A Comparison of Confidence Intervals of Negative Binomial Parameter p by Maximum Likelihood, Bayesian and Markov Chain Monte Carlo Methods

Unyamanee Kummaraka[1*] and Autcha Araveeporn[2]

[1,2]Department of statistics, Faculty of Science,

King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

*Corrresponding Author
Email Address : akasaka.jun23@gmail.com

Abstract

The objective of this research is to estimate the confidence interval of population parameter (p) or probability of success in each experiment based on negative binomial distribution. The interval estimation is evaluated by Maximum Likelihood (ML), Bayesian, and Markov Chain Monte Carlo (MCMC) methods. The performance of these methods is considered by Confidence Coefficients (CC) and Average Width (AW). When the CC values are greater than the fixed confidence interval, the AW of the confidence interval will focus the performance of these methods. The data is generated from negative binomial distribution as follows : the population parameter as small (0.2), medium (0.5) ,and large (0.8), the sample sizes and parameter r or called number of success as small sample sizes (n =30) r =10, 15, 20, medium sample sizes (n =50) r =10, 20, 30, and large sample sizes (n =70) r =15, 30, 45 and the 90%, 95%, and 99% confidence interval. The results show that the ML method exhibits poor interval estimation in most all cases, but the Bayesian method performs very satisfactorily in most all cases especially when true parameters are medium or large values for all sample sizes and parameter r . For the small true parameters, the MCMC method is a good performance in most cases. However, the confidence coefficient and average width of MCMC and Bayesian are equal in some case, so MCMC are reasonable working as good as Bayesian.

Keywords: Negative Binomial, Maximum Likelihood, Bayesian, Markov Chain Monte Carlo

1. Introduction

Statistics inference is one of the most importance and crucial aspects of the decision making process in economics, business, and science. The part of statistics inference is the estimation and hypothesis testing. Estimation is the process of inferring or estimating a population parameter from the corresponding statistic of a sample drawn from the population.

An estimation of a population parameter may be expressed in two ways: point estimation and interval estimation. A point estimation is focused on a single value of a statistic. This estimation will be inaccurate if the random sample is not a good representation of the population. The interval estimation is a formula that tells us how to use sample data to calculate an interval that estimates a population parameter. Confidence intervals have some measure of reliability of the estimation. Confidence Coefficient is the probability that an interval estimator encloses the population parameter.

In this case, we interested the discrete random variable in term of negative binomial distribution. Negative binomial have been found to provide useful representations that may be mentioned accident statistics, in birth-and-death processes, in psychological data, in demand for "frequently bought product", observed distributions of consumer expenditure, and as weights for times series in economics. [1]

At the first process of interval estimation is to estimate the point estimation from ML, Bayesian, and Markov Chain Monte Carlo (MCMC). Next the standard deviations are measured dispersion for these methods. The ML method is the most popular techniques because it is simple to understand and to calculate the estimators. The Bayesian method uses both a probability distribution and prior probability distribution to approach a posterior probability distribution [2]. However, it is difficult to demonstrate a posterior distribution from a probability distribution and prior probability distribution. However, the MCMC method [3] can approximate the estimator from Gibbs sampling algorithm [4] based on the posterior distribution.

For this reason, we propose the ML, Bayesian, and MCMC methods to estimate confidence interval with negative binomial distribution using simulation studies.

2. Scope of Research

The scope of this research is considered as follows :

2.1 Let random variables X be independent and identically distributed (iid) random variables of a negative binomial with parameter r and p, and the probability distribution function is

$$f\left(x_i \middle| r,p\right) = \begin{cases} \binom{x_i+r-1}{x_i} p^r\left(1-p\right)^{x_i} & ,x_i=0,1,2,\dots \\ 0 & ,\text{otherwise} \end{cases}$$

2.2 Let the prior distribution of p be a beta distribution with parameter a,b ($Beta(a,b)$), and the probability density function is

$$g\left(p \middle| a,b\right) = \begin{cases} \dfrac{\Gamma(a+b)\,p^{a-1}(1-p)^{b-1}}{\Gamma(a)\Gamma(b)} & ,0\le p\le 1 \\ 0 & ,\text{otherwise} \end{cases},$$

where mean is $E(p)=\dfrac{a}{a+b}$, and variance is $Var(p)=\dfrac{ab}{(a+b+1)(a+b)^2}$.

The beta distribution shows the parameter format as the negative binomial or called conjugate distribution. The prior distribution is defined with parameter (a,b) as (6,2) shown in Figure 1.

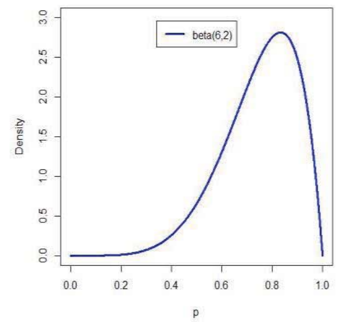

Figure 1 : The curve of beta distribution with parameter (a,b) as (6,2).

2.3 The sample sizes are considered at n = 30, 50, and 70.

2.4 The true parameters of negative binomial are defined as p = 0.2, 0.5, and 0.8, and r = 10, 15, and 20 (n = 30), r = 10, 20, and 30 (n = 50), and r = 15, 30, and 45 (n = 70).

2.5 The significant confidence level (α) is defined 3 levels at 0.1, 0.05, and 0.01 following the confidence interval $(1-\alpha)100\%$ as 90%, 95%, and 99%.

2.6 The R program is used to generate data at 1,000 replications for each cases.

3. Research Methodology

3.1 The random variables X is generated in a class of negative binomial distribution following the sample sizes, the true parameter, and the significance confidence level.

3.2 The methods for computing the confidence interval consist of following 3 methods:

3.2.1 Maximum Likelihood Method (ML)

The ML estimator of p can be computed by $\hat{p}_{ML} = \dfrac{rn-1}{rn+\sum_{i=1}^{n} x_i - 1}$, and

$$Var(\hat{p}_{ML}) = \frac{p^2(1-p)}{2}\left(\frac{2\sum_{i=1}^{n} x_i + 2 - p}{\sum_{i=1}^{n} x_i\left(\sum_{i=1}^{n} x_i - p + 2\right)}\right) \text{ [5]}$$

The confidence interval $(1-\alpha)100\%$ of p is approximated by

$$\left(\hat{p}_{ML} - z_{\alpha/2}\sqrt{Var(\hat{p}_{ML})}, \hat{p}_{ML} + z_{\alpha/2}\sqrt{Var(\hat{p}_{ML})}\right)$$

3.2.2 Bayesian Method

Let $X_1,\ldots,X_n$ is the random variable of negative binomial distribution with parameter r and p while beta distribution is considered the prior distribution with parameter $a,$ and b. From the Bayesian theorem [2] , let $h\left(p|x_i\right)$ is a posterior distribution function, $L\left(p\,|\,r,x_i\right)$ is the likelihood function, and $g\left(p|a,b\right)$ is a prior distribution function. The form of posterior distribution can be written as

$$h\left(p|x_i\right) \propto L\left(p\,|\,r,x_i\right) g\left(p|a,b\right),$$

$$\propto \left(\prod_{i=1}^{n}\binom{x_i+r-1}{x_i} p^{rn}(1-p)^{\sum_{i=1}^{n}x_i}\right) \times \left(\frac{\Gamma(a+b)\,p^{a-1}(1-p)^{b-1}}{\Gamma(a)\Gamma(b)}\right),$$

$$\propto \prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{rn+a-1}(1-p)^{\sum_{i=1}^{n}x_i+b-1}.$$

We reach to conclude that

$$p|r,x_i \sim Beta\left(rn+a,\sum_{i=1}^{n}x_i+b\right).$$

Therefore the beta distribution is conjugate to the negative binomial distribution. The posterior mean or called Bayesian estimator is given by

$$\hat{p}_{Bayesian} = E\left(p|x_i\right) = \frac{rn+a}{rn+a+\sum_{i=1}^{n}x_i+b},$$

while the posterior variance is computed by

$$Var(\hat{p}_{Bayesian}) = Var\left(p|x_i\right) = \frac{(rn+a)\left(\sum_{i=1}^{n}x_i+b\right)}{\left(rn+a+\sum_{i=1}^{n}x_i+b\right)^2\left(rn+a+\sum_{i=1}^{n}x_i+b+1\right)}.$$

The confidence interval $(1-\alpha)100\%$ of p is approximated by

$$\left(\hat{p}_{Bayesian} - z_{\alpha/2}\sqrt{Var(\hat{p}_{Bayesian})}\,,\,\hat{p}_{Bayesian} - z_{\alpha/2}\sqrt{Var(\hat{p}_{Bayesian})}\right).$$

3.2.3 Markov Chain Monte Carlo Method (MCMC)

The Markov Chain Monte Carlo (MCMC) [6] method is enable simulation from a distribution by embedding it as limiting distribution of a Markov chain and simulating from the chain until it approached equilibrium. The Gibbs sampling [4] is a algorithm for MCMC computing that can generate random values from posterior distribution function for which a widely variety of computational tools . We carry out the RJAGS package of R program to obtain the estimator from the posterior distribution function.

The MCMC estimators can be computed by

$$\hat{p}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T} p^{(t)},$$

$$Var\left(\hat{p}_{MCMC}\right) = \frac{1}{T-1}\sum_{t=1}^{T}\left(p^{(t)} - \overline{p}\right)^2,$$

where $p^{(t)}$ is generated from the posterior distribution based on the beta distribution at parameter $a^{(t)}$ and $b^{(t)}$, $\overline{p}$ is a sample mean of independent observation, and T is a iteration of posterior distribution function.

The confidence interval $(1-\alpha)100\%$ of p is written by

$$\left(\hat{p}_{MCMC} - z_{\alpha/2}\sqrt{Var\left(\hat{p}_{MCMC}\right)}, \hat{p}_{MCMC} + z_{\alpha/2}\sqrt{Var\left(\hat{p}_{MCMC}\right)}\right).$$

3.3 The estimating confidence coefficient

The confidence interval is approximated by ML, Bayesian, and MCMC methods at significance level 0.1, 0.05, and 0.01. If the confidence intervals cover the true parameters, we will count the number and compute the proportion denoted the confidence coefficient $\left(1-\hat{\alpha}\right)$.

3.4 The comparison of the confidence coefficient and fixed confidence interval

The confidence coefficient $\left(1-\hat{\alpha}\right)$ is to compare with the fixed confidence interval $\left(1-\alpha_0\right)$ that we define the significance level at 0.05. If the confidence coefficient is more than the fixed confidence interval, we will perform these methods. The comparison is given by

$$1-\hat{\alpha} \geq 1-\alpha_0 = P_0 - z_{\alpha/2}\sqrt{\frac{P_0\left(1-P_0\right)}{m}},$$

where P_0 is the fixed probability given by 0.9, 0.95, and 0.99, m is the number of replications.

The fixed confidence intervals are computed by $P_0 = 0.9$,

$$1-\alpha_0 = 0.9 - 1.96\sqrt{\frac{0.9(1-0.9)}{1,000}} = 0.8814,$$

$P_0 = 0.95$,

$$1-\alpha_0 = 0.95 - 1.96\sqrt{\frac{0.95(1-0.95)}{1,000}} = 0.9365,$$

$P_0 = 0.99$,

$$1-\alpha_0 = 0.99-1.96\sqrt{\frac{0.99(1-0.99)}{1,000}} = 0.9838.$$

3.5 The average width of confidence interval

The average width of confidence interval is evaluated by computing the average if difference values between upper limit and lower limit or written as $\sum_{j=1}^{1,000}\frac{(U_j - L_j)}{1,000}$, where U_j is the upper confidence interval, and L_j is the lower confidence interval.

4. Results

The estimating confidence interval of p with negative binomial distribution is presented by the Confidence Coefficient (CC) and the Average Width (AW) at Table 1-3. The first column to the third column of these tables are shown the true parameters, sample sizes, and r. The Confidence Coefficient (CC) and the Average Width (AW) are presented in the next six columns for three methods. The minimizing AW values are illustrated the performance of these methods, but some AW values are in the blank because the confidence coefficient is less than the fixed confidence interval. By observing the CC and AW, the results appear as follows:

4.1 *A 90% confidence interval*

Table1: The Confidence Coefficient (CC) and Average Width (AW) obtained via 90% confidence interval.

p	n	r	Methods					
			ML		Bayesian		MCMC	
			CC	AW	CC	AW	CC	AW
0.2	30	10	0.563	-	0.878	-	0.885	**0.0340**
		15	0.553	-	0.887	0.0278	0.896	**0.0277**
		20	0.607	-	0.889	**0.0240**	0.898	**0.0240**
	50	10	0.608	-	0.882	**0.0264**	0.881	-
		20	0.561	-	0.884	**0.0186**	0.890	**0.0186**
		30	0.600	-	0.904	**0.0152**	0.905	**0.0152**
	70	15	0.570	-	0.877	-	0.883	**0.0182**
		30	0.601	-	0.898	0.0129	0.895	**0.0128**
		45	0.584	-	0.904	**0.0105**	0.906	**0.0105**
0.5	30	10	0.913	0.0673	0.897	**0.0667**	0.912	0.0670
		15	0.909	0.0550	0.897	**0.0546**	0.910	0.0548
		20	0.886	0.0475	0.888	**0.0473**	0.889	0.0474
	50	10	0.896	0.0519	0.903	**0.0517**	0.897	0.0519
		20	0.917	0.0368	0.916	**0.0367**	0.919	**0.0367**
		30	0.907	0.0301	0.910	**0.0300**	0.905	**0.0300**
	70	15	0.898	0.0359	0.900	**0.0358**	0.904	**0.0358**
		30	0.895	**0.0254**	0.895	**0.0254**	0.896	**0.0254**
		45	0.892	**0.0207**	0.895	**0.0207**	0.893	**0.0207**
0.8	30	10	0.997	0.1361	0.907	**0.0671**	0.908	0.0677
		15	1	0.1113	0.894	**0.0550**	0.895	0.0553
		20	0.998	0.0960	0.895	**0.0478**	0.894	0.0480
	50	10	0.999	0.1055	0.898	**0.0522**	0.904	0.0525
		20	1	0.0744	0.904	**0.0371**	0.899	0.0372
		30	0.999	0.0607	0.903	**0.0303**	0.907	0.0304
	70	15	0.995	0.0727	0.902	**0.0362**	0.897	0.0363
		30	0.999	0.0514	0.891	**0.0256**	0.887	0.0257
		45	0.999	0.0420	0.920	**0.0209**	0.919	0.0210

From Table 1, the AW of ML method is a minimum values when $p = 0.5$ with $n = 70$ and $r = 30$ and 45. For Bayesian method, the AW appears the minimum values at most all cases especially when $p = 0.5$, and 0.8. The AW of MCMC method outperforms at $p = 0.2$, but the minimum of AW by MCMC is the same value of AW by Bayesian at $p = 0.5$.

4.2 *A 95% confidence interval*

Table2: The Confidence Coefficient (CC) and Average Width (AW) obtained via 95% confidence interval.

p	n	r	Methods					
			ML		Bayesian		MCMC	
			CC	AW	CC	AW	CC	AW
0.2	30	10	0.698	-	0.947	**0.0405**	0.945	**0.0405**
		15	0.653	-	0.933	-	0.949	**0.0332**
		20	0.652	-	0.948	**0.0287**	0.961	**0.0287**
	50	10	0.688	-	0.961	0.0315	0.962	**0.0314**
		20	0.673	-	0.961	**0.0222**	0.959	**0.0222**
		30	0.651	-	0.940	**0.0181**	0.944	**0.0181**
	70	15	0.677	-	0.944	**0.0217**	0.948	**0.0217**
		30	0.686	-	0.962	**0.0153**	0.961	**0.0153**
		45	0.666	-	0.945	**0.0125**	0.950	**0.0125**
0.5	30	10	0.956	0.0801	0.958	**0.0794**	0.959	0.0798
		15	0.687	-	0.954	**0.0331**	0.951	**0.0331**
		20	0.953	0.0566	0.947	**0.0563**	0.954	0.0565
	50	10	0.956	0.0619	0.952	**0.0617**	0.954	0.0618
		20	0.949	0.0438	0.948	**0.0437**	0.950	0.0438
		30	0.946	0.0358	0.940	**0.0357**	0.944	0.0358
	70	15	0.939	0.0428	0.940	**0.0427**	0.939	0.0428
		30	0.949	**0.0302**	0.950	**0.0302**	0.949	**0.0302**
		45	0.947	**0.0247**	0.942	**0.0247**	0.945	**0.0247**
0.8	30	10	0.999	0.1620	0.958	**0.0800**	0.958	0.0806
		15	1	0.1326	0.954	**0.0655**	0.951	0.0659
		20	1	0.1148	0.950	**0.0568**	0.949	0.0571
	50	10	1	0.1254	0.952	**0.0623**	0.944	0.0626
		20	1	0.0887	0.958	**0.0442**	0.955	0.0443
		30	1	0.0724	0.947	**0.0361**	0.948	0.0362
	70	15	1	0.0866	0.951	**0.0431**	0.950	0.0432
		30	1	0.0613	0.959	**0.0305**	0.961	0.0306
		45	1	0.0500	0.966	**0.0250**	0.965	**0.0250**

From Table 2, the AW of ML method is a minimum values when $p = 0.5$ with $n = 70$ and $r = 30$ and 45. For Bayesian method, the AW appears the minimum values in all cases except when $p = 0.2$, $n = 30$, $r = 15$ and $n = 50$, $r = 10$. The AW of MCMC method appears the minimum values at all cases on $p = 0.2$, but some values are similar the AW of Bayesian.

4.3 *A 99% confidence interval*

Table3: The Confidence Coefficient (CC) and Average Width (AW) obtained via 99% confidence interval.

p	n	r	Methods					
			ML		Bayesian		MCMC	
			CC	AW	CC	AW	CC	AW
0.2	30	10	0.812	-	0.989	0.0536	0.994	**0.0534**
		15	0.779	-	0.987	0.0436	0.987	**0.0435**
		20	0.820	-	0.995	**0.0376**	0.995	**0.0376**
	50	10	0.806	-	0.991	**0.0412**	0.991	**0.0412**
		20	0.802	-	0.988	0.0292	0.991	**0.0291**
		30	0.815	-	0.988	**0.0238**	0.987	**0.0238**
	70	15	0.826	-	0.987	**0.0285**	0.991	**0.0285**
		30	0.816	-	0.989	**0.0201**	0.988	**0.0201**
		45	0.807	-	0.987	**0.0164**	0.987	**0.0164**
0.5	30	10	0.990	0.1053	0.993	**0.1043**	0.993	0.1049
		15	0.985	0.0860	0.982	-	0.985	**0.0857**
		20	0.991	0.0745	0.988	**0.0741**	0.989	0.0743
	50	10	0.989	0.0814	0.990	**0.0810**	0.989	0.0813
		20	0.987	0.0577	0.989	**0.0575**	0.987	0.0576
		30	0.991	**0.0470**	0.991	**0.0470**	0.990	**0.0470**
	70	15	0.992	0.0562	0.993	**0.0561**	0.991	**0.0561**
		30	0.990	0.0398	0.988	**0.0397**	0.988	**0.0397**
		45	0.989	0.0325	0.989	**0.0324**	0.989	**0.0324**
0.8	30	10	1	**0.2136**	0.983	-	0.983	-
		15	1	0.1738	0.988	**0.0862**	0.985	0.0866
		20	1	0.1507	0.990	**0.0747**	0.990	0.0751
	50	10	1	0.1652	0.988	**0.0817**	0.989	0.0822
		20	1	0.1168	0.989	**0.0580**	0.990	0.0582
		30	1	0.0953	0.994	**0.0474**	0.994	0.0475
	70	15	1	0.1138	0.995	**0.0567**	0.995	0.0568
		30	1	0.0804	0.989	**0.0402**	0.990	**0.0402**
		45	1	0.0656	0.993	**0.0328**	0.993	**0.0328**

From Table 3, the AW of ML method is a minimum values when $p = 0.5$ with $n = 50, r = 30$ and $p = 0.8$ with $n = 80$, $r = 10$. For Bayesian method, the AW appears the minimum values at most all cases. The AW of MCMC method outperforms at $p = 0.2$.

5. Conclusions

In this research, data is generated from a negative binomial distribution and estimated the confidence interval which is obtained confidence coefficient and average interval width to perform ML, Bayesian, and MCMC methods. Through a simulation study, the ML method is a poor method to estimated interval because the ML method is used just real data, then it's not consistent estimator. The Bayesian method is depended on the prior distribution to create the posterior distribution, so it makes a good performance method in most all cases especially when true parameters are medium or large values for all sample sizes and or parameter r. For the small true parameters, the MCMC method is a good performance in most all cases. However, the confidence coefficient and average width of Bayesian and MCMC are equal in some case, so we would recommend user to use MCMC method especially small true parameters because the Bayesian method depends on parameter of prior distribution while it is hard to find the appropriate parameter. For real data set, MCMC method outperforms the other methods to estimate confidence interval as same as the simulation study data.

References

[1] Norman LJ and Samuel K, Distributions in Statistics: Discrete Distributions, New York: John Wiley & Sons; 1969.

[2] Christensen R, Johnson W, Branscum A, Hanson TE, Bayesian ideas and data analysis: An introduction for scientists and statisticians, Florida: Taylor & Francis Group; 2011.

[3] Laskey P, Mayer J, Population Markov Chain Monte Carlo, Machine Learning, 2003; 50 : 175-196.

[4] Geman S, Geman D, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984; 6: 721-41.

[5] Krishnamoorthy K, Handbook of Statistical Distributions with Applications, London: Chapman & Hall/CRC; 2016.

[6] Gamerman D, Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference, London: Chapman & Hall; 1997.