The ESTIMATION OF THE NILE PROBLEM WITHIN JEFFREYS PRIOR DISTRIBUTION

Yada Pornpakdee[1], Tammarat Kleebmek[1,*]

[1]*Department of Applied Mathematics and Statistics, Rajamangala University of Technology, Nakhon Ratchasima, Thailand*

[*]For correspondence; Tel. + (66) 850074144, E-mail: tammarat.kl@rmuti.ac.th

ABSTRACT: *The aim of this study is to develop the parameter estimation for the Fisher information "the Nile Problem." This paper will compare the new idea estimator in the sense of Bayesian estimation with Jeffreys's prior distribution and maximum likelihood method. This research simulates two independent data sets using the Mote Carlo Simulation. The results demonstrate that the proposed estimators show less mean squared error than maximum likelihood estimators. Monte Carlo simulations are illustrated to compare the efficiency of the estimators.*

Keywords: the Nile problem, Bayes's estimator, Maximum likelihood, Jeffreys prior distribution

1. INTRODUCTION

The Nile problem was prepared by Fisher [1] in content about statistical inference for a special curved exponential family if the minimal sufficient statistic is incomplete. The original statement of the problem in Fisher's unique style is in Fisher: The agricultural land of a pre-dynamic Egyptian village is if unequal fertility. Given the height to which the Nile rise, the fertility of every portion of it is known with exactitude, but the height of the flood affects different parts of the territory unequally. It is required to divide the area, between the several households of the village, so that the yields of the lots assigned to each shall be in pre-determined proportion, whatever may be the height to which that portion of the river rises.

For such a model the inferential structure is fully specified by three-element $\{S,\Omega, f\}$, the sample $S=\{x\}$, the parameter space $\Omega=\{\theta\}$, and the probability function $f:\Omega\to R$. The importance of the Nile problem lies in the fact that inference θ is based on the conditional distribution of the observations. There are many research papers written on this topic [2-6]. Let X Y and be two independent random variables. When X is an exponential distribution with parameter θ Y and is an exponential distribution with parameter $1/\theta$ X Y? So, and have a joint probability density function as:

$$f(x,y:\theta)=\exp\{-(\theta x+\theta^{-1}y)\},\quad x>0,y>0,\theta>0 \tag{1}$$

Where, $\bar{x}=\sum_{i=1}^{n}x_i/n$ and $\bar{y}=\sum_{i=1}^{n}y_i/n$. The pair $(\bar{x},\bar{y})$ of the sample means it is an incomplete sufficient statistic of θ. So this probability density function (PDF) was called the Nile problem [3]. In parameter estimation, the maximum likelihood estimator (MLE) is a solution for estimating the parameters of statistical inference. It is used with a popular statistical analysis, which is a method that finds the most likely value for the parameter based on the data set collected. The properties of MLE are consistent, if the data was generated by $f(x;\theta_0)$ and we have a sufficiently large number of observations, then it is possible to find the value of θ_0 with arbitrary precision. For mathematics, this means that as observations go to infinity the estimator $\hat{\theta}$ converges in probability to its true value $\hat{\theta}_{MLE}\to\theta_0$: This MLE we created is called a likelihood function and is written as:

$$L(\theta)=\prod_{i=1}^{n}f_x(x_i;\theta). \tag{2}$$

For a likelihood function $L(\theta)$, where θ is an unknown parameter? Let θ_e be a value of the parameter such that $L(\theta_e)\geq L(\theta)$ for all possible value of θ. Th θ_e is called an MLE [7]. In this paper, we have the MLE of θ for the Nile problem as follows:

$$\hat{\theta}_{MLE}=\sqrt{\bar{y}/\bar{x}}. \tag{3}$$

Bayes' estimators differ from all traditional estimators studied so far in that they consider the parameters as random variables instead of unknown constants. It is based on Bayes' theorem for conditional probability. The Bayesian analysis starts with little to no information about the parameter to be estimated. Any data collected can be used to adjust the function of the parameter, thereby improving the estimation of the parameter. This process of refinement can continue as new data is collected until a satisfactory estimate is found. For events *A* and *B*, recall that the conditional probability is:

$$P(A\,|\,B)P(B)=P(A\cap B)=P(B\,|\,A)P(A) \tag{4}$$

or

$$P(A\,|\,B)=\frac{P(B\,|\,A)P(A)}{P(B)} \tag{5}$$

Now, if it set $A=\{\theta=\theta_0\}$ and $B=\{X=x\}$, then

$$P\{\theta=\theta_0,X=x\}=\frac{P(X=x\,|\,\theta=\theta_0\}P\{\theta=\theta_0\}}{P\{X=x\}}. \tag{6}$$

If the appropriate density exists, then we can write Bayes' formula as:

$$f_{\Theta|X}(\theta_0\,|\,X)=\left(\frac{f_{X|\Theta}(X\,|\,\theta_0)}{\int f_{X|\Theta}(X\,|\,\theta_0)\pi(\theta)d\theta}\right)\pi(\theta), \tag{7}$$

To compute the posterior density $f_{\Theta|X}(\theta_0\,|\,X)$ as the product of the Bayes' factor and the prior density. One would often like to have a reference prior distribution, a roughly noninformative prior distribution against whose results inference, that is based on more subjective priors, can be

compared. Since its introduction by Aldric [8], the Jeffreys [9] prior has been one of the most intensively studied reference priors in Bayesian statistics and econometrics. The Jeffreys prior is defined in terms of the Fisher information matrix as:

$$\pi(\theta) \propto I(\theta)^{\frac{1}{2}} \tag{8}$$

where the Fisher information $I(\theta)$ is given by

$$I(\theta) = -E_{\theta}\left[\frac{d^2 \log p(X \mid \theta)}{d\theta^2}\right]. \tag{9}$$

Example of the Jeffrey prior, suppose X was binomially distribution: $X \sim Bi(n,\theta), 0 \le \theta \le 1$ and X had probability mass function as

$$f(x \mid \theta) = \binom{n}{x}\theta^x (1-\theta)^{n-x} \tag{10}$$

They choose a prior $\pi(\theta)$ that is invariant under reparameterizations. So they saw previously that a flat prior $\pi(\theta) \propto 1$ does not have this property. Let's derive a Jeffreys prior for θ. Ignoring terms that don't depend on θ, we have

$$\log p(x \mid \theta) = x\log\theta + (n-x)\log(1-\theta)$$

$$\frac{d}{d\theta}\log p(x \mid \theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$\frac{d^2}{d\theta^2}\log p(x \mid \theta) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \tag{11}$$

Since $E_{\theta}(X) = n\theta$ under $Bi(n,\theta)$, we have

$$\begin{aligned} I(\theta) &= -E_{\theta}\left[\frac{d^2 \log p(x \mid \theta)}{d\theta^2}\right] \\ &= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} \\ &= \frac{n}{\theta} + \frac{n}{1-\theta} \\ &= \frac{n}{\theta(1-\theta)}. \end{aligned} \tag{12}$$

Therefore $\pi_J(\theta) = I(\theta)^{\frac{1}{2}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$, which is the form of a Beta$\left(\frac{1}{2},\frac{1}{2}\right)$ density. Jeffreys priors work well for single parameter models, but not for models with multidimensional parameters. A lot of research has been conducted on estimating the parameter θ of the Nile Problem and presented several estimation methods. Abram and Yaakov [6] displayed the existence of uniformly minimum variance unbiased estimators in the models Eq.(1) as an open problem. Van der Geer [10] showed to find two estimators that Minimum Risk Scale Equivariant (MRE) to give the least risk function which the estimator as:

$$\hat{\theta}_{MRE} = \hat{\theta}_{MLE}\frac{K_1(2u)}{K_2(2u)}, \tag{13}$$

where $K_r(v) = \int_0^{\infty} t^{-r}\exp\{-(t+\frac{v}{t})\}dt, -\infty < r < \infty$ $u = \sqrt{xy}$ and.

And Bayes' Estimator within the prior distribution was the inverses gamma function which prior distribution was

$$\psi_{c,w}(\theta) = \frac{\theta^{-c-1}e^{-w/\theta}w^c}{\Gamma(c)}, \theta > 0, w > 0, c \in I^+. \tag{14}$$

And Posterior Distribution of θ obtained

$$h(\theta \mid \theta_{MLE}, u) = \frac{\psi_{c,w}(\theta)g(\theta_{MLE}, u \mid \theta)}{\int_0^{\infty}\psi_{c,w}(\theta)g(\theta_{MLE}, u \mid \theta)d\theta}. \tag{15}$$

Hence Bayes's estimators were

$$\begin{aligned} \hat{\theta}_B(c,w) &= E(\theta \mid \hat{\theta}_{MLE}, u) \\ &= \frac{\int_0^{\infty}\theta^{c-2}\exp(\frac{-w\theta}{t})\exp\{-u(\theta+\frac{1}{\theta})\}d\theta}{\int_0^{\infty}\theta^{c-1}\exp(\frac{-w\theta}{t})\exp\{-u(\theta+\frac{1}{\theta})\}d\theta}. \end{aligned} \tag{16}$$

A comparison of both methods, like mean square error (MSE). Joshi and Nabar [11] displayed a linear estimator which is an unbiased projection to construct in the form of a linear equation. Let $Z_1 = (n-1)/\sum_{i=1}^{n} x_i$ and $Z_2 = \sum_{i=1}^{n} y_i$ was unbiased estimators of θ. Then them estimators were:

$$\hat{\theta}_{JN} = \beta Z_1 + (1-\beta)Z_2 \tag{17}$$

They calculated the coefficients that made the minimum error. After that compared the MSE between the linear estimator and maximum likelihood estimator (MLE). The results have been found not significant from MSE. Nayak and Singha invented a new theory of estimator by minimum variance unbiased estimator (MVUE). The Jeffreys prior distribution Joshi and Nabar [3] constructed the prior distribution for unknown distribution by:

$$I(\theta) = -E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta^2}\right], \tag{18}$$

where $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$.

We present the new proposed estimator with Bayes' method by using Jeffreys's prior distribution function. Then the estimated values are compared with the MLE by minimum MSE criteria.

2. PARAMETER ESTIMATION

We find the estimator using the Bayes' theory. By providing prior distribution function as the distribution of Jeffrey. Estimates are proposed as follows.

Step 1: To compute the Jeffreys prior to the equation

$$\pi(\theta) = \sqrt{-E\frac{d^2 \ln L(\theta)}{d\theta^2}} \tag{19}$$

And the MLE θ can be found such that

$$L(\theta) = \prod_{i=1}^{n} f(x_i, y_i \mid \theta). \tag{20}$$

When $f(x,y:\theta)=\exp\{-(\theta x+\theta^{-1}y)\}$ th $L(\theta)$ is is

$$L(\theta)=\exp(-(\theta\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i/\theta)). \quad (21)$$

So logarithm of Eq.(21) is

$$\ln L(\theta)=-\theta\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}y_i/\theta \quad (22)$$

We calculate Eq.(22) with respect to θ lead to

$$\frac{d^2\ln L(\theta)}{d\theta^2}=-\frac{2\sum_{i=1}^{n}y_i}{\theta^3}. \quad (23)$$

We obtain the expectation of Eq. (23) from

$$E(\frac{d^2\ln L(\theta)}{d\theta^2})=-\frac{2n}{\theta^3}E(y) \quad (24)$$

Substituting $E(\frac{d^2\ln L(\theta)}{d\theta^2})$ in Eq.(24), it can be written as follows

$$\pi(\theta)=\sqrt{-(-\frac{2n}{\theta^2})}=\frac{\sqrt{2n}}{\theta}. \quad (25)$$

Step 2: To calculate posterior distribution of θ as follows

$$h(\theta\mid x,y)=\frac{\sqrt{2n}\exp(-(\theta\sum x+\sum y/\theta))/\theta}{\int_0^\infty\sqrt{2n}\exp(-(\theta\sum x+\sum y/\theta))/\theta d\theta} \quad (26)$$

So Bayes' estimator $(\hat{\theta}_{bayes})$ is

$$\hat{\theta}_{bayes}=E[\theta\mid x,y]$$

$$=\frac{\int_0^\infty\sqrt{2n}\exp(-(\theta\sum x_i+\sum y_i/\theta))d\theta}{\int_0^\infty\frac{\sqrt{2n}}{\theta}\exp(-(\theta\sum x_i+\sum y_i/\theta))d\theta}$$

$$=\frac{\int_0^\infty\exp(-(\theta\sum x_i+\sum y_i/\theta))d\theta}{\int_0^\infty\frac{1}{\theta}\exp(-(\theta\sum x_i+\sum y_i/\theta))d\theta}$$

$$=\frac{\int_0^\infty\exp(-(\theta\sum x+\sum y/n))d\theta}{\int_0^\infty\exp(-(\theta\sum x+\sum y/\theta))/\theta d\theta} \quad (27)$$

3. NUMERICAL ANALYSIS AND SIMULATION

This research has simulated two independent data set by the R program. Consider $X=(X_1,...,X_n)$ $Y=(Y_1,...,Y_n)$ and where X as an exponential with parameter θ Y and has an exponential distribution with parameter $1/\theta$, respectively. We set parameter 0.1, 0.3 and 0.5 and A sample size (n) is 10, 20, 30 40 and 50. We estimate the parameter from the data obtained 1,000 times to calculate MSE. Of each estimator compare MSE of the estimator. The results are shown in TABLE 1-3 and Fig.1-3 below

Table 1. The mean square error between the estimator and sample size when $\theta=0.1$

Estimator	Sample size ($\theta=0.1$)				
	10	**20**	**30**	**40**	**50**
θ_{MLE}	1.43E-06	1.43E-07	4.67E-07	2.69E-08	3.53E-08
$\hat{\theta}_{bayes}$	9.69E-07	2.17E-08	3.52E-07	1.43E-07	2.38E-08

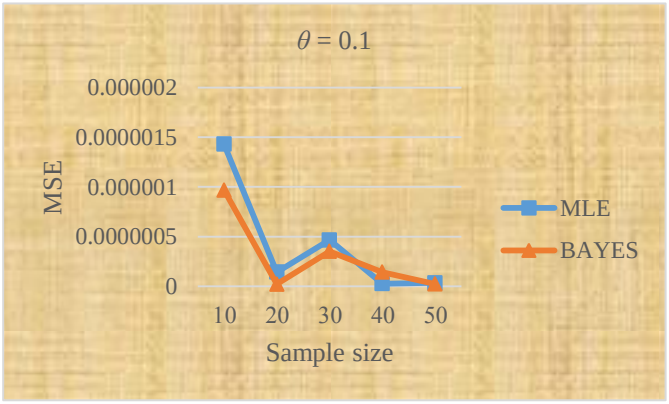

Fig (1) MSE of $\hat{\theta}_{MLE}$ and $\hat{\theta}_{BAYES}$ when $\theta=0.1$

From Table 1. and Fig. 1. on above for $\theta=0.1$, we see that when sample size increases, MSE of both $\hat{\theta}_{Bayes}$ and $\hat{\theta}_{MLE}$ are decreased to zero. It shows that the sample size influences all estimation methods. In addition, when the sample size is small, the MSE $\hat{\theta}_{BAYES}$ is less than $\hat{\theta}_{MLE}$.

Table 2. The mean square error between the estimator and sample size when $\theta=0.3$

Estimator	Sample size ($\theta=0.3$)				
	10	**20**	**30**	**40**	**50**
θ_{MLE}	2.83E-06	1.23E-07	1.69E-07	3.52E-09	1.72E-06
$\hat{\theta}_{bayes}$	2.14E-06	4.53E-08	1.55E-07	2.14E-08	1.69E-06

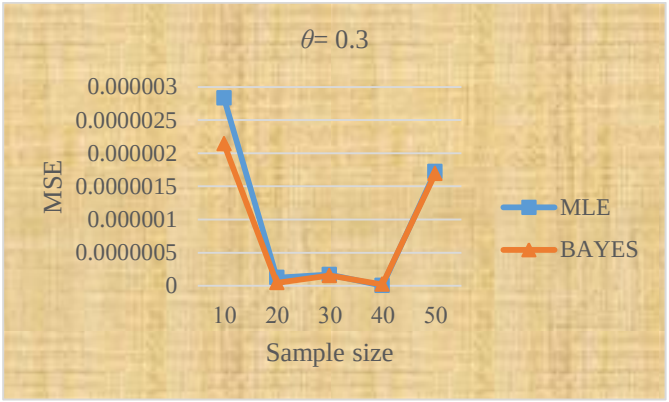

Fig (2) MSE of $\hat{\theta}_{MLE}$ and $\hat{\theta}_{BAYES}$ when $\theta=0.3$

From Table 2. and Figure. 2. Above, for $\theta=0.3$, we found that when sample size increased, MSE of both $\hat{\theta}_{Bayes}$ and $\hat{\theta}_{MLE}$

are decreased, except for sample size 50. In addition, when the sample size is small, the MSE $\hat{\theta}_{BAYES}$ is less than $\hat{\theta}_{MLE}$

Table 3. The mean square error between the estimation and sample size when $\theta = 0.5$

Estimator	Sample size ($\theta = 0.5$)				
	10	**20**	**30**	**40**	**50**
θ_{MLE}	9.31E-06	1.95E-05	1.61E-05	1.56E-06	5.07E-07
$\hat{\theta}_{bayes}$	1.12E-06	1.80E-05	1.75E-05	1.35E-06	6.28E-07

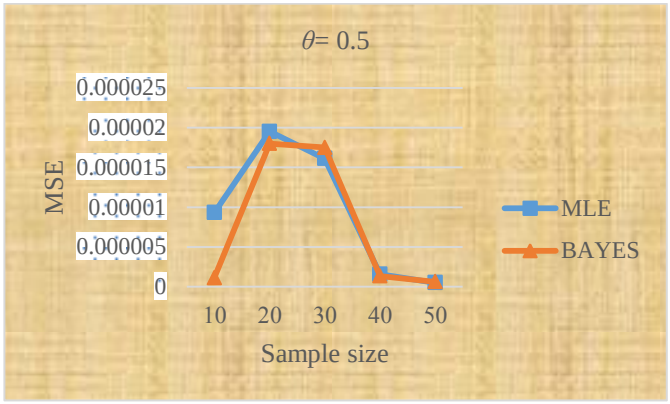

Fig (3) MSE of $\hat{\theta}_{MLE}$ and $\hat{\theta}_{BAYES}$ when $\theta = 0.5$

From Table 3, and Figure. 3 above, for $\theta = 0.5$, the sample size set, it was found that sample size 20 had an MSE higher than that of the other sample.

In addition,, when the sample size is small, the MSE $\hat{\theta}_{BAYES}$ is less than $\hat{\theta}_{MLE}$. In these figures we simulated data when the parameters were 0.1, 0.3 and 0.5, it was found that $\hat{\theta}_{BAYES}$ has the lowest MSE at a sample size of 10 and 20 but when the sample size increases, the MSE is similar

4. CONCLUSION

The results show that Bayes' estimator $(\hat{\theta}_{bayes})$ using Jeffreys's prior distribution function which is the estimator from Eq.(17). The minimum mean square error is achieved when the sample size is 10 or 20 at all levels, The Bayes' Method is used but when the sample increases, the best approximation will depend on the sample size and the parameters. When the sample size is 30, 40, and 50, it is difficult to determine which method is the most accurate because it is due to both the sample size and parameter values. In this article, we are not interested in the features of the very estimated. The criteria we use are MSE. We just want to get new ideas for estimating parameters so that the values are close to the very parameters.

5. REFERENCES

[1] Van der Geer, J., Hanraads, J. A. J., & Lupton R. A. "The art of writing a scientific article" *Journal of Scientific Communications*, **163**:51-59 (2000)

[2] Joshi, S. M., & Nabar, S. P. "Estimation of the parameter in the problem of the Nile" *Communications in Statistics-Theory and Method*, **16**:3149-3156 (1987)

[3] Joshi, S. M., & Nabar, S. P. "Linear Estimation for the parameter in the problem of the nile" *American Statistical Association*, **43**:40-41 (1989)

[4] Andrew, G. Bayes, Jeffreys, "Prior Distribution and the Philosophy of Statistics" *Statistical Science*, **24**:176-178 (2009)

[5] Barnard G.A.& Sprott D. A. "The generalized problem of the Nile: Robust confidence sets for parametric functions" *The Annals of Statistics*, **11**:104-113(1983)

[6] Fisher, R. A. "Uncertain inference" *Proc. Amer. Acad. Arts and Sciences*, **71**:245-258 (1939)

[7] David, G. Fisher's "Problem of the Nile"*nature International journal of science*, 158:452-453 (1946)

[8] Cobb, G. "The Problem of the Nile: Conditional Solution to a Changepoint Problem" *Biometrika*, **65**:243-251 (1978).

[9] Abram M. Kagan & Yaakov Malonovsky. On the Nile problem by Sir Ronald Fisher. *Electric Journal of Statistics*, **7**:1968-1982 (2013)

[10] Nayak, T.K. and Sinha, B. "Some aspects of minimum variance unbiased estimation in presence of ancillary statistics" *Stat. Prob. Letters*, 82:1129-1135 (2012).

[11] Aldric, J. "R. A. Fisher and the making of maximum likelihood 1912-1922" *Statistical Science*, **12**(3):162-176(1997)

[12] Jeffreys, H. "An invariance form for the prior probability in estimation problems" *Proceedings of the Royal Society of London*, **186**:453-461 (1946)

*For correspondence; Tel. + (66) 850074144, E-mail: tammarat.kl@rmuti.ac.th