Original

The Comparison of Point Estimation for Parameter for Geometric Distribution Data in Small Sample Size

Yadpirun Supharokonsakun[1]

Received: 18 February 2013 Accepted: 15 June 2013

Abstract

The objective of this research is to compare two point estimation methods: Maximum Likelihood Method (MLE) and Bayesian Method (Baye). When data is Geometric distribution, the parameters (p) are 0.1, 0.3, 0.5, 0.7 and 0.9, whereas the sample sizes (n) are 3, 5, 8, 10, 12, 15, 18, 23, 25, 28 and 30. In each situation, the data has been simulated and repeated for 1,000 times. The Mean Absolute Error is used as a criterion for comparison.

According to the results, when the sample sizes are 3, 5, 8 and 10, on overall MLE yields the least mean absolute error when parameter equal to 0.1. Whereas parameter larger than or equal to 0.3, on overall Baye yields the least mean absolute error. when the sample size equal to 12, 15, 18, 20, 23, 25, 28 and 30, on overall MLE yields the least mean absolute error when parameter equal to 0.1 and 0.3. Whereas parameter larger than or equal to 0.5, on overall Baye yields the least mean absolute error.

Keywords: Geometric Distribution, Point Estimation

Introduction

In general, when we do research, it is impossible to investigate every unit or population. We are interested in multiple limitations, such as restrictions on budget and time. Thus, we can study the characteristics of population by simple random sampling from the mentioned population. The features of sample come from statistics calculated from sample data. It can be said that statistical inference is to infer the statistics of the sample data to the population itself.

According to the statistical principle, statistics can be classified into two parts: descriptive and inferential statistics. The descriptive statistics is used for planning, designing operating, collecting and presenting data including evaluating and calculating primarily the collected data. The inferential statistics is concerned with data analysis which comprises estimation and hypothesis testing and may include forecasting or prediction and model building. Therefore, the inferential statistics must be used to analyze the data collected in order to sum up not only the facts or characteristics of parameter as a whole, but also the available data.

Statistical hypothesis testing will be used to test the value of data or interesting matters in order to prove that it is true or not. The estimation will be done when we want to know the amount of unknown value, such as the estimated average cost per month of Phetchabun Rajabhat University's students. The estimation can be divided into two methods: point and interval estimation. The purpose of these two methods is to estimate a parameter which is the same. However, we can get only a single value through the point estimation while the interval estimation provides us a set in the range of estimation which is commonly referred to "Confidence Interval."

Moreover, the concepts used in statistical inference can be divided into two categories: classical and Bayes inferences. In the past, many researches were investigated through the comparison of various estimations. The most popular category is the maximum likelihood which is easy to compute and provides a good estimation. In addition, the Bayesian inference is also a popular approach because it provides an approximation which is close to a true parameter value. Furthermore, it is a new concept to use prior knowledge to assist in estimating values.

[1] *Dept. Mathematics, Faculty of Science and Technology Phetchabun Rajabhat University, PCRU Phetchabun, Thailand E-mail:tammakulpitak_zcc_7@hotmail.com*

The researcher would like to study and compare how to estimate a parameter by using the point estimation through the maximum likelihood method and the Bayes inference as the data are geometrically distributed. Anyway, such a kind of research has never done before.

Literature Review

P. Arkhom[1] compared the method of estimating parameter interval by using three methods: Pivotal Quantity Method, Bayes' Estimator Method and Minimize Method with small sample sizes of Poisson Distribution. The findings revealed Bayes' Estimator Method indicated the smallest interval.

P. Krittaya[2] compared two interval estimation methods for parameter of Poisson distribution. There were Maximum Likelihood Method and Baysian Estimation Method which used Gamma prior distribution when the data were small sample size. The findings revealed the Baysian yieded an average width of interval was less than the Maximum Likelihood method for each case studied. T. Manlika[3] to compare point estimation methods for parameter of binomial distribution by using three methods: Maximum likelihood method, Bayesian method and Minimax method. The findings revealed the Bayesian method should be used for small sample size. For parameter p between 0.30 to 0.50 all three method gave similar result. Anyway, the maximum likelihood method should be considered because it is easier and more convenient that the others.

The Scope of Research

In this study, researcher has determined the scope of the research is follows.

1. Using the Beta Distribution is the prior distribution.
2. Setting the sample size (n) are 3, 5, 10, 12, 15, 18, 20, 23, 25, 28 and 30.
3. Setting the parameter p are 0.1, 0.3, 0.5, 0.7, and 0.9.
4. Simulating and repeating for 1,000 times.

Objective

To compare two point estimation methods: Maximum Likelihood Method (MLE) and Bayesian Method (Baye).

The Point Estimator

A. Abbreviations and Acronyms The Maximum Likelihood Estimator of Eeometric Distribution

Let $x_1, x_2, ..., x_n$ are random samples with a geometric distribution is defined as $X \sim Geo(p)$, the distribution function of X were following (1)

$$f(x \mid p) = p(1-p)^{x-1} \tag{1}$$

By x = $1, 2, 3, 0 < p \le 1$ and the Maximum Likelihood Estimator were following (2)

$$\hat{p}_{MLE} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}} \tag{2}$$

B. The Posterior of Geometric Distribution.

Let $x_1, x_2, ..., x_n$ are random samples with a geometric distribution is defined as $X \sim Geo(p)$, the distribution function of X from (1). When $p \sim Beta(\alpha, \beta)$, we have the posterior function distribution were following (3)

$$p \mid \underline{x} \sim Beta\left(\alpha + n, \sum_{i=1}^{n} X_i - n + \beta\right) \tag{3}$$

Criteria Used In The Comparison

The Mean Absolute Error are used as criteria for comparison. The criteria for a selection from these methods were their performance on the lowest mean absolute error in each of the simulation following (4)

$$|e_i| = |p_i - \hat{p}_i| \tag{4}$$

When e_i is absolute error, p_i is parameter and $\hat{p}_i$ is estimates of the parameters.

Research Methodology

The data were generated through the Monte Carlo simulation technique with the following steps.

1. Set the sample size (n) and parameter (p).
2. Generated data.
3. Calculates estimation 2 methods for parameter.
4. Calculates the mean absolute error for each estimation method.
5. Compare the mean absolute error for each estimation method.
6. Concludes the result in each case.

Result

Table 1 showed that for the sample size equal to 3, 5, 8 and 10, on overall MLE yields the least mean absolute error when parameter equal to 0.1. Whereas parameter larger than or equal to 0.3, on overall Baye yields the least mean absolute error. when the sample size equal to 12, 15, 18, 20, 23, 25, 28 and 30, on overall MLE yields the least mean absolute error when parameter equal to 0.1 and 0.3. Whereas parameter larger than or equal to 0.5, on overall Baye yields the least mean absolute error.

Table 1 The mean absolute error.

n	p	Method	
		MLE	Bayes
3	0.1	0.113854[a]	0.205382
	0.3	0.652112	0.236896[a]
	0.5	0.469286	0.126384[a]
	0.7	1.275000	0.117540[a]
	0.9	1.800000	0.346725[a]
5	0.1	0.066676[a]	0.167442
	0.3	0.223697	0.176494[a]
	0.5	1.324819	0.143075[a]
	0.7	2.650000	0.079432[a]
	0.9	2.433333	0.133552[a]
8	0.1	0.042066[a]	0.100081
	0.3	0.255551	0.191366[a]
	0.5	0.842521	0.166149[a]
	0.7	2.456127	0.100564[a]
	0.9	4.033333	0.066017[a]

a. Least mean absolute error

Table 2 The mean absolute error. (Cont.)

n	p	Method	
		MLE	Bayes
10	0.1	0.052990[a]	0.118936
	0.3	0.201694	0.180234[a]
	0.5	0.759002	0.187358[a]
	0.7	2.011905	0.099167[a]
	0.9	7.766667	0.042836[a]
12	0.1	0.039149[a]	0.122186
	0.3	0.180784[a]	0.186662
	0.5	0.783106	0.205226[a]
	0.7	2.455247	0.107428[a]
	0.9	7.100000	0.112296[a]
15	0.1	0.030522[a]	0.094951
	0.3	0.175930[a]	0.178810
	0.5	0.670108	0.180213[a]
	0.7	2.355366	0.078875[a]
	0.9	8.100000	0.059839[a]
18	0.1	0.027007[a]	0.123648
	0.3	0.179597[a]	0.182667
	0.5	0.628504	0.121993[a]
	0.7	2.254313	0.052554[a]
	0.9	9.270000	0.138387[a]
20	0.1	0.025606[a]	0.114608
	0.3	0.162722[a]	0.172315
	0.5	0.592239	0.125814[a]
	0.7	2.163981	0.058969[a]
	0.9	12.100000	0.092711[a]
23	0.1	0.024960[a]	0.107718
	0.3	0.163972[a]	0.174340
	0.5	0.606365	0.144193[a]
	0.7	2.089743	0.072220[a]
	0.9	10.216670	0.040991[a]
25	0.1	0.022557[a]	0.098156
	0.3	0.162848[a]	0.172678
	0.5	0.584174	0.150842[a]
	0.7	2.182243	0.092137[a]
	0.9	12.873810	0.032082[a]
28	0.1	0.021242[a]	0.088453
	0.3	0.154588[a]	0.172630
	0.5	0.577110	0.163865[a]
	0.7	2.000246	0.112911[a]
	0.9	9.236242	0.036289[a]
30	0.1	0.020811[a]	0.093885
	0.3	0.154426[a]	0.169324
	0.5	0.573016	0.151895[a]
	0.7	1.955711	0.089182[a]
	0.9	10.361900	0.025562[a]

a. Least mean absolute error

Conclusion

The findings revealed the of point estimation for parameter for geometric distribution data in small sample size, when the parameter is a small size, it will show MLE better than Baye. If the parameter is big, should use Baye better than MLE.

Acknowledgment

This research has been successfully completed by the financial support of Research and Development Institution of Phetchabun Rajabhat University.

Most of all, the researcher would appreciate all professors in the Mathematics Program, Faculty of Science and Technology, Phetchabun Rajabhat University.

Thanks for inspiring my family and help to achieve everything.

References

1. P. Arkhom, “A Comparison on Small Sample Interval Estimation of Poisson Parameter”, Thesis M.S., Chiang Mai University, 2002.
2. P. Krittaya, “Small Sample Estimation of Poisson Parameter Using Bayesian Method”, Thesis M.S., King Mongkut’s Institute of Technology North Bangkok, 2006.
3. T. Manlika, “A Comparative Study of Parameter Estimation Method for Binomial Distribution”, Thesis M.S., Kasetsart University, 2008.