Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Asymptotically Unbiased, Efficient, and Consistent Properties of the Bayes estimator in the Binomial Distribution with Prior Beta

Asymptotically Unbiased, Efficient, and Consistent Properties of the Bayes estimator in the Binomial Distribution with Prior Beta

E G Simanjuntak[1], Widiarti[1,a], D Kurniasari[1], and Amanto[1]

[1]Departement of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia, Jl. Soemantri Brojonegoro No. 1, Bandar Lampung, Indonesia.

email: widiarti08@gmail.com[a]

Abstract. This study will examine the characteristics of the Bayes estimator in the Binomial distribution with prior Beta theoretically and empirically. The theoretical result shows that the Bayes estimator in this distribution is an asymptotically unbiased and consistent, but inefficient estimator. Meanwhile, empirically, Bayes's estimator is an unbiased estimator, efficient, and consistent.

Keyword: Bayes estimator, unbiased, efficient, consistent, binomial distribution.

1. Introduction

The estimation method is a statistical tool that is useful for estimating the value of population parameters based on sample data. Basically, the estimation is divided into two, namely point estimation and interval estimation [1]. There are two methods to estimate the point, namely the classical method and the Bayes method. In the classical method, conclusions are based on information from a random sample drawn from the population. Meanwhile, the Bayes method uses or combines subjective knowledge about the probability distribution of the unknown parameter with the information obtained from the sample data. Subjective knowledge about the probability distribution of the unknown parameter is an initial distribution that provides information about a parameter called a prior. After the observations are made, the information in the prior distribution is combined with the information with the sample data through the Bayes theorem, and the results are expressed in a distribution called the posterior distribution which then becomes the basis for inference in the Bayes method [2]. In the Bayes method, all parameters in the model are treated as variables while in the classical method the parameters are considered as constants. So that if a case occurs, that is, in different situations and places of observation, the parameters change, then with the Bayes principle this problem can be solved [3]. The Bayes method has been developed to estimate small area parameters such as the Empirical Bayes (EB), Empirical Best Linear Unbiased Prediction (EBLUP), Hierarchical Bayes (HB), and Spatial EBLUP. Several studies related to the application of this method are Widiarti et.al. [4] using Spatial EBLUP method for estimating per capita expenditure in Lampung Province, Zou et.al. [5] using EB method to be applied to highway safety, Pusponegoro et.al. [6] using Spatial EBLUP method for estimating poverty, Najera [7] using HB method for estimating stunting, Buil-Gil [8] using Spatial EBLUP method for estimating perceived neighbourhood disorder, and Li [9] using EB method for estimate road safety.

This study will examine the characteristics of the Bayes estimator such as unbiased, minimum variance (efficiency), and consistency of the Binomial distribution with the prior Beta. The empirical study was carried out through a simulation using RStudio-1.1.463.

2. Method

2.1. Research Data
The research data were simulation data generated by RStudio-1.1.463. Data generated using the Binomial distribution (m = 5, and m = 10) with prior Beta. The sample sizes selected were 50, 100, 500, and 1000. These various sample sizes were used to prove whether the sample size had an effect on the characteristics studied. The simulation was carried out by selecting 3 pairs (α, β) which had different variance, namely (1, 3), (10, 6), and (20, 24) with each of the variations, namely 0.0375, 0.0138, 0, 0054. The pairs (α, β) do not have a significant difference in variance because the variance of the beta distribution is not more than 1 and 0 <x <1.

2.2. Bayes Method
In the classical approach, the parameter θ is an unknown fixed quantity. Whereas in the Bayesian approach, θ is seen as a quantity whose variation is described by the probability distribution (called the prior distribution). It is a subjective distribution, based on a person's beliefs and formulated before data is retrieved. Then, the sample is taken from the population indexed θ and the prior distribution is adjusted according to this sample information. The adjusted prior is called the posterior distribution. This adjustment is made using the Bayes rule [10].

2.3. Estimator Characteristics
An estimator is called best estimator if it has the following characteristics:

1. Unbiased Estimator

   An estimator is unbiased if the mean of its sampling distribution is the true parameter value. That is, an estimator $\hat{p}$ is unbiased if and only if
   $$E(\hat{p}) = \int \hat{p}\, f(\hat{p}|p)\, d\hat{p} = p$$
   where $f(\hat{p}|p)$ is the sampling distribution of the estimator $\hat{p}$ given the parameter p [11].

2. Efficient Estimator

   Let $\hat{p}$ be an unbiased estimator of a parameter p in the case of point estimation. $\hat{p}$ is called an efficient estimator of p if and only if the variance of $\hat{p}$ attains the Rao-Cramér lower bound. Let $X_1, X_2, \dots, X_n$ be iid with common pdf $f(x;p)$ for $p \in \Omega$, and $\hat{p} = u(X_1, X_2, \dots, X_n)$ be a statistic with mean $E(\hat{p}) = E[u(X_1, X_2, \dots, X_n)] = k(p)$, then
   $$Var(\hat{p}) \geq \frac{[k'(p)]^2}{n\, I(p)}$$
   If $\hat{p} = u(X_1, X_2, \dots, X_n)$ is an unbiased estimator of p, so that $k(p) = p$, then the Rao-Cramér inequality becomes [12]
   $$Var(\hat{p}) \geq \frac{1}{n\, I(p)}$$

3. Consistent Estimator
   An estimator is consistent if the sample size (n) is enlarged to near infinity, the estimator value will tend to approach the population parameter value [1]. An estimator $\hat{p}$ of p based on a random sample of size n is said to be consistent if for any small $\varepsilon > 0$, then

$$\lim_{n\to\infty} P(|\hat{p}-p| < \varepsilon) = 1$$
$$\lim_{n\to\infty} P(|\hat{p}-p| \geq \varepsilon) = 0$$

   The following two conditions are sufficient to define consistency [13].
   1. $\lim_{n\to\infty} E(\hat{p}) = p$
   2. $\lim_{n\to\infty} Var(\hat{p}) = 0$

3. Result and Discussion

In the Bayes estimator, the probability function $f(x_i, p)$ is expressed by the conditional probability function $f(x_i|p)$, so that for $X_1, \dots, X_n$ a random sample of a population with a Binomial distribution with parameter p can be written as follows:

$$X_i \sim Bin(m,p) \Leftrightarrow f(x_i|p) = \begin{cases} \binom{m}{x_i} p^{x_i}(1-p)^{m-x_i}, x_i = 0,1,2,\dots,n \; i = 1,2,\dots,n \\ 0 \qquad , \qquad x_i \; lainnya \end{cases}$$

The prior distribution for $X_i \sim Bin(m,p), i = 1, 2, \dots, n$ is $p \sim Beta(\alpha, \beta)$. So the probability function of the prior distribution is

$$\pi(p) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, & 0 < p < 1 \\ 0 \qquad\qquad , & p \; lainnya \end{cases}$$

The likelihood function of $X_i \sim Bin(m,p)$:

$$f(x_1, x_2, \dots, x_n|p) = \prod_{i=1}^{n} \binom{m}{x_i} p^{x_i}(1-p)^{m-x_i} = \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] p^{\sum_{i=1}^{n} x_i}(1-p)^{mn-\sum_{i=1}^{n} x_i}$$

Joint probability density function of $X_1, \dots, X_n$ and p is:

$$f(x_1, x_2, \dots, x_n, p) = \pi(p)\, f(x_1, x_2, \dots, x_n|p)$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] p^{\alpha+\sum_{i=1}^{n} x_i - 1}(1-p)^{mn+\beta-\sum_{i=1}^{n} x_i - 1}$$

Marginal function of X_i is:

$$m(x_i) = \int_{-\infty}^{\infty} f(x_i, p)\, dp$$
$$= \int_{0}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] p^{\alpha+\sum_{i=1}^{n} x_i - 1}(1-p)^{mn+\beta-\sum_{i=1}^{n} x_i - 1}\, dp$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] B\left[\left(\alpha + \sum_{i=1}^{n} x_i\right), \left(mn + \beta - \sum_{i=1}^{n} x_i\right)\right]$$

The posterior distribution can be written as follows:

$$\pi(p|x_i) = \frac{f(x_i, p)}{m(x_i)}$$

$$= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left[\prod_{i=1}^{n}\binom{m}{x_i}\right] p^{\alpha+\sum_{i=1}^{n}x_i-1}(1-p)^{mn+\beta-\sum_{i=1}^{n}x_i-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left[\prod_{i=1}^{n}\binom{m}{x_i}\right] B\left[\left(\alpha+\sum_{i=1}^{n}x_i\right),\left(mn+\beta-\sum_{i=1}^{n}x_i\right)\right]}$$

$$= \frac{p^{\alpha+\sum_{i=1}^{n}x_i-1}(1-p)^{mn+\beta-\sum_{i=1}^{n}x_i-1}}{B\left[\left(\alpha+\sum_{i=1}^{n}x_i\right),\left(mn+\beta-\sum_{i=1}^{n}x_i\right)\right]}$$

So, the Bayes estimator for p is:

$$\hat{p} = E(p|x_i) = \int_0^1 p\,\pi(p|x_i)\,dp$$

$$= \int_0^1 p\,\frac{p^{\alpha+\sum_{i=1}^{n}x_i-1}(1-p)^{mn+\beta-\sum_{i=1}^{n}x_i-1}}{B\left[\left(\alpha+\sum_{i=1}^{n}x_i\right),\left(mn+\beta-\sum_{i=1}^{n}x_i\right)\right]}\,dp$$

$$= \frac{\alpha+\sum_{i=1}^{n}x_i}{mn+\alpha+\beta}$$

3.1. Estimator Characteristics

The study of the characteristics of the Bayes estimator will be carried out theoretically and empirically. Based on the above calculations, the Bayes estimator for the p parameter in the Binomial distribution with prior Beta is $\hat{p} = (\alpha + \sum_{i=1}^{n} X_i)/(mn + \alpha + \beta)$. Furthermore, the characteristics of $\hat{p}$ will be evaluated, namely unbiased, efficient, and consistent.

1. Unbiased

The estimator $\hat{p}$ is unbiased for p if $E(\hat{p}) = p$. The following shows whether $\hat{p}$ is an unbiased estimator for p.

$$E(\hat{p}) = E\left(\frac{\sum_{i=1}^{n}X_i+\alpha}{mn+\alpha+\beta}\right)$$

$$= \frac{\alpha+mnp}{mn+\alpha+\beta}$$

Since $E(\hat{p}) \neq p$, it can be conluded that $\hat{p}$ is a biased estimator for p. However, asymptotically, $\hat{p}$ is an unbiased estimator because $\lim_{n\to\infty} E(\hat{p}) = p$.

$$\lim_{n\to\infty} E(\hat{p}) = \lim_{n\to\infty}\left(\frac{\alpha+mnp}{mn+\alpha+\beta}\right) = \lim_{n\to\infty}\left(\frac{1}{mn+\alpha+\beta}\right)\alpha + \lim_{n\to\infty}\left(\frac{mn}{mn+\alpha+\beta}\right)p = 0 + p = p$$

2. Efficient

The estimator $\hat{p}$ is said to be efficient if the variance of $\hat{p}$ reaches the lower bound *Rao-Cramer*. Next we will show the variance of Bayes estimator as follows

$$var(\hat{p}) = var\left(\frac{\sum_{i=1}^{n}X_i+\alpha}{mn+\alpha+\beta}\right)$$

Since $var(aX) = a^2 var(X)$, so

$$var(\hat{p}) = \frac{var(\sum_{i=1}^{n}X_i)}{(mn+\alpha+\beta)^2}$$

To solve for $var(\sum_{i=1}^{n}X_i)$, it is necessary to first know the distribution of $\sum_{i=1}^{n}X_i$. Let $Y = \sum_{i=1}^{n}X_i$, then the probability density function of Y will be sought.

$X_i \sim Binomial(m,p)$

Let: $Y = \sum_{i=1}^{n}X_i$

$$M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^{n}x_i})$$

$$= [M_X(t)]^n, \text{ since } X_i \sim Binomial(m,p)$$

$$= \{[(1-p)+pe^t]^m\}^n$$
$$= [(1-p)+pe^t]^{mn}$$
$$Y \sim Binomial(mn, p)$$
$$f(y) = \binom{mn}{x} p^x (1-p)^{mn-x}$$
So $var(\sum_{i=1}^{n} X_i) = mnp(1-p)$, then
$$var(\hat{p}) = \frac{var(\sum_{i=1}^{n} X_i)}{(mn+\alpha+\beta)^2} = \frac{mnp(1-p)}{(mn+\alpha+\beta)^2}$$

It can be seen that the variance of $\hat{p}$ is $\frac{mnp(1-p)}{(mn+\alpha+\beta)^2}$, then the Rao-Cramer lower bound will be sought as follows.

$$\ln f(x|p) = \ln \binom{m}{x} + x \ln p + (m-x)\ln(1-p)$$
$$\frac{\partial \ln f(x|p)}{\partial p} = \frac{x}{p} - \frac{m-x}{1-p}$$
$$\left(\frac{\partial \ln f(x|p)}{\partial p}\right)^2 = \left(\frac{x}{p} - \frac{m-x}{1-p}\right)^2$$
$$= \frac{(x-mp)^2}{p^2(1-p)^2}$$
$$var(X) = E\big(X - E(X)\big)^2 = E(X-mp)^2 = mp(1-p)$$
$$I(p) = E\left(\frac{\partial \ln f(x|p)}{\partial p}\right)^2$$
$$= E\left(\frac{(x-mp)^2}{p^2(1-p)^2}\right)$$
$$= \frac{var(X)}{p^2(1-p)^2}$$
$$= \frac{mp(1-p)}{p^2(1-p)^2}$$
$$= \frac{m}{p(1-p)}$$

$$HB = \frac{1}{n\,I(p)} = \frac{1}{n\,\frac{m}{p(1-p)}} = \frac{p(1-p)}{mn}$$

Since the variance of $\hat{p}$ doesn't reach the lower bound Rao-Cramer, it can be concluded that $\hat{p}$ is an inefficient estimator.

3. Consistent

The estimator $\hat{p}$ is consistent if $\lim_{n\to\infty} var(\hat{p}) = 0$. The following shows whether $\hat{p}$ is a consistent estimator.

$$\lim_{n\to\infty} var(\hat{p}) = \lim_{n\to\infty} \frac{mnp(1-p)}{(mn+\alpha+\beta)^2}$$
$$= \lim_{n\to\infty} \frac{mnp - mnp^2}{m^2n^2 + 2mn\alpha + 2mn\beta + 2\alpha\beta + \alpha^2 + \beta^2}$$
$$= \lim_{n\to\infty} \frac{\frac{mnp - mnp^2}{n^2}}{\frac{m^2n^2}{n^2} + \frac{2mn\alpha}{n^2} + \frac{2mn\beta}{n^2} + \frac{2\alpha\beta}{n^2} + \frac{\alpha^2}{n^2} + \frac{\beta^2}{n^2}}$$

$$= \frac{0}{m^2} = 0$$

Since $\lim_{n\to\infty} var(\hat{p}) = 0$, it can be concluded that $\hat{p}$ is a consistent estimator.

3.2. Simulation

Based on 3 pairs (α, β), sample size (n), and number of experiments (m), it will be seen the effect of (α, β), n, and m on bias, variance and MSE value. Based on Table 1 and Figure 1, it can be seen that when 5 independent trials are carried out with $(\alpha, \beta) = (1,3), (10,6)$ will produce a bias value that is getting smaller / closer to 0 when the sample size is enlarged. Meanwhile, when $(\alpha, \beta) = (20, 24)$ will produce in a fluctuating bias value. When conducted 10 independent trials, the three pairs (α, β) will produce a fluctuating bias value. So it can be concluded that the Bayesian estimator is asymptotically unbiased when it is carried out 5 independent trials with (α, β)= (1, 3) and (10, 6), and is biased for the others.

Table 1. The bias value of the Bayes estimator.

α	β	n	Bias	
			m=5	m=10
1	3	50	0,001014	0,000138
		100	0,000942	0,000554
		500	0,00028	3,3E-05
		1000	4,6E-05	7,61E-05
10	6	50	0,001154	0,00017
		100	0,001092	0,000172
		500	0,000309	0,000488
		1000	0,000123	0,000224
20	24	50	2,84E-06	0,001516
		100	0,000318	5,46E-06
		500	0,000618	5,92E-05
		1000	0,000262	9,04E-05

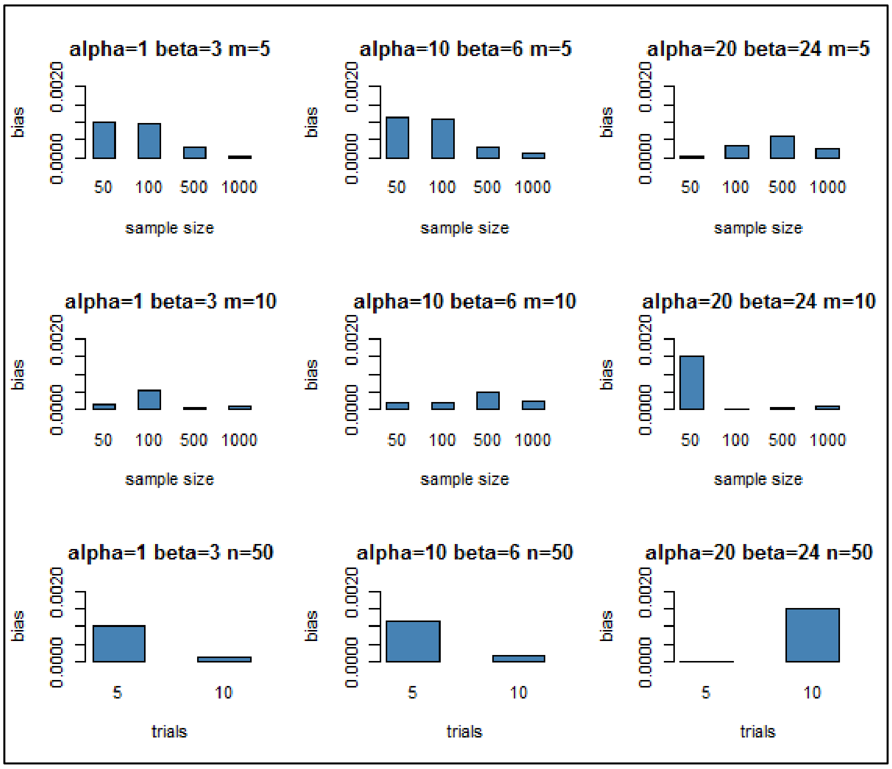

Figure 1. Histogram of the bias value on several sample sizes and parameter values.

Based on Table 2 and Figure 2, for all pairs (α, β) will result in a smaller variance value when the sample size and trials is enlarged. When the pairs (α, β) fluctuate, the variance values tend not to change significantly. So it can be concluded that the Bayesian estimator is an efficient estimator.

Table 2. Variance of the Bayes estimator.

α	β	n	Variance	
			m=5	m=10
1	3	50	0,000774	0,000328
		100	0,000356	0,000194
		500	7,37E-05	3,74E-05
		1000	4,11E-05	1,9E-05
10	6	50	0,000864	0,000429
		100	0,000473	0,000232
		500	9,58E-05	4,67E-05
		1000	4,66E-05	2,45E-05
20	24	50	0,000711	0,000426
		100	0,00041	0,000222
		500	9,27E-05	4,99E-05
		1000	4,82E-05	2,39E-05

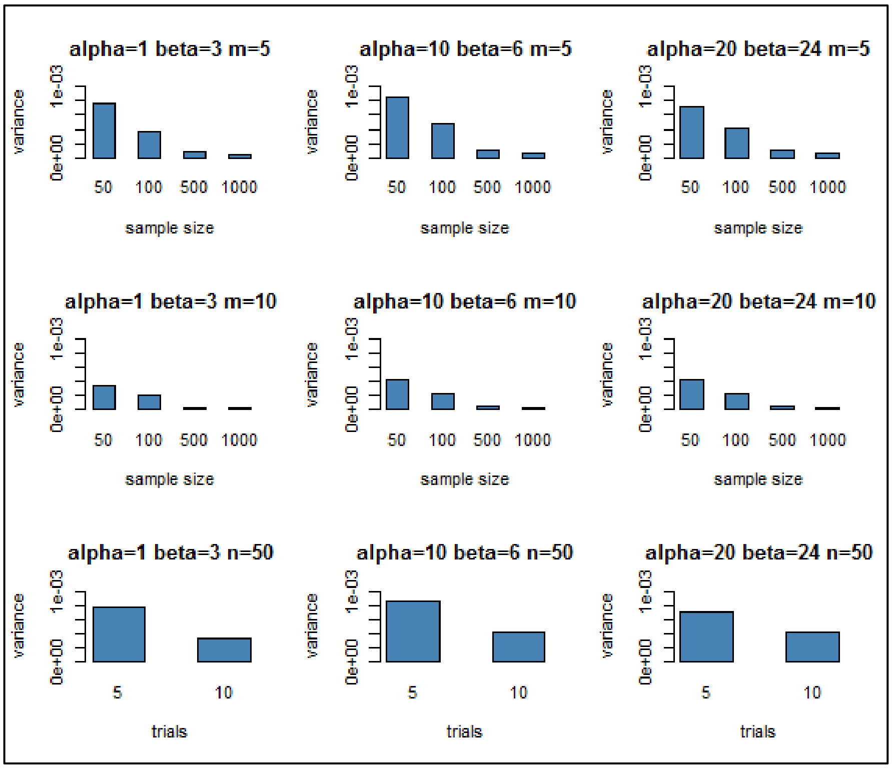

Figure 2. Histogram of Bayes estimator's variance on several sample sizes and parameter values.

Based on Table 3 and Figure 3, for all pairs (α, β) the MSE value will be smaller when the sample size and trials is enlarged. When the pairs (α, β) fluctuate, the MSE values tend not to change significantly. So it can be concluded that Bayes estimator is a consistent estimator.

Table 3. MSE value of Bayes estimator.

α	β	n	MSE	
			m=5	m=10
1	3	50	0,000775	0,000328
		100	0,000357	0,000195
		500	7,38E-05	3,74E-05
		1000	4,11E-05	1,9E-05
10	6	50	0,000865	0,000429
		100	0,000474	0,000232
		500	9,59E-05	4,7E-05
		1000	4,66E-05	2,46E-05
20	24	50	0,000711	0,000428
		100	0,00041	0,000222
		500	9,31E-05	4,99E-05
		1000	4,83E-05	2,39E-05

IOP Publishing
1751 (2021) 012019 doi:10.1088/1742-6596/1751/1/012019

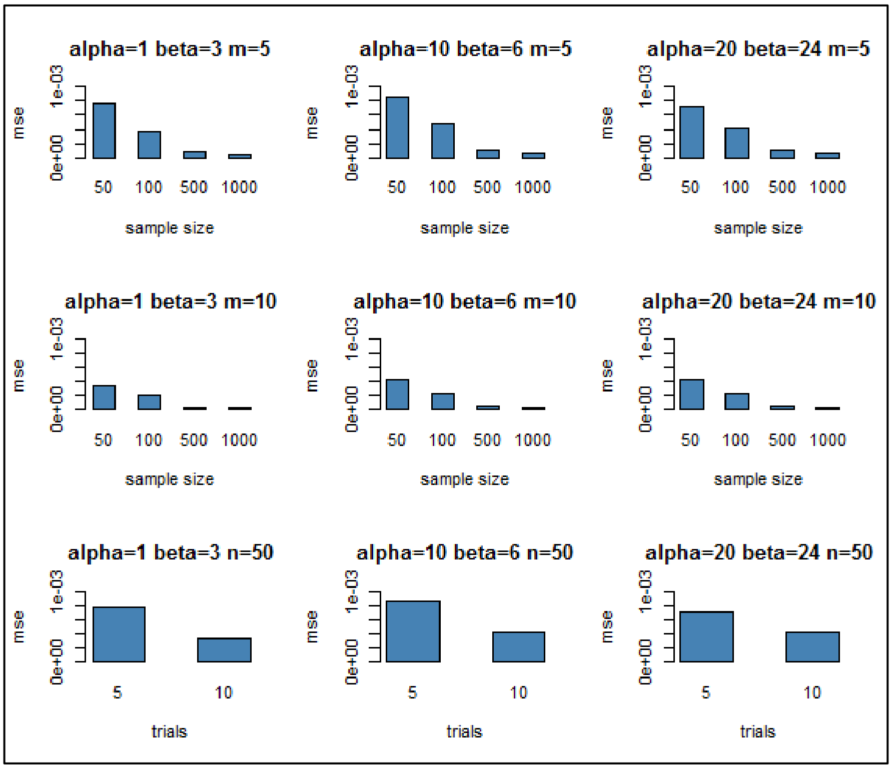

Figure 3. Histogram of MSE value Bayes estimator on several sample sizes and parameter values.

4. Conclusion

The result of this study show that the Bayes estimator in the Binomial distribution with prior Beta is $\hat{p} = \frac{\alpha+\sum_{i=1}^{n} X_i}{mn+\alpha+\beta}$ which theoretically has asymptotically unbiased and consistent, but not efficient. But empirically, Bayes estimator is asymptotically unbiased, efficient, and consistent estimator.

Acknowledgments

The authors thank to Universitas Lampung for financial support for this study through Basic Research Grant 2020.

References

[1] Asra A and Rudiansyah 2014 *Statistika Terapan* 2nd Edition (In Media, Indonesia)
[2] Casella G and Berger R L 2002 *Statistical Inference* 2nd Edition (Duxbury, United States of America)
[3] Kismiantini and Lestari H P Seminar Nasional IV, Universitas Negeri Semarang
[4] Widiarti, D D Oktafiani, M Usman, and D Kurniasari 2020 *J. Phys: Conf. Ser.* **1567** 022082
[5] Yajie Zou, John E. Ash, Byung-Jung Park, Dominique Lord and Lingtao Wu 2018 *Journal of Applied Statistics* 45:9, 1652-1669, DOI: 10.1080/02664763.2017.1389863
[6] Pusponegoro N H and Rachmawati N R 2018 *Procedia Computer Science* **135** 712
[7] H E C Najera 2019 *Spatial and Spatio-temporal Epidemiology* **29** 1-11
[8] Buil-Gil D, Medina J, and Shlomo N 2019 *Applied Geography* **109** 102037
[9] Li H, Graham D J, Ding H, and Ren G2019 *Accident Analysis & Prevention* **129** 148.
[10] Subanar 2013 *Statistika Matematika* (Graha Ilmu, Yogyakarta)
[11] Bolstad W M 2007 *Introduction to Bayesian Statistics* 2nd Edition (John Wiley & Sons United States of America)

[12] Hogg R V, McKean J W and Craig A T 2019 *Introduction to Mathematical Statistics* 8th Edition (Pearson, United States of America)
[13] Hsu H P 2011 *Schaum's Outline's of Probability, Random Variables, and Random Processes* 2nd Edition (Mc Graw Hill, United States of America)