THE EXTENDED JEFFREY PRIOR BASED ON GENERALIZED FISHER INFORMATION

Muhammad Fajar

Badan Pusat Statistik Jl. Dr. Sutomo 6-8 Jakarta, Indonesia: mfajar@bps.go.id

Abstract

The Jeffrey prior is noninformative prior formed from the Fisher information. However, nowadays, Fisher information is extended to generalized fisher information, which has various forms of specifications. Therefore, this paper aims to establish a theoretical framework of the extended Jeffrey prior based on generalized Fisher information as a first step of investigation for future research. The findings of this study are that the form of the extended Jeffrey prior varies depending on the type of generalized fisher information used. However, the posterior formed using the extended Jeffrey prior is not closed-formed and requires a more complex estimation algorithm than the standard algorithm.
Keyword: Jeffrey prior, fisher information, Bayesian, posterior

1. Introduction

In the Bayesian Statistics paradigm, priors play an important role as information on parameters as random variables. This is what distinguishes it from the Frequentist approach. The prior can be interpreted as a form of frequency distribution, a form of normative and objective representation of the pattern of a parameter change, and a subjective representation of the researcher in seeing the pattern of parameter change. This means that the value of a parameter can be generated from the mode pattern of the prior data (either symmetry or asymmetry), and the prior usually has a physical meaning according to the frequency of occurrence in the data. Although the prior distribution is unknown, in order for the selected prior to having meaning, the selection of the prior distribution must represent the conditions of the phenomenon in the field. Suppose θ only has values in a limited range. In that case, it is reasonable to use a uniformly distributed prior so that each component in the prior domain is given an equal chance to be selected as support in forming the posterior. The prior can have a meaning that is not under the problem being modeled if the selection process is incorrect, resulting in a posterior that is formed differently from the observed data pattern (Iriawan, 2020).

The prior distribution is one of the keys to Bayesian inference and displays information about the uncertainty of the parameter θ. Determining the prior distribution can be done by considering various information that will be used as a prior distribution and the nature of the posterior distribution that will be generated. Various prior distributions have been developed in many kinds of literature. However, they can be broadly grouped based on three types of priors, namely conjugate prior, noninformative prior, and subjective prior. A prior is said to be conjugate prior if the posterior distribution of the resulting parameters comes from the same family of probability distributions as the prior (Hogg et al., 2005). Examples of conjugate priors are Normal distribution priors, Beta distribution, and Gamma distribution.

Noninformative priors are priors that are used when the prior knowledge of the model parameters is absent or very little and uncertain. One approach to this prior is to choose a prior that is an approximation uniformly distributed in the domain of the parameter space under study. Another example of a noninformative prior is the diffuse prior, where a Normal distribution with a large enough variance is chosen as the prior distribution. Noninformative priors are generally obtained from the sample data and the model under study to construct the prior distribution.

One approach to this noninformative prior is to use Jeffrey's rule, which uses Fisher's information. The Jeffrey prior is the square root of the Fisher information (Box & Tiao, 1992). Fisher information is mathematically the variance of the partial derivation of the log-likelihood function concerning a parameter contained in its likelihood. Thus philosophically, Fisher information can be interpreted as how much information about the parameters contained in the sample data, where the sample data follows a particular probability distribution. Then, suppose it is associated with the formulation of the Jeffrey prior. In that case, the Jeffrey prior is the standard deviation of the partial derivation of the log-likelihood function of a parameter. With that in mind, a Jeffrey prior means a prior based on the distribution of candidate parameters that contains the posterior parameters in its range.

Fisher information in its development has a more general form or what is called Generalized fisher information (Boekee, 1977; Furuichi, 2010; Lutwak et al., 2012; Bercher, 2013) proposes a more general form of Fisher information. The purpose of the generalized formulation of Fisher information is to relate the mean absolute error to the higher order power of an estimator (Boekee, 1977), to make Fisher information invariant under all entropies concerning Stam's Inequality (Lutwak et al., 2012), and to accommodate the relationship between maximizing entropy and minimizing Fisher information in a more general form (Furuichi, 2010 and Bercher, 2013). Based on the description of the last two paragraphs, an extended Jeffrey prior can be formed by replacing the original form of Fisher information with generalized Fisher information, which is the goal of this paper.

2. Method
Fisher Information
The Fisher information of a random variable X that follows a probability function $f(X; \theta)$ is defined as follows:

$$I(\theta) = E\left[\left(\frac{\partial}{\partial \theta} \ln f(X; \theta)\right)^2\right] \qquad (1).a$$

$$= -E\left[\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta)\right] \qquad (1).b$$

The Fisher information based on Eqs.(1).a and (1).b is a means to quantify how much information an observable random variable X has about a parameter that is unknown but affects its likelihood. The probability density function (or probability mass

function) for X conditioned on the value of is denoted by $f(X; \theta)$. It describes the likelihood that, given a known value of, we will witness a specific outcome of X. It is simple to determine the "correct" value of θ from the data, or, put another way, that the data X contains a lot of information about the parameter, if $f(X; \theta)$ is strongly peaked with regard to changes in. The "true" value of θ, which would be determined using the complete population being sampled, would need numerous samples of X if $f(X; \theta)$ is flat and spread-out. This implies researching a particular variant with regard to θ.

Generalized Fisher Information
Generalized Fisher Information-Boekee

Boekee (1977) proposed Generalized fisher information as follows:

$$I_s(\theta) = \left[E\left[\left| \frac{\partial}{\partial \theta} \ln f(X; \theta) \right|^{\frac{s}{s-1}} \right] \right]^{s-1} , s \geq 1. \tag{2}$$

If $I_s(\theta)$ is of order $s \geq 2$, $I_s(\theta)$ is a convex function of $f(x; \theta)$. If $I_s(\theta)$ is of order $s = 2$, $I_2(\theta)$ becomes Fisher information, $I(\theta)$.

Generalized Fisher Information-Furuichi
Furuichi (2010) proposed q-fisher information, which is formulated as follows:

$$I_q(\theta) = E_q\left[\left(\frac{\partial}{\partial \theta} \ln_{(q)} f(X; \theta) \right)^2 \right], \tag{3}$$

where

$$\ln_{(q)}\big(f(x; \theta)\big) = \frac{(f(x; \theta))^{1-q} - 1}{1 - q}, \qquad q \in \mathbb{R}, q \neq 1, x > 0.$$

$$E_q[g(X)] = \int_{R_X} \big(f(x)\big)^q g(x) dx$$

R_X is the domain of the random variable X, $g(x)$ is a continuous function, and $f(x)$ is the probability density function.

Generalized Fisher Information-Lutwak

Lutwak et al. (2012) proposed a form of generalized fisher information formulated as follows:

$$I_{p,\lambda}(\theta) = E\left[\left| (f(X; \theta))^{\lambda-1} \left(\frac{\partial}{\partial \theta} \ln f(X; \theta) \right) \right|^p \right], p > 0, \lambda > 0 \tag{4}$$

Generalized Fisher Information-Bercher
Bercher (2013) proposed two forms of generalized fisher information:
 a. Generalized fisher information-Bercher Type 1

$$I_{\beta,m}(\theta) = E\left[(f(X;\theta))^{\beta(m-1)}\left|\frac{\partial}{\partial\theta}\ln f(X;\theta)\right|^{\beta}\right], \beta > 1, m \geq 0 \qquad (5)$$

Eq.(5) is an extension of Eq.(4).

b. Generalized fisher information-Bercher Type 2

$$I_{\beta,m}^{*}(\theta) = \frac{k}{M_q(f)}I_{\beta,m}(\theta), \qquad (6)$$

where

$$M_k(f) = \int_{R_X}(f(x))^k dx, \qquad k \geq 0$$

The Jeffrey Prior

The Jeffrey prior (Jeffrey, 1967) proposes the following prior form:

$$\pi(\theta) \propto \sqrt{I(\theta)} \qquad (7)$$

Eq.(7) implicitly states the existence of a Jeffrey prior to the standard error θ. The relation between standard error and Fisher information is formulated as follows:

$$V(\theta) = \frac{1}{I(\theta)} \rightarrow s.e(\theta) = \frac{1}{\sqrt{I(\theta)}}, \qquad (8)$$

where $s.e(\theta)$ is the standard error of the parameter θ. Based on Eq. (8), Eq. (7) becomes:

$$\pi(\theta) \propto \frac{1}{s.e(\theta)} \qquad (9)$$

Based on Eq.(9), the Jeffrey prior is also proportional to the root of the inverse standard error θ.

3. Results and Discussion
The Extended Jeffrey Prior

In this study, the author proposed an extended form of Jeffrey prior based on generalized Fisher informations (2) - (6) by adopting the formulation (7) to obtain several types of priors as follows:

$$\pi_s(\theta) \propto \sqrt{I_s(\theta)} \qquad (10)$$

$$\pi_q(\theta) \propto \sqrt{I_q(\theta)} \qquad (11)$$

$$\pi_{p,\lambda}(\theta) \propto \sqrt{I_{p,\lambda}(\theta)} \qquad (12)$$

$$\pi_{\beta,m}(\theta) \propto \sqrt{I_{\beta,m}(\theta)} \qquad (13)$$

$$\pi_{\beta,m}^{*}(\theta) \propto \sqrt{I_{\beta,m}^{*}(\theta)} \qquad (14)$$

Eqs. (10) - (14) give a prior that is not close form. This contrasts Jeffrey's prior, which uses original fisher information that gives a close form to some probability distributions. For this reason, the posterior formulation using the extended Jeffrey priors (10) − (14) provides a more complex form than the posterior with the Jeffrey

prior so that the additional parameters contained in the extended of the Jeffrey prior can be assumed to be random variables that under a particular distribution.

Example Studies
The Exponential Distribution
Suppose $X_1, \dots, X_n$ under an exponential distribution with rate θ,
$$f(x; \theta) = \theta e^{-\theta x}; \theta > 0, x \geq 0 \tag{15}$$
$$\ln f(x; \theta) = -\theta x + \ln \theta$$
$$\frac{\partial}{\partial \theta} \ln f(X; \theta) = -x + \frac{1}{\theta}$$

So the Generalized fisher information-Boekee of Eq.(15) is

$$I_s(\theta) = \left[E \left[\left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} \right] \right]^{s-1}, s \geq 1. \tag{16}$$

Based on Eq.(16), the extended Jeffrey priors can be formed:

a. $\pi_s(\theta)$

$$\pi_s(\theta) \propto \sqrt{I_s(\theta)}$$

$$\propto \sqrt{\left[E \left[\left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} \right] \right]^{s-1}}$$

$$\propto \left[E \left[\left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} \right] \right]^{\frac{s-1}{2}}, \frac{s-1}{2} \geq 0, \frac{s}{s-1} \geq 0 \tag{17}$$

where:

$$E \left[\left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} \right] = \int_0^\infty (\theta e^{-\theta x}) \left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} dx \tag{18}$$

b. $\pi_q(\theta)$

$$\pi_q(\theta) \propto \sqrt{I_q(\theta)}$$

$$\propto \sqrt{E_q \left[\left(\frac{\partial}{\partial \theta} \left[\frac{(\theta e^{-\theta x})^{1-q} - 1}{1 - q} \right] \right)^2 \right]}$$

$$\propto \sqrt{E_q \left[\left(\frac{\partial}{\partial \theta} \left[\frac{\theta^{1-q} e^{-\theta x(1-q)} - 1}{1 - q} \right] \right)^2 \right]}$$

$$\propto \sqrt{E_q[(\theta^{-q}(1 - \theta x)e^{x(-1+q)\theta})^2]} \tag{19}$$

where

$$E_q\left[\left(\theta^{-q}(1-\theta x)e^{x(-1+q)\theta}\right)^2\right] = \int_0^\infty (\theta e^{-\theta x})^q \left(\theta^{-q}(1-\theta x)e^{x(-1+q)\theta}\right)^2 dx \quad (20)$$

c. $\pi_{p,\lambda}(\theta)$

$$\pi_{p,\lambda}(\theta) \propto \sqrt{I_{p,\lambda}(\theta)}$$

$$\propto \sqrt{E\left[\left|(\theta e^{-\theta x})^{\lambda-1}\left(-x+\frac{1}{\theta}\right)\right|^p\right]}$$

$$\propto \sqrt{E\left[\left|-x(\theta e^{-\theta x})^{\lambda-1}+\frac{(\theta e^{-\theta x})^{\lambda-1}}{\theta}\right|^p\right]} \quad (21)$$

where

$$E\left[\left|-x(\theta e^{-\theta x})^{\lambda-1}+\frac{(\theta e^{-\theta x})^{\lambda-1}}{\theta}\right|^p\right]$$

$$= \int_0^\infty (\theta e^{-\theta x})\left|-x(\theta e^{-\theta x})^{\lambda-1}+\frac{(\theta e^{-\theta x})^{\lambda-1}}{\theta}\right|^p dx \quad (22)$$

4. $\pi_{\beta,m}(\theta)$

$$\pi_{\beta,m}(\theta) \propto \sqrt{I_{\beta,m}(\theta)}$$

$$\propto \sqrt{E\left[(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]}, \quad (23)$$

where

$$E\left[(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]$$

$$= \int_0^\infty (\theta e^{-\theta x})(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta dx \quad (24)$$

5. $\pi_{\beta,m}^*(\theta)$

$$\pi_{\beta,m}(\theta) \propto \sqrt{I_{\beta,m}^*(\theta)}$$

$$\propto \sqrt{\left(\frac{k}{\int_0^\infty (\theta e^{-\theta x})^k dx}\right)I_{\beta,m}(\theta)}$$

$$\propto \sqrt{\left(\frac{k}{\int_0^\infty (\theta e^{-\theta x})^k dx}\right)E\left[(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]} \quad (25)$$

where

$$E\left[(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]$$

$$= \int_0^\infty (\theta e^{-\theta x})(\theta e^{-\theta x})^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta dx \quad (26)$$

The Form of Posteriors

Suppose $X_1, \dots, X_n$ follows an exponential distribution with rate θ. The likelihood function is:

$$f(x_1, \dots, x_n | \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \tag{27}$$

Posterior of θ using Jeffrey's extended prior

a. Posterior based on $\pi_s(\theta)$

$$\propto f(x_1, \dots, x_n | \theta) \pi_s(\theta) \qquad \begin{array}{c} f_s(\theta | x_1, \dots, x_n) \\ (28) \end{array}$$

Insert Eqs. (17), (18), and (27) into Eq.(28) to obtain:

$$f_s(\theta | x_1, \dots, x_n) \propto f(x_1, \dots, x_n | \theta) \pi_s(\theta)$$

$$f_s(\theta | x_1, \dots, x_n) \propto \theta^n e^{-\theta \sum_{i=1}^n x_i} \left[E\left[\left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} \right] \right]^{\frac{s-1}{2}}$$

$$\propto \theta^n e^{-\theta \sum_{i=1}^n x_i} \left[\int_0^\infty (\theta e^{-\theta x}) \left| -x + \frac{1}{\theta} \right|^{\frac{s}{s-1}} dx \right]^{\frac{s-1}{2}}, s$$

$$\geq 1 \qquad (29)$$

b. Posterior based on $\pi_q(\theta)$

$$\propto f(x_1, \dots, x_n | \theta) \pi_q(\theta) \qquad \begin{array}{c} f_q(\theta | x_1, \dots, x_n) \\ (30) \end{array}$$

Insert Eqs. (19), (20), and (27) into Eq. (30) to obtain:

$$f_q(\theta | x_1, \dots, x_n) \propto f(x_1, \dots, x_n | \theta) \pi_q(\theta)$$

$$\propto \theta^n e^{-\theta \sum_{i=1}^n x_i} \sqrt{E_q[(\theta^{-q}(1 - \theta x)e^{x(-1+q)\theta})^2]}$$

$$\propto \theta^n e^{-\theta \sum_{i=1}^n x_i} \left[\int_0^\infty (\theta e^{-\theta x})^q (\theta^{-q}(1 \right.$$

$$\left. - \theta x)e^{x(-1+q)\theta})^2 dx \right]^{\frac{1}{2}} \qquad (31)$$

c. Posterior based on $\pi_{p,\lambda}(\theta)$

$$\propto f(x_1, \dots, x_n | \theta) \pi_{p,\lambda}(\theta) \qquad \begin{array}{c} f_{p,\lambda}(\theta | x_1, \dots, x_n) \\ (32) \end{array}$$

Insert Eqs.(21), (22), and (27) into Eq. (32) to obtain:

$$f_{p,\lambda}(\theta|x_1,\dots,x_n) \propto f(x_1,\dots,x_n|\theta)\pi_{p,\lambda}(\theta)$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \sqrt{E\left[\left|-x(\theta e^{-\theta x})^{\lambda-1} + \frac{(\theta e^{-\theta x})^{\lambda-1}}{\theta}\right|^p\right]}$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \left[\int_0^\infty (\theta e^{-\theta x})\left|-x(\theta e^{-\theta x})^{\lambda-1}\right.\right.$$

$$\left.\left. + \frac{(\theta e^{-\theta x})^{\lambda-1}}{\theta}\right|^p dx\right]^{\frac{1}{2}} \qquad (33)$$

d. Posterior based on $\pi_{\beta,m}(\theta)$

$$f_{\beta,m}(\theta|x_1,\dots,x_n)$$
$$\propto f(x_1,\dots,x_n|\theta)\pi_{\beta,m}(\theta) \qquad (34)$$

Insert Eqs.(23), (24), and (27) into Eq. (34) to obtain:

$$f_{\beta,m}(\theta|x_1,\dots,x_n) \propto f(x_1,\dots,x_n|\theta)\pi_{\beta,m}(\theta)$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \sqrt{E\left[(\theta e^{-\theta x}))^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]}$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \left[\int_0^\infty (\theta e^{-\theta x})(\theta e^{-\theta x}))^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta dx\right]^{\frac{1}{2}} \qquad (35)$$

e. Posterior based on $\pi_{\beta,m}^*(\theta)$

$$f_{\beta,m}^*(\theta|x_1,\dots,x_n)$$
$$\propto f(x_1,\dots,x_n|\theta)\pi_{\beta,m}^*(\theta) \qquad (36)$$

Insert Eqs. (25), (26), and (27) into Eq. (36) to obtain:

$$f_{\beta,m}^*(\theta|x_1,\dots,x_n) \propto f(x_1,\dots,x_n|\theta)\pi_{\beta,m}^*(\theta)$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \sqrt{\left(\frac{k}{\int_0^\infty (\theta e^{-\theta x})^k dx}\right)E\left[(\theta e^{-\theta x}))^{\beta(m-1)}\left|-x+\frac{1}{\theta}\right|^\beta\right]}$$

$$\propto \theta^n e^{-\theta \sum_{=1}^n x_i} \left[\left(\frac{k}{\int_0^\infty (\theta e^{-\theta x})^k dx}\right)\int_0^\infty (\theta e^{-\theta x})(\theta e^{-\theta x}))^{\beta(m-1)}\left|-x\right.\right.$$

$$\left.\left. + \frac{1}{\theta}\right|^\beta dx\right]^{\frac{1}{2}} \quad (37)$$

Eq.(37) adalah bentuk posterior yang diturunkan berdasarkan prior $\pi_{\beta,m}^*(\theta)$ (36).

4. Conclusions

The extended of Jeffrey prior form (10) - (14), which is based on several types of generalized fisher information, produces a form that is not closed-form, so the MCMC approach is an alternative. The implication is that if the extended of the Jeffrey prior is used to form a posterior so that the posterior form is not closed-form, then the estimation process involves parallel MCMC in one process. The algorithms that can be applied are Metropolis-Hastings, Gibbs sampling, and Hamiltonian Monte Carlo.

References

Bercher, J-F. 2013. Some properties of generalized fisher information in the context of nonextensive thermostatistics. *Physica A: Statistical Mechanics and its Applications*, Elsevier, 392(15), 3140-3154.

Boekee, D.E. 1977. Generalized fisher information with Application to Estimation Problems. *IFAC Proceedings*, 10(12), 75-82.

Box, G.E.P., dan Tiao, G.C. 1992. *Bayesian Inference in Statistical Analysis*. United States: John Wiley and Sons, Inc.

Furuichi, S. 2010. On generalized fisher informations and Cramér-Rao type inequalities. *Journal of Physics: Conference Series* 201, 012016.

Hogg, R.V., McKean, J.W., dan Craig, A.T. 2005. *Introduction to Mathematical Statistics, Sixth Edition*. USA: Pearson Prentice Hall.

Iriawan, N. 2020. *Materi kuliah Analisis Bayesian*.

Jeffreys, S. H. 1967. *Theory of Probability, 3th edition*. London: Oxford University Press.

Lutwak, E., Lv, S., Yang, D., dan Zhang, G. 2012. Extensions of Fisher information and Stam's inequality. IEEE Transactions on Information Theory, 58(3), 1319 –1327.