

# Prior Information in a Jeffreys Prior: Distinction between Invariance and Non-informativeness<sup>1)</sup>

OKAMOTO Yasuharu  
岡本安晴

[Summary] Most Bayesians favor a Jeffreys prior as a non-informative one. This interpretation of non-informativeness of a Jeffreys prior is based on invariance. It is pointed out here, however, that a Jeffreys prior is not non-informative. The condition of invariance of probability under parameter transformation was examined, and Jeffreys priors were characterized to be related to each other in a way that they satisfy the condition of invariance. But invariance and non-informativeness must be distinguished. The fact that a Jeffreys prior in item response theory depends on information about an examinee was pointed out. This fact indicates that a Jeffreys prior is not non-informative, although the probability that it defines is invariant.

## Introduction

The distinction between invariance and non-informativeness is important because Jeffreys priors are invariant but not non-informative. Invariance in this case concerns probability in Bayesian analysis. The basic framework of Bayesian analysis is given by a joint distribution  $p(\theta, D)$  of parameter  $\theta$  and data  $D$ . The parameter  $\theta$  is a vector when a multiparameter model is employed. This joint distribution is given by Eq. (1) (Gelman, Carlin, Stern and Rubin, 2004):

$$p(\theta, D) = p_0(\theta)p(D|\theta) \quad (1)$$

where  $p_0(\theta)$  is a probability distribution of parameter  $\theta$  and  $p(D|\theta)$  is a conditional probability distribution of data  $D$  given parameter  $\theta$ .  $p(D|\theta)$  is a model of generating data  $D$  with parameter  $\theta$ . In the Bayesian framework (1), this model  $p(D|\theta)$  is combined with knowledge about parameter  $\theta$ , which is represented by  $p_0(\theta)$ .  $p_0(\theta)$  represents knowledge before data  $D$  is given, so it is called a prior distribution.

When data  $D$  is given, knowledge about  $\theta$  is updated based on data  $D$ . That is, the conditional probability distribution of  $\theta$  given data  $D$  is calculated as follows:

$$p(\theta|D) = \frac{p(\theta, D)}{p(D)} = \frac{p_0(\theta)p(D|\theta)}{p(D)}, \quad (2)$$

where  $p(D)$  is probability of data  $D$  and given by

$$p(D) = \int p(\theta, D)d\theta$$

Equation (2) shows that a posterior distribution  $p(\theta|D)$  depends both on a data generating model  $p(D|\theta)$  and a prior distribution  $p_0(\theta)$ . If no prior information about parameter  $\theta$  is given, the prior distribution  $p_0(\theta)$  should not affect the posterior distribution  $p(\theta|D)$ . In this case, the prior distribution

$p_0(\theta)$  is called non-informative, and the most favored one is a Jeffreys prior (Marin and Robert, 2007). The non-informativeness of a Jeffreys prior is explained by invariance as to scale transformation (Gelman et al., 2004). But invariance is not equivalent to non-informativeness. A Jeffreys prior may contain prior information about parameter  $\theta$ . Before discussing the informativeness of a Jeffreys prior, the invariance of a Jeffreys prior will be examined in the next section to clarify its meaning.

### Invariance

Invariance concerning transformation of parameters is the key characteristic of a Jeffreys prior. To clarify the meaning of the invariance, consider a one-to-one onto transformation  $\tau=h(\theta)$ . The invariance means that for any measurable set  $S=\{(\theta, D)\}$ , the following equation holds:

$$\int_{h(S)} p(\tau, D)dDd\tau = \int_S p(\theta, D)dDd\theta, \text{ where } h(\theta, D)=(h(\theta), D),$$

i.e.

$$\int_{h(S)} p(\tau)p(D|\tau)dDd\tau = \int_S p(\theta)p(D|\theta)dDd\theta \quad (3)$$

The right hand side of Eq. (3) can be rewritten as follows:

$$\int_S p(\theta)p(D|\theta)dDd\theta = \int_{h(S)} p(\theta)p(D|\theta) \left| \frac{d\theta}{d\tau} \right| dDd\tau \quad (4)$$

From Eqs. (3) and (4), we have

$$\int_{h(S)} p(\tau)p(D|\tau)dDd\tau = \int_{h(S)} p(\theta)p(D|\theta) \left| \frac{d\theta}{d\tau} \right| dDd\tau$$

Because  $S$  in the above equation is any event, we have

$$p(\tau)p(D|\tau) = p(\theta)p(D|\theta) \left| \frac{d\theta}{d\tau} \right| \quad (5)$$

Since

$$p(D|\tau)=p(D|h(\theta))=p(D|\theta),$$

we have

$$p(\tau) = p(\theta) \left| \frac{d\theta}{d\tau} \right| \quad (6)$$

Any transformation  $\tau=h(\theta)$  that satisfies Eq. (6) preserves the probabilities of corresponding events by transformation  $h$ , i.e., Eq. (3) holds (invariance). One of these transformations that satisfy Eq. (6) is given by Jeffreys priors. A Jeffreys prior is given to be proportional to the square root of the Fisher information  $I(\theta)$ , i.e.,

$$p(\theta) \propto \sqrt{I(\theta)}.$$

Under transformation  $\tau=h(\theta)$ , the Fisher informations satisfy (Gelman et al., 2004)

$$I(\tau) = I(\theta) \left| \frac{d\theta}{d\tau} \right|^2. \quad (7)$$

From Eq. (7), we have

$$p(\tau) \propto \sqrt{I(\tau)} = \sqrt{I(\theta)} \left| \frac{d\theta}{d\tau} \right| \propto p(\theta) \left| \frac{d\theta}{d\tau} \right|, \quad (8)$$

that is, Jeffreys priors are related by the following equation:

$$p(\tau) = p(\theta) \left| \frac{d\theta}{d\tau} \right|.$$

This equation shows that Jeffreys priors are linked to each other in such that they satisfy Eq. (6). Hence, Jeffreys priors assure invariance of probabilities under parameter transformation. But invariance is something other than non-informativeness. Eq. (3) does not say anything about prior information. Actually, in item response theory, a Jeffreys prior reflects information about an examination of a test constructor. This point will be discussed in the next section.

### Non-informativeness

A case where a Jeffreys prior contains prior information despite the popular interpretation that a Jeffreys prior is non-informative can be seen in item response theory (IRT) (Okamoto, 2010). In IRT, the Fisher information is called the test information and is given as follows:

$$I(\theta) = \sum \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (9)$$

where  $P_i(\theta)$  is the function that defines the item characteristic curve of item  $i$  and  $Q_i(\theta)=1-P_i(\theta)$  (Baker and Kim, 2004).

The inverse of the test information  $I(\theta)$  gives an asymptotic estimation of variance of the maximum likelihood estimator of  $\theta$ , so test items can be selected to minimize variance of the estimator of  $\theta$  by maximizing  $I(\theta)$  around a probable value of  $\theta$  for an examinee. In this sense,  $I(\theta)$  reflects information about an anticipated value of the examinee and this information can be called a prior one. A Jeffreys prior is given as proportional to the square root of  $I(\theta)$ :

$$p(\theta) \propto \sqrt{I(\theta)}.$$

Hence this prior  $p(\theta)$  reflects information about examinee's  $\theta$ , and its density is concentrated around the anticipated value of  $\theta$ , where  $I(\theta)$  is constructed so that  $I(\theta)$  becomes large around the anticipated value of  $\theta$ .

### Conclusion

A Jeffreys prior is favored by most Bayesians (Marin and Robert, 2007) and is interpreted to be non-informative. This interpretation of non-informativeness is based on invariance (Eq. (3)) concerning transformation of parameters, and this invariance holds when prior distributions satisfy Eq. (6). Jeffreys priors are ones that satisfy Eq. (6). But invariance is one thing and non-informativeness another. A Jeffreys prior assures invariance of probabilities but may accompany some information about an object to be analyzed. In fact, it was pointed out in this paper that a Jeffreys prior in calculating a posterior distribution

of ability parameter  $\theta$  in IRT contains prior information about  $\theta$ . A Jeffreys prior is invariant and also informative at least in the case of analyzing parameter  $\theta$  for a battery of test items that are chosen optimally to make the variance of the estimator small.

#### References

- Baker, F. B. and Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques, 2<sup>nd</sup> ed., Revised and expanded*. Marcel Dekker, Inc.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis, 2<sup>nd</sup> ed.* Chapman & Hall/CRC.
- Marin, J.-M. and Robert, C. P. (2007). *Bayesian core: A practical approach to computational Bayesian statistics*. Springer.
- Okamoto, Y. (2010). Some problems in using a Fisher information matrix. *Proceedings of the 8<sup>th</sup> Annual Conference of the Japanese Association for Research on Testing*. 71-72(in Japanese).

---

1) This research was supported in part by Grants-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science and Technology, 20530668.