ISBN No. 978-602-98559-3-7

Parameter Estimation of Bernoulli Distribution using Maximum Likelihood and Bayesian Methods

Nurmaita Hamsyiah[1)], Khoirin Nisa[1)], &Warsono[1)]
[1)]Departmentof Mathematics, Faculty of Mathematics and Science, University of Lampung
Jl. Prof. Dr. Sumantri Brojonegoro No. 1 Bandar Lampung
Phone Number +62 721 701609 Fax +62 721 702767
E-mail: itamath98@gmail.com

ABSTRACT
The term parameter estimation refers to the process of using sample data to estimate the parameters of the selected distribution.There are several methodsthat can be used to estimate distributionparameter(s).In this paper,the maximum likelihood andBayesian methodsare usedfor estimating parameter ofBernoulli distribution, i.e. θ*, which isdefined asthe probability of success event for two possible outcomes.The maximum likelihood and Bayesian estimators of Bernoulli parameter are derived,for the Bayesian estimator the Beta prior is used. The analytical calculation shows that maximum likelihood estimator is unbiased while Bayesian estimator is asymptotically unbiased. However, empirical analysis by Monte Carlo simulation shows that the mean square errors (MSE) of the Bayesian estimatorare smaller than maximum likelihood estimator for large sample sizes.*

Keywords*: Bernoulli distribution, beta distribution, conjugate prior, parameter estimation.*

1. PENDAHULUAN

Parameter estimation is a way to predict the characteristics of a population based on the sample taken. In general, parameter estimation is classified into two types, namely point estimation and interval estimation. The point estimation of a parameter is a value obtained from the sample and is used as a parameter estimator whose value is unknown.

Several point estimation methods are used to calculate the estimator, such as moment method, maximum likelihood method, and Bayesian method. The moment method predicts the parameters by equating the values of sample moments to the population moment and solving the resulting equation system [1]. The maximum likelihood (ML)methoduses differential calculus to determine the maximum of the likelihoodfunction to obtain theparameters estimates. The Bayesianmethod differs from the traditional methods by introducing a frequency function for the parameter being estimated namely prior distribution. The Bayesian method combines the prior distribution and sample distribution. The prior distribution is the initial distribution that provides information about the parameters. The sample distribution combined with the prior distribution provides a new distribution i.e. the posterior distribution that expresses a degree of confidence regarding the location of the parameters after the sample is observed [2].

Researches on parameter estimation using various methods of various distributions have been done, for example: Bayesian estimation of exponential distribution [3], [4], ML and Bayesianestimations of Poisson distribution [5], Bayesianestimation of Poisson-Exponential distribution [6], and Bayesianestimation of Rayleigh distribution [7].

The difference between the ML and the Bayesian methods is that the ML method considers that the parameter is an unknown quantity of fixed value and the inference is based only on the information in the sample; while the Bayesian method considers the parameter as a variable that describes the initial knowledge of the parameters before the observation is performed and expressed in a distribution called the prior distribution. After the observation is performed, the information in the prior distribution is combined with the sample data information through Bayesian theorem, and the result is expressed in a distribution form called the posterior distribution, which further becomes the basis for inference in the Bayesian method [8].

The Bayesian method has advantages over other methods, one of which is the Bayesian method can be used for drawing conclusions in complicated or extreme cases that cannot be handled by other methods, such as in complex hierarchical models. In addition, if the prior information does not indicate complete and clear information about the distribution of the prior, appropriate assumptions may be given to its distribution characteristics. Thus, if the prior distribution can be determined, then a posterior distribution can be obtained which may require mathematical computation [8].

This paper examines the parameter estimation of Bernoulli distribution using ML and Bayesianmethods. A review of Bernoulli distribution and Beta distribution is presented in Section 2. The research methodologyis described in Section 3. Section 4 provides the results and discussion. Finally, the conclusion is given in Section 5.

2. THEORETICAL FRAMEWORK

2.1 Bernoullli Distribution

Bernoulli distribution was introduced by Swiss mathematician,Jacob Bernoulli (1654-1705). It is the probability distribution resulting from two outcomes or events in a given experiment, i.e. success ($X = 1$) and fail ($X = 0$), with the probability of the success is θ and the probability of failure is $1 - \theta$.

Definition

A random variable $X$is called a Bernoulli random variable (or X is Bernoulli distributed) if and only if its probability distribution is given by

$$f(x;\theta) = \theta^x(1-\theta)^{1-x}, \text{ for} x = 0,1.$$

Proposition 1

Bernoulli distribution$f(x;\theta)$ has mean and variance as follows:

$$\mu = \theta \text{and} \sigma^2 = \theta(1-\theta).$$

Proof :

The mean of Bernoulli random variable X is

$$\begin{aligned}\mu &= E(X)\\ &= \sum_{x=0}^{1} x f(x;\theta)\\ &= \sum_{x=0}^{x=1} x\theta^{x}(1-\theta)^{1-x}\\ &= 0\cdot\theta(1-\theta)^{1-0} + 1\cdot(1-\theta)^{1-1} = \theta.\end{aligned}$$

The variance, i.e. $\sigma^2 = E(X-\mu)^2 = E(X^2) - [E(X)]^2$, of Bernoulli distribution is obtained as follows:

$$\begin{aligned}E(X^2) &= \sum_{x=0}^{1} x^2 f(x;\theta)\\ &= \sum_{x=0}^{1} x^2 f(x;\theta)\\ &= \sum_{x=0}^{1} x^2\theta(1-\theta)^{1-x}\\ &= 0^2\cdot\theta^0(1-\theta)^{1-0} + 1^2\cdot(1-\theta)^{1-1} = \theta.\end{aligned}$$

Then,

$$\sigma^2 = E(X-\mu)^2 = \theta - \theta^2 = \theta(1-\theta).$$

2.2. Beta Distribution

Definition

A random variable X is called a betarandom variable with parameters*a* and b if the density function of X is given by

$$f(x) = \begin{cases}\dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, & 0 < x < 1\\ 0 & , lainnya\end{cases}$$

where$B(a,b)$is betafunction defined as

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx\ ;\ a > 0, b > 0. \quad (1)$$

Proposition 2

The beta function and gamma function is connected by

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \quad (2)$$

Proof :

$$\Gamma(a)\Gamma(b) = \int_{x=0}^{\infty} x^{a-1}e^{-x}dx \cdot \int_{y=0}^{\infty} y^{b-1}e^{-y}dy$$

$$= \int_{y=0}^{\infty} \int_{x=0}^{\infty} x^{a-1} y^{b-1} e^{-x-y} \, dxdy.$$

Let$x = f(z,t) = zt$and $y = g(z,t) = z(1-t)$,

$$\mathbb{\Gamma}(a)\mathbb{\Gamma}(b) = \int_{z=0}^{\infty} \int_{t=0}^{1} (zt)^{a-1} [z(1-t)]^{b-1} e^{-z} \, |J(z,t)| dtdz$$

$$= \int_{z=0}^{\infty} \int_{t=0}^{1} (zt)^{a-1} [z(1-t)]^{b-1} e^{-z} \, zdtdz$$

$$= \int_{z=0}^{\infty} \int_{t=0}^{1} z^{a-1+b-1+1} e^{-z} t^{a-1} (1-t)^{b-1} \, dtdz$$

$$= \int_{z=0}^{\infty} z^{a+b-1} e^{-z} dz \cdot \int_{t=0}^{1} t^{a-1} (1-t)^{b-1} dt$$

$$= \mathbb{\Gamma}(a+b) B(a,b).$$

Then,

$$B(a,b) = \frac{\mathbb{\Gamma}(a)\mathbb{\Gamma}(b)}{\mathbb{\Gamma}(a+b)}.$$

Proposisi 3

The mean and variance of beta distribution with parameters$a$and$b$are

$$\mu = \frac{a}{a+b} \qquad \text{and} \qquad \sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}.$$

Proof :

The proposition can be proved by using the moment ofbeta distribution as follows:

$$E(X^n) = \frac{1}{B(a,b)} \int_0^1 x^n x^{a-1} (1-x)^{b-1} dx$$

$$= \frac{1}{B(a,b)} \int_0^1 x^{(a+n)-1} (1-x)^{b-1} dx.$$

From equations (1) and (2) we obtain

$$E(X^n) = \frac{B(a+n,b)}{B(a,b)}$$

$$= \frac{\frac{\mathbb{\Gamma}(a+n)\mathbb{\Gamma}(b)}{\mathbb{\Gamma}(a+b+n)}}{\frac{\mathbb{\Gamma}(a)\mathbb{\Gamma}(b)}{\mathbb{\Gamma}(a+b)}}$$

$$= \frac{\Gamma(a+n)\Gamma(b)}{\Gamma(a+b+n)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$= \frac{\Gamma(a+n)\Gamma(a+b)}{\Gamma(a+b+n))\Gamma(a)}. \qquad (3)$$

Thus the mean and variance of beta distribution will be obtained by substituting $n = 1$and$n = 2$to equation (3), then

$$Mean(X) = E(X^1) = \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a+b+1))\Gamma(a)}$$

$$= \frac{a\Gamma(a)\Gamma(a+b)}{(a+b)\Gamma(a+b))\Gamma(a)}$$

$$= \frac{a}{a+b}$$

and $Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$.

Since

$$E(X^2) = \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a+b+2))\Gamma(a)}$$

$$= \frac{(a+1)\Gamma(a+1)\Gamma(a+b)}{(a+b+1)\Gamma(a+b+1)\Gamma(a)}$$

$$= \frac{(a+1)a\Gamma(a)\Gamma(a+b)}{(a+b+1)(a+b)\Gamma(a+b)\Gamma(a)}$$

$$= \frac{(a+1)a}{(a+b+1)(a+b)},$$

then

$$Var(X) = \frac{(a+1)a}{(a+b+1)(a+b)} - \left(\frac{a}{a+b}\right)^2$$

$$= \frac{(a+1)a}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2}$$

$$= \frac{(a+b)(a^2+a) - a^2(a+b+1)}{(a+b)^2(a+b+1)}$$

$$= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2b - a^2}{(a+b)^2(a+b+1)}$$

$$= \frac{ab}{(a+b)^2(a+b+1)}.$$

3. RESEARCH METHOD

The research method for estimating the parameter of Bernoulli distribution in this paper can be described as follows. For ML estimation, the parameter estimation is done by differentiatingpartially the log of the likelihood function and equation it by zero,

$$\frac{\partial lnL(\theta)}{\partial\theta} = 0$$

to obtain ML estimator($\hat{\theta}_{ML}$). The second derivation assessment is performed to show that the resulted$\hat{\theta}$ truly maximize the likelihood function. For the Bayesian method, the parameter estimation is done through the following steps:

1. Form the likelihood function of Bernoulli distribution as follows:

$$L(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^{n} f((x_i)|\theta).$$

2. Calculate the joint probability distribution, which is obtained by multiplying the likelihood function and the prior distribution,

$$H(x_1, x_2, \cdots, x_n; \theta) = L(x_1, x_2, \cdots, x_n|\theta) \cdot \pi(\theta).$$

3. Calculate the marginal probability distribution function,

$$p(x_1, x_2, \cdots, x_n) = \int H(x_1, x_2, \cdots, x_n; \theta) d\theta.$$

4. Calculatethe posterior distributionby dividing the joint probability distribution function by the marginal function,

$$\pi(\theta|x_1, x_2, \cdots, x_n) = \frac{H(x_1, x_2, \cdots, x_n; \theta)}{p(x_1, x_2, \cdots, x_n)}.$$

The Bayesian parameter estimate of $\theta$is then produced as the mean of the posterior distribution.

After the parameter estimate of $\theta$is obtained by MLE and Bayesian methods, the evaluation of the estimators is performed by assessing their bias, variance, and mean square error.

4. RESULT AND DISCUSSION

4.1. The ML Estimator of the Bernoulli Distribution Parameter (θ)

Let $X_1, X_2, \dots, X_n$ be Bernoulli distributed random sample with $X_i \sim Bernoulli(\theta)$, where$\theta \in \Omega = (0,1)$. The probability function of X_i is

$$f(x_i; \theta) = \theta^{x_i}(1-\theta)^{1-x_i} \text{with} x_i \in \{0,1\}.$$

The likelihood function of Bernoulli distribution is given by

$$\begin{aligned} L(\theta) &= f(x_1, x_2, \cdots, x_n; \theta) \\ &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \end{aligned}$$

$$= \theta^{x_1}(1-\theta)^{1-x_1} \cdot \theta^{x_2}(1-\theta)^{1-x_2} \cdots \theta^{x_n}(1-\theta)^{1-x_n}$$
$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}. \tag{4}$$

The natural logarithm of the likelihood function is then

$$lnL(\theta) = \ln\left[\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}\right]$$
$$= ln\theta^{\sum_{i=1}^{n} x_i} + ln(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$
$$= \sum_{i=1}^{n} x_i \, ln\theta + \left(n - \sum_{i=1}^{n} x_i\right) ln(1-\theta). \tag{5}$$

The ML estimate value of $\theta$is obtained by differentiating equation (5)with respect toθandequating the differential result to zero, i.e.

$$\frac{\partial}{\partial\theta} lnL(\theta) = \frac{\partial}{\partial\theta}\left[\sum_{i=1}^{n} x_i \, ln\theta + \left(n - \sum_{i=1}^{n} x_i\right) ln(1-\theta)\right] = 0$$
$$\frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} = 0$$
$$(1-\theta)\sum_{i=1}^{n} x_i - \theta\left(n - \sum_{i=1}^{n} x_i\right) = 0$$
$$\sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i - n\theta + \theta\sum_{i=1}^{n} x_i = 0$$
$$\sum_{i=1}^{n} x_i = n\theta,$$

then we obtain

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

To show that $\hat{\theta}$ is the value that maximizes the likelihood function$L(\theta)$, it must be confirmed that the second derivative of the likelihood function for $\theta = \hat{\theta}$ is negative:

$$\frac{\partial^2}{\partial\theta^2} lnL(\theta) = \frac{\partial^2}{\partial\theta^2}\left[\sum_{i=1}^{n} x_i \, ln\theta + \left(n - \sum_{i=1}^{n} x_i\right) ln(1-\theta)\right]$$
$$= \frac{-\sum_{i=1}^{n} x_i}{\theta^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-\theta)^2}$$
$$= \frac{-(1-\theta)^2\sum_{i=1}^{n} x_i - \theta^2\left(n - \sum_{i=1}^{n} x_i\right)}{\theta^2(1-\theta)^2}$$
$$= \frac{-\sum_{i=1}^{n} x_i + 2\theta\sum_{i=1}^{n} x_i - \theta^2\sum_{i=1}^{n} x_i - n\theta^2 + \theta^2\sum_{i=1}^{n} x_i}{\theta^2(1-\theta)^2}$$

$$= \frac{-n\theta^2 + 2\theta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i}{\theta^2(1-\theta)^2} < 0.$$

Since $\hat{\theta}$ maximizes the likelihood function, we conclude that the ML estimator of θ is given by

$$\hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

4.2. The Bayesian Estimator of the Bernoulli Distribution Parameter($\boldsymbol{\theta}$)

To estimate θ using Bayesian method, it is necessary to choose the initial information of a parameter called the prior distribution, denoted by $\pi(\theta)$, to be applied to the basis of the method namely the conditional probability. In this paper, the prior selection for Bernoulli distribution refers to the formation of its likelihood function. From equation (4) we have

$$\pi(\theta) \propto \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{1-\sum_{i=1}^{n} x_i}.$$

A distribution having probability function in the same form as the above expression is the beta distribution with density function

$$f(\theta; a, b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}, 0 < \theta < 1$$

where $a - 1 = \sum_{i=1}^{n} x_i$, $b - 1 = n - \sum_{i=1}^{n} x_i$, and $\frac{1}{B(a,b)}$ are factors requiredfor the density function tobe satisfied. The prior distributionis combinedwith the sample distribution to produce a new distribution calledposterior distributionand denoted by $\pi(\theta|x_1, x_2, \cdots, x_n)$. Posterior distributionis obtained by dividing the joint density distribution by the marginal distribution.

Joint probability density function of $(x_1, x_2, \cdots, x_n)$ is given by:

$$H(x_1, x_2, \cdots, x_n; \theta) = L(x_1, x_2, \cdots, x_n|\theta) \cdot \pi(\theta)$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \cdot \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$= \frac{1}{B(a,b)}\theta^{a+\sum_{i=1}^{n} x_i - 1}(1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1} \qquad (6)$$

and the marginal function of $(x_1, x_2, \cdots, x_n)$ is obtained as follows:

$$p(x_1, x_2, \cdots, x_n) = \int_0^1 H(x_1, x_2, \cdots, x_n; \theta)d\theta.$$

Using equation (6) we have

$$p(x_1, x_2, \cdots, x_n) = \int_o^1 \frac{1}{B(a,b)}\theta^{a+\sum_{i=1}^{n} x_i - 1}(1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1}d\theta$$

$$= \frac{1}{B(a,b)} \int_{o}^{1} \theta^{a+\sum_{i=1}^{n} x_i - 1} (1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1} d\theta$$

$$= \frac{1}{B(a,b)} B\left(a + \sum_{i=1}^{n} x_i, b + n - \sum_{i=1}^{n} x_i\right). \quad (7)$$

Then from equation (6) and (7) the posterior distribution can be written as follows:

$$\pi(\theta | x_1, x_2, \cdots, x_n) = \frac{H(x_1, x_2, \cdots, x_n; \theta)}{p(x_1, x_2, \cdots, x_n)}$$

$$= \frac{\frac{1}{B(a,b)} \theta^{a+\sum_{i=1}^{n} x_i - 1} (1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1}}{\frac{1}{B(a,b)} B\left(a + \sum_{i=1}^{n} x_i, b + n - \sum_{i=1}^{n} x_i\right)}$$

$$= \frac{\theta^{a+\sum_{i=1}^{n} x_i - 1} (1-\theta)^{b+n-\sum_{i=1}^{n} x_i - 1}}{B\left(a+\sum_{i=1}^{n} x_i, b+n-\sum_{i=1}^{n} x_i\right)}. \quad (8)$$

The posterior distribution expressed in equation (8) is obviously following beta distribution also with parameter $(a + \sum_{i=1}^{n} x_i)$and$(b + n - \sum_{i=1}^{n} x_i)$, or

$$\hat{\theta} \sim Beta\left(a + \sum_{i=1}^{n} x_i, b + n - \sum_{i=1}^{n} x_i\right).$$

Since the prior and posterior distribution of Bernoulli follows the same distribution, i.e. the Beta distribution, beta distribution is called as the conjugate prior of the Bernoulli distribution. The posterior mean is used as the parameter estimate θ in Bayesian method. Using Proposition 2, the Bayesian estimator of parameter θ is obtained as follows:

$$\hat{\theta}_B = \frac{a + \sum_{i=1}^{n} x_i}{a + \sum_{i=1}^{n} x_i + b + n - \sum_{i=1}^{n} x_i}$$

$$= \frac{a + \sum_{i=1}^{n} x_i}{a + b + n}.$$

4.3. Evaluation of the Estimators Properties

The parameter estimationof the Bernoulli distribution isobtained by the MLE and Bayesian methods yields different estimates. The best estimator has to meet the following properties:

1. Unbiased

An estimator is called to be unbiased if its expected values is equal to the estimated parameter, i.e.$\hat{\theta}$is an unbiased estimator of θif$E(\hat{\theta}) = \theta$.The bias of an estimator is then given by:

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta. \quad (9)$$

Let $X_1, X_2, \ldots, X_n$are Bernoulli(θ) random sample observations. Since $\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$is the ML estimator of θ, its expected value is as follows:

$$E(\hat{\theta}_{ML}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$
$$= \frac{1}{n}E\left(\sum_{i=1}^{n} x_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$
$$= \frac{1}{n}\cdot n\theta = \theta. \quad (10)$$

Since $E(\hat{\theta}_{ML}) = \theta$,$\hat{\theta}_{MLE}$ is an unbiased estimator of θ.

Now consider the Bayesian estimator of θi.e. $\hat{\theta}_B = \frac{a+\sum_{i=1}^{n} x_i}{a+b+n}$. The expexted value of Bayesian estimator is given by

$$E(\hat{\theta}_B) = E\left(\frac{a+\sum_{i=1}^{n} x_i}{a+b+n}\right)$$
$$= \frac{1}{a+b+n}E\left(a+\sum_{i=1}^{n} x_i\right)$$
$$= \frac{1}{a+b+n}\left[E(a)+E\left(\sum_{i=1}^{n} x_i\right)\right]$$
$$= \frac{1}{a+b+n}\left[E(a)+\sum_{i=1}^{n} E(x_i)\right]$$
$$= \frac{1}{a+b+n}(a+n\theta). \quad (11)$$

Since $E(\hat{\theta}_B) \neq \theta$,$\hat{\theta}_B$ is an biased estimator of θ. The bias value of $\hat{\theta}_B$is:

$$Bias(\hat{\theta}_B) = E(\hat{\theta}_B) - \theta$$
$$= \frac{a+n\theta}{a+b+n} - \theta. \quad (12)$$

Although $\hat{\theta}_B$is an biased estimator of θ, it can be shown that $\hat{\theta}_B$is asymptotically unbiased.The proof is given as follows:

$$\lim_{n\to\infty} E(\hat{\theta}_B) = \lim_{n\to\infty}\frac{a+n\theta}{a+b+n}$$
$$= \lim_{n\to\infty}\frac{\frac{a}{n}+\frac{np}{n}}{\frac{a}{n}+\frac{b}{n}+\frac{n}{n}}$$
$$= \lim_{n\to\infty}\frac{\frac{a}{n}+p}{\frac{a}{n}+\frac{b}{n}+1}$$

$$= \frac{\theta}{1} = \theta. \qquad (13)$$

Since $\lim_{n\to\infty} E(\hat{\theta}_B) = \theta$, $\hat{\theta}_B$ is an asymptotically unbiased estimator of θ.

2. Efficiency

The efficiency of an estimator is observed from its variance. The best parameter estimator is the one that has the smallest variance. This is because the variance of an estimator is a measure of the spread of the estimator around its mean.

The variance of ML estimator $\hat{\theta}_{ML}$ is:

$$\begin{aligned} Var(\hat{\theta}_{ML}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) \\ &= \frac{1}{n^2} n\theta(1-\theta) \\ &= \frac{1}{n}\theta(1-\theta). \end{aligned} \qquad (14)$$

While the variance of the Bayesian estimator $\hat{\theta}_B$ is given by:

$$\begin{aligned} Var(\hat{\theta}_B) &= Var\left(\frac{a+\sum_{i=1}^{n} x_i}{a+b+n}\right) \\ &= \frac{1}{(a+b+n)^2} Var\left(a+\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{(a+b+n)^2}\left[Var(a)+\sum_{i=1}^{n} Var(x_i)\right]. \end{aligned}$$

Since $Var(a) = 0$ and $Var(x_i) = \theta(1-\theta)$, we obtain

$$Var(\hat{\theta}_B) = \frac{1}{(a+b+n)^2} n\theta(1-\theta). \qquad (15)$$

From equation (10), it is shown that the ML estimator is unbiased, whereas from equations (11) and (12) it is shown that Bayesian estimator is biased. As a result, the efficiency of the two methods cannot be compared because the efficiency of estimators applies to unbiased estimators.

3. Consistency

The consistency of the estimators is evaluated from their mean square error (MSE). The MSE can be expressed as

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + (bias\hat{\theta})^2. \tag{16}$$

If the sample size grows infinitely, a consistent estimator will give a perfect point estimate to θ. Mathematically, θ is a consistent estimator if and only if

$$E(\hat{\theta} - \theta)^2 \to 0 \text{when} n \to \infty,$$

which means that the bias and the variance approaches to 0 if$n \to \infty$.

Substituting equation (10) and (14) to equation (16), the MSE of ML estimator$\hat{\theta}_{MLE}$ is then

$$E(\hat{\theta}_{MLE} - \theta)^2 = Var(\hat{\theta}_{MLE}) + (bias\hat{\theta}_{MLE})^2$$

$$E(\hat{\theta}_{MLE} - \theta)^2 = Var(\hat{\theta}_{MLE}) = \frac{1}{n}\theta(1-\theta).$$

For $n \to \infty$, we have

$$\lim_{n\to\infty} E(\hat{\theta}_{MLE} - \theta)^2 = \lim_{n\to\infty} \frac{1}{n}\theta(1-\theta) = 0. \tag{17}$$

In the same manner, by substituting equation (12) and (15) the MSE of Bayesian estimator $\hat{\theta}_B$is:

$$E(\hat{\theta}_B - \theta)^2 = Var(\hat{\theta}_B) + (bias\hat{\theta}_B)^2$$

$$E(\hat{\theta}_B - \theta)^2 = \left[\frac{1}{(a+b+n)^2} n\theta(1-\theta)\right] + \left(\frac{a+n\theta}{a+b+n} - \theta\right)^2.$$

For $n \to \infty$, we have

$$\lim_{n\to\infty} (\hat{\theta}_B - \theta)^2 = \lim_{n\to\infty} \left[\left\{\frac{1}{(a+b+n)^2} n\theta(1-\theta)\right\} + \left(\frac{a+n\theta}{a+b+n} - \theta\right)^2\right] = 0. \tag{18}$$

From equation (17) and (18), we can conclude that ML and Bayesian estimators are consistent estimators of θ.

4.4. Empirical Comparison of the Properties of ML and Bayesian Estimators

To compare the ML and Bayesian estimators of θ, a Monte Carlo simulation using R program was conducted. The simulation was performed by generating Bernoulli distributed data with θ = 0.1, 0.3, and 0.5 and eight different sample sizes, i.e. n = 20, 50, 100, 300, 500, 1000, 5000, and 10000. The simulation was repeated 1000 times for each combination of θ and n. The generated data were used to estimate parameter θ using the two methods. Furthermore, the bias and MSE of both estimators were calculated using the formulas in equations (9) and (16) andthe results are presented in Table 1.

Table 1. *The biasand MSEofML and Bayesian estimators ofθ*

θ	N	Bias		MSE	
		ML	Bayesian ($\alpha = 1, \beta = 1$)	ML	Bayesian ($\alpha = 1, \beta = 1$)
0,1	20	0,001200	0,223084	0,031478	0,558149
	50	0,002180	0,036602	0,015264	0,041740
	100	0,000270	0,009328	0,007058	0,008843
	300	0,000413	0,001075	0,001901	0,002906
	500	0,000210	0,000364	0,001183	0,000551
	1000	0,000195	0,000091	0,000503	0,000128
	5000	0,000003	0,000003	0,000143	0,000004
	10000	0,000114	0,000001	0,000184	0,000002
0,3	20	0,003100	0,536609	0,001652	0,493287
	50	0,001300	0,090000	0,003113	0,142692
	100	0,000630	0,021341	0,001583	0,067615
	300	0,000003	0,002205	0,000327	0,002307
	500	0,000312	0,000845	0,000111	0,000899
	1000	0,000545	0,000207	0,000444	0,000217
	5000	0,000384	0,000008	0,000403	0,000018
	10000	0,000023	0,000002	0,000013	0,000004
0,5	20	0,001450	0,714234	0,023000	1,453419
	50	0,002260	0,097036	0,011566	0,147045
	100	0,003860	0,024341	0,008602	0,024833
	300	0,000143	0,002856	0,001792	0,003374
	500	0,000146	0,001004	0,001139	0,003040
	1000	0,000432	0,000252	0,000929	0,000253
	5000	0,000132	0,000009	0,000232	0,000056
	10000	0,000066	0,000002	0,000016	0,000003

Table 1 shows the bias and MSE values of ML and Bayesian estimates for a successful probability of $\theta = 0.1$, 0.3 and 0.5. From the table it can be seen ML estimator produces smaller biases than Bayesian estimates for finite sample (i.e. $n < 1000$). However, when the sample size equal or larger than 1000 (i.e. 5000 and 10000, the biasesof the Bayesian estimator are smaller than the ML estimator. Even though the bias values of ML estimates changes inconsistently throughout the sample sizes, analytically it has been proved that ML estimator is an unbiased estimator.This appears to be different from the bias values for Bayesian estimator. This is because for all the considered success probabilitiesof the bias values become smaller whenthe sample size increases, although analytically it is found that Bayesian estimatoris a biased estimator. As a result the efficiency of the two estimators cannot be compared. Therefore, to compare the best estimators we use MSE of both estimators. This is because MSE considers both the bias and variance values.

The MSE values of MLand Bayesian estimators that have been shown in Table 1 have similarities, i.e. the MSE value decreases as the sample size increases and it closes to 0. Thus, both estimators are consistent estimators. This also corresponds to the results obtained analytically. Based on the simulation results in this study, it can be seen that for the larger sample sizes Bayesian estimatoris better than ML estimator. This is because the MSE value of Bayesian estimator is smaller than the ML estimator. As shown in Table 1, when $\theta = 0.1$, the MSE value of the Bayesian estimator is smaller than the ML estimator for *n*= 500, 1000, and 10000; and when θ = 0.3 and 0.5, the MSE values of the Bayesian estimator are smaller than the ML estimator for *n* = 1000, 5000, and 10000.

5. CONCLUSION

In this paper, we derived the ML and Bayesian estimator (using beta prior) of Bernoulli distribution parameter. Analytically we show that the ML estimator is an unbiased estimator and Bayesian estimator is a biased estimator for parameter θ. However, Bayesian estimator is asymptotically unbiased. Based on the simulation result, both ML and Bayesian estimator are consistent estimators of θ because the two estimators satisfy the property of consistency, i.e.$E(\hat{\theta} - \theta)^2 \to 0$ when$n \to \infty$.The simulation result also shows that the Bayesianestimator usingbeta prioris better than the MLE method for large sample sizes (n ≥ 1000).

REFERENCES

[1]. Bain, L.J. and Engelhardt, M. (1992).*Introduction to Probability and Mathematical Statistics*. Duxbury Press, California.

[2]. Walpole, R.E dan Myers, R.H. (1995). *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*. ITB, Bandung.

[3]. Al-Kutubi H. S., Ibrahim N.A. (2009). Bayes Estimator for Exponential Distribution with Extensionof Jeffery Prior Information. *Malaysian Journal of MathematicalSciences*. 3(2):297-313.

[4]. Nurlaila, D., Kusnandar D.,& Sulistianingsih, E. (2013). Perbandingan Metode *Maximum Likelihood Estimation* (MLE) dan Metode Bayes dalam Pendugaan Parameter Distribusi Eksponensial. *Buletin Ilmiah Mat. Stat. dan Terapannya*. 2(1):51-56.

[5]. Fikhri, M., Yanuar, F., & Yudiantri A. (2014). Pendugaan Parameter dari Distribusi Poisson dengan Menggunakan Metode *Maximum Likelihood Estimation* (MLE) dan Metode Bayes. *Jurnal Matematika UNAND*. 3(4):152-159.

[6]. Singh S. K., Singh, U., &Kumar, M. (2014). Estimation for the Parameter of Poisson-Exponential Distribution under Bayesian Paradigm. *Journal of Data Science*.12:157-173.

[7]. Gupta, I. (2017). Bayesian and E-Bayesian Method of Estimation of Parameter of Rayleigh Distribution-A Bayesian Approachunder Linex Loss Function. *International Journal of Statistics and Systems*.12(4):791-796.

[8]. Box, G.E.P& Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addision-Wesley Publishing Company, Philippines.