Proceedings of Joint Conference of Korean Data And Information Science Society and The Korean Data Analysis Society,
April 28 - 29, 2006, pp. 1-6

Matching Conditions for Predicting the Random Effects in ANOVA Models

In Hong Chang[1]

Abstract

We consider the issue of Bayesian prediction of the unobservable random effects, And we characterize priors that ensure approximate frequentist validity of posterior quantiles of unobservable random effects. Finally we show that the probability matching criteria for prediction of unobservable random effects in one-way random ANOVA model.

1. Introduction

Consider random variables $Y_1, \cdots, Y_n$ and $\xi_1, \cdots, \xi_n$, with their joint distribution indexed by an unknown parameter $\theta = (\theta_1, \cdots, \theta_p)^T$, such that

(a) given $\xi_1, \cdots, \xi_n$, conditionally $Y_1, \cdots, Y_n$ are independent and Y_i has a density $f(y_i|\xi_i, \theta)$;

(b) marginally $\xi_1, \cdots, \xi_n$ are independent and identically distributed (iid), each with density $g(.;\theta)$;

(c) $Y_1, \cdots, Y_n$ are observable but $\xi_1, \cdots, \xi_n$ are unobservable and interest lies in predicting the random effects ξ_i on the basis of $Y_1, \cdots, Y_n$.

The above setup is motivated by ANOVA models with random effects though it can arise in other situations too. Thus in the one-way random ANOVA setting with n classes and $k(\geq 2)$ observations $Y_{i1}, \cdots, Y_{ik}$ in the ith class $(1 \leq i \leq n)$, we have the model

$$Y_{ij} = \theta_1 + \xi_i + e_{ij} \quad (1 \leq j \leq k,\ 1 \leq i \leq n), \tag{1.1}$$

where θ_1 is the unknown general mean, ξ_i is the unobservable random effect associated with the ith class and e_{ij} is random error. As usual, suppose

(i) $\xi_1, \cdots, \xi_n$ are iid, each normal with mean zero and unknown variance θ_2,

---

[1]Assistant professor, Department of Computer Science and Statistics, Chosun University, Gwangju 501-759, Korea. E-mail : ihchang@chosun.ac.kr

(ii) the $\{e_{ij}\}$ are iid, each normal with mean zero and unknown variance θ_3,

(iii) the $\{\xi_i\}$ and the $\{e_{ij}\}$ are independent.

Write $Y_i=(Y_{i1},\cdots,Y_{ik})^T$. Then the setup of the previous paragraph arises. Here $f(y_i|\xi_i,\theta)$ represents the k-variate normal density with mean vector $(\theta_1+\xi_i)1_k$ and covariance matrix $\theta_3 I_k$, and $g(.;\theta)$ stands for the univariate normal density with mean zero and variance θ_2. We write I_k for the identity matrix of order k and 1_k for the $k\times 1$ vector with each element unity.

Returning to the general setup as described in the beginning of this section, we consider the issue of Bayesian prediction of the unobservable random effect ξ_i on the basis of $Y_1,\cdots,Y_n$ using a prior density $\pi(.)$ on θ. Specifically, as in the ANOVA setting stated above, we consider scalar ξ_i but allow $Y_1,\cdots,Y_n$ to be possibly vector-valued, and characterize prior $\pi(.)$ which ensure approximate frequentist validity, as $n\to\infty$, of the posterior quantiles of ξ_i. Prior of this kind are known as probability matching priors. As noted in Tibshirani (1989), such priors are in a sense noninformative; moreover, they provide a means for getting accurate frequentist confidence or predictive regions that enjoy a Bayesian interpretation as well.

Probability matching priors have been studied extensively in the context of estimation of parameters; see Ghosh and Mukerjee (1998) and Mukerjee and Reid (1999) for references. The exploration of such priors for prediction, rather than estimation, is on the hand of much recent origin. Datta et al., (2000); initiated work in this area considering the prediction of a future observation in a sequence of iid random variables. Datta and Mukerjee (2003) reported further results for the situation where, with an independent random variable X and a dependent random variable Y, interest lies in predicting Y in a current pair on the basis of past pairs of observations on both X and Y and knowledge of X in the current pair. None of these settings, however, covers the situation considered in the present article. For example, here the random effect ξ_i is unobservable for each i, whereas is Datta and Mukerjee (2003) both X and Y are known in the past of observations. Because of this reason, we first indicate a characterization for probability matching priors in the present setup in Section 2. The characterization is then applied to one-way ANOVA model with random effects in Section 3.

2. Probability matching conditions

We continue with the general setup described in the first paragraph of Section 1. Note that marginally $Y_1,\cdots,Y_n$ are iid with common density

$$\psi(y_i;\ \theta) = \int_{-\infty}^{\infty} f(y_i|\xi_i,\theta)g(\xi_i;\theta)d\xi_i. \tag{2.1}$$

Let $h(\xi_i|y_i,\theta)$ be the conditional density of ξ_i given $Y_i\ (1\le i\le n)$. Along the line of Ghosh and Mukerjee (1993), we work essentially under the assumptions of Johnson (1970) and also need the Edgeworth assumptions of Bickel and Ghosh (1990). The parameter space for θ is an open set in R^p. The per observation (i.e., per Y_i, as given by (2.1)) Fisher information matrix $I(\theta)$ is assumed to be positive definite for all θ. We work with a prior density $\pi(.)$ that is positive and thrice continuously differentiable. Formal expansions for the posterior, as used here, are valid for sample points in a set S with P_θ-probability $1+o(n^{-1})$ uniformly over compact subsets of θ; cf. Bickel and Ghosh (1990).

Let $\lambda(\theta) = n^{-1}\sum_{i=1}^{n}\log\psi(y_i;\theta)$ and $\hat{\theta}$ be the maximum likelihood estimate of θ based on $y_1,\cdots,y_n$, where y_i is the realized value of Y_i. With $D_j\equiv d/d\theta_j$, let

$$c_{jr} = -\{D_jD_r\lambda(\theta)\}_{\theta=\hat{\theta}},\ a_{jrs} = \{D_jD_rD_s\lambda(\theta)\}_{\theta=\hat{\theta}},\ \pi_j(\theta) = D_j\pi(\theta),$$
$$h_j(\xi_i|y_i,\theta) = D_jh(\xi_i|y_i,\theta),\ h_{jr}(\xi_i|y_i,\theta) = D_jD_rh(\xi_i|y_i,\theta).$$

The matrix $C=((c_{jr}))$ is positive definite over S. Let $C^{-1}=((c^{jr}))$.

Let $\tilde{\pi}(\xi_i|d)$ be the posterior predictive density of ξ_i, given the data $d=(y_1,\cdots,y_n)$, under the prior $\pi(.)$. One can check that $\tilde{\pi}(\xi_i|d)$ is the expectation of $h(\xi_i|y_i,\theta)$ with reference to the posterior density of θ given d under $\pi(.)$. Now, an expansion for the posterior density of $n^{1/2}(\theta-\hat{\theta})$ is available in Ghosh and Mukerjee (1993). Using this together with an expansion for $h(\xi_i|y_i,\theta)$ about $\hat{\theta}$, one gets after some algebra

$$\tilde{\pi}(\xi_i|d) = h(\xi_i|y_i,\hat{\theta}) + \frac{1}{2n}\left[c^{st}\left\{c^{jr}a_{jrs} + \frac{2\pi_s(\hat{\theta})}{\pi(\hat{\theta})}\right\}h_t(\xi_i|y_i,\hat{\theta}) + c^{jr}h_{jr}(\xi_i|y_i,\hat{\theta})\right] + o(n^{-1}). \tag{2.2}$$

In (2.2) and elsewhere we follow the summation convention with all implicit sums ranging from 1 to p. Eq. (2.2) resembles Eq. (2.1) of Datta et al., (2000) although our setting is quite different from theirs. This, however, enables us to follow the line of Datta et al.,(2000), via the use of a shrinkage argument popular in Bayesian asymptotics, to obtain the probability matching condition for the present problem. To save space, we omit the details and present only the final result after stating the requisite notation.

Let $\{I(\theta)\}^{-1}=((I^{st}))$. For $0<\alpha<1$, let $q(\theta,\alpha,y_i)$ be such that

$$\int_{q(\theta,\alpha,y_i)}^{\infty} h(\xi_i|y_i,\theta)d\xi_i=\alpha \tag{2.3}$$

and define

$$V_i(\theta,\alpha)=E_\theta\left\{\int_{q(\theta,\alpha,Y_i)}^{\infty} h(\xi_i|y_i,\theta)d\xi_i\right\} \tag{2.4}$$

Then it can be seen that a prior $\pi(.)$ is probability matching in the present setup, in the sense of ensuring frequentist validity of the posterior quantiles of ξ_i with margin of error $o(n^{-1})$ if and only if it satisfies the partial differential equation

$$D_s\{I^{st}V_t(\theta,\alpha)\pi(\theta)\}=0. \tag{2.5}$$

Eq. (2.5) again formally resembles the corresponding matching condition in Datta and Mukerjee (2003). However, this similarity is superficial since the meaning of our notation is different from theirs. For example, in (2.5) I^{st} refers to the per observation information matrix based on Y alone (vide (2.1)) whereas in their setup it refers to the per observationl pair information matrix with both the independent and dependent random variables included in the pair. Moreover, as seen in the rest of the present article, (2.5) facilitates new applications that could not be handled by the previously available results.

3. Example: application to one-way ANOVA model

We now consider ANOVA models with random effects that were the main motivation for considering the present problem. We first consider the one-way random ANOVA model (1.1) with fixed k, the notation being as in the second paragraph of the Introduction. Then marginally $Y_1,\cdots,Y_n$ are iid, each k-variate normal with mean vector $\theta_1 1_k$ and covariance matrix $\theta_2 J_k+\theta_3 I_k$, where $J_k=1_k 1_k^T$. Hence it can be seen that

$$\begin{aligned}&I^{11}=(k\theta_2+\theta_3)/k,\ I^{12}=I^{13}=0,\ I^{22}=2\{(k\theta_2+\theta_3)^2+(k-1)^{-1}\theta_3^2\}/k^2,\\&I^{23}=-2\theta_3^2\{k(k-1)\},\ I^{33}=2\theta_3^2/(k-1).\end{aligned} \tag{3.1}$$

Here $h(\xi_i|y_i,\theta)$ is the univariate normal density with mean $\mu(y_i,\theta)$ and variance $M(\theta)$, where

$$\mu(y_i,\theta)=\{k\theta_2/(k\theta_2+\theta_3)\}(\overline{y_i}-\theta_1),\ M(\theta)=\theta_2\theta_3/(k\theta_2+\theta_3) \tag{3.2}$$

and $\overline{y_i}=(y_{i1}+\cdots+y_{ik})/k$. Hence, as in Example 2, $q(\theta,\alpha,y_i)=\mu(y_ij,\theta)+z\{M(\theta)\}^{1/2}$, and

(2.8) holds. Hence by (2.4) and (3.2), $V_1(\theta, \alpha)$ is free from θ_1, and

$$V_2(\theta, \alpha)=\frac{1}{2} z\phi(z)\theta_3/\{\theta_2(k\theta_2+\theta_3)\},\quad V_3(\theta, \alpha)=\frac{1}{2} z\phi(z)\theta_2/\{\theta_3(k\theta_2+\theta_3)\} \tag{3.3}$$

We consider a natural class of priors of the form

$$\pi(\theta)=\theta_2^r\theta_3^s(k\theta_2+\theta_3)^t, \tag{3.4}$$

where r, s and t are any real number. By (3.1), (3.3) and the fact that $V_1(\theta, \alpha)$ is free from θ_1, it can be seen after considerable algebra that the unique prior of the form (3.4) that satisfies the matching condition (2.5) is $\pi(\theta)=\{\theta_2/\theta_3\}(k\theta_2+\theta_3)^{(3-k)/(k-1)}$. It can also be checked that this prior guarantees the propriety of the posterior. As in Example 2, this prior is different form probability matching priors for interval estimation of θ_2 or θ_3 (vide Peers, 1965). In addition, it is different from the prior $\{\theta_3(k\theta_2+\theta_3)\}^{-1}$, which also has the form (3.4) and uniquely ensures approximate frequentist validity of simultaneous Bayesian inference on θ_1, θ_2/θ_3 and θ_3 (vide Datta, 1996). This again shows that the matching criteria for prediction of unobservable random effects and estimation of parameters can yield different results.

References

[1] Bickel, P. J., Ghosh, J. K. (1990). A decomposition for the likelihood ratio statistic and Bartlett correction—a Bayesian argument, *Ann. Statist.,* 18, 1070-1090.

[2] Datta, G. S. (1996). On priors providing frequentist validity of Bayesian inference for multiple parametric functions, *Biometrika,* 83, 287-298.

[3] Datta, G. S., Mukerjee, R. (2003). Probabaility matching priors for predicting a dependent variable with application to regression models, *Ann. Inst. Statist. Math.,* to appear.

[4] Datta, G. S., Ghosh, M., Mukerjee, R. (2000). Some new results on probability matching priors, *Calcutta Statist. Assoc. Bull.* (Golden Jubilee Issue), 50, 179-192.

[5] Datta, G. S., Mukerjee, R., Ghosh, M., Sweeting, T. J. (2000). Bayesian prediction with approximate frequentist validity, *Ann. Statist.,* 28, 1414-1426.

[6] Ghosh, J. K., Mukerjee, R. (1993). Frequentist validity of highest posterior density regions in the multiparameter case, *Ann. Inst. Statist. Math.,* 45, 293-302.

[7] Ghosh, M., Mukerjee, R. (1998). Recent developments on probability matching priors, In: Ahmed, S.E., et al. (Ed.), *Applied Statistical Science* Ⅲ. Nova Science Publishers, New York,

pp. 227-252.

[8] Johnson, R. A. (1970). Asymptotic expansions associated with posterior distribution, *Ann. Math. Statist,* 41, 851-864.

[9] Mukerjee, R., Reid, N. (1999). On a property of probability matching priors: matching the alternative coverage probabilities, *Biometrika,* 86, 333-340.

[10] Peers, H. W. (1965). On confidence sets and Bayesian probability points in the case of several parameters, *J. Roy. Statist. Soc.,* B, 27, 9-16.

[11] Tibshirani, R. (1989). Noninformative priors for one parameter of many, *Biometrika,* 76, 604-608.

[12] Welch, B., Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods, *J. Roy. Statist. Soc.*, B, 25, 318-329.