Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Bootstrap Method in Estimation of Mean Squared Error of Beta-Bernoulli Model

To cite this article: Widiarti *et al* 2019 *J. Phys.: Conf. Ser.* **1338** 012039

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012039 doi:10.1088/1742-6596/1338/1/012039

Bootstrap Method in Estimation of Mean Squared Error of Beta-Bernoulli Model

Widiarti[1,a], N Adityawati[1,b] and Nusyirwan[1,c]

[1]Mathematics Department, Faculty of Mathematics and Natural Sciences, Lampung University, Indonesia

[a]widiarti08@gmail.com; [b]nadityawati@gmail.com;
[c]nusyirwan.1966@fmipa.unila.ac.id

Abstract. Small area estimation is defined as a statistical technique to estimate small sub population of a certain area. There are several method of small area including Empirical Bayes (EB), Hierarchical Bayes (HB) and Empirical Best Linear Unbiassed Prediction (EBLUP). EB method is one of methods in small area estimation for count or binary data. Estimation with EB method is based on posterior which its parameter is estimated by data. One application of EB method for binary data is Beta-Bernoulli Model. In this study, Mean Squared Error (MSE) of the EB estimator was evaluated by Bootstrap method by theory and empirical though simulation by using Ri386 3.4.3 software. The results of this study showed that EB estimator is biased, and the Bootstrap MSE becomes smaller when the amount of areas get greater.

1. Introduction

Small area estimation is a statistical technique to estimate parameters of sub populations with small sample sizes. Estimating a small area based on the application of a design model is referred to a direct estimation. This estimation is not able to provide sufficient accuracy if the sample size is small, so the statistics obtained will have a large variety or even the estimation can not be made because they are not represented in the survey. Therefore, alternative estimation techniques have been developed to increase the effectiveness of sample size and reduce standard errors, namely indirect estimation. Indirect estimation borrows the power from the sample's observation of adjacent areas by using additional information, from censuses and administrative records [1].

Various indirect estimation methods have been developed to obtain small area estimators. They are Empirical Best Linear Unbiased Prediction (EBLUP) for continuous data, Empirical Bayes (EB) and Hierarchical Bayes (HB) for count and empirical data.

Estimation using EB method is based on posterior spread which parameter is estimated by data. Momen method is used to estimate the parameter estimation used in the EB method.The concept of estimating small areas is to utilize additional information known as prior distributions.Thus, a prior distribution is needed to accommodate this additional information.In this research, the prior distribution used is the Beta distribution, while the sample distribution used is the Bernoulli distribution which belongs to the exponential family. If the distribution of the sample used comes from an exponential family, then the way to determine the prioris by using the prior conjugate [2].The probability distribution function of Beta has the same functional form with Bernoulli likelihood distribution or prior conjugate and comes from exponential families.

Bayes estimator is usually biased, so in this research the EB quality obtained will be evaluated through the Mean Squared Error (MSE) criteria. MSE is a quantity to measure the estimation of small area estimators. The smaller the MSE of an estimator, the more accurate the estimator, the method that could correct bias is Bootstrap [3]. Research on the use of bootstrap methods has been conducted in the case of small area estimation by [1]. In the research, Rao explained about the standard error estimation for the Jackniffe and Bootstrap methods in small areas usage. Previously, he had explained the point and interval accuracy using bootstrap samples [3]. This is because bootstrap estimation is closer to its population compared to other sampling. The accuracy of the bootstrap sample for point and intervals estimation is evaluated using standard errors. Based on the description, the bootstrap method will be used in this research to evaluate MSE in Beta-Bernoulli model.

2. Materials and Methods

This research investigates the parameter estimation of Beta-Bernoulli model using empirical Bayes method with Bernoulli-distributed sample and Beta prior-distribution,i.e.y_i ~ Bernoulli (p_i) and as a prior distribution p_i ~ Beta (α, $\frac{1}{\alpha}$).

The steps in this research are as follows:

1. Determining posterior density function from Beta-Bernoulli model
2. Determining empirical bayes from Beta-Bernoulli model
3. Determining α estimator with method of moments based on probability density function of the marginal
4. Proving the unbiasedness of Empirical Bayes estimator
5. Determining variance of Empirical Bayes estimator
6. Evaluating Mean Square Error for Empirical Bayes with Bootstrap method

Evaluating bias and Mean Squared Error (MSE) empirically through simulation studies with the following steps:

1. Determining variety between areas for Beta-Bernoulli model 1, 5, and 10 as a representation of small, medium, and large variance
2. Determining the number of areas 10, 50, and 100 as a representation of the number of areas small, medium, and large
3. Generating p_i data with p_i ~Beta(α,1/ α)
4. Generating y_i data with y_i ~Bernoulli(p_i)
5. Calculating the estimated value for $\hat{\alpha}$ with method of moments
6. Calculating the estimated value for p_i with Emprical Bayes method ($\hat{p}_i^{EB}$)
7. Calculating the bias value
8. Calculating MSE $\hat{p}_i^{EB}$ with bootstrap method

3. Results and Discussion

The empirical Bayes estimator is $\hat{p}_i^{EB}= \frac{(\hat{\alpha} y_i+\hat{\alpha}^2)}{(\hat{\alpha}^2+\hat{\alpha}+1)}$. The assumption for the α parameter is based on the marginal density function Y. According to [4] the expectation value of Beta-Bernoulli distribution is E(y_i | $\alpha, \frac{1}{\alpha}$) = $\frac{\alpha}{\alpha+\frac{1}{\alpha}}$ then equalizing it with the 1st moment and we get $\hat{\alpha} = \frac{\bar{y}_i}{(1-\bar{y}_i)}$. The data used in this research is simulation data generated by Ri386 3.4.3 software. Data generation in simulation is assumed to be the Beta-Bernoulli distribution that is y_i~Bernoulli (p_i) with p_i~Beta (α, $1/\alpha$,). Simulations are carried out to see the estimators quality produced by the Empirical Bayes method empirically.This simulation study is carried out to see the number of areas and variety between areas effects on the EB estimator quality. Variations between areas are predetermined as are 1, 5, and 10 as a representations of the variety between small, medium and large areas. Then, the data are generated with the number of areas that is 10, 50, and 100.The difference of various areas and the number areas representations aims to see the size of bias value, MSE, and bootstrap MSE from p_i estimator. The resampling size on MSE bootstrap follows the number of areas.

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012039 doi:10.1088/1742-6596/1338/1/012039

Table 1. Bias value, MSE and bootstrap MSE for empirical bayes estimation

Variety between areas	Number of areas	Bias	MSE	MSE with Bootstrap
	10	0.23602098	0.07354150	0.0591361
1	50	0.20849206	0.06807047	0.0350651
	100	0.20277110	0.06715636	0.0315577
	10	0.03805107	0.00158787	0.0013615
5	50	0.03621019	0.00131828	7.13135x10^{-5}
	100	0.03604356	0.00131214	2.82921x10^{-5}
	10	0.01033848	0.00012975	0.0013599
10	50	0.00987204	9.28303x10^{-5}	4.03085x10^{-6}
	100	0.00958387	9.16224x10^{-5}	1.03983x10^{-6}

Based on table 1 it can be seen that the bias value of the EB estimator gets smaller with the greater number of areas. Likewise, MSE and MSE bootstrap values are also getting smaller with the increasing number of areas. In general, the bootstrap method produces a smaller MSE value compared to MSE. Thus, the bootstrap method is able to correct the bias in estimation of the EB method.

4. Conclusions

From the results of this research, the estimated parameter value of Beta-Bernoulli model using the method of moment is obtained, namely $\hat{\alpha} = \frac{\bar{y_i}}{(1-\bar{y_i})}$. The Bootstrap Mean Square Error (MSE) estimates show a decreasing trend and approaching zero when the number of areas and the range between areas increase. Thus, it is said that the Bootstrap method can be used to correct bias.

References

[1] Rao J N 2003 *Small Area Estimation* (New York: John Willey and Sons)

[2] Bolstad W M 2007 *Introduction to Bayesian Statistics Second Edition* (USA: A John Wiley & Sons Inc)

[3] Efron B and Tibshirani R J 1993 *An Introduction to the Bootstrap* (New York: Chapman and Hall)

[4] Martinez E Z, Achcar JA and Aragon D C 2015 Paramater estimation of the beta-binomial distribution: Anapllication using the SAS software *Cienciae Natural* **37** 4 pp 12-19