Statistics and Probability Letters 137 (2018) 292–296

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Objective Bayesian inference for the intraclass correlation coefficient in linear models

Duo Zhang [a], Min Wang [b,*]

[a] Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, United States
[b] Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, United States

ARTICLE INFO

Article history:
Received 26 June 2017
Received in revised form 18 December 2017
Accepted 2 February 2018
Available online 21 February 2018

Keywords:
Bayes factor
Divergence-based prior
Hypothesis testing
Intraclass model
Objective priors

ABSTRACT

We outline objective Bayesian testing procedure for the intraclass correlation coefficient in linear models. For it, we derive the Bayes factors based on the divergence-based priors, which have unidimensional integral expressions and can thus be easily approximated numerically.

1. Introduction

Consider the intraclass model of the form $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,\ i = 1, 2, \ldots, n$, where $\mathbf{y}_i$ is a $k \times 1$ ($k \geq 2$) vector of response variables, $\mathbf{X}_i$ is a $k \times p$ design matrix of $(p-1)$ regressors (assuming the first column is ones) with $p < k$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters. We assume that the random error $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N_k(\mathbf{0}_k, \sigma^2\mathbf{V})$, where $\overset{iid}{\sim}$ stands for "independent and identically distributed", $\mathbf{0}_k$ is a $k \times 1$ vector of zeros, and $\mathbf{V} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ with $\mathbf{I}_k$ being a $k \times k$ identity matrix and $\mathbf{J}_k$ being a $k \times k$ matrix containing only ones. The parameter ρ is often referred as the intraclass correlation coefficient (for short, ICC). Note that $\rho \in (-(k-1)^{-1}, 1)$ is the necessary and sufficient condition for positive-definiteness of $\mathbf{V}$.

By letting $\mathbf{y}' = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)$, $\mathbf{X}' = (\mathbf{X}'_1, \ldots, \mathbf{X}'_n)$, and $\mathbf{y}' = (\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_n)$, the intraclass model can be represented as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_{nk}(\mathbf{0}_{nk}, \sigma^2\mathbf{W})$, where $\mathbf{W} = \mathbf{I}_n \otimes \mathbf{V}$ with $\otimes$ being the Kronecker product. Let $\boldsymbol{\nu} = (\sigma^2, \boldsymbol{\beta})$. We are interested in testing $H_1 : \rho = 0$ versus $H_2 : \rho \neq 0$, or equivalently, using the model selection notation, in comparing two competing models

$$M_1 : f_1(\mathbf{y} \mid \boldsymbol{\nu}) = f(\mathbf{y} \mid 0, \boldsymbol{\nu}) \quad \text{versus} \quad M_2 : f_2(\mathbf{y} \mid \rho, \boldsymbol{\nu}) = f(\mathbf{y} \mid \rho, \boldsymbol{\nu}). \tag{1}$$

The ICC has a lengthy history of practical applications as a coefficient of reliability. For example, in the multilevel modeling, ICC is often adopted to measure the strength of correlation in a hierarchical data, which helps researchers determine if the uncorrelatedness assumption is violated in the data. Another practical example is the following, extracted from Chapter 5.2 of Frees (2004): twenty-seven individuals including 16 boys and 11 girls were measured for distances from the pituitary to the pteryomaxillary fissure in millimeters, at ages 8, 10, 12, and 14. In this case, the distance y_{ij} measured in millimeters is the response for individual i measured at age j, the design matrix consists of two columns with the first being age and

* Corresponding author.
E-mail address: min.wang@ttu.edu (M. Wang).

https://doi.org/10.1016/j.spl.2018.02.004
0167-7152/

the second being gender, and $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_4, \sigma^2 \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_4 + \rho \mathbf{J}_4$. We are interested in studying how strong the individuals resemble each other (i.e., $\rho = 0$, where ρ represents the resemblance among individuals).

As suggested by Berger and Pericchi (2001), we adopt the Bayesian approach to address the model selection problem in (1). Although there exist several Bayesian alternatives (see, for example, Ghosh and Heo, 2003; Lee and Kim, 2006), the hypothesis testing of the ICC has not well been studied from an objective Bayesian perspective. We here consider the Bayes factor (Kass and Raftery, 1995), because it has an intuitive meaning of "measure of evidence" in favor of a model under the hypotheses. The Bayes factor (BF) in favor of M_2 and against M_1 can be expressed as

$$\mathrm{BF}_{21} = \frac{p(\mathbf{y} \mid M_2)}{p(\mathbf{y} \mid M_1)} = \frac{\int f_2(\mathbf{y} \mid \rho, \boldsymbol{\nu}) \pi_2(\rho, \boldsymbol{\nu})\, d\rho\, d\boldsymbol{\nu}}{\int f_1(\mathbf{y} \mid \boldsymbol{\nu}) \pi_1(\boldsymbol{\nu})\, d\boldsymbol{\nu}}, \tag{2}$$

where $\pi_1(\boldsymbol{\nu})$ and $\pi_2(\rho, \boldsymbol{\nu})$ are the prior probabilities under M_1 and M_2, respectively. When $\mathrm{BF}_{21} > (<)1$, it indicates the data are more likely to have occurred under M_2 (M_1). For instance, $\mathrm{BF}_{21} = 5$ indicates that the data are 5 times more likely under M_2 than under M_1 ($\mathrm{BF}_{12} = 1/\mathrm{BF}_{21} = .2$). A set of verbal labels to categorize the evidential impact in terms of the values of the BF was provided by Jeffreys (1961) and further illustrated by Kass and Raftery (1995). The posterior probability of M_1 given the data is $p(M_1 \mid \mathbf{y}) = \left[1 + \mathrm{BF}_{21} p(M_2)/p(M_1)\right]^{-1}$, where $p(M_2)/p(M_1)$ is the prior model odds between two models, which is assumed to be 1 in this paper.

A critical ingredient of deriving the BF is to specify priors for the unknown parameters. Direct use of noninformative priors, such as the Jeffreys prior (Jeffreys, 1961), often result in the BF containing undefined constants. Bayarri and García-Donato (2008) proposed an attractive way to obtain noninformative while proper priors, (so-called the divergence-based (DB) priors). Since then, the DB priors have been implemented for Bayesian hypothesis testing; see, for example, García-Donato and Sun (2007) for the one-way random-effects model, Kim et al. (2017) for linear models with first-order autoregressive residuals. We here derive the DB priors and their resulting BFs for the model selection problem in (1). Numerical results show that they perform very well in terms of the sum of two error probabilities, i.e., the probability of incorrectly choosing M_2 while M_1 is true and the probability of incorrectly choosing M_1 while M_2 is true, respectively.

The remainder of this paper is organized as follows. In Section 2, we derive the DB priors and their resulting BFs. In Section 3, we conduct simulations to evaluate the performance of the BFs. Some concluding remarks are provided in Section 4, with proofs given in the supplementary file.

2. The DB priors and the resulting BFs

2.1. Objective priors for the unknown parameters

We here used the orthogonal reparameterization to the parameters, which means that if the parameters are orthogonal if their expected Fisher information matrix is diagonal. This would justify the use of same (even improper) priors for the orthogonal parameters (see Kass and Vaidyanathan, 1992). We follow the orthogonal reparameterization of Ghosh and Heo (2003) and let $\theta_1 = \rho,\ \theta_2 = \frac{1}{\sigma^2}(1-\rho)^{-(k-1)/k}\big(1+(k-1)\rho\big)^{-1/k}$, and $\boldsymbol{\theta}_3 = \boldsymbol{\beta}$. The model selection problem in (1) becomes comparing two models

$$\begin{aligned} &M_1 : f_1(\mathbf{y} \mid \theta_0 = 0, \theta_2, \boldsymbol{\theta}_3) = N_{nk}\big(\mathbf{X}\boldsymbol{\theta}_3, \theta_2^{-1}\mathbf{I}_{nk}\big), \\ &M_2 : f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) = N_{nk}\big(\mathbf{X}\boldsymbol{\theta}_3, \theta_2^{-1}\boldsymbol{\Sigma}\big), \end{aligned} \tag{3}$$

where $\boldsymbol{\Sigma} = (1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k}\mathbf{W}$ and $\mathbf{V} = (1-\theta_1)\mathbf{I}_k + \theta_1\mathbf{J}_k$. We here focus on the second-order matching prior for $(\theta_1, \theta_2, \boldsymbol{\theta}_3)$ due to its nice frequentist coverage probability (Datta and Mukerjee, 2004). This prior (Theorem 1 of Ghosh and Heo, 2003) under M_2 is given by

$$\pi^{\mathrm{N}}(\theta_1, \theta_2, \boldsymbol{\theta}_3) \propto (1-\theta_1)^{-1}\big(1+(k-1)\theta_1\big)^{-1}\theta_2^{-1}.$$

Under the orthogonal transformation (θ_1 is orthogonal to $\boldsymbol{\theta}_2$ and θ_3), $\boldsymbol{\nu} = (\boldsymbol{\theta}_2, \theta_3)$ can be viewed as common parameters of both models in (1) and assumed to have the same meaning to both models. This allows us to adopt the improper prior $\pi^{\mathrm{N}}(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1}$, which shows that a noninformative prior for θ_1 can be written as

$$\pi^{\mathrm{N}}(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) \propto (1-\theta_1)^{-1}\big(1+(k-1)\theta_1\big)^{-1}. \tag{4}$$

For the unknown parameters of two models in (3), we consider $\pi^{\mathrm{N}}(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1}$ under M_1 and $\pi^{\mathrm{N}}(\theta_1, \theta_2, \boldsymbol{\theta}_3) \propto \pi^{\mathrm{N}}(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)\pi^{\mathrm{N}}(\theta_2, \boldsymbol{\theta}_3)$ under M_2.

2.2. The DB priors

The DB priors are designed to use other formal rules to construct objective priors of the new parameters under the alternative hypothesis and are derived based on the measure of the direct Kullback–Leibler (KL) divergence of two competing models, raised to a negative power. The KL divergence between M_1 and M_2 in (3) is given by

$$\mathrm{KL}[\theta_0 : \theta_1] = \int \log \frac{f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)}{f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)} f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) d\mathbf{y}. \tag{5}$$