DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 18 | Issue 1 Article 3

2-13-2020

Assessing the Accuracy of Approximate Confidence Intervals Proposed for the Mean of Poisson Distribution

Alireza Shirvani
Velayat University of Iranshahr, Iran, stat.shirvani@gmail.com

Malek Fathizadeh
Vali-e-Asr University of Rafsanjan, Iran, m.fathi@vru.ac.ir

Recommended Citation

Shirvani, A., & Fathizadeh, M. (2019). Assessing the accuracy of approximate confidence intervals proposed for the mean of Poisson distribution. Journal of Modern Applied Statistical Methods, 18(1), eP2322. doi: 10.22237/jmasm/1556668800

Journal of Modern Applied Statistical Methods
May 2019, Vol. 18, No. 1, eP2322
doi: 10.22237/jmasm/1556668800

ISSN 1538 − 9472

Assessing the Accuracy of Approximate Confidence Intervals Proposed for the Mean of Poisson Distribution

Alireza Shirvani
Velayat University of Iranshahr
Iranshahr, Iran

Malek Fathizadeh
Vali-e-Asr University of Rafsanjan
Rafsanjan, Iran

The Poisson distribution is applied as an appropriate standard model to analyze count data. Because this distribution is known as a discrete distribution, representation of accurate confidence intervals for its distribution mean is extremely difficult. Approximate confidence intervals were presented for the Poisson distribution mean. The purpose of this study is to simultaneously compare several confidence intervals presented, according to the average coverage probability and accurate confidence coefficient and the average confidence interval length criteria.

Keywords: Poisson distribution, confidence interval, average coverage probability, confidence coefficient, confidence interval length

Introduction

Generally, there are several cases that are confronted with counting an event occurring in fixed interval of time and/or space. For example, the number of telephone calls linked to a call center per hour, the number of customers referred to a shopping center per day, the number of traffic accidents per hour in a city, the number of fish in a particular part of the Pacific Ocean and the number of births per month in a country, etc. In these cases, the Poisson distribution will be an appropriate and standard model to analyze count data.

It is remarkable to create a confidence interval for the Poisson distribution mean. Suppose that X has a Poisson distribution with λ mean. As this distribution is known as a discrete distribution, representation of accurate confidence intervals for its distribution mean is extremely difficult. So far, several approximate confidence intervals have been presented for the Poisson distribution mean. For

doi: 10.22237/jmasm/1556668800 | Accepted: September 30, 2016; Published: February 13, 2020.
Correspondence: Alireza Shirvani, stat.shirvani@gmail.com

example see Crow and Gardner (1959), Kabaila and Byrne (2001), Kabaila and Lloyd (1997), Schwertman and Martinez (1994), Guan (2011), Khamkong (2012), Ross (2003) and Sahai and Khurshid (1993). In this paper, the confidence intervals proposed by Schwertman and Martinez (1994), Guan (2011) and Khamkong (2012) is considered for λ parameter and by simulating the methodology proposed by Wang (2009), the average coverage probability and confidence coefficient is computed accurately. Moreover, the confidence intervals lengths are determined and finally a comparison is drawn between these intervals.

Accurate Computation Method of Average Coverage Probability and Confidence Coefficient

Assume X is a one-dimensional discrete random variable with probability density function $f_\theta(x)$, where θ is unknown parameter and $x \in S = \{0, 1, \ldots, n\}$. Moreover, $\Omega = (l, u)$ is considered as parameter space of θ. If a confidence interval for parameter θ is shown as $(L(X), U(X))$, its coverage probability is equal to $P_\theta\,(\theta \in (L(X), U(X)))$ which means the value of probability that this random interval contain the actual value of θ. The confidence coefficient of this interval is equal to the infimum of coverage probabilities deployed in the parameter space, which is obtained as following representation:

$$\text{confidence coefficient} = \inf_{\theta \in \Omega} P_\theta\left(\theta \in \left(L\left(X\right), U\left(X\right)\right)\right)$$

If $\eta(\theta)$ is a prior density function on Ω, then the average coverage probability of $(L(X),U(X))$ interval under the prior density function $\eta(\theta)$ is defined as:

$$\int_\Omega P_\theta\left(\theta \in \left(L\left(X\right), U\left(X\right)\right)\right) \eta\left(\theta\right) d\theta$$

In continuous distributions the coverage probability function may be the same for all points of the parameter space, but in discrete distributions, the coverage probability function is varied by changing the truth values of unknown parameters in the parameter space. Calculating the accurate confidence coefficient and average coverage probability in these distributions is too difficult. The exact method of calculating these two criteria for confidence intervals that have certain conditions in discrete distributions that satisfy the condition of Assumption 1 is proposed by Wang (2009), who considered some distributions for which Assumption 1 is valid.

Assumption 1. If the probability density function of the desired distribution is declared as $f_\theta(x)$, then $\sum_{x=h_1}^{h_2} f_\theta(x)$ will be a unimodal, descending or ascending function in terms of θ (for each h_1, h_2 that $0 \leq h_1 \leq h_2 \leq n$).

Definition 2. According to Casella and Berger (2002), a family of pdfs or pmfs $\{g(t|\theta) : \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a monotone likelihood ratio (MLR) if, for every $\theta_2 > \theta_1$, $g(t|\theta_2) / g(t|\theta_1)$ is a monotone (nonincreasing or nondecreasing) function of t on $\{t$: $g(t|\theta_1) > 0$ or $g(t|\theta_2) > 0\}$. Note that $c/0$ is defined as ∞ if $0 < c$.

Note 3. The exponential families that have MLR property in x can satisfy the conditions of Assumption 1.

Note 3 shows the method presented here can be used for exponential families including Poisson distribution. To calculate the confidence coefficient and average coverage probability, the confidence interval must satisfy various conditions and requirements. The requirements will be stated on several assumptions and then the approaches applied for calculating confidence coefficient and the average coverage probability is detailed with no proofs.

Assumption 4. Assume for confidence interval $(L(X),U(X))$:

1. If $X_1 < X_2$, then $L(X_1) < L(X_2)$, $U(X_1) < U(X_2)$ which means $L(x)$ and $U(x)$ are two increasing functions respect to x.
2. $L(0) \leq l \leq U(0)$, $L(n) \leq u \leq U(n)$

Assumption 5. Assume for confidence interval $(L(X),U(X))$:

1. For $X_1 > 0$, $X_2 < n$, If $X_1 < X_2$, then $L(X_1) < L(X_2)$, $U(X_1) < U(X_2)$
2. $L(0) = U(0) = l$, $L(n) = U(n) = u$

Assumption 6. Assume for confidence interval $(L(X),U(X))$:

For each point θ that are included in parameter space, there exist one x_0 in sampling space so that $\theta \in (L(x_0), U(x_0))$ and $P_\theta (X = x_0) > 0$ are satisfied.

Before representation of the main results, demarcations are defined as follows:

For a confidence interval $(L(X),U(X))$, there exist $2(n+1)$ end points corresponds to $X = 0, 1, ..., n$ that are shown as follows:

$$L(0), L(1), \ldots, L(n), U(0), U(1), \ldots, U(n)$$

Assume a set of selected end points between l and u which is g points. These points are increasingly sequenced and they are named $v_1, \ldots, v_g$. Applying this defined set the parameter space can be divided into $(g+1)$ sub intervals. Ω^o is considered as the subset of internal points of Ω and then we can define the following set:

$$W = \left\{ w \mid w = l, w = u, w = L\left(X\right), w = U\left(X\right), X = 0,1,\ldots,n, w \in \Omega^o \right\} \tag{1}$$

which means W is a type of set that contains the lower and upper limits of parameter space and end points (belonged to Ω^o) of confidence intervals.

Theorem 7. Assume $f_\theta(x)$ can satisfy the condition of Assumption 1. The confidence coefficient for confidence interval $(L(X),U(X))$ for θ that is valid on the condition of Assumption 6 and satisfy one of the Assumptions 4 or 5, equals minimum coverage probability for points belong to W.

Theorem 8. Consider a discrete random variable X with $f_\theta(x)$ as a mass probability function that can satisfy the condition of Assumption 1, Moreover, considering a confidence interval $(L(X),U(X))$ for θ that is valid on the condition of Assumption 6 and satisfy one of Assumptions 4 or 5 and from $2(n+1)$ end points corresponds to this distance, g points belong to Ω^o set. These g points $v_1, \ldots, v_g$. can divide the parameter space to $(g+1)$ sub intervals. The first sub interval is (l,v_1), hence the lower limit of first sub interval is equaled to lower limit of parameter space (l). The lower limits of other sub intervals, is either a lower end point or upper end point of confidence intervals. For all θ belonging to the first sub interval, when the confidence interval satisfies the Assumptions 4 or 5, the coverage probability function equals $\sum_{i=0}^{L^{-1}(l)} f_\theta\left(i\right)$ or $\sum_{i=1}^{L^{-1}(l)} f_\theta\left(i\right)$, respectively. For all θ belonging to the other sub intervals, (these intervals are considered as $(v_i, v_{(i+1)})$, where $v_{g+1} = u$) bear one of the following conditions:

1. If the lower limit of sub interval (v_i) is a lower end point of confidence intervals, then the coverage probability function for θ belong to the sub interval equals $\sum_{i=w_1(x)}^{x-1} f_\theta(i)$.

2. If the lower limit of sub interval (v_i) is an upper end point of confidence intervals, then the coverage probability function for θ belong to the sub interval equals $\sum_{i=x+1}^{w_2(x)} f_\theta(i)$, where:

$$w_1(x) = \max\left(\left[U^{-1}\left(L(x)\right)\right] + 1, 0\right) \text{ and } w_2(x) = \min\left(\left[L^{-1}\left(U(x)\right)\right], n\right) \qquad (2)$$

Accurate Calculation of Confidence Coefficient

Step 1: It must be evaluated whether or not the confidence interval is satisfied in the Assumption 6. If this assumption is not valid, confidence coefficient will be equal to zero. If the Assumption 6 is valid, we must ensure that either of the assumptions 4 or 5 is satisfied, otherwise, the next step should not be evaluated.

Step 2: The end points of confidence intervals corresponds to X = 0, 1, ..., *n* which are included in the parameter space are considered.

Step 3: The coverage probabilities corresponds to the points obtained in the second step and the lower and upper limits of parameter space are determined. The minimum of these coverage probabilities are equal to accurate confidence coefficient.

Calculation of Average Coverage Probability Considering the Prior Density Function $\eta(\theta)$

Step 1: It must be evaluated the confidence interval satisfies the conditions of the Assumptions 4 or 5.

Step 2: If the conditions of the Assumptions 4 or 5 are met and there exist g end points belonging to Ω^o, these points are sequenced increasingly and they are named as $v_1, \ldots, v_g$.

Step 3: By using the Theorem 8 the coverage probability for $g + 1$ sub intervals of parameter space is determined. The coverage probability for each sub interval $(v_i, v_{(i+1)})$; $i = 1, ..., g$ is shown by e_i and also the coverage probability for interval (l, v_1) is shown by e_0.

Step 4: The accurate value of average coverage probability by considering the prior density function $\eta(\theta)$ is obtained as follows:

$$\int_l^{v_1} e_0(\theta)\eta(\theta)d\theta + \int_{v_1}^{v_2} e_1(\theta)\eta(\theta)d\theta + \cdots + \int_{v_1}^{v_{i+1}} e_i(\theta)\eta(\theta)d\theta + \cdots + \int_{v_g}^{u} e_g(\theta)\eta(\theta)d\theta$$

Note 9. If the parameter space is limited, the above approaches for calculating confidence coefficient and average coverage probability are still applicable. In order to calculate the confidence coefficient in the second step, the end points deployed in the limited space are considered, then the minimum of coverage probabilities corresponds to these points in the second step and the lower and upper limits in the limited space is determined and also in order to calculate the average coverage probability, the limits of restricted parameter space are considered as limits of parameter space.

Introducing Some of the Confidence Intervals Defined for Mean of Poisson Distribution

Assume Z_α as the upper cutoff point of standard normal distribution, so by this definition we have the following considerations.

Wald Confidence Interval

Use the center limit theorem to calculate the $1 - \alpha$ Wald interval for parameter λ as the following representation:

$$X \pm Z_{\frac{a}{2}}\sqrt{X} \tag{3}$$

This confidence interval is proposed in Schwertman and Martinez (1994) and also the condition of Assumptions 4 and 6 are met (to see, the lower and upper bounds of the confidence interval must be derived with respect to x). Therefore, calculate accurate confidence coefficient and average coverage probability. Because the normal approximation is used to Poisson distribution (the approximation of discrete random variable to continuous type) for calculating Wald interval, it is recommended to use continuous correction.

Improved Wald Interval with Continuous Correction

The $1 - \alpha$ Wald confidence interval for parameter λ by utilizing continuous correction equals:

$$X \pm Z_{\frac{\alpha}{2}} \sqrt{X + 0.5} \tag{4}$$

This confidence interval given in Khamkong (2012) also meets the conditions of Assumptions 4 and 6 (to see, the lower and upper bounds of the confidence interval must be derived with respect to x).

SC Confidence Interval (Score Interval)

The $1 - \alpha$ SC confidence interval for parameter λ equals:

$$X + \frac{Z_{\frac{\alpha}{2}}^2}{2} \pm Z_{\frac{\alpha}{2}} \sqrt{X + \frac{Z_{\frac{\alpha}{2}}^2}{4}} \tag{5}$$

This confidence interval given in Guan (2011) also meets the conditions of Assumptions 4 and 6.

MSC Confidence Interval (Moved Score Confidence Interval)

The $1 - \alpha$ MSC confidence interval for parameter λ equals:

$$X + 0.46 Z_{\frac{\alpha}{2}}^2 \pm Z_{\frac{\alpha}{2}} \sqrt{X + \frac{Z_{\frac{\alpha}{2}}^2}{4}} \tag{6}$$

This confidence interval given in Guan (2011) also satisfies the conditions of Assumptions 4 and 6.

FNCC Confidence Interval (Wald CC)

The $1 - \alpha$ FNCC confidence interval for parameter λ equals:

$$\left(X - 0.5 + Z_{1-\frac{\alpha}{2}}\sqrt{X - 0.5},\ X + 0.5 + Z_{\frac{\alpha}{2}}\sqrt{X + 0.5}\right) \tag{7}$$

This confidence interval given in Schwertman and Martinez (1994) also meets the conditions of Assumptions 4 and 6.

Comparison between Confidence Intervals for the Mean of Poisson Distribution

Now, confidence intervals introduced for the mean of Poisson distribution are compared considering confidence coefficients, average coverage probability and average length criteria. It is known that the parameter space for λ equals $(0,+\infty)$ interval. But, it must be noted that in particular applications according to information given about sampling data, the parameter space may have upper and lower limits. As mentioned in Note 9, the confidence coefficients and the average coverage probability for limited parameter space can be calculated. Consider several limited parameter spaces and compare optimality of intervals in each. Shown in Table 1 is the average coverage probability for confidence intervals by considering (0,5) as parameter space and gamma prior density function with parameters α and β for different values of α and β.

Table 1. The average coverage probability of 0.95 confidence intervals by considering (0,5) as parameter space and gamma prior density function for different values of α,β parameters.

(*α*,*β*)	**Wald**	**Improved Wald**	**SC**	**MSC**	**FNCC**
(3,2)	0.8641425	0.9143622	0.9571222	0.9645696	0.9449194
(2,2)	0.8072475	0.9133294	0.9566735	0.9668320	0.9504911
(2,3)	0.8297575	0.9140922	0.9567840	0.9658480	0.9482634
(1,1)	0.5627519	0.9105985	0.9494315	0.9745308	0.9450845
(1,2)	0.6598992	0.9119293	0.9525660	0.9715013	0.9489584
(2,0.25)	0.3971195	0.8985933	0.9446330	0.9753967	0.9285543
(3,0.25)	0.4913916	0.9315413	0.9499122	0.9722384	0.9643795

According to the Table 1, when the parameter space is (0,5), the average coverage probability of Wald intervals, improved Wald, FNCC, SC, MSC have minimum and maximum values, respectively. Hence, by considering the average coverage probability criterion, the optimal intervals are sequenced as MSC, SC, FNCC, improved Wald and Wald. The confidence coefficients of Wald intervals, improved Wald, FNCC, SC, MSC are valued as 0, 0.7493219, 0.9728728, 0.838182 and 0.924353, respectively. Hence, by considering the confidence coefficients criterion, the optimal intervals are sequenced as FNCC, MSC, SC, improved Wald and Wald.

Shown in Table 2 is the average coverage probability for confidence intervals by considering (0,10) as parameter space and gamma prior density function with parameters α and β for different values of α and β.

Table 2. The average coverage probability of 0.95 confidence intervals by considering (0,10) as parameter space and gamma prior density function for different values of α,β parameters.

(α,β)	**Wald**	**Improved Wald**	**SC**	**MSC**	**FNCC**
(3,2)	0.8888037	0.9204003	0.9552390	0.9604853	0.9473020
(2,2)	0.8347166	0.9165118	0.9558278	0.9641060	0.9503650
(2,3)	0.8652953	0.9194707	0.9552924	0.9617933	0.9489643
(1,1)	0.5650555	0.9106648	0.9494641	0.9744146	0.9451272
(1,2)	0.6790454	0.9128737	0.9526446	0.9703599	0.9490552
(2,0.25)	0.3971196	0.8985933	0.9446330	0.9753967	0.9285543
(3,0.25)	0.4913918	0.9315413	0.9499122	0.9722384	0.9643795

According to the Table 2, when the parameter space is (0,10), the average coverage probability of Wald intervals, improved Wald, FNCC, SC, MSC have minimum and maximum values, respectively. Hence, by considering the average coverage probability criterion, the optimal intervals are sequenced as MSC, SC, FNCC, improved Wald and Wald. The confidence coefficients of Wald intervals, improved Wald, FNCC, SC, MSC are valued as 0, 0.7493219, 0.8475551, 0.838182 and 0.9343535, respectively. Hence, by considering the confidence coefficients criterion, the optimal intervals are sequenced as MSC, FNCC, SC, improved Wald and Wald.

Shown in Table 3 is the average coverage probability for confidence intervals by considering (0,15) as parameter space and gamma prior density function with parameters α and β for different values of α and β.

According to the Table 3, when the parameter space is (0,15), the average coverage probability of Wald intervals, improved Wald, FNCC, SC, MSC have minimum and maximum values, respectively. Hence, by considering the average coverage probability criterion, the optimal intervals are sequenced as MSC, SC, FNCC, improved Wald and Wald. The confidence coefficients of Wald intervals, improved Wald, FNCC, SC, MSC are valued as 0, 0.7493219, 0.8475551, 0.838182 and 0.9343535, respectively. Hence, by considering the confidence coefficients criterion, the optimal intervals are sequenced as MSC, FNCC, SC, improved Wald and Wald.

Table 3. The average coverage probability of 0.95 confidence intervals by considering (0,15) as parameter space and gamma prior density function for different values of α,β parameters.

(α,β)	**Wald**	**Improved Wald**	**SC**	**MSC**	**FNCC**
(3,2)	0.8933807	0.9222194	0.9548630	0.9596592	0.9476260
(2,2)	0.8381964	0.9172590	0.9556815	0.9636963	0.9503577
(2,3)	0.8732198	0.9216328	0.9548621	0.9607176	0.9491470
(1,1)	0.5650720	0.9106660	0.9494642	0.9744146	0.9451274
(1,2)	0.6806082	0.9130242	0.9526394	0.9702502	0.9490608
(2,0.25)	0.3971196	0.8985933	0.9446330	0.9753967	0.9285543
(3,0.25)	0.4913918	0.9315413	0.9499122	0.9722374	0.9643795

Shown in Table 4 is the average coverage probability for confidence intervals by considering (0,20) as parameter space and gamma prior density function with parameters α and β for different values of α and β.

Table 4. The average coverage probability of 0.95 confidence intervals by considering (0,20) as parameter space and gamma prior density function for different values of α,β parameters.

(α,β)	**Wald**	**Improved Wald**	**SC**	**MSC**	**FNCC**
(3,2)	0.8941198	0.9225174	0.9548003	0.9595360	0.9477236
(2,2)	0.8386047	0.9173505	0.9556630	0.9636500	0.9503699
(2,3)	0.8751611	0.9221872	0.9547505	0.9604636	0.9492689
(1,1)	0.5650721	0.9106660	0.9494642	0.9744136	0.9451274
(1,2)	0.6807374	0.9130373	0.9526387	0.9702413	0.9490630
(2,0.25)	0.3971196	0.8985933	0.9446330	0.9753967	0.9285543
(3,0.25)	0.4913918	0.9315413	0.9499122	0.9722384	0.9643795

According to Table 4, when the parameter space is (0,20), the average coverage probability of Wald intervals, improved Wald, FNCC, SC, MSC have minimum and maximum values, respectively. Hence, by considering the average coverage probability criterion, the optimal intervals are sequenced as MSC, SC, FNCC, improved Wald and Wald. The confidence coefficients of Wald intervals, improved Wald, FNCC, SC, MSC are valued as 0, 0.7493219, 0.8475551, 0.838182 and 0.9343535, respectively. Hence, by considering the confidence coefficients criterion, the optimal intervals are sequenced as MSC, FNCC, SC, improved Wald and Wald.

Then average length of confidence intervals is calculated for $x = 0, 1, 2, \ldots, 10$. The average length of confidence intervals of Wald interval, improved Wald, SC, MSC and FNCC are valued as 8.006877, 8.686906, 9.151103, 9.151103, and 8.892629, respectively. Hence, by considering the average length criterion, the optimal intervals are sequenced as Wald, improved Wald, FNCC, SC, and MSC.

Conclusion

In general, with respect to the previous section:

- by considering the average coverage probability criterion, the optimal intervals are sequenced as MSC, SC, FNCC, improved Wald and Wald.
- by considering the confidence coefficients criterion, the optimal intervals are sequenced as MSC, FNCC, SC, improved Wald and Wald.
- by considering the average length of confidence intervals criterion, the optimal intervals are sequenced as Wald, improved Wald, FNCC, SC, and MSC.

Because there are no significant differences between the lengths of intervals for large values of *X*, the comparison can be evaluated regarding to average coverage probability or confidence coefficient. So, between these confidence intervals the MSC confidence interval is optimal.

References

Casella, G. & Berger, R. L. (2002). *Statistical Inference* (2nd Ed.). Pacific Grove, CA: Duxbury.

Crow, E. L. & Gardner, R. S. (1959). Confidence intervals for the expectation of a Poisson variable. *Biometrika, 46*(3/4), 441-453. doi: 10.1093/biomet/46.3-4.441

Guan, Y. (2011). Moved score confidence intervals for means of discrete distributions. *Open Journal of Statistics*, *1*(2), 81-86. doi: 10.4236/ojs.2011.12009

Kabaila, P. & Byrne, J. (2001). Exact short Poisson confidence intervals. *Canadian Journal of Statistics*, *29*(1), 99-106. doi: 10.2307/3316053

Kabaila, P. & Lloyd, C. J. (1997). Tight upper confidence limits from discrete data. *Australian Journal of Statistics*, *39*(2), 193-204. doi: 10.1111/j.1467-842x.1997.tb00535.x

Khamkong, M. (2012). Approximate confidence interval for the mean of Poisson distribution. *Open Journal of Statistics*, *2*(2), 204-207. doi: 10.4236/ojs.2012.22024

Ross, T. D. (2003). Accurate confidence intervals for binomial proportion and Poisson rate estimation. *Computers in Biology and Medicine*, *33*(6), 509-531. doi: 10.1016/s0010-4825(03)00019-2

Sahai, H. & Khurshid, A. (1993). Confidence intervals for the mean of a Poisson distribution: a review. *Biometrical Journal*, *35*(7), 857-867. doi: 10.1002/bimj.4710350716

Schwertman, N. C. & Martinez, R. A. (1994). Approximate Poisson confidence limits. *Communications in Statistics–Theory and Methods*, *23*(5), 1507-1529. doi: 10.1080/03610929408831336

Wang, H. (2009). Exact average coverage probabilities and confidence coefficients of confidence intervals for discrete distributions. *Statistics and Computing*, *19*(2), 139-148. doi: 10.1007/s11222-008-9077-8