Advances and Applications in Statistics
Volume 14, Number 1, 2010, Pages 49-55
Published Online: December 19, 2009
This paper is available online at http://www.pphmj.com

A COMPARATIVE STUDY OF PARAMETER INFERENCE METHODS FOR POISSON DISTRIBUTION: SMALL SAMPLE SIZES

MANLIKA TANUSIT

Department of Mathematics and Statistics
Faculty of Science
Maejo University
Chiang Mai 50290, Thailand

Abstract

The objective of this study is to compare point estimation methods and interval estimation methods for the Poisson distribution with small sample sizes. Three methods of point estimation: maximum likelihood method, Bayesian method and minimax method, and three methods of interval estimation: normal method, normal-Bayesian method and score-Bayesian method are considered. The lowest mean absolute error and the lowest average width are used as the criteria of selection for point estimation and interval estimation, respectively. The scopes of this study consist of sample sizes: 5, 6, 7, 8, 9, 10 and the parameter λ is equal to 0.02, 0.04, 0.06, 0.08 and 0.1. Data is simulated 1,000 times generated by using the JAVA software. The results of this research are as follows: For point estimation, we recommend that for all sample sizes and parameter λ, the Bayesian method should be used. In case of interval estimation, normal-Bayesian method is recommended for sample sizes are 5 to 8, λ between 0.2 to 0.4 and sample sizes are 9 to 10, λ is equal to 0.2 whereas score-Bayesian method should be considered for sample sizes are 5 to 8, values of λ ranging from 0.6 to 1 and sample sizes are 9 to 10, λ between 0.4 to 1.0.

2010 Mathematics Subject Classification: 62-XX.

Keywords and phrases: Poisson distribution, interval estimation, sample size.

Received August 24, 2009

Introduction

Estimation is one of the most important methods used in statistics. The value of the chosen parameters is depending on the estimation or conclusion from the sample data. There are 2 methods used to estimate value of parameters. They are point estimation and interval estimation. Point estimation is parameter estimation by focusing only on some particular values such as average estimation, variance estimation and population proportion estimation. Interval estimation is the estimation of population parameters at some given point. For example, estimation of population mean within the range of A and B at 90% confidence level. For this method, the highest and lowest values for this estimation will be varied depend on level of confidence.

In this work, we are interested in comparative study of parameter inference methods for Poisson distribution when small sample sizes. Point estimation: maximum likelihood method, Bayesian method and minimax method are considered. Interval estimation: normal method, normal-Bayesian method and score-Bayesian method are considered to find appropriate method under the scope of same study.

Materials and Methods

Materials

JAVA Software.

Methods

1. Create random variable that has Poisson distribution.

2. Compare the methods in point estimation by using the lowest mean absolute error as the selection criteria. The method that gives the lower means absolute error will be the most suitable method to be implemented for this situation. Methods of point estimation are as follows:

2.1. Maximum likelihood method [4].

Maximum likelihood estimator for λ is $\hat{\lambda} = \frac{X}{n}$.

2.2. Bayesian method, in this case, determines prior distribution of λ is exponential distribution $(X; \beta)$ [1].

Bayesian estimator for λ is $\hat{\lambda} = \frac{(X+1)\beta}{n\beta + 1}$ when $\beta = \frac{1}{\lambda}$.

2.3. Minimax method [6].

Minimax estimator of λ is $\hat{\lambda} = \frac{2(X+1)}{3n}$.

3. Compare the method of interval estimation by using the lowest average wide in making decision and the approximate confidence coefficient that was computed in each trial must be greater than 0.8930, 0.9449 and 0.9877 [5] at level of confidence 90%, 95% and 99%, respectively then we will consider that the methods of estimation gives the approximate confidence coefficients not less than given confidence coefficient. After that the methods of estimation that give the approximate confidence coefficients not less than given confidence coefficient will be chosen to do further comparison with the average width of interval estimation. Any methods that are able to give lowest average width of interval estimation then we consider that method as the most suitable for that situation. Methods of interval estimation are as follows:

3.1. Normal method [2] with confidence interval of $(1-\alpha)100\%$ λ is

$$\left(\hat{\lambda} - Z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{n}}, \hat{\lambda} + Z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{n}}\right),$$

when

$$\hat{p} = \frac{X}{n}.$$

3.2. Normal-Bayesian method with confidence interval of $(1-\alpha)100\%$ λ is

$$\left(\hat{\lambda} - Z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{n}}, \hat{\lambda} + Z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{n}}\right),$$

when

$$\hat{\lambda} = \frac{(X+1)\beta}{n\beta + 1}.$$

3.3. Score-Bayesian method [3] with confidence interval of $(1-\alpha)100\%$ p is

$$\left(\hat{\lambda}+\frac{Z_{1-\alpha/2}^{2}}{2n}-\frac{Z_{1-\alpha/2}}{n^{1/2}}\sqrt{\hat{\lambda}+\frac{Z_{1-\alpha/2}^{4}}{4n}},\ \hat{\lambda}+\frac{Z_{1-\alpha/2}^{2}}{2n}+\frac{Z_{1-\alpha/2}}{n^{1/2}}\sqrt{\hat{\lambda}+\frac{Z_{1-\alpha/2}^{4}}{4n}}\right),$$

when

$$\hat{\lambda}=\frac{2(X+1)}{3n}.$$

4. Summary.

Result and Discussion

Point estimation from comparing the mean absolute error of parameter estimator estimated from parameter λ shown by 3 methods are summarized in Table 1.

Table 1. Mean absolute error separate by sample size and parameter

n	λ	Maximum likelihood method	Bayesian method	Minimax method
5	0.2	0.1696	0.0420*	0.0693
	0.4	0.3220	0.1591*	0.2147
	0.6	0.4866	0.3273*	0.3895
	0.8	0.6410	0.4921*	0.5604
	1.0	0.7400	0.6167*	0.6933
6	0.2	0.1710	0.0561*	0.0782
	0.4	0.3331	0.1918*	0.2439
	0.6	0.5007	0.3593*	0.4222
	0.8	0.6645	0.5334*	0.5986
	1.0	0.8000	0.6857*	0.7556
7	0.2	0.1736	0.0664*	0.0879
	0.4	0.3396	0.2100*	0.2642
	0.6	0.5151	0.3905*	0.4482
	0.8	0.6286	0.5179*	0.5905
	1.0	0.8571	0.7500*	0.8095
8	0.2	0.1771	0.0748*	0.1021
	0.4	0.3469	0.2316*	0.2812
	0.6	0.5253	0.4142*	0.4668
	0.8	0.6994	0.5949*	0.6496
	1.0	0.8736	0.7766*	0.8324

9	0.2	0.1775	0.0804*	0.1113
	0.4	0.3571	0.2530*	0.2972
	0.6	0.5352	0.4352*	0.4827
	0.8	0.7143	0.6193*	0.6688
	1.0	0.8900	0.8010*	0.8526
10	0.2	0.1802	0.0865*	0.1201
	0.4	0.3571	0.2628*	0.3047
	0.6	0.5422	0.4511*	0.4948
	0.8	0.7217	0.6349*	0.6811
	1.0	0.9009	0.8190*	0.8673

• The lowest mean absolute error for each case.

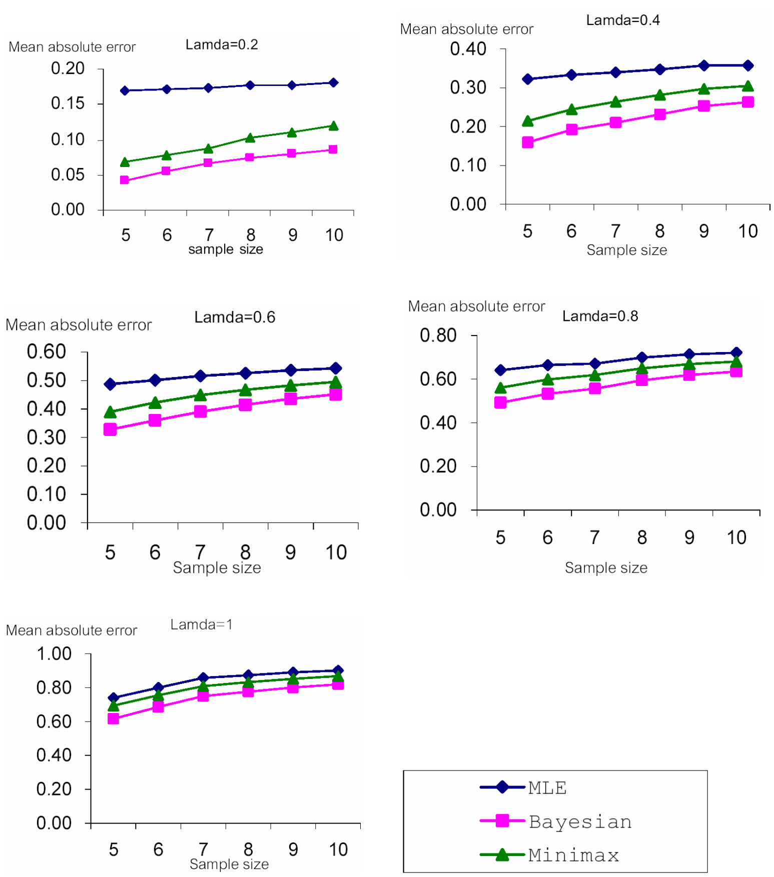

Figure 1. Examples of mean absolute error separate by sample size and parameter.

From Table 1 and Figure 1, we can conclude that for all sample sizes and parameter λ, Bayesian method should be used because these gave the lowest mean absolute error.

Interval estimation considering level of confidence at 90%, 95% and 99%, we have found that each method that gives the approximate confidence coefficient not less than given confidence coefficient yield are as follows:

Normal-Bayesian method: sample sizes are 5 to 8, λ between 0.2 to 0.4 and sample sizes are 9 to 10, λ is equal to 0.2.

Score-Bayesian method with all parameter λ and all sample sizes.

When taking all given cases that give the approximate confidence coefficient not less than given confidence coefficient into comparison for average width, we have found that the method that gives the lowest average width are as follows:

Normal-Bayesian method: sample sizes are 5 to 8, λ between 0.2 to 0.4 and sample sizes are 9 to 10, λ is equal to 0.2.

Score-Bayesian method: sample sizes are 5 to 8, values of λ ranging from 0.6 to 1 and sample sizes are 9 to 10, λ between 0.4 to 1.

Conclusion

This work can be concluded as follows:

For point estimation, considering the mean absolute error of the Maximum likelihood, Bayesian and Minimax methods found for all sample sizes and parameter λ, Bayesian method should be used because these gave the lowest mean absolute error.

For interval estimation, using the lowest average wide as method selection criteria, we conclude that we should use the appropriate methods for each situation as follows. Normal-Bayesian method is recommended for sample sizes are 5 to 8, λ between 0.2 to 0.4 and sample sizes are 9 to 10, λ is equal to 0.2 whereas score-Bayesian method should be used for sample sizes are 5 to 8, values of λ ranging from 0.6 to 1 and sample sizes are 9 to 10, λ between 0.4 to 1.0.

References

[1] William M. Bolstad, Introduction to Bayesian Statistics, John Wiley & Sons, Inc., New York, 2004.

[2] Lawrence D. Brown, T. Tony Cai and Anirban Dasgupta, Interval estimation in exponential families, Statist. Sinica. 13(1) (2003), 19-49.

[3] T. Tony Cai, One-sided confidence intervals in discrete distributions, J. Statist. Plann. Inference. 131(1) (2005), 63-88.

[4] G. Casella and R. L. Berger, Statistical Inference, John Wiley & Sons, Inc., New York, 2002.

[5] B. K. Ghosh, Comparison of some approximate confidence intervals for the binomial parameter, J. Amer. Statist. Assoc. 74(368) (1979), 894-900.

[6] Roger V. Hogg and Elliot A. Tanis, Probability and Statistical Inference, Pearson Education, Inc., New Jersey, 2006.