Vol.9.Issue.3.2021 (July-Sept)

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

THE NORMAL APPROXIMATION TO THE POISSON DISTRIBUTION – HOW LARGE λ SHOULD BE?

RAMNATH TAKIAR
Scientist G – (Retired)
National Centre for Disease Informatics and Research (NCDIR),
Indian Council of Medical Research (1978-2013)
Bangalore – 562110, Karnataka, India.
Email: ramnathtakiar@gmail.com, ramnath_takiar@yahoo.co.in
DOI:10.33329/bomsr.9.3.1


RAMNATH TAKIAR

ABSTRACT

The Poisson distribution is a discrete distribution. It has only one parameter namely "λ", the mean number of successes per given unit of time and space. In literature, there are conflicting reports about the size of the λ that should be acceptable for the initiation of the Normal approximation. It ranges from 5-20. The present paper examines empirically the relationship between the Poisson and the Normal distribution. The paper also attempts to find out the values of "λ", above which the Poisson distribution can safely be approximated by the Normal distribution? For the different combinations of λ (2,3,5 and 10) and the number of successes (X = 0,1,2,3…n), the probabilities are calculated using the Poisson distribution approach and the Normal distribution approach and comparisons are made. If X follows the Poisson distribution with mean = λ, then Z = $\frac{x-\lambda}{\sqrt{\lambda}}$ is assumed to follow the standard Normal variate. For the same set of "λ" and X values, two types of Normal probabilities are calculated, termed as Normal approx-1 and Normal approx-2 and comparisons are made among them. In calculation of Normal probabilities, in case of Normal approx -1, the selected λ values are considered while for Normal approx-2, a correction 0f -0.35 was applied to the selected values of λ.

A high correlation of 0.95 or above was observed between the Poisson probabilities and the Normal approximated probabilities. The results of the K-S test also supported the view that both the Normal approximations to the Poisson distribution are valid even when λ is below 5 and equal to 2. This is in much deviation from the opinion expressed in the literature by the various authors that at least $\lambda > 5$ to go for the Normal Approximation of the Poisson distribution. Further, it is apparent from the results that a correction of -0.35 to λ values improve the Normal approximation (Normal approx-2) to the Poisson distribution in general and especially when $\lambda < 10$.

Keywords: Poisson distribution, Normal approximation, Correlation, K-S test, $\lambda < 5$

INTRODUCTION

Poisson distribution is a discrete distribution. It deals with the average number of successes observed per unit of time or space. Unlike in the case of a Binomial distribution, for the Poisson distribution, the knowledge of the number of trials to get the observed number of successes is not required. Poisson distribution has only one parameter namely "λ", the mean number of successes per given unit of time and space. Further, for the Poisson distribution, the whole interest lies in the random occurring of the event termed as the success. The typical examples of this distribution may include, the number of suicides occurring in an area, the number of persons losing jobs due to Covid conditions in a month, the number of patients dying on the surgery table in a hospital.

The Probability mass function of the Poisson distribution is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x\,!}$$ where x = 0,1,2,3……… n and denotes the number of successes

λ = Average number of successes based on the historical data

When "X" is a single number, the calculations involved with the above formula are simple. However, when we are interested in a range of outcomes as our success then the computations become complex. Suppose that in a city, on an average, 60 patients are reporting, every day, for Covid 19. In case, we are interested in assessing the probability of at least 50 persons reporting with Covid 19 conditions in the city on a given day, we require the calculations of 50 probabilities to answer the question. In such a situation, the use of Normal approximation to the Poisson distribution is recommended.

According to few text books, as $\lambda \to \infty$, the standard Poisson variate tends to the standard normal variate. However, it is not mentioned that when $\lambda \to \infty$ means how large it should be (Gupta and Kapoor 2001, Gupta 2012). According to Dinov if $\lambda > 20$, then the Normal approximation to the Poisson can be attempted and the Normal approximation improves as λ increases (Dinov 2019). Another author states that when $\lambda \geq 10$, the Normal distribution, with an appropriate continuity correction, is a good approximation to the Poisson distribution (AP Statistics 2014.) Few authors also suggest that to use the Normal approximation for the Poisson distribution, it is sufficient if $\lambda \geq 5$ (Calcworkshop 2021, Chaudhary 2017). So, in literature, we do not find a uniform suggestion as to how large λ should be to follow the Normal approximation for the Poisson distribution. In the present study, therefore, an attempt is made to find out empirically, how large λ should be so that a valid Normal approximation for the Poisson distribution can be carried out?

METHODOLOGY

For the study purposes, the Poisson and the Normal probabilities are calculated for different combinations of λ and X values until the calculated probabilities reached closed to zero. The values of X, generally, ranged from 0 to 20 and the values of λ considered were 2, 3, 5 and 10.

If X follows the Poisson distribution with mean = λ, then Z = $\frac{x-\lambda}{\sqrt{\lambda}}$ follows N (0,1). The probability of Z so calculated, can be found out using the table of Normal probabilities. As the Poisson distribution is a discrete distribution and the Normal distribution is a continuous distribution, an application of appropriate continuity correction is advocated. The scheme of continuity corrections is shown below:

- P (X = x) = P (x – 0.5 < X < x + 0.5)
- P (X < x) = P (X < x – 0.5)
- P (X ≥ x) = P (X ≥ x – 0.5)
- P (X ≤ x) = P (X ≤ x + 0.5)
- P (X > x) = P (X > x + 0.5)

For the study purposes, the Poisson and the Normal probabilities for given X values are calculated using the Excel functions (Microsoft Corporation 2019) as shown below:

POISSON.DIST(x, mean, cumulative)

Example: When λ = 2, x = 3, to find P (X = 3), use the Excel function, as follows:

POISSON.DIST(3, 2, False) will give you 0.1804 which is the required probability. As we are not interested in cumulative probability, we enter False for cumulative.

NORM.DIST(x, mean, standard_dev, cumulative)

For X = 3, on applying the continuity correction, we get, P (X = 3) = P (2.5 < X < 3.5)

CALCULATION OF Normal approx-1:

NORM.DIST(2.5, 2, 1.4142, TRUE) = 0.6382

NORM.DIST(3.5, 2, 1.4142, TRUE) = 0.8556

So, P (X=3) = P (2.5 < X < 3.5) = P (X < 3.5) – P (X < 2.5) = 0.8556 – 0.6382 = 0.2174

CALCULATION OF Normal approx-2

A correction of - 0.35 was applied to the value of λ, under consideration. This implies that when λ = 2, after the correction, λ would be λ = 2 – 0.35 = 1.65 and the Excel function used is

NORM.DIST(2.5, 1.65, 1.2845, cumulative) = 0.7459

NORM.DIST(3.5, 1.65, 1.2845, cumulative) = 0.9251

So, P (X=3) = P (2.5 < X < 3.5) = P (X = 3.5) – P (X = 2.5) = 0.9251 – 0.7459 = 0.1792

The probability due to Normal-approx-2 = 0.1792 is much closer to the Poisson probability of 0.1804 than the probability value calculated by Normal-Approx-1 = 0.2174.

In the present study, for the Poisson distribution, two Normal approximations are provided. To assess which approximation is better, the absolute deviations are considered. The approximation with the least sum of absolute deviations, is considered as the best.

TEST FOR GOODNESS OF FIT

It is interesting to see, if we approximate the Poisson distribution with the Normal approx-1 and Normal approx-2, whether they confirm to a goodness of fit or not? To test this, the D-statistic was calculated for each fit and according to Kolmogorov-Smirnov test, its significance or otherwise, was established (Charles 2021, Dransfield and Brightwell 2012). For Critical D values, appropriate K-S table values are used (Charles 2021).

CORRELATION AND REGRESSION EQUATION

To assess the closeness between the Poisson and the Normal approx-1, the correlation between their probabilities is calculated. A correlation of above 0.95, a slope close to 1 and the intercept close to zero, is taken to indicate the goodness of fit. Same exercise was also attempted between the Poisson and Normal approx-2 probabilities.

PROBABILITY DISTRIBUTION CURVES

Apart from the Correlation and K-S test, to have the visual impressions about the comparability of the Poisson distribution, Normal Approx-1 and Normal Approx-2 distribution, their distribution curves are plotted and displayed in figures.

RESULTS

The Comparison of Normal approximated Probabilities with that of the Poisson Probabilities for the feasible values of X and (λ = 2), is shown in Table 1.

Table 1: The Comparison of the Normal approximated Probabilities with that of the Poisson Probabilities for the Feasible values of X - (λ = 2)

X	Poisson	Normal approx - 1	Normal approx - 2	Absolute Deviation-1	Absolute Deviation-2
0	0.135	0.106	0.138	0.029	0.003
1	0.271	0.217	0.268	0.054	0.003
2	0.271	0.276	0.292	0.005	0.021
3	0.18	0.217	0.179	0.037	0.001
4	0.09	0.106	0.062	0.016	0.028
≥ 5	0.053	0.078	0.061	0.025	0.008
Total	1	1	1	0.166	0.064
Average Absolute Deviation				0.028	0.011
Standard Deviation (SD)				0.017	0.0112

Based on the absolute deviations, it can be seen that the Normal approx-2 has the least sum of absolute deviations as compared to that seen in the case of Normal approx-1, suggesting that the

Normal approx-2 is a better fit than the Normal approx-1. It can be further confirmed from the Fig.1 that the curve due to Normal approx-2 is much closer to the Poisson distribution curve than the curve of Normal approx-1.

The Comparison of Normal approximated Probabilities with that of Poisson Probabilities for the feasible values of X and (λ = 3), is shown in Table 2.

Fig. 1: The Probability distribution of the Poisson, Normal approx-1 and the Normal approx-2 for the Feasible values of X - (l = 2)

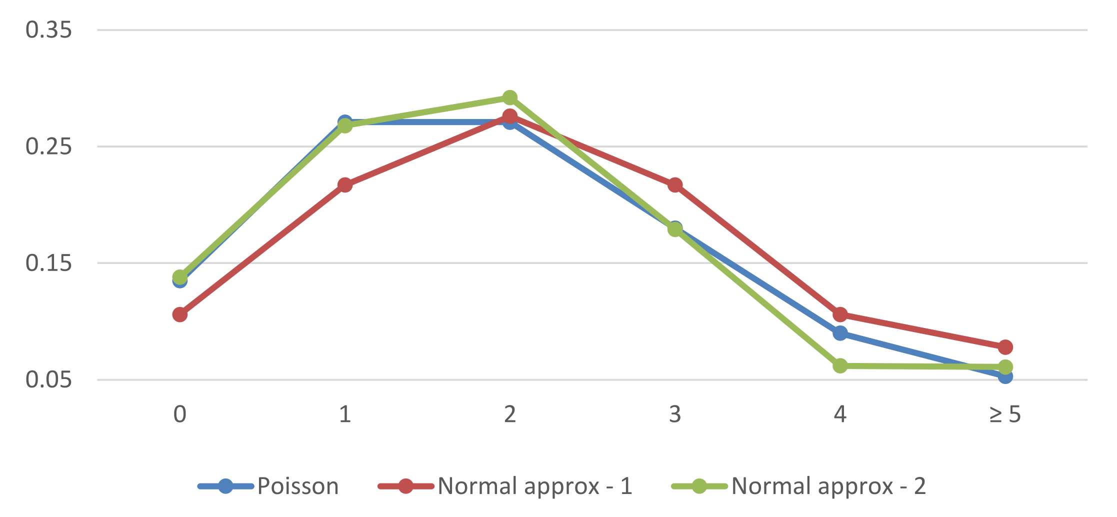


Table 2: The Comparison of the Normal approximated Probabilities with that of the Poisson Probabilities for the Feasible values of X - (λ = 3)

X	Poisson	Normal approx - 1	Normal approx - 2	Absolute Deviation-1	Absolute Deviation-2
0	0.050	0.053	0.067	0.003	0.017
1	0.149	0.119	0.147	0.030	0.002
2	0.224	0.193	0.223	0.031	0.001
3	0.224	0.227	0.236	0.003	0.012
4	0.168	0.193	0.173	0.025	0.005
5	0.101	0.119	0.088	0.018	0.013
6	0.050	0.053	0.031	0.003	0.019
7	0.022	0.017	0.008	0.005	0.014
≥ 8	0.012	0.026	0.027	0.014	0.015
Total	1.000	1.000	1.000	0.132	0.098
Average Abs Deviation				0.015	0.011
Standard Deviation (SD)				0.0118	0.0066

It can be seen from the Table 2, that the Normal approx-2 has the least sum of absolute deviations as compared to Normal approx-1, suggesting that the Normal approx-2 is a better fit for the Poisson distribution than the Normal approx-1. Accordingly, in the Fig. 2, the Normal approx-2 curve, as compared to that of Normal approx -1 curve, appears to be closer to the curve of the Poisson distribution.

Fig. 2: The Probability distribution of Poisson, Normal approx-1 and Normal approx-2 for the Feasible values of X - (λ = 3)

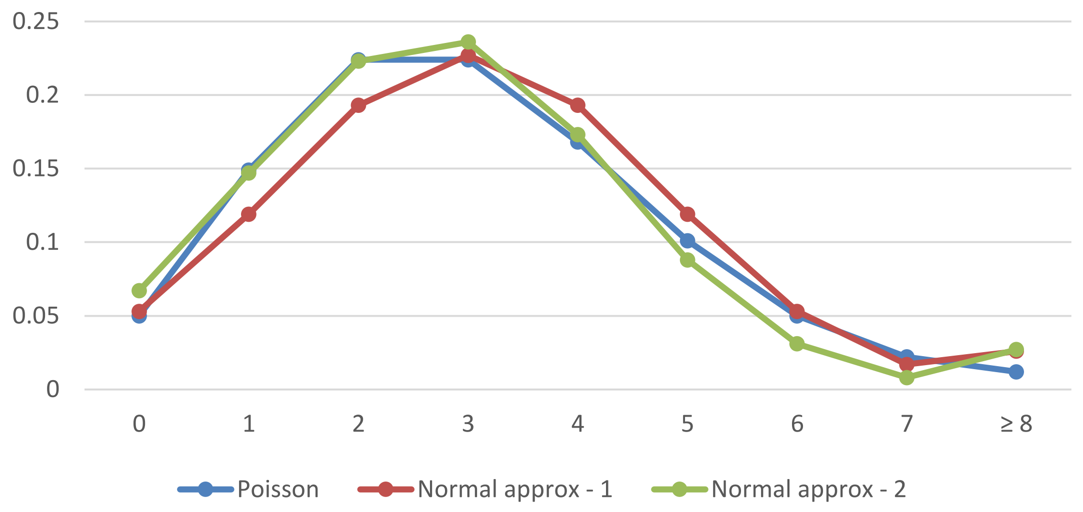
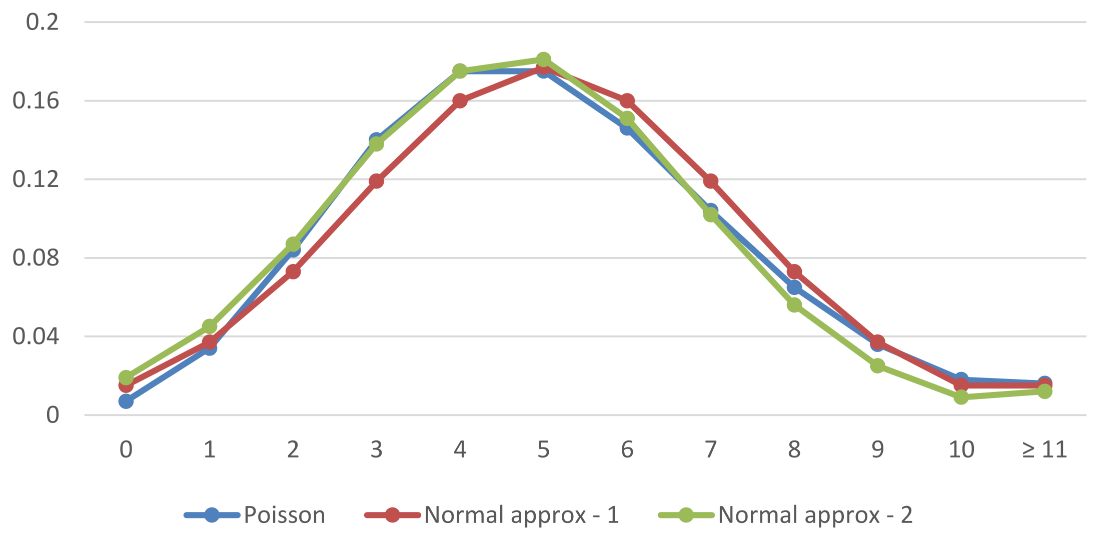


Fig. 3: The Probability distribution of Poisson, Normal approx-1 and Normal approx-2 for the Feasible values of X - (λ = 5)



It is apparent from the Table 3 that the Normal approx-2 has the least sum of absolute deviations as compared to the Normal approx -1, suggesting that the Normal approx-2 is relatively a better fit as compared to Normal approx -1. From the Fig. 3, it is further confirmed that the Normal approx-2 curve is closer to the Poisson distribution curve than that of the Normal approx-1.

Table 3: The Comparison of the Normal approximated Probabilities with that of the Poisson Probabilities for the Feasible values of X - (λ = 5)

X	Poisson	Normal approx - 1	Normal approx - 2	Absolute Deviation-1	Absolute Deviation-2
0	0.007	0.015	0.019	0.008	0.012
1	0.034	0.037	0.045	0.003	0.011
2	0.084	0.073	0.087	0.011	0.003
3	0.14	0.119	0.138	0.021	0.002
4	0.175	0.16	0.175	0.015	0
5	0.175	0.177	0.181	0.002	0.006
6	0.146	0.16	0.151	0.014	0.005
7	0.104	0.119	0.102	0.015	0.002
8	0.065	0.073	0.056	0.008	0.009
9	0.036	0.037	0.025	0.132	0.099
10	0.018	0.015	0.009	0.132	0.099
≥ 11	0.016	0.015	0.012	0.132	0.099
Average Abs Deviation				0.011	0.006
Standard Deviation (SD)				0.0062	0.0043

The Comparison of the Normal approximated Probabilities with that of the Poisson Probabilities for the feasible values of X and (λ = 10), is shown in Table 4 and the Fig. 4 is also enclosed.

It can be seen from the Table 4 that the sum of absolute deviations is the least for Normal approx.-2 as compared to that seen for the Normal approx -1 suggesting that the former is relatively a better fit as compared to Normal approx -1. Similar conclusion can be drawn from the Fig.4.

Table 4: The Comparison of the Normal approximated Probabilities with that of Poisson Probabilities, for the Feasible values of X - (λ = 10)

X	Poisson	Normal approx - 1	Normal approx - 2	Absolute Deviation-1	Absolute Deviation-2
≤ 3	0.01	0.015	0.019	0.005	0.009
4	0.019	0.021	0.025	0.002	0.006
5	0.038	0.036	0.042	0.002	0.004
6	0.063	0.057	0.065	0.006	0.002
7	0.09	0.080	0.089	0.010	0.001
8	0.113	0.103	0.111	0.010	0.002
9	0.125	0.12	0.125	0.005	0
10	0.125	0.126	0.127	0.001	0.002
11	0.114	0.120	0.116	0.006	0.002
12	0.095	0.103	0.096	0.008	0.001
13	0.073	0.08	0.072	0.007	0.001
14	0.052	0.057	0.048	0.005	0.004
15	0.035	0.036	0.029	0.001	0.006
16	0.022	0.021	0.016	0.001	0.006
17	0.013	0.011	0.008	0.002	0.005

≥ 18	0.013	0.014	0.012	0.001	0.001
Average Abs Deviation				0.005	0.003
Standard Deviation (SD)				0.0032	0.0025

The results of K-S test and the correlation coefficients for different values of λ are shown in Table 5.

The K-S test, for the selected values of λ, revealed no significant differences between the Normal fits and the Poisson distribution, suggesting that the Normal approximation can safely be used when λ value is 2 or more. The correlation between the Poisson and the Normal probabilities are found to be more than 0.95 in all the cases. The intercept values closer to zero and slope being closer to 1 in the case of Normal-Approx-2, as compared to Normal-Approx-1, indicate that in general, the Normal approx-2 is a better fit for the Poisson distribution as compared to the fit of Normal approx-1.

Fig. 4: The Probability distribution of the Poisson, the Normal approx-1 and the Normal approx-2 for the Feasible values of X - (λ = 10)

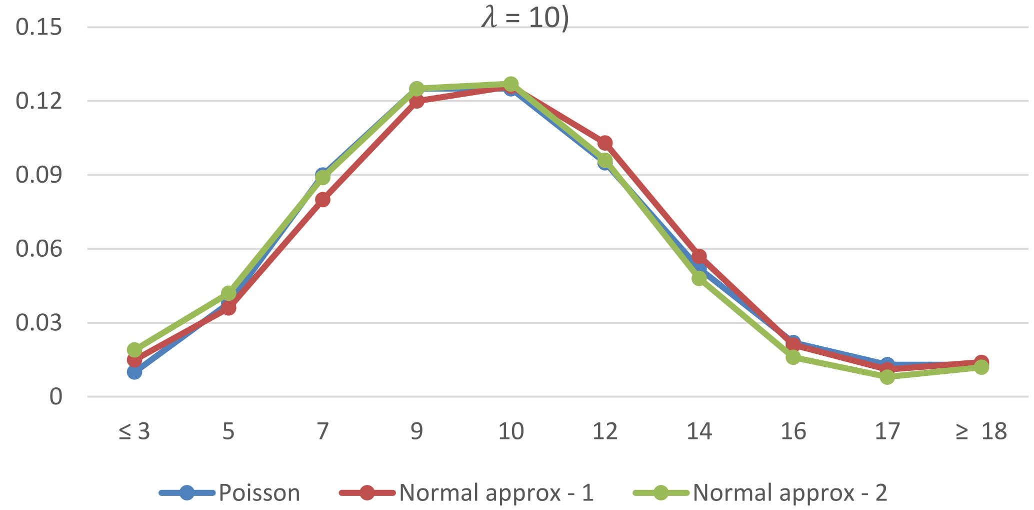


Table 5: The Goodness of fit test for the Normal Approx-1, Normal Aprox-2 to the Poisson distribution, as Judged by K-S test and by the Correlation

λ	Fitted Distribution	D Statistic	Significance	r with Poisson	Significance	Slope	Intercept
2	Norm-Approx-1	0.083	NS	0.978	< 0.001	0.814	0.031
	Norm-Approx-2	0.021	NS	0.989	< 0.001	1.072	-0.012
3	Norm-Approx-1	0.058	NS	0.972	< 0.001	0.921	0.009
	Norm-Approx-2	0.031	NS	0.988	< 0.001	1.026	-0.003
5	Norm-Approx-1	0.036	NS	0.985	< 0.001	0.957	0.004
	Norm-Approx-2	0.035	NS	0.993	< 0.001	1.018	-0.002
10	Norm-Approx-1	0.026	NS	0.991	< 0.001	0.975	0.002
	Norm-Approx-2	0.023	NS	0.995	< 0.001	1.001	0

DISCUSSION

In literature, generally, the normal approximation for the Poisson distribution is advocated when $\lambda > 5$ but the results of the present study have clearly indicated that the Normal approximation to the Poisson holds good even when λ is as low as 2. Further, it is apparent from the results that a correction of -0.35 to λ values improve the Normal approximation to the Poisson distribution as evident by the intercept values closer to zero and slopes being closer to 1 in the case of Normal Approx-2.

The results of the K-S test supported the view that the Normal approximation to the Poisson distribution is valid even when λ is below 5 and equal to 2. This is in much deviation from the opinion expressed in the literature by the various authors that $\lambda > 5$ to go for the Normal Approximation of the Poisson distribution. With defined λ values, random samples of 50 each are also generated for Poisson, Normal approx-1 and Normal approx-2 and then comparisons are made. Results of the samples, again showed that both the Normal approximations are good enough for the Poisson distribution.

A quick inspection of the Fig. 1 to Fig. 4 will reveal that the fits due to Normal approx-1 and Normal approx-2 are slightly skewed. It is possible that for lower values of λ, the Normal approximation may not yield an absolute symmetric curve. However, it is to be noted that our primary interest was to see whether is it possible for us to estimate the Poisson probabilities using the Normal approximation. The results of the study have amply shown that unlike indicated in most of the articles on the internet, the normal approximation for the Poisson distribution is possible even when $\lambda \leq 5$.

CONCLUSION

The results of the present study suggests that the Normal approximation to the Poisson distribution is valid for $\lambda \leq 5$ and even when equal to 2 which is in much deviation from the conditions mentioned in the literature. It is further suggested that a correction factor of - 0.35 in λ values is useful in improving the normal approximation of the Poisson distribution especially when $\lambda \leq 5$.

REFERENCES

[1]. AP Statistics Curriculum 2007 Limits Norm2Poisson (2014): Normal Approximation to Poisson Distribution,

[2]. http://wiki.stat.ucla.edu/socr/index.php/AP_Statistics_Curriculum_2007_Limits_Norm2Poisson Accessed on 26/06/2021.

[3]. CalcWorkshop (2001): Normal Approximation – w/5 Step-by-Step Examples - Tenth revised Edition – Reprint 2001, Sultan Chand & Sons, 7.57

[4]. https://calcworkshop.com/continuous-probability-distribution/normal-approximation/

[5]. Accessed on 26/09/2021

[6]. Charles Z (2021): Two Sample Kolmogorov and Smirnov test, Real Statistics using Excel https://www.real-statistics.com/non-parametric-tests/goodness-of-fit-tests/two-sample-kolmogorov-smirnov-test/ - Accessed on 28-06-2021.

[7]. Charles Z (2021): Two Sample Kolmogorov and Smirnov Table, Real Statistics using Excel - Accessed on 28-06-2021.

[8]. https://www.real-statistics.com/statistics-tables/two-sample-kolmogorov-smirnov-table/

[9]. Chaudhary CR (2017): Normal Approximation to Poisson(λ) Distribution

[10]. https://vrcacademy.com/tutorials/normal-approximation-poisson-distribution/

[11]. Accessed on 26/06/2021

[12]. Dinov ID (2019): Normal Approximation to Poisson(λ) Distribution

[13]. http://www.socr.ucla.edu/Applets.dir/NormalApprox2PoissonApplet.html

[14]. Accessed on 26/06/2021

[15]. Dransfield RD, Brightwell R. (2012): How to: Kolmogorov-Smirnov test 1-sample KS test - https://influentialpoints.com/Training/kolmogorov-smirnov_test.htm

[16]. Accessed on 28/06/2021

[17]. Gupta SC (2012): Fundamental of Statistics, 7th Edition, Himalaya Publishing House, 14.38.

[18]. Gupta SC and Kapoor VK (2001): Fundamentals of Mathematical Statistics, Tenth revised Edition – Reprint 2001, Sultan Chand & Sons, 7.57

[19]. Microsoft Corporation, 2019. Microsoft Excel, Available at: https://office.microsoft.com/excel.

Biography of Dr. Ramnath Takiar

I am a Post graduate in Statistics from Osmania University, Hyderabad. I did my Ph.D. from Jai Narain Vyas University of Jodhpur, Jodhpur, while in service, as an external candidate. I worked as a research scientist (Statistician) for Indian Council of Medical Research from 1978 to 2013 and retired from the service as a Scientist G (Director Grade Scientist). I am quite experienced in large scale data handling, data analysis and report writing. I have 56 research publications in national and International Journals related to various fields like Nutrition, Occupational Health, Fertility and Cancer epidemiology. During the tenure of my service, I attended three International conferences namely in Goiana (Brazil-2006), Sydney (Australia-2008) and Yokohoma (Japan-2010) and presented a paper in each. I also attended the Summer School related to Cancer Epidemiology (Modul I and Module II) conducted by International Agency for Research in Cancer (IARC), Lyon, France from 19th to 30th June 2007. After my retirement, I joined my son at Ulaanbaatar, Mongolia. I worked in Ulaanbaatar as a Professor and Consultant from 2013-2018 and was responsible for teaching and guiding Ph.D. students. I also taugth Mathematics to undergraduates and Econometrics to MBA students. During my service there, I also acted as the Executive Editor for the in-house Journal “International Journal of Management”. I am still active in research and have published 15 research papers in past 7 years.