Vol.9.Issue.2.2021 (April-June)
©KY PUBLICATIONS

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

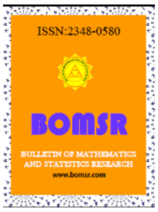
BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

AN EMPIRICAL APPROACH TO ASSESS THE RELATIONSHIP BETWEEN THE BINOMIAL AND POISSON DISTRIBUTION

RAMNATH TAKIAR
Scientist G – (Retired)
National Centre for Disease Informatics and Research (NCDIR),
Indian Council of Medical Research (1978-2013)
Bangalore – 562110, Karnataka, India.
DOI:10.33329/bomsr.9.2.9



ABSTRACT

The present paper explore empirically the relationship between the Binomial and Poisson distribution. The paper also attempts to find out the values of “n” and “p” above which the Poisson distribution can safely be approximated to Binomial distribution? For different combinations of n (10, 15, 20 and 50), p (0.01, 0.05, 0.1 and 0.2) and the number of successes (X = 0, 1, 2, 3…10), the probabilities are calculated using both the Binomial and Poisson approaches and comparisons are made. For n ≤ 50 (10, 15, 20, 50) and P=0.01, in more than 95% of the cases, the difference between the Binomial and Poisson probabilities are found to be less than 0.005. For P = 0.05, in 93.8% of the cases and when P = 0.10, 84.3% of the cases, the differences between the probabilities are observed to be below 0.015. However, when p=0.2, this percentage decreased to 71.9%. A very high correlation, above 0.99, is observed between the Binomial and Poisson probabilities suggesting the existence of a good linear relationship among them. In the light of the results of the present study, the concept that the Poisson distribution can be approximated by Binomial distribution when n is large, and p tends to zero need to be reexamined. The present study clearly proves that even when 10 ≤ n ≤50 and p ≤ 0.2, the Poisson distribution can be a good approximation to Binomial distribution.

Keywords: Binomial distribution, Poisson distribution, Relationship, n = 10, p=0.01

INTRODUCTION

One of the common discrete distribution which is used in Statistics is the Binomial distribution. The Binomial distribution often deals with two types of situations, first, the occurring of an event and second, the non-occurring of an event termed respectively as success and failure. The Binomial distribution can be typically characterized by the two parameters namely “n” and “p” where “n” represents the number of trials and “p” represents the fixed probability of occurring of the desired event. Further, the Binomial distribution can be visualized as an approach to determine the probability of observing a specified number of successes in a specified number of trials. The typical examples of Binomial distribution in our day-to-day life may include the survival of number of cancer patients, from the treatment offered, over a specified period of time; Winning the number of seats against the number of candidates filed by a certain political party for an election; Number of people benefitting from a certain house scheme when applied; Number of covid patients discharged within three days from the hospital after their admission to a hospital for the treatment.

There is another discrete distribution which is equally popular as the Binomial distribution is the Poisson distribution. While the Binomial distribution deals with the number of trials and the probability of a success, Poisson distribution deals with the average number of successes per day or the number of successes observed per unit of time. Unlike, in the case of a Binomial distribution, for a Poisson distribution, the knowledge of number of trials to get the observed number of successes is not required. The Poisson distribution has only one parameter namely “m” the mean number of successes per given unit of time. The typical examples of this distribution may include, the number of accidents occurring during a day on a certain busy road, the number of typographical errors found in a page of a certain book, or the number of bacteria growing per unit of time in an experimental solution.

The Poisson distribution is often taken as a limiting case of Binomial distribution, under the following conditions (Gupta and Kapoor 2001, Gupta 2012):

- The number of trails is indefinitely large i.e., $n \to \infty$
- The probability of success, “p” is very small i.e., $p \to 0$
- np is finite i.e., $np = m$ where $p = m/n$ and $q = 1 - m/n$ and $m > 0$.

However, it is not defined if “n” is large means how large and “p” is closed to zero means how small? It is left to the readers to assume how large n should be and how small p should be? As mentioned earlier, the Poisson distribution is often described as a limiting case of Binomial distribution. One article state that whenever $n \geq 100$ and $np \leq 10$, the Poisson distribution can provide a very good approximation to Binomial distribution (Oxford College, accessed in 2021). Another article says that it is sufficient if $n > 50$ and $np < 5$, then the Poisson distribution can provide a good approximation to Binomial distribution (JBstatistics accessed in 2021).

The present papers explore empirically the relationship between the Binomial and Poisson distribution? The paper also attempts to find out the values of “n” and “p” above which the Poisson distribution can safely be approximated to Binomial distribution?

METHODOLOGY:

Let X be a Bernoulli variable. The probability of X = x successes can be obtained by the probability distribution of Binomial distribution, given by

$P(X = x) = {}^nC_x p^x q^{1-x}$ where

"p" is the fixed probability of getting a success and "n" is the number of trials made.

If we assume that X follows a Poisson distribution, then the probability of "x" successes can be given as follows:

$P(X = x) = e^{-m} m^x / x!$

For the study purposes, first the "p" value was selected as P = 0.01, then for n = 10, and for varying values of x = 0,1,2, 3,….10, the Binomial and Poisson probabilities are calculated until one of the probability becomes 0. This exercise was repeated for each n = 15, 20, and 50. As a next step, for other values of P = 0.05, 0.1 and 0.2, again for selected combination of n and x, the Binomial and Poisson probabilities are calculated. For the calculation of Poisson probabilities np = m was assumed.

All the probabilities obtained for the Binomial and Poisson distribution for varying x (1,2,3….10) and n (10, 15, 20, 50) are pooled for each P level, separately and correlation (r), slope (b) is obtained, and the regression equation is formulated. The scatter graph along with the trend line is also plotted for different "P" levels, to show the linear relationship between the Binomial and Poisson probabilities.

RESULTS

For P = 0.01, the Binomial and Poisson probabilities are calculated for selected n, X and shown in Table 1. The second and the third column provide the probabilities of the Poisson and Binomial distribution. For example, for the calculation of Poisson probabilities np = m = 10*0.01 = 0.1 is considered while for the binomial n=10 and p= 0.01 is considered.

Table 1: The Binomial and Poisson Probabilities for the Selected n and X when P = 0.01

	n=10		n=15		n=20		n=50	
X	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial
0	0.9048	0.9044	0.8607	0.8601	0.8187	0.8179	0.6065	0.6050
1	0.0905	0.0914	0.1291	0.1303	0.1637	0.1652	0.3033	0.3056
2	0.0045	0.0042	0.0097	0.0092	0.1465	0.1369	0.0758	0.0756
3	0.0002	0.0001	0.0005	0.0004	0.0164	0.0159	0.0126	0.0122
4	0.0000	0.0000	0.0000	0.0000	0.0011	0.0010	0.0016	0.0015
5					0.0001	0.0000	0.0002	0.0001

Table 2: The Binomial and Poisson Probabilities for the Selected n and X when P = 0.05

n=10		n=15		n=20		n=50	
Poisson	Binomial	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial
0.6065	0.5987	0.4724	0.4633	0.3679	0.3585	0.0821	0.0769
0.3033	0.3151	0.3543	0.3658	0.3679	0.3774	0.2052	0.2025
0.0758	0.0746	0.1329	0.1348	0.1839	0.1887	0.2565	0.2611
0.0126	0.0105	0.0332	0.0307	0.0613	0.0596	0.2138	0.2199
0.0016	0.001	0.0062	0.0049	0.0153	0.0133	0.1336	0.136
0.0002	0.0001	0.0009	0.0006	0.0031	0.0022	0.0668	0.0658
-	-	0.0001	0	0.0005	0.0003	0.0278	0.026
-	-	-	-	0.0001	0	0.0099	0.0086
-	-	-	-	-	-	0.0031	0.0024
-	-	-	-	-	-	0.0009	0.0006
-	-	-	-	-	-	0.0002	0.0001

In 81.8% (18 out of 22) of cases, the Poisson probabilities are higher than Binomial probabilities. To form an idea about how close the Binomial and Poisson probabilities are there, the absolute differences among them are obtained. For example, when n=10 and X = 0, the Poisson and Binomial probabilities calculated are 0.9048 and 0.9044, respectively. Similarly, when n = 15 and X = 0, the Poisson and Binomial probabilities are 0.8607 and 0.8606, respectively. Clearly, the difference among the above two probabilities is 0.0004 and 0.0001, respectively. Considering the distribution of the differences of the probabilities, it is seen that in more than 95% of the cases, the differences are observed to be less than 0.005.

On the line of the Table1, Table 2, Table 3 and Table 4 are generated for P = 0.05, 0.1 and 0.2, respectively.

Table 3: The Binomial and Poisson Probabilities for given n and X when P = 0.10

n=10		n=15		n=20		n=50	
Poisson	Binomial	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial
0.3679	0.3487	0.2231	0.2059	0.1353	0.1216	0.0067	0.0052
0.3679	0.3874	0.3347	0.3432	0.2707	0.2702	0.0337	0.0286
0.1839	0.1937	0.251	0.2669	0.2707	0.2852	0.0842	0.0779
0.0613	0.0574	0.1255	0.1285	0.1804	0.1901	0.1404	0.1386
0.0153	0.0112	0.0471	0.0428	0.0902	0.0898	0.1755	0.1809
0.0031	0.0015	0.0141	0.0105	0.0361	0.0319	0.1755	0.1849
0.0005	0.0001	0.0035	0.0019	0.012	0.0089	0.1462	0.1541
0.0001	0	0.0008	0.0003	0.0034	0.002	0.1044	0.1076
		0.0001	0	0.0009	0.0004	0.0653	0.0643
				0.0002	0.0001	0.0363	0.0333
						0.0181	0.0152

More than observing the pair of probability values of the Poisson and Binomial distribution, the differences among them is expected to throw more light on their comparability. Lower difference values like 0.01 or 0.015 will support the view that both the probabilities are comparable while high differences that is more than 0.015 will support the view that both the probabilities are differing probably to a great extent. Hence, a table is generated showing the frequency distribution of differences for each of the selected P level (0.01, 0.05, 0.1 and 0.2) and shown in Table 5.

For P = 0.01, in more than 95% cases, the difference between the Poisson and Binomial probabilities are observed to be less than 0.005. When P = 0.05, in 93.8% of the cases and when P = 0.10, 84.3% of the cases, the differences between the Poisson and Binomial probabilities are observed to be below 0.015. However, when P = 0.20, this percentage decreased to 71.9%.

To study the linear relationship between the Binomial and Poisson probabilities, the correlation (r) is seen separately, and the scatter plot is plotted between the Binomial and Poisson probabilities for each P value and shown in Fig.1 to Fig. 4.

From the Fig. 1, it can be observed that the Poisson and Binomial probabilities are very close to each other. The fact that $r^2$is 0.99995 clearly establishes the close linear relationship between the two sets of probabilities. The regression line can easily be seen to be passing through the origin and slope is 1.00047.

Similar conclusions can be drawn from the Fig. 2 where r^2 is 0.99546 which is sufficient to show that the model fitted is quite good and there exist a very strong linear relationship between the Binomial and Poisson probabilities.

Table 4: The Binomial and Poisson Probabilities for given n and X when P = 0.20

	n=10		n=15		n=20		n=50	
X	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial	Poisson	Binomial
0	0.1353	0.1074	0.0498	0.0352	0.0183	0.0115	0	0
1	0.2707	0.2684	0.1494	0.1319	0.0733	0.0576	0.0005	0.0002
2	0.2707	0.302	0.224	0.2309	0.1465	0.1369	0.0023	0.0011
3	0.1804	0.2013	0.224	0.2501	0.1954	0.2054	0.0076	0.0044
4	0.0902	0.0881	0.168	0.1876	0.1954	0.2182	0.0189	0.0128
5	0.0361	0.0264	0.1008	0.1032	0.1563	0.1746	0.0378	0.0295
6	0.012	0.0055	0.0504	0.043	0.1042	0.1091	0.0631	0.0554
7	0.0034	0.0008	0.0216	0.0138	0.0595	0.0545	0.0901	0.087
8	0.0009	0.0001	0.0081	0.0035	0.0298	0.0222	0.1126	0.1169
9	0.0002	0	0.0027	0.0007	0.0132	0.0074	0.1251	0.1364
			0.0008	0.0001	0.0053	0.002	0.1251	0.1398

Table 5: The frequency distribution of difference in Binomial and Poisson Probabilities for the selected n and X and Varying P values

Difference Range	p = 0.05		p = 0.10		p = 0.20	
	Frequency	%	Frequency	%	Frequency	%
0.005 - 0.010	24	75	24	63.2	18	41.8
0.010 - 0.015	6	18.8	8	21.1	13	30.2
0.015 - 0.02	2	6.2	2	5.3	3	7
0.020 -0.025	0	0	4	10.5	4	9.3
0.025-0.030	-	-	-	-	2	4.7
0.030 - 0.035	-	-	-	-	2	4.7
0.035-0.040	-	-	-	-	1	2.3
Total	32	100	38	100	43	100

The linear relationship between the Binomial and Poisson probabilities when P = 0.1, is shown in Fig. 3. The r^2 in this case was observed to 0.99571 which is quite high by any standards. The slope is 0.97369 which is again close to 1.

In Fig. 4, when P = 0.2, the r^2 is observed to be 0.83078, quite less as compared to when P was less than 0.20. The slope of 0.91147 suggests that when the Binomial probability changes by 0.1-unit, the Poisson probability changes by 0.09115 unit.

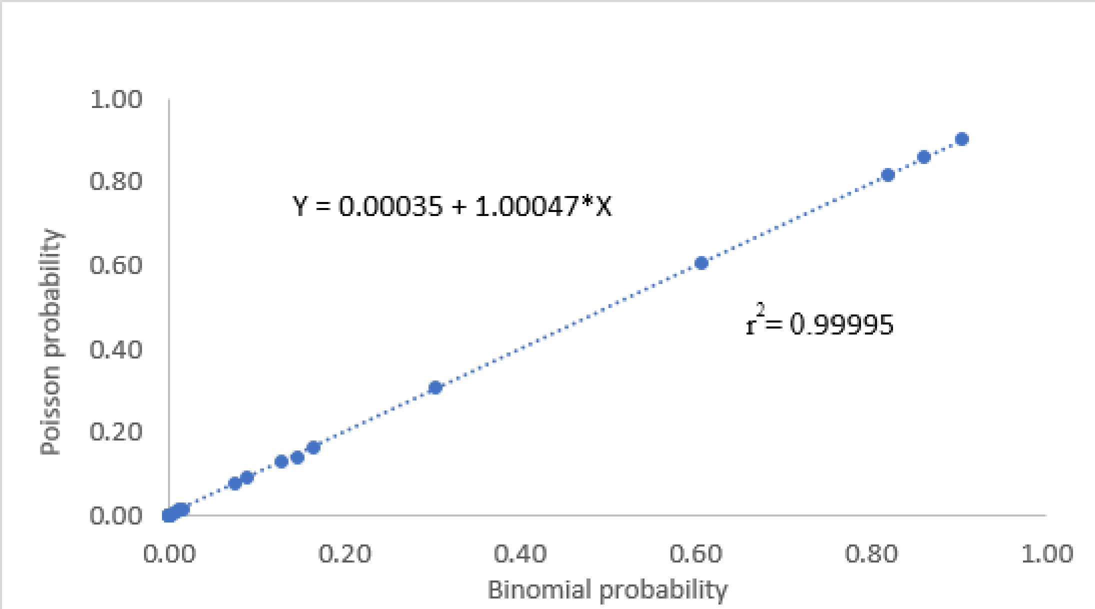


Fig. 1: The Linear Relationship between the Binomial and Poisson probabilities - (p=0.01; n=22)

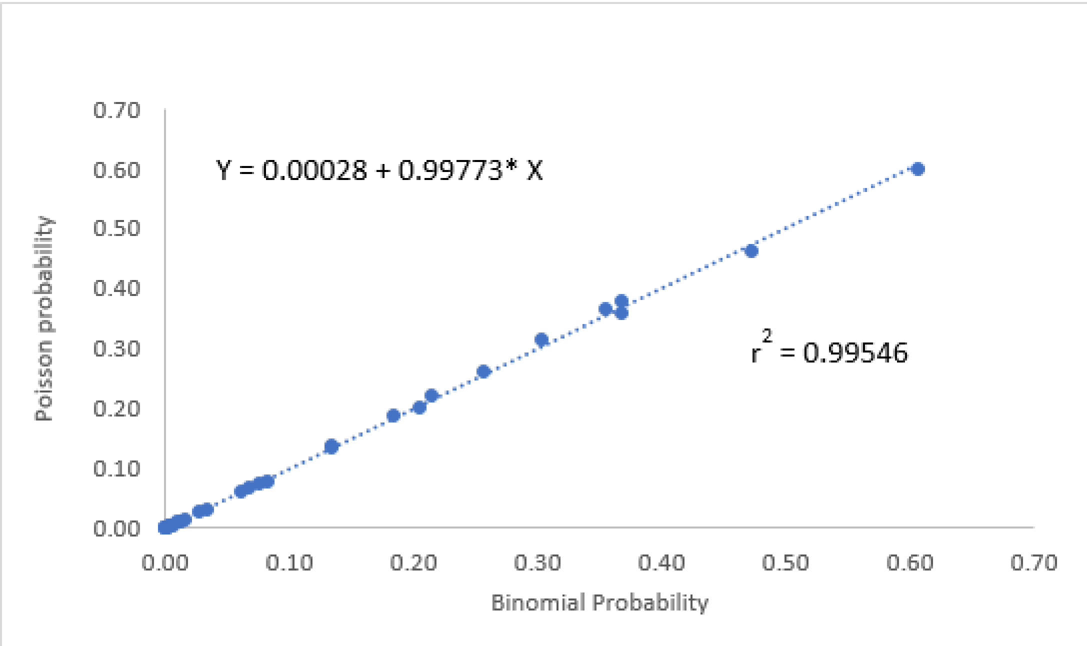


Fig. 2: The Linear relationship between the Binomial and Poisson probabilities (p = 0.05; n=32)

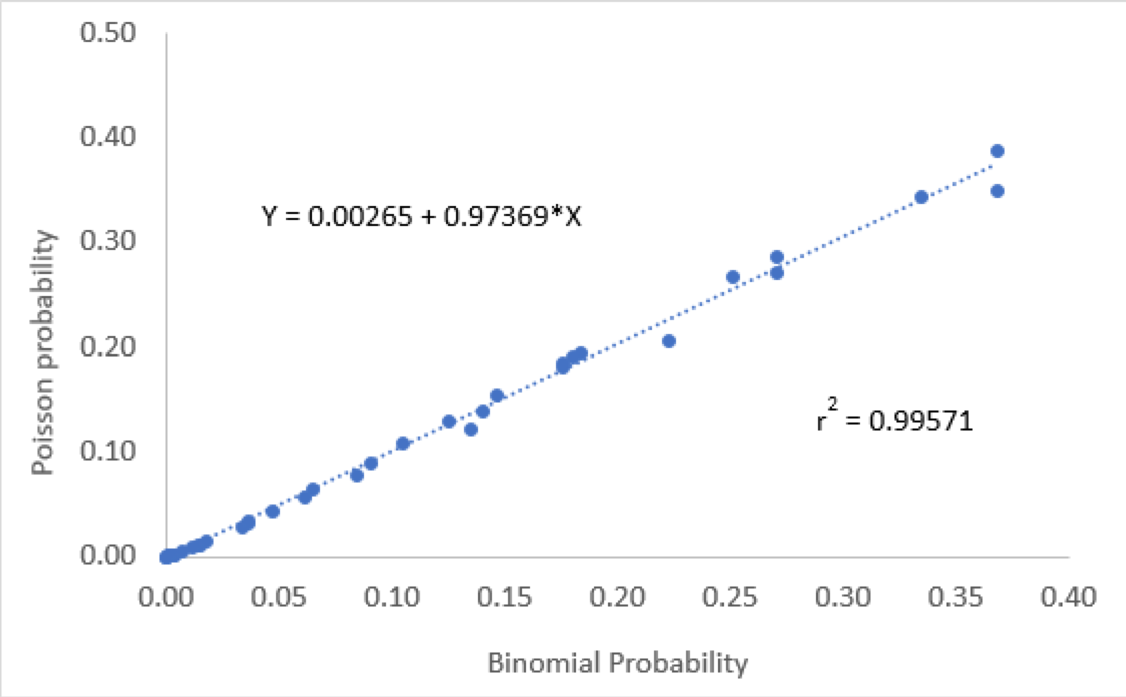


Fig. 3: The Linear Relationship between the Binomial and Poisson Probabilities (p = 0.1; n=38)

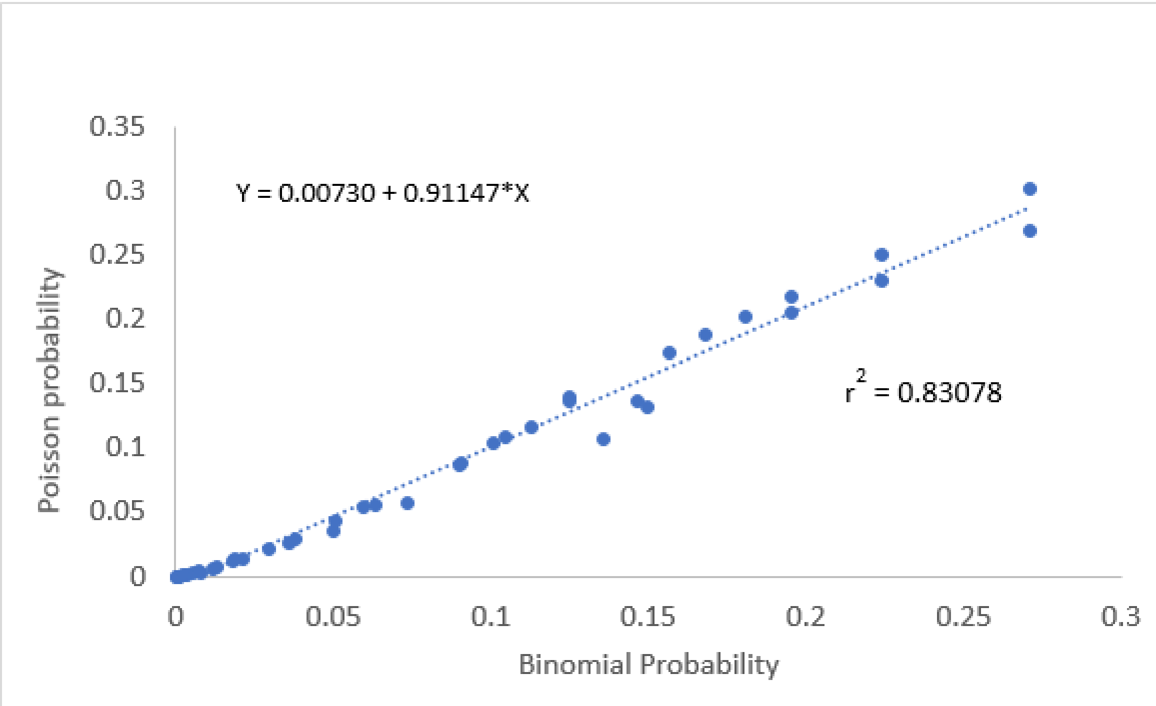


Fig. 4: The Relationship between the Binomial and Poisson probabilities (p = 0.20; n=43)

DISCUSSION

Most of the textbooks maintain that the Poisson distribution is a limiting case of Binomial distribution when n is large, and p tends to zero. However, we have seen that when n=10 and P = 0.01, even then the Poisson probabilities are quite close to corresponding Binomial probabilities. This contradicts the prevailing concept that the Poisson probabilities are close only when n is large, and p is close to zero. The study results strongly support that the Poisson probabilities can be close even when n is as low as 10 and P is as high as 0.20. I claim that this is a new finding. The correlation graphs also amply support the above claim. It is also seen that as P goes above 0.2, the relationship between the Binomial and Poisson distribution distorts and do not remain comparable. There is a need to carry out more studies to see why the relationship between the two distributions is distorted as P goes closer to 0.5. There is need to change the old concept that when n is large and p is close to zero, the Poisson distribution can be a good approximation to Binomial distribution. This statement should be modified suitably in the light of the results of the present study.

CONCLUSION

The concept that the Poisson distribution can be approximated by Binomial distribution when n is large, and p tends to zero need to be re-examined in the light of the results of the present study. The present study adequately proves that even when $10 \leq n \leq 50$ and $P \leq 0.2$, the Poisson distribution can be a good approximation to Binomial distribution.

REFERENCES

[1]. Gupta SC, Kapoor VK (2001): Fundamental of Mathematical Statistics, 10th edition, Sultan Chand & Sons, 7.40.

[2]. Gupta SC (2012): Fundamental of Statistics, 7th Edition, Himalaya Publishing House, 14.20.

[3]. JBstatistics: The Relationship between Binomial and Poisson distribution, *jbstatistics.com/the-relationship-between-the-binomial-and-poisson-distributions/*Accessed on 16/04/2021

[4]. Oxford College: The Connection between Poisson and Binomial Distribution, *http://mathcenter.oxford.emory.edu/site/math117/connectingPoissonAndBinomial/*ACCESSED ON 17/04/2021