families are unlikely, the number of persons that lives together is unlimited in a household. Secondly, occurrences are independent and random. The size of a particular household is independent and randomly determined. If two households are neighboring, the number of persons within their family is completely independent. Thirdly, each event is counted only once and is not relevant to the number of cases that don’t take place. Households satisfy the rules that each household is calculated only once, and the number of cases that are not counted is not relevant. Lastly, the average frequency (or population mean) is known. Average household size = household population households. When the four key criteria are met, the Poisson distribution is a good candidate to estimate the size distribution of households. Therefore, it is proved that Poisson distribution can be used under household and household size situations: The Poisson function can be written as:

$$f(n; M) = \frac{M^n e^{-M}}{n!} \tag{7}$$

In this function, n is the number of occurrences (where $n= 0 \to\infty$), M is the population mean, and f (n; M) is the probability of case n, when average=M.

To test the model, this research compares 10-year census data with the model and survey estimates of the ACS.

For mean error bias, ACS estimates reveal fewer biases within each size class. For counties and tracts, size one households are typically underestimated since the average error is negative by using the Poisson model, while size no.3 are in general over-estimated because of the positive average error.

For mean absolute error, both models are reasonably accurate, and neither the models nor the survey estimates perform substantially better than the other.

The Poisson model and ACS estimates are both suitable for larger households. Although the model is well suited for large families, there should be more treatment to improve the model to decrease the error for size one and three.

2.3. Application in Infectious Disease Research

Numerous scholars have conducted extensive research in describing the characteristics of public health emergencies of infectious diseases by using relevant indicators of dynamic series [7], [8]. This section concerned with Poisson distribution model of the occurrence frequency of infectious disease public health emergencies in Shanxi province, and obtained the corresponding probability distribution characteristics of the event frequency, so as to provide reference for guiding the prediction and emergency preparedness of infectious disease public health events in Shanxi province [9].

Table 1. Data collection of Poisson distribution model and ACS for the census area in California [10]

	Mean error					
	Size 1	Size 2	Size 3	Size 4	Size 5	Size 6
Model	-0.049	-0.015	0.076	0.007	0.000	-0.002
ACS	0.011	0.01	0.001	0.003	-0.004	-0.005
	Mean absolute error					
	Size 1	Size 2	Size 3	Size 4	Size 5	Size 6
Model	0.055	0.035	0.076	0.027	0.012	0.009
ACS	0.037	0.042	0.036	0.034	0.026	0.019

Table 2. Data collection of Poisson distribution model and ACS for counties [10]

	Mean error					
	Size 1	Size 2	Size 3	Size 4	Size 5	Size 6
Model	-0.044	-0.012	0.088	-0.004	-0.011	-0.007
ACS	0.006	0.008	-0.004	-0.003	-0.003	-0.002
	Mean absolute error					
	Size 1	Size 2	Size 3	Size 4	Size 5	Size 6
Model	0.044	0.022	0.088	0.011	0.011	0.007
ACS	0.016	0.015	0.013	0.011	0.008	0.005

We will use the Excel to establish a database and conduct data mining analysis to describe the characteristics of public health emergencies of infectious diseases by using relevant indicators of dynamic series. We will use R to carry out relevant statistical analysis (test level $\alpha = 0.05$).

Poisson distribution model will be used to analyze the distribution of the incidence rate or the number of sick people with very low incidence of non-communicable diseases in the population. The establishment of Poisson distribution model is conducting by statistical analysis on the average number of public cattle incidents with bovine infectious diseases in each month from 2013 to 2020. Then we obtain the parameters by calculating the probability of K public health emergencies of infectious diseases in a specific month through formula of the probability of various rare times x, P{X=k}. Taking P{X=k} as the prediction coefficient, it can be calculated that the theoretical frequency of K public health emergencies of infectious diseases per month is 8P{X=K}, and K is 0,1,2,3…. Finally, the Poisson distribution model of the frequency of infectious disease public health emergencies in Shanxi province can be established by putting the parameters of 12 months into the separate substitution formula.

The Table 3 shows the Poisson distribution probability of infectious disease public health emergencies in Shanxi province (Probability of number of emergencies%) in 12 months. The model prediction shows that the frequency of infectious disease public health emergencies in Shanxi province is mainly 1~2 times, with the highest probability in May and the lowest in March, August and September. The top five probabilities are may (6753%), November (58.32%), June and December (52.76%) and January (46.47%). The above results are basically consistent with the monitoring information. Therefore, May, November, June, December and January should be the key periods of infectious disease prevention and control in Shanxi province, and make full preparations for the corresponding emergency work.

This section further confirms the frequency distribution characteristics of infectious disease public health emergencies in Shanxi province through the practice of model imitation and identification. In conclusion, Poisson distribution model can predict and analyze the probability characteristics of infectious disease public health emergencies.

3. Conclusion

In this article, the definition and properties of Poisson distribution are reviewed. Then two applications of Poisson distribution are listed based on their definition and properties. At first, this study presents the definition by describing the P.M.F formula and how it can be substituted. Next, the three properties of Poisson distribution are introduced by an example of flipping coins. The first application is about how Poisson distribution is used to estimate the household. By analyzing the data and developing the model, the result is acquired that besides ACS, Poisson distribution can also become a usable method to predict households. The second application is about how Poisson distribution is applied in infectious disease research. Throughout a data table that shows the Poisson distribution probability of infectious disease public health, the result is concluded that the Poisson distribution model can predict and analyze the probability characteristics of infectious disease public health emergencies. According to the results of two applications, this research finds out the practicability and universality of Poisson distribution and the researchers believe it will become more significant in the future.

Table 3. Probability of infectious disease public health emergencies in Shanxi province

	0	1	2	3	4	≥ 1
Jan.	53.52	33.45	10.45	2.28	0.34	46.47
Feb.	77.88	19.47	2.43	0.20	0.01	22.12
Mar.	88.24	11.03	0.68	0.03	0.00	11.75
Apr.	60.65	30.33	7.58	1.26	0.16	39.35
May	32.47	36.52	20.54	7.70	2.17	67.53
June	47.24	35.43	13.28	3.32	0.62	52.76
Jul.	77.88	19.47	2.43	0.20	0.01	22.12
Aug.	88.24	11.03	0.68	0.03	0.00	11.75
Sept.	88.24	11.03	0.68	0.03	0.00	11.75
Oct.	60.65	30.33	7.58	1.26	0.16	39.35
Nov.	41.69	36.48	15.96	4.65	1.02	58.32
Dec.	47.24	35.43	13.28	3.32	0.62	52.76

References

[1] F. Dexter, A. Macario, Optimal number of beds and occupancy to minimize staffing costs in an obstetrical unit. Can Journal of Anesthesia, vol. 48, 2001, pp. 295 - 301. DOI: 10.1007/BF03019762.

[2] J. Jiang, X. Xiong, Y. Ou, H. Wang, An Improved Bacterial Foraging Optimization with Differential and Poisson Distribution Strategy and its Application to Nurse Scheduling Problem. In: Tan, Y., Shi, Y., Tuba, M. (eds) Advances in Swarm Intelligence. ICSI 2020. Lecture Notes in Computer Science, vol 12145, Springer, Cham, 2020, pp. 312 – 324. DOI: 10.1007/978-3-030-53956-6_28.

[3] K. G. Janardan, H.W. Kerster, D. J. Schaeffer, Biological Applications of the Lagrangian Poisson Distribution, BioScience, vol. 29, 1979, pp. 599 - 602, DOI: 10.2307/1307766.

[4] A. Gandy, L. A. M. Veraart, Compound Poisson models for weighted networks with applications in finance, LSE Research Online Documents on Economics, vol.15, 2021, pp.131 - 153. DOI: 10.1007/s11579 - 020 - 00268 - 9.

[5] A. Katz, A. Hayes, T. Suresh, et al. Poisson Distribution, Brilliant.org. 2022, https://brilliant.org/wiki/poisson-distribution/.

[6] S. Ross, A First Course in Probability 9th Edition, Pearson, 2014, pp. 146 - 158.

[7] R Routledge, Poisson distribution, Encyclopedia Britannica, 2020, https: //www.britannica.com/topic/Poisson-distribution.

[8] S. Song, H. Ding, C. Huang, et al, Application of poisson distribution in early warning of HFMD, Zhejiang journal of preventive medicine, vol. 27, 2015, pp. 32-35, DOI: 1007 - 0931(2015) 01 - 0032 - 04.

[9] J. Chen, S. Mu, L. Fan, S. Zuo, et al, Study on probability characteristics of Poisson distribution of public health emergencies of infectious diseases, in Shanxi Province, Chinese Remedies & Clinics, vol. 21, 2021, pp. 3505 - 3506. DOI: 10.11655/zgywylc2021.21.001.

[10] B. Jarosz, Poisson distribution: A model for estimating households by household size, Population Research and Policy Review, vol. 40, 2021, pp. 149-162. DOI: 10.1007/s11113-020-09575-x.