รายงานการวิจัยฉบับสมบูรณ์

การประมาณพารามิเตอร์ของการแจกแจงปกติโดยวิธีความควรจะเป็นสูงสุด วิธีมาร์คอฟ เชน มอนติคาร์โล และวิธีของเบส์

The Parameter Estimation of Normal Distribution by Using Maximum Likelihood, Markov Chain Monte Carlo ,and Bayesian Methods

นางสาวอัชฌา อระวีพร

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2559

คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

Research Title : The Parameter Estimation of Poisson Distribution by Using Maximum Likelihood, Bayesian, and Markov Chain Monte Carlo Methods

Research : Autcha Araveeporn

Department : Statistics

Faculty : Science

ABSTRACT

This paper is to compare the parameter estimation of the mean in normal distribution by Maximum Likelihood (ML), Bayes, and Markov Chain Monte Carlo (MCMC) methods. The ML estimator is estimated by the average of data, the Bayesian method is considered from the prior distribution to estimate Bayes estimator, and MCMC estimator is approximated by Gibbs sampling from posterior distribution. These methods are also to estimate a parameter then the hypothesis testing is used to check a robustness of the estimators. Data is simulated from normal distribution with the true parameter of mean 2, and variance 4, 9, and 16 when the sample sizes is set as 10, 20, 30, and 50. For the results, it can be seen that the estimation of ML, and MCMC are perceivably different from the true parameter when the sample size is 10 and 20 with variance 16. Furthermore the Bayes' estimator is estimated from the prior distribution when mean is 1, and variance is 12 which shown the significant difference in mean with variance 9 at the sample size 10 and 20.

The objective of this research is to estimate the interval estimation for the mean of normal distribution by Markov Chain Monte Carlo (MCMC) method. The interval estimation of MCMC method is compared with the common method such as Maximum Likelihood (ML) and Bayesian methods. The MCMC and Bayesian methods can join for estimating Bayes'-MCMC estimator. The MCMC, ML, Bayesian, Bayesian-MCMC methods are approximated by the point estimation for mean and variance before computing the confidence interval at 90%, 95%, and 99%. The performance of these methods is considered by Confidence Coefficients (CC) and Average Width (AW). When the CC values are covered the true parameters, the minimum of AW is used as a criterion to check the performance of these methods. In this case, the data is simulated by the Monte Carlo process of normal distribution with the true parameter of mean 2 and variance 2, 4, 9, and 16, and the sample sizes is 5,10, 25, and 30. The results show that the Bayes'-MCMC estimator is a good performance at small sample sizes (n=5). Overall, Bayesian method outperforms the other methods and sample sizes. The MCMC method presents the maximum values of AW because the estimating variance gives a high values, so it makes the AW's are larger than the other methods. However, the MCMC method is of benefit to approximate the prior estimator that can be used to estimate the Bayes'-MCMC estimator.

Keywords: Bayes Method; Maximum Likelihood Method; Markov Chain Monte Carlo Method; Normal Distribution.

CONTENTS

Chapter 1

Point Estimation

1.1 Introduction

The statistical inference is used to the estimator from sampling data to describe the characteristic population or called parameter. The estimation is a part of statistical inference that consisted of point estimation and interval estimation. The point estimation is an one estimator from sample that referred to parameter. The interval estimation presents a range which is constructed along with point estimation to show the reliability of the estimation.

Normal distribution is an important distribution in the field of statistics and is often used in tremendous data especially in social and science. The most data are presented in terms of continuous probability distribution such as income, weight, and height of a person especially normal distribution. The parameters of normal distribution consist of mean and standard deviation that remarkably useful to explain any characteristics of a population. The mean determines the location of the center of population and standard deviation determines the dispersion from the mean of population.

The several methods of parameter estimation are common known such as the moments method, the maximum likelihood method, the minimum chi-square method, the least square method, and the Bayes method. The estimators are obtain from these methods that are shown a good estimator following: an unbiasedness, sufficiency, completeness, and minimum variance unbiased estimator.

In this case, we interest the maximum likelihood method because the estimator is shown in a class of minimum variance unbiased estimator [1]. The Bays method depends on a prior probability distribution to estimate a posterior distribution which is obtained a Bayes' estimator. Moreover, the posterior distribution can be used with Markov Chain Monte Carlo (MCMC) method [2] to estimate MCMC estimator.

1.2 Methods for Parameter Estimation

The parameter estimation of the mean in normal distribution consists of the following three methods.

1.2.1 Maximum Likelihood (ML) Method

The ML method corresponds to many well-known estimation in statistics because it is easy to understand and calculate the estimators. The basic idea of ML estimation is to treat the likelihood function as a function of parameter, and find the value of parameter that maximizes it.

Suppose we have the random variables $X_1,...,X_n$ be independent and identically distributed (iid) random variables following a normal distribution with parameter μ,σ^2, and $f(x_i \mid \mu,\sigma^2)$ denote the probability density function of $x_i \mid \mu,\sigma^2$ as

$$f(x_i \mid \mu,\sigma^2)=\left(\sqrt{2\pi\sigma^2}\right)^{-1}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right), -\infty < x_i < \infty.$$

Hence, the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} f\left(x_i \mid \mu,\sigma^2\right).$$

The ML estimator of the parameter μ is solved as follows:

$$L(\mu) = \left(\sqrt{2\pi\sigma^2}\right)^{-n}\exp\sum_{i=1}^{n}\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right),$$

$$\ln L(\mu) = -n\ln\left(\sqrt{2\pi\sigma^2}\right)+\sum_{i=1}^{n}\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right),$$

$$\frac{\partial \ln L(\mu)}{\partial\mu} = \frac{-2\sum_{i=1}^{n}(x_i-\mu)(-1)}{2\sigma^2} = 0,$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}, \text{ and } \quad \hat{\lambda} = \bar{x}.$$

Therefore, the ML estimator is $\hat{\lambda}_{MLE} = \bar{x}$.

1.2.2 Bayesian Method

In Bayesian probability theory, if the posterior distributions are in the same distribution as the prior probability distribution, the prior and posterior will be called conjugate distributions, and the prior is called a conjugate prior for the likelihood function.

The likelihood function can be written as

$$\prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) = \prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right\}$$

$$= (\sqrt{2\pi}\sigma^2)^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

Considering in term of exp as

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2,$$

$$\sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})^2+2(X_i-\overline{X})(\overline{X}-\mu)+(\overline{X}-\mu)^2\right],$$

$$= \sum_{i=1}^{n}(X_i-\overline{X})^2+2\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)+n(\overline{X}-\mu)^2,$$

when $\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu) = (\sum_{i=1}^{n}X_i-n\overline{X})(\overline{X}-\mu) = 0$,

then

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2.$$

The new form of the likelihood function can be shown by

$$\prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) = (\sqrt{2\pi}\sigma^2)^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$= (\sqrt{2\pi}\sigma^2)^{-n}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2\right]\right).$$

The prior distribution of μ is a normal distribution with parameters μ_0, σ_0^2 defined as $Normal(\mu_0, \sigma_0^2)$ or rewritten as

$$g(\mu \mid \mu_0, \sigma_0^2) = (\sqrt{2\pi}\sigma_0^2)^{-1} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) \quad , -\infty < \mu < \infty .$$

The posterior distribution of μ given x_i is

$$h(\mu, \sigma^2 \mid x_1, x_2, .., x_n) = \frac{\prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) g(\mu \mid \sigma^2)}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) g(\mu \mid \sigma^2) d\mu},$$

and

$$\prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) g(\mu \mid \sigma^2) = (\sqrt{2\pi}\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X}-\mu)^2\right]\right) (\sqrt{2\pi}\sigma_0^2)^{-1} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right),$$

$$\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) g(\mu \mid \sigma^2) d\mu = \int_{-\infty}^{\infty} (\sqrt{2\pi}\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X}-\mu)^2\right]\right) (\sqrt{2\pi}\sigma_0^2)^{-1} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) d\mu,$$

$$= \frac{(\sqrt{2\pi}\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X}-\mu)^2\right]\right) (\sqrt{2\pi}\sigma_0^2)^{-1} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right)}{\int_{-\infty}^{\infty} (\sqrt{2\pi}\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X}-\mu)^2\right]\right) (\sqrt{2\pi}\sigma_0^2)^{-1} \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) d\mu},$$

$$= \frac{\exp\left(-\frac{1}{2\sigma^2}\left[n(\overline{X}-\mu)^2\right]\right) \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[n(\overline{X}-\mu)^2\right]\right) \exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right) d\mu},$$

$$= \frac{\exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\sigma_0^2(\overline{X}-\mu)^2 + \frac{\sigma^2}{n}(\mu-\mu_0)^2\right]\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\sigma_0^2(\overline{X}-\mu)^2 + \frac{\sigma^2}{n}(\mu-\mu_0)^2\right]\right) d\mu}.$$

Considering in term of exp as

$$\exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\sigma_0^2(\overline{X}-\mu)^2+\frac{\sigma^2}{n}(\mu-\mu_0)^2\right]\right),$$

$$=\ \exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\sigma_0^2(\overline{X}^2-2\overline{X}\mu+\mu^2)\ +\frac{\sigma^2}{n}(\mu^2-2\mu\mu_0+\mu_0^2)\right]\right),$$

$$=\ \exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\left(\sigma_0^2+\frac{\sigma^2}{n}\right)\mu^2-2\mu\left(\sigma_0^2\overline{X}+\frac{\sigma^2}{n}\mu_0\right)+\left(\sigma_0^2\overline{X}^2+\frac{\sigma^2}{n}\mu_0^2\right)\right]\right).$$

The term of μ is formed as normal distribution by

$$=\ \exp\left(-\frac{1}{2\frac{\sigma^2\sigma_0^2}{n}}\left[\left(\frac{n\sigma_0^2+\sigma^2}{n}\right)\mu^2-2\mu\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{n}\right)\right]\right),$$

$$=\ \exp\left(-\frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left[\mu^2-2\mu\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}\right)\right]\right),$$

$$=\ \exp\left(-\frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left[\mu-\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}\right)\right]^2\right),$$

$$=\ \exp\left(-\frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left[\mu^2-2\mu\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}\right)+\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}\right)^2\right]\right)$$

$$\exp\left(\frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left(\frac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}\right)^2\right).$$

Let $\mu_1=\dfrac{n\sigma_0^2\overline{X}+\sigma^2\mu_0}{\sigma^2+n\sigma_0^2}$ and $\sigma_1^2=\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}$, we get

$$h(\mu,\sigma^2 \mid x_1,x_2,..,x_n) = \frac{\exp\left(-\frac{1}{2\sigma_1^2}[\mu-\mu_1]^2\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2\sigma_1^2}[\mu-\mu_1]^2\right)d\mu}.$$

$$= \frac{(\sqrt{2\pi}\sigma_1^2)^{-1}\exp\left(-\frac{1}{2\sigma_1^2}[\mu-\mu_1]^2\right)}{\int_{-\infty}^{\infty}(\sqrt{2\pi}\sigma_1^2)^{-1}\exp\left(-\frac{1}{2\sigma_1^2}[\mu-\mu_1]^2\right)d\mu}.$$

The parameter μ can write in form of normal distribution as

$$h(\mu,\sigma^2 \mid x_1,x_2,..,x_n) = (\sqrt{2\pi}\sigma_1^2)^{-1}\exp\left(-\frac{1}{2\sigma_1^2}[\mu-\mu_1]^2\right) \sim N(\mu_1,\sigma_1^2).$$

In this case, the normal distribution is conjugate distribution with respect to a normal likelihood function: if the likelihood function is normal, choosing a normal prior over the mean will ensure that the posterior distribution is also normal. The resulting posterior distribution is

$$f(\mu \mid \sigma^2, \underset{\sim}{x}) = f(\underset{\sim}{x} \mid \mu,\sigma^2)g(\mu \mid \mu_0,\sigma_0^2)$$

$$\propto (\sqrt{2\pi}\sigma^2)^{-n}\exp\left(-\frac{1}{2\sigma^2}(\sum_{i=1}^{n}x_i^2+n\mu^2-2n\mu\bar{x})\right)$$

$$\times(\sqrt{2\pi}\sigma_0^2)^{-1}\exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right),$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(n\mu^2-2n\mu\bar{x})-\frac{1}{2\sigma_0^2}(\mu^2-2\mu\mu_0)\right),$$

$$\propto \exp\left(-\frac{1}{2}\left\{\left[\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right]\mu^2-2\mu\left[\frac{n\,\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right]\right\}\right),$$

$$\propto \exp\left(-\frac{1}{2}\left[\frac{n\sigma_0^2+\sigma^2}{\sigma^2\sigma_0^2}\right]\left\{\mu^2-2\mu\left[\frac{n\,\bar{x}\sigma_0^2+\mu_0\sigma^2}{n\sigma_0^2+\sigma^2}\right]\right\}\right).$$

The parameter μ can write in form of normal distribution as

$$\mu \mid \mu_0, \sigma_0^2, \sigma^2, \underset{\sim}{x} \sim Normal\left(\frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \right),$$

or
$$\mu \mid \mu_0, \sigma_0^2, \sigma^2, \underset{\sim}{x} \sim Normal\left(\frac{\bar{x}\sigma_0^2 + \mu_0 \dfrac{\sigma^2}{n}}{\sigma_0^2 + \dfrac{\sigma^2}{n}}, \frac{\sigma_0^2 \dfrac{\sigma^2}{n}}{\sigma_0^2 + \dfrac{\sigma^2}{n}} \right).$$

Then the Bayes' estimator or $\hat{\mu}_{Bayes}$ and $\hat{\sigma}^2_{Bayes}$ can compute by

$$\hat{\mu}_{Bayes} = \frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \text{ and } \hat{\sigma}^2_{Bayes} = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}.$$

In this case, the prior is denoted as $\mu_0 = 1$ and $\sigma_0^2 = 12$.

1.2.3 Markov Chain Monte Carlo Method

The MCMC method was the first introduced by [3] as a method to simulate values from posterior distribution that developed from the Bayes method. The Gibbs sampling ([4] and [5]) is a popular method that generated values from the posterior distribution which are approximated a MCMC estimator.

Therefore, we carry out the WinBUGS program [6] which is a statistical software for Bayesian analysis to estimate MCMC estimator. In order to construct a Gibbs sampling from MCMC, the posterior distribution from Bayes method is to calculate following

$$\mu \mid \mu_0, \sigma_0^2, \sigma^2, \underset{\sim}{x} \sim Normal\left(\frac{n\,\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2} \right),$$
$$\sim Normal(\hat{\mu}_{Bayes}, \hat{\sigma}^2_{Bayes}).$$

Using this result, the Gibbs sampling algorithm proceed as follows.

1. Set prior parameter : μ_0 and σ_0^2.
2. Set $\mu = \mu^{(t-1)}$.
3. Calculate $\hat{\mu}_{Bayes}$ and $\hat{\sigma}^2_{Bayes}$.
4. Generate μ from $Normal(\hat{\mu}_{Bayes}, \hat{\sigma}^2_{Bayes})$.
5. Set $\mu^{(t)} = \mu$, $t = 1, 2, \ldots, T$.

Finally the MCMC estimator is approximated by

$$\hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T}\mu^{(t)}.$$

1.3 Simulation Study

The simulation study is to generate the data in terms of normal distribution with true parameter $\mu=2$ and $\sigma^2 = 4, 9, 16$ at the sample sizes n = 10, 20,30, and 50. The data is generated 500 replications in each situations by R program [7]. To investigate the performance of ML, Bayesian, and MCMC methods, these estimators are computed by

$$\hat{\mu}_{MLE} = \bar{x},$$

$$\hat{\mu}_{Bayes} = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2},$$

$$\hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T}\mu^{(t)}.$$

Next, we obtain these estimators from 3 methods. The hypothesis testing is used to test the mean of estimator in normal distribution which is different from the true parameters. In this case, the hypotheses are

$H_0 : \mu_{\hat{\mu}} = \mu$ and $H_1 : \mu_{\hat{\mu}} \neq \mu$.

The t statistic is computed as follows:

$$t = \frac{\bar{\hat{\mu}} - \mu}{s_{\hat{\mu}}/\sqrt{m}},$$

where $s_{\hat{\mu}} = \sqrt{\frac{\sum_{j=1}^{m}(\hat{\mu}_j - \bar{\hat{\mu}})^2}{m-1}}$, and df = m-1.

For the level of significance at α, we will reject H_0 if $|t| > t_{\alpha/2,m-1}$.

1.4 The Results

The parameter estimations of normal distribution by ML, Bayesian, and MCMC methods are given to illustrate at Table 1-3. The first and the second columns of these tables present the sample sizes and the true parameters from simulated data. A mean and a standard deviation are shown in the next two columns. The last two columns of these tables list the t statistics and p-values for hypothesis testing.

The p-values of the ML and MCMC from the Table 1 and 3 are indicated that the means of the estimated parameter are different from the true parameters with $\mu=2$, and $\sigma^2= 16$ at the sample sizes n = 10 and 20. For Table 2, the Bayes' estimator is shown the significant difference in the mean with $\mu=2$, $\sigma^2= 9$ at the sample sizes n = 10 and 20.

Figure 1-3 show the histograms of the estimated parameter with ML method that followed a normal distribution. The Bayesian method shows the histogram in normal distribution at $\sigma^2= 4,9,16$ on Figure 4-6. For MCMC method, the histograms follow a normal distribution at $\sigma^2= 4,9,16$ on Figure 7-9.

Table 1. The mean, standard deviation (S.D.), t statistics (t), and p-values obtained via the ML method.

n	$\mu=2$	mean	S.D.	t	P-values
n=10	$\sigma^2=4$	1.9799	0.6474	-0.6933	0.4844
	$\sigma^2=9$	1.9830	0.9136	-0.4140	0.6790
	$\sigma^2=16$	2.1312	1.2436	2.3589	0.0187*
n=20	$\sigma^2=4$	2.0141	0.4648	0.6806	0.4964
	$\sigma^2=9$	1.9678	0.6889	-1.0440	0.2970
	$\sigma^2=16$	2.1036	0.9016	2.5760	0.0104*
n=30	$\sigma^2=4$	2.0189	0.3527	1.2013	0.2302
	$\sigma^2=9$	1.9979	0.5237	-0.0891	0.9290
	$\sigma^2=16$	1.9860	0.7354	0.4237	0.6719
n=50	$\sigma^2=4$	1.9990	0.2846	-0.0673	0.9463
	$\sigma^2=9$	2.0112	0.3991	0.6305	0.5287
	$\sigma^2=16$	2.0250	0.6074	0.9215	0.3572

* indicates significance level at 5%

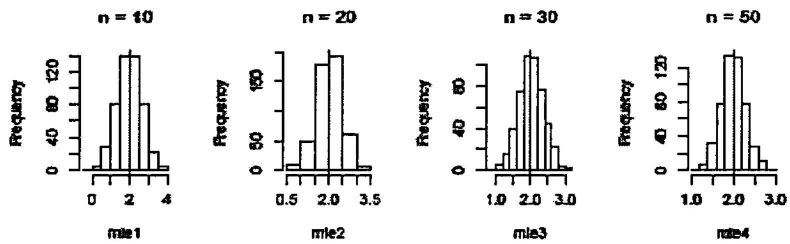

Fig. 1. Histograms of estimated parameters μ with ML method when $\mu = 2$ and $\sigma^2 = 4$.

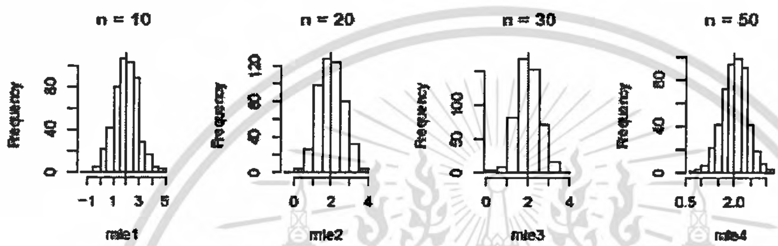

Fig. 2. Histograms of estimated parameters μ with ML method when $\mu = 2$ and $\sigma^2 = 9$.

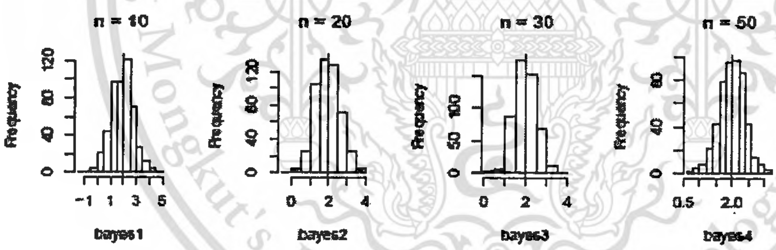

Fig. 3. Histograms of estimated parameters μ with ML method when $\mu = 2$ and $\sigma^2 = 16$.

Table 2. The mean, standard deviation (S.D), t statistic (t), and p-values obtained via the Bayesian method.

n	μ=2	mean	S.D.	t	P-values
n=10	σ^2=4	1.9489	0.6259	-1.8233	0.0688
	σ^2=9	1.9145	0.8531	-2.2396	0.0255*
	σ^2=16	2.000	1.1042	0.0018	0.9985
n=20	σ^2=4	1.9976	0.4571	-0.1140	0.9093
	σ^2=9	1.9328	0.6634	-2.2642	0.0239*
	σ^2=16	2.0363	0.8458	0.9619	0.3365
n=30	σ^2=4	2.0076	0.3488	0.4913	0.6234
	σ^2=9	1.9738	0.5116	-1.1414	0.2542
	σ^2=16	1.9444	0.7049	-1.7617	0.0787
n=50	σ^2=4	1.9924	0.2827	-0.5935	0.5531
	σ^2=9	1.9964	0.3934	-0.2008	0.8409
	σ^2=16	1.9986	0.5921	-0.0507	0.9596

* indicates significance level at 5%

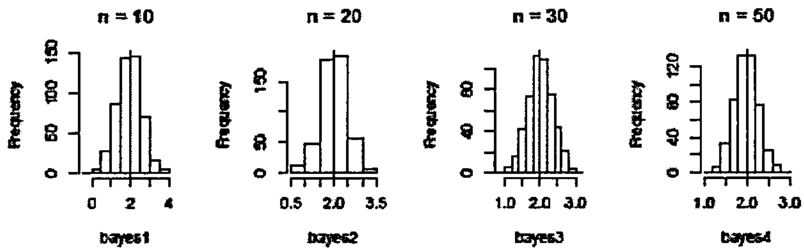

Fig. 4. Histograms of estimated parameters μ with Bayes method when $\mu = 2$ and $\sigma^2 = 4$.

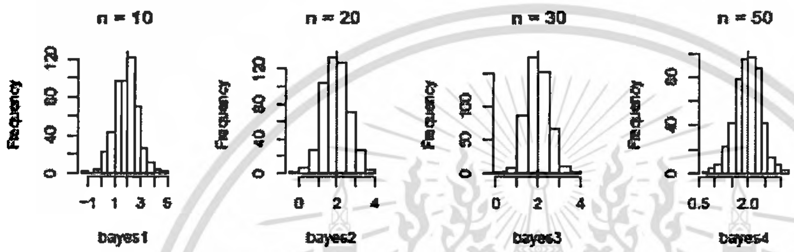

Fig. 5. Histograms of estimated parameters μ with Bayes method when $\mu = 2$ and $\sigma^2 = 9$.

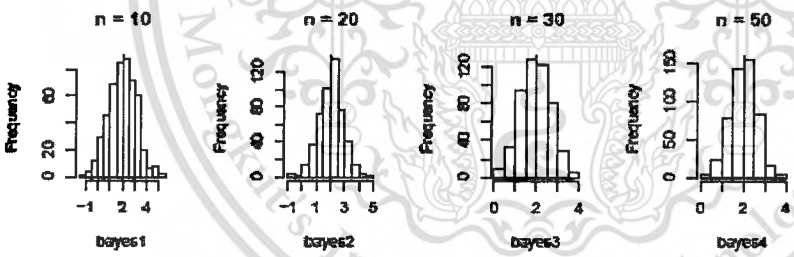

Fig. 6. Histograms of estimated parameters μ with Bayes method when $\mu = 2$ and $\sigma^2 = 16$.

Table 3. The mean, standard deviation (S.D.), t statistic (t), and p-values obtained via the MCMC method.

n	μ=2	mean	S.D.	t	P-values
n=10	σ^2=4	1.9489	0.6259	-1.8233	0.0688
	σ^2=9	1.9145	0.8531	-2.2396	0.0255*
	σ^2=16	2.000	1.1042	0.0018	0.9985
n=20	σ^2=4	1.9976	0.4571	-0.1140	0.9093
	σ^2=9	1.9328	0.6634	-2.2642	0.0239*
	σ^2=16	2.0363	0.8458	0.9619	0.3365
n=30	σ^2=4	2.0076	0.3488	0.4913	0.6234
	σ^2=9	1.9738	0.5116	-1.1414	0.2542
	σ^2=16	1.9444	0.7049	-1.7617	0.0787
n=50	σ^2=4	1.9924	0.2827	-0.5935	0.5531
	σ^2=9	1.9964	0.3934	-0.2008	0.8409
	σ^2=16	1.9986	0.5921	-0.0507	0.9596

* indicates significance level at 5%

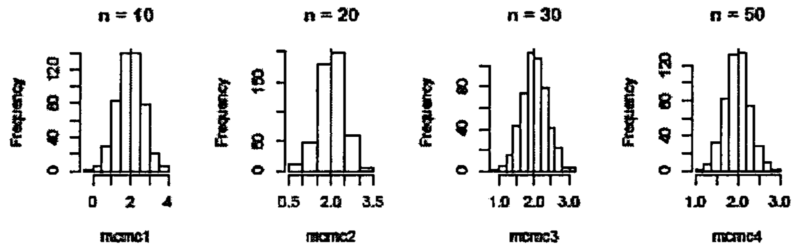

Fig. 7. Histograms of estimated parameters μ with MCMC method when $\mu = 2$ and $\sigma^2 = 4$.

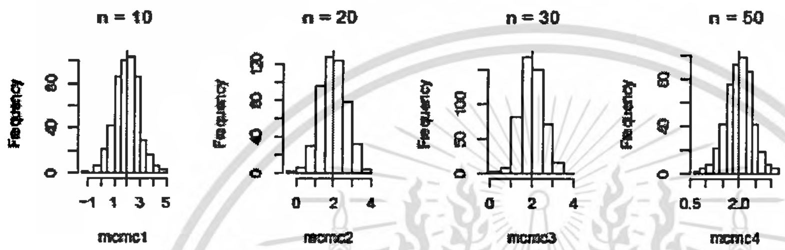

Fig. 8. Histograms of estimated parameters μ with MCMC method when $\mu = 2$ and $\sigma^2 = 9$.

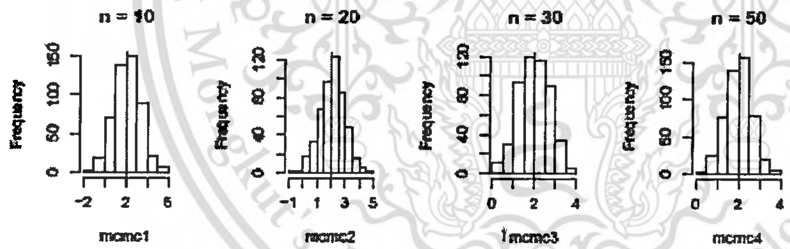

Fig. 9. Histograms of estimated parameters μ with MCMC method when $\mu = 2$ and $\sigma^2 = 16$.

1.5 Conclusions

The mean of estimated parameter from ML and MCMC methods are not different from the true parameters in most case except the large variance and small sample sizes. However, Bayesian method are not different from the true parameters in most case except the moderate variance and small sample sizes, but the Bayesian method is depended on the parameter of prior distribution so the output may be changed from this case. If we did not identify the prior distribution, the ML method will work the good performance for estimating parameter of normal distribution.

Chapter 2

Interval Estimation

2.1 Introduction

Confidence interval is an interval estimation to approximate the range of parameter. Interval estimation consists of lower confidence interval and upper confidence interval which depended on the level of confidence and standard deviation. The level of confidence interval is a range of probability that captured this population parameter. When the population parameter hold on the narrow confidence intervals, it can be concluded that the estimator is a high accuracy [8].

In this case, we interested the continuous random variable in form of normal distribution. The normal distribution is the most widely used distribution in statistics. All normal distributions are symmetric and show the bell-shaped curves with a single peak. The parameter of normal distribution defined by mean and variance. The mean is the peak of distribution, and variance indicates the spread of the bell-shaped.

At the first process of interval estimation is to estimate the point estimation from the Markov Chain Monte Carlo (MCMC), Maximum Likelihood (ML), Bayesian, Bayesian-MCMC methods. The MCMC [9] method can approximate the estimator from the Gibbs sampling algorithm [4] based on the posterior distribution. The ML method is a basic method to approximate the mean of the parameter because the estimator is to be a class of uniformly minimum variance unbiased estimator [1]. To perform Bayesian method, the parameter is identified on the prior distribution. For this paper, the prior and the likelihood function are the normal distribution, then the posterior distribution is a normal distribution or called conjugate distribution which is used to approximate Bayes' estimator[10]. Moreover, the Gibbs sampling algorithm produces the estimator of prior distribution, so we use this process to combine the Bayesian method for estimating Bayes'-MCMC estimator.

For this reason, we propose the MCMC, ML, Bayesian, and Bayesian-MCMC methods to estimate confidence interval with normal distribution using simulation studies.

2. Research Scope

2.1 Random variables X are independent and identically distributed (iid) random variables following a normal distribution with parameter μ and σ^2, hence the probability density function is

$$f(x_i \mid \mu,\sigma^2)=\left(\sqrt{2\pi\sigma^2}\right)^{-1}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right),-\infty<x_i<\infty.$$

2.2 The prior distribution of μ is a normal distribution with parameters μ_0,σ_0^2, and the probability density function is

$$g(\mu \mid \mu_0,\sigma_0^2)=(\sqrt{2\pi\sigma_0^2})^{-1}\exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right),-\infty<\mu<\infty.$$

2.3 The sample sizes are considered at n = 5, 10, 25 , and 30.

2.4 The true parameter of normal distribution is defined as μ = 2, and σ^2 = 2, 4, 9 , and 16.

2.5 The significance confidence level (α) is defined 3 levels at 0.1, 0.05, and 0.01 following the confidence interval (1-α)100% as 90%, 95%, and 99%.

2.6 The R program [7] is used to generate data at 500 replicates for each cases.

3. Methodology

3.1 The random variable X is generated in a class of normal distribution following the sample sizes, the true parameter, and the significance confidence level.

3.2 The methods for computing the confidence interval consist of following 4 methods:

3.2.1 Markov Chain Monte Carlo (MCMC) Method

The Markov Chain Monte Carlo (MCMC) method was the first introduced by [3] as a method to generate parameter values from the posterior distribution. The Gibbs sampling ([5], [2]) is a algorithm for MCMC computing that can generate random values from unidimensional distribution for which a widely variety of computational tools [2]. We carry out the WinBUGS Program [6] to obtain the estimating estimator from the posterior distribution function. The MCMC estimators can be computed by

$$\hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T}\mu^{(t)},$$

$$Var\left(\hat{\mu}_{MCMC}\right) = \frac{1}{T-1}\sum_{t=1}^{T}\left(\mu^{(t)}-\bar{\hat{\mu}}\right)^2,$$

where $\mu^{(t)}$ is generated from the posterior distribution based on the normal distribution at parameter $\mu_0^{(t)}$ and $\sigma_0^{2(t)}$, $\bar{\hat{\mu}}$ is a sample mean of independent observation, and T is a iteration of posterior distribution function .

The confidence interval (1-α)100% of μ is written by

$$\mu = \hat{\mu}_{MCMC} \pm t_{\alpha/2,(n-1)}\sqrt{Var\left(\hat{\mu}_{MCMC}\right)}$$

3.2.2 Maximum Likelihood (ML) Method

The $\hat{\mu}_{ML}$ is ML estimator of μ [1] or written as $\hat{\mu}_{ML} = \bar{x}$, and the variance of ML estimator is given by

$$Var(\hat{\mu}_{ML}) = Var(\bar{x}) = \frac{1}{n}Var(x) \because Var(x) = \sigma^2$$

$$= \frac{\sigma^2}{n}.$$

The population parameter (μ) is estimated by $\hat{\mu}_{ML} = \bar{x}$.

The confidence interval (1-α)100% of μ is approximated by

$$\mu = \hat{\mu}_{ML} \pm t_{\alpha/2,(n-1)}\sqrt{Var(\hat{\mu}_{ML})}$$

$$= \bar{x} \pm t_{\alpha/2,(n-1)}\sqrt{\frac{\sigma^2}{n}}$$

If the variance of population σ^2 is unknown value, we will use the variance of sample (s^2).

3.2.3 Bayesian Method

Let $X_1,...,X_n$ is the random variable of normal distribution with parameter μ and σ^2 while normal distribution is considered the prior distribution with parameter μ_0 and σ_0^2. The resulting posterior distribution is depended on the likelihood function $\left(L(x_i \mid \mu,\sigma^2)\right)$ and prior distribution $\left(g(\mu \mid \mu_0,\sigma_0^2)\right)$, and written as

$$f(\mu|\sigma^2, x_i) \propto L(x_i|\mu,\sigma^2)g(\mu|\mu_0,\sigma_0^2),$$

$$\propto (\sqrt{2\pi\sigma^2})^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\times(\sqrt{2\pi\sigma_0^2})^{-1}\exp\left(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right),$$

$$\propto \exp\left(-\frac{1}{2}\left[\frac{n\sigma_0^2+\sigma^2}{\sigma^2\sigma_0^2}\right]\left\{\mu^2-2\mu\left[\frac{n\bar{x}\sigma_0^2+\mu_0\sigma^2}{n\sigma_0^2+\sigma^2}\right]\right\}\right).$$

We reach to conclusion that

$$\mu|\sigma^2, x_i \sim Normal\left(\frac{n\bar{x}\sigma_0^2+\mu_0\sigma^2}{n\sigma_0^2+\sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}\right).$$

Therefore the normal distribution is conjugate to the normal distribution. The posterior mean or called Bayes' estimator is given by

$$E(\mu|x_i) = \frac{n\bar{x}\sigma_0^2+\mu_0\sigma^2}{n\sigma_0^2+\sigma^2},$$

while the posterior variance is given by

$$Var(\mu|x_i) = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}.$$

The confidence interval (1-α)100% of μ is written by

$$\mu = \frac{n\bar{x}\sigma_0^2+\mu_0\sigma^2}{n\sigma_0^2+\sigma^2} \pm t_{\alpha/2,(n-1)}\sqrt{\frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}}.$$

Hence, the $\mu_0 = 1$ and $\sigma_0^2 = 12$ are chosen for parameter of prior distribution because these values show the good performance estimator.

3.3.4 Bayesian-MCMC Method

For the MCMC method, the parameter μ_0 and σ_0^2 are approximated by

$$\hat{\mu}_{MCMC} = \frac{1}{T}\sum_{t=1}^{T}\mu^{(t)} \quad ,$$

$$\hat{\sigma}^2_{MCMC} = \frac{1}{T}\sum_{t=1}^{T}\sigma^{2(t)} \quad ,$$

where $\mu^{(t)}$ is generated from normal distribution , and $\sigma^{2(t)}$ is generated from inverse-gamma distribution.

Furthermore, the Bayes'-MCMC estimator are approximated from the MCMC method following

$$\hat{\mu}_{Bayes'-MCMC} = \frac{n\,\bar{x}\hat{\sigma}^2_{MCMC} + \hat{\mu}_{MCMC}\sigma^2}{n\hat{\sigma}^2_{MCMC} + \sigma^2},$$

and variance is estimated by

$$Var(\hat{\mu}_{Bayes'-MCMC}) = \frac{\sigma^2\hat{\sigma}^2_{MCMC}}{n\hat{\sigma}^2_{MCMC} + \sigma^2}.$$

The confidence interval (1-α)100% of μ is shown by

$$\mu = \frac{n\,\bar{x}\hat{\sigma}^2_{MCMC} + \hat{\mu}_{MCMC}\sigma^2}{n\hat{\sigma}^2_{MCMC} + \sigma^2} \pm t_{\alpha/2,(n-1)}\sqrt{\frac{\sigma^2\hat{\sigma}^2_{MCMC}}{n\hat{\sigma}^2_{MCMC} + \sigma^2}}.$$

3.3 The estimating confidence coefficient

The confidence interval is approximated by MCMC, ML, Bayes', Bayes'-MCMC estimators at significance level 0.1, 0.05, and 0.01. If the confidence intervals cover the true parameters, we will count the number and compute the proportion denoted the confidence coefficient $(1-\hat{\alpha})$.

3.4 The comparison of the confidence coefficient and the fixed confidence interval [1]

The confidence coefficient $(1-\hat{\alpha})$ is to compare with the fixed confidence interval $(1-\alpha_0)$ that we define the significance level at 0.05. If the confidence coefficient is more than the fixed confidence interval , we will perform these methods. The comparison is given by

$$1-\hat{\alpha} \geq 1-\alpha_0 = P_0 - z_{\alpha/2}\sqrt{\frac{P_0(1-P_0)}{M}},$$

where P_0 is the fixed probability given by 0.9, 0.95, and 0.99, and M is the number of replications.

The fixed confidence intervals are computed by:

$$P_0 = 0.9, 1-\alpha_0 = 0.9 - 1.96\sqrt{\frac{0.9(1-0.9)}{500}} = 0.8737,$$

$$P_0 = 0.95, 1-\alpha_0 = 0.95 - 1.96\sqrt{\frac{0.95(1-0.95)}{500}} = 0.9308,$$

$$P_0 = 0.99,\ 1-\alpha_0 = 0.99-1.96\sqrt{\frac{0.99(1-0.99)}{500}} = 0.9812.$$

3.5 The average width of confidence interval

If the confidence coefficients are larger than the fixed confidence interval, then the average width of confidence interval will be considered instead. The average width of confidence interval is evaluated by computing the average of difference values between upper limit and lower limit or written as $\sum_{j=1}^{500}\frac{(U_j-L_j)}{500}$, where U_j is the upper confidence interval, and L_j is the lower confidence interval.

4. Results

The estimating confidence interval of population parameter with normal distribution is presented by the Confidence Coefficient (CC) and the Average Width (AW) at Table 1-3. The first column and the second columns of these tables are shown the sample sizes and the true parameters. The Confidence Coefficient (CC) and the Average Width (AW) are presented in the next eight columns for four methods. The minimizing AW values are illustrated the performance of these methods ,but some AW values are in the blank because the confidence coefficient is less than the fixed confidence interval. By observing the CC and AW, the results appear that the CC and AW values of Bayes'-MCMC estimator is a minimum values for all variance at sample sizes 5. For Bayesian method, the AW outperforms at $n = 10, 25,$ and 30 for all variances and confidence intervals. When sample sizes increase, the Bayesian method is approached to the smallest of AW.

Table 6 : The Confidence Coefficient (CC) and Average Width (AW) obtained via 90% confidence interval.

n	μ=2	Methods							
		MCMC		ML		Bayesian		Bayesian-MCMC	
		CC	AW	CC	AW	CC	AW	CC	AW
5	σ^2=2	1	4.328	0.916	1.779	0.920	1.762	0.904	**1.616**
	σ^2=4	0.988	4.871	0.898	3.589	0.904	3.459	**0.830**	-
	σ^2=9	0.956	5.652	0.910	5.488	0.992	5.037	**0.824**	-
	σ^2=16	0.924	6.592	0.904	7.278	0.920	6.316	**0.786**	-
10	σ^2=2	1	4.332	0.930	1.3316	0.930	**1.325**	0.978	1.7151
	σ^2=4	1	5.367	0.992	2.636	0.926	**2.589**	0.968	3.046
	σ^2=9	1	6.824	0.942	4.030	0.946	**3.866**	0.960	4.363
	σ^2=16	0.996	8.439	0.946	5.402	0.958	**5.024**	0.956	5.655
25	σ^2=2	1	4.306	0.970	0.841	0.970	**0.840**	1	1.721
	σ^2=4	1	6.313	0.952	1.708	0.954	**1.696**	0.998	3.266
	σ^2=9	1	8.580	0.95	2.535	0.952	**2.497**	0.998	4.729
	σ^2=16	1	11.216	0.966	3.444	0.972	**3.349**	1	6.349
30	σ^2=2	1	4.299	0.956	0.769	0.956	**0.768**	1	1.724
	σ^2=4	1	6.423	0.958	1.534	0.960	**1.526**	1	3.243
	σ^2=9	1	8.920	0.970	2.308	0.970	**2.279**	1	4.770
	σ^2=16	1	11.509	0.956	3.070	0.966	**3.003**	1	6.296

Table 7: The Confidence Coefficient (CC) and Average Width (AW) obtained via 95% confidence interval.

n	$\mu=2$	Methods							
		MCMC		ML		Bayesian		Bayesian-MCMC	
		CC	AW	CC	AW	CC	AW	CC	AW
5	$\sigma^2=2$	1	5.644	0.956	2.345	0.958	2.321	0.948	**2.124**
	$\sigma^2=4$	0.998	6.269	0.954	4.486	0.954	4.327	**0.926**	-
	$\sigma^2=9$	0.994	7.347	0.948	7.040	0.954	6.477	**0.898**	-
	$\sigma^2=16$	0.976	6.592	0.950	7.278	0.958	6.316	**0.886**	-
10	$\sigma^2=2$	1	5.640	0.980	1.727	0.980	**1.719**	0.996	2.225
	$\sigma^2=4$	1	6.946	0.976	3.390	0.976	**3.331**	1	3.925
	$\sigma^2=9$	1	8:739	0.974	5.115	0.976	**4.917**	0.986	5.556
	$\sigma^2=16$	1	10.738	0.980	6.835	0.982	**6.389**	0.986	7.175
25	$\sigma^2=2$	1	5.633	0.992	1.111	0.992	**1.109**	1	2.269
	$\sigma^2=4$	1	8.211	0.986	2.221	0.986	**2.206**	1	4.248
	$\sigma^2=9$	1	11.029	0.984	3.251	0.984	**3.203**	1	6.069
	$\sigma^2=16$	1	14.345	0.984	4.397	0.986	**4.280**	1	8.109
30	$\sigma^2=2$	1	5.579	0.992	0.994	0.992	**0.992**	1	2.228
	$\sigma^2=4$	1	8.432	0.990	2.019	0.990	**2.008**	1	4.265
	$\sigma^2=9$	1	11.666	0.994	3.020	0.994	**2.982**	1	6.247
	$\sigma^2=16$	1	15.170	0.990	4.052	0.992	**3.960**	1	8.305

Table 8: The Confidence Coefficient (CC) and Average Width (AW) obtained via 99% confidence interval.

n	$\mu=2$	Methods							
		MCMC		ML		Bayesian		Bayesian-MCMC	
		CC	AW	CC	AW	CC	AW	CC	AW
5	$\sigma^2=2$	1	9.353	0.978	-	0.978	-	0.974	-
	$\sigma^2=4$	1	10.543	0.994	7.835	0.994	7.528	0.99	6.146
	$\sigma^2=9$	0.9812	12.215	0.992	11.751	0.992	10.809	0.986	5.323
	$\sigma^2=16$	1	14.300	0.992	15.871	0.994	13.779	0.976	-
10	$\sigma^2=2$	1	9.324	1	2.816	1	2.803	1	3.635
	$\sigma^2=4$	1	11.562	0.996	5.664	0.996	5.564	1	6.548
	$\sigma^2=9$	1	14.535	1	8.522	1	8.191	1	9.250
	$\sigma^2=16$	1	17.792	1	11.312	1	10.568	1	11.879
25	$\sigma^2=2$	1	9.319	1	1.830	1	1.827	1	3.743
	$\sigma^2=4$	1	13.335	1	3.574	1	3.550	1	6.851
	$\sigma^2=9$	1	18.346	1	5.411	1	5.329	1	10.098
	$\sigma^2=16$	1	23.507	1	7.195	1	7.008	1	13.276
30	$\sigma^2=2$	1	9.289	1	1.664	1	1.662	1	3.729
	$\sigma^2=4$	1	13.981	1	3.349	1	3.330	1	7.073
	$\sigma^2=9$	1	19.396	1	5.024	1	4.960	1	10.390
	$\sigma^2=16$	1	25.049	1	6.688	1	6.538	1	13.710

2.5 Conclusions

In this research, data is generated from a normal distribution and estimated the confidence interval which is obtained confidence coefficient and average width to perform MCMC, ML, Bayesian, and Bayesian-MCMC methods. The MCMC method is proposed to estimate the prior parameter that can be used for estimating Bayes'-MCMC estimator. The posterior distribution function is related with the MCMC method to evaluate MCMC estimator. Through a simulation study, the Bayes'-MCMC estimator is a good performance when sample sizes is small values for all variances. For the other sample sizes, Bayesian method outperforms the confidence interval for all variances. The large variance of prior distribution is affected to the mean and variance of Bayes' estimator, so the interval estimation is given the smallest average width. The confidence coefficients are closed to one when the sample sizes is large. However, if the user can not to estimate the prior parameter from Bayesian method, we would like to recommend Bayes'-MCMC estimator because the MCMC method can estimate the prior parameters.

Acknowledgements

This Research was financially supported by the Faculty of Science of King Mongkut's Institute of Technology Ladkrabang.

References

[1] Rohatgi V.K, and Saleh A.K, E, An Introduction to Probability and Statistics, Canada: A Wieley-Interscience Publication; 2001.

[2] Gilks, W, Richardson, S, and Speigelhalter, D, Markov Chain Monte Carlo in Practice, Interdisciplinary Statistics, Chapman & Hall, Suffolk, UK; 1996.

[3] Metropolis, N, Rosenbluth, A, Rosenbluth, M, Teller, A, and Teller, E, Equations of State Calculations by Fast Computing Machine, Journal of Chemical Physics, 1953; 21: 1087-1092.

[4] Gemam S, Geman D, Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. IEEE Transactions on Pattern analysis and Machine Intelligence. 1984; 6: 721-741.

[5] Gelfand A, Hills S, Racine-Poon A., Smith A, Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association, 1990; 85: 972-985.

[6] Ntzoufras I, Bayesian Modeling Using WinBUGS, New Jersey: John Wiley & Sons; 2009.

[7] R Development Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2004.

[8] Araveeporn A, The Interval Estimation of Poisson Distribution by Using Maximum Likelihood, Markov Chain Monte Carlo, Bayes methods. Proceeding of International Conference on Applied Statistics 2015. 2015; 36-42.

[9] Laskey, P, and Mayers, J, Population Markov Chain Monte Carlo, Machine Learning, 2003; 50 : 175- 196.

[10] Christensen R, Johnson W, Branscum A, Hanson TE, Bayesian ideas and data analysis : an introduction for scientists and statisticians. Florida: Taylor & Francis Group; 2011.