Listy Biometryczne - Biometrical Letters
Vol. 36(1999), No. 1, 61-67

Bayesian estimation of the intraclass correlation coefficients in the multivariate one-way model

Teresa H. Jelenkowska

Institute of Applied Mathematics University of Agriculture,
Akademicka 13, 20-934 Lublin, Poland

SUMMARY

The paper presents a procedure to estimate the matrix of intraclass correlation coefficients from the Bayesian point of view. Multivariate measures of intraclass correlations are introduced to assess the degree of resemblance between family members with respect to more than one characteristic. Unified estimators for the multivariate measures are proposed as the posterior means of the multivariate inverted beta distribution. The proposed method is illustrated with the one-way model. An illustrative example is included.

KEY WORDS: Baesian estimation, MANOVA model, one-way layout, multivariate inverted beta distribution, intraclass correlations.

1. Introduction

In this paper we develop a Bayesian procedure for problems in the multivariate analysis of variance (MANOVA). We study the one-way classification, and we adopt a random-effect model (the extension to multiway layouts is straightforward). We develop posterior analysis for the intraclass correlation coefficients. A related problem for the univariate case was discussed in Box and Tiao (1973), Palmer and Broemeling (1990), Jelenkowska (1996) and Srivastawa (1993). The last paper concerns the estimation of intraclass correlation coefficients, but not from the Bayesian point of view. Recent theory and classical methodology for inference concerning the intraclass correlation coefficient in the one-way model was reviewed by Donner (1986). Non-Bayesian multivariate measures of the intraclass correlation in the one-way model were introduced by Konishi et al. (1991). Unified estimators for the multivariate measures were proposed as the eigenvalues of certain random matrices constructed by the matrices of the weighted sums of squares and products of observations.

In this paper, the main point in the posterior analysis is the approximate posterior distribution of the covariance matrices associated with the random effects obtained by Jelenkowska (1995). Section 2 describes the one-way model, the likelihood function and the prior density function for the unknown parameters. In section 3 we obtain the marginal posterior density for the function of the intraclass correlation coefficients (called generalized heritability in genetics, see Narain, 1990). The mean of this posterior density, which is a multivariate inverted beta distribution, will be used as its estimator. Bayesian estimator for the matrix of the intraclass correlation coefficients will be derived in Section 4. In Section 5, we provide an application of our approach to a data set.

2. Model and assumptions

Adopt the model

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \mathbf{b}_i + \mathbf{e}_{ij}, \quad i = 1, ..., m, \ j = 1, ..., n_i, \tag{2.1}$$

where $\mathbf{y}_{ij}$ denotes a p-vector of observations representing the j-th replication in the i-th population, $\boldsymbol{\mu}$ denotes the grand mean, $\mathbf{b}_i$ denotes the random effect due to population i ans $\mathbf{e}_{ij}$ denotes a disturbance term. This model can be written in the matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{UB} + \mathbf{E}, \tag{2.2}$$

where $\mathbf{Y}' = (\mathbf{y}_{11}, ..., \mathbf{y}_{1n_1}; ...; \mathbf{y}_{mn_1}, ..., \mathbf{y}_{mn_m})$ denotes the $p \times n$ matrix of observations, $\mathbf{X} = \mathbf{J}_n$, a $(n \times 1)$ vector of ones, with $n = \sum_{i=1}^{m} n_i$, $\boldsymbol{\theta}' = \boldsymbol{\mu}$, $\mathbf{U} = diag(\mathbf{J}_{n_1}, ..., \mathbf{J}_{n_m})$, $\mathbf{B}' = (\mathbf{b}_1, ..., \mathbf{b}_m)$ and $\mathbf{E}' = (\mathbf{e}_{11}, ..., \mathbf{e}_{1n_1}; ...; \mathbf{e}_{mn_1}, ..., \mathbf{e}_{mn_m})$. The random effects $(\mathbf{b}_1, ..., \mathbf{b}_m)$ represent effects associated with the random input factor. Since one's interest is often focused on the variances of the random effects, the mixed linear model is often called a variance components model. If $n_1 = n_2 = ... = n_m$ then the design is said to be a balanced design. Otherwise, the design is unbalanced. Further, it will be assumed that $\mathbf{b}_i$, $i = 1, 2, ..., m$, have a Normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_1$, and it will be assumed that $\mathbf{e}_{ij}$, $i = 1, 2, ..., n$, $j = 1, 2, ..., n_i$, all have a Normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_0$. The population mean, $\boldsymbol{\theta}$, is treated as a nuisance parameter and its estimates will not be reported. So, the basic assumptions are

a) $\mathbf{e}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$,

b) $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1)$,

c) the $\mathbf{e}_{ij}$ and $\mathbf{b}_i$ are independent,

d) the matrix of the intraclass correlation coefficients is $\mathbf{H}_1 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1$.

The likelihood function is then given by

$$p(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}_0|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} tr\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta}-\mathbf{U}\mathbf{B})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta}-\mathbf{U}\mathbf{B})\boldsymbol{\Sigma}_0^{-1}\right]\right\} \quad (2.3)$$

Note that the likelihood function (2.3) is not the usual one which is employed with non-Bayesian as well as Bayesian studies. In the Bayesian approach $\mathbf{B}$ is regarded as a random parameter, but since random parameters are not permitted in the sampling approach, the likelihood (2.3) is averaged over the distribution of $\mathbf{B}$.

A Jeffrey's prior pdf will be used for $\boldsymbol{\theta}$, $\mathbf{B}$ and $\boldsymbol{\Sigma}_0$. The Jeffrey's prior is chosen so that the data may speak for themselves as described in Box and Tiao (1973). The prior knowledge about $(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\Sigma}_0)$ is expressed by

$$p(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}(p+1)}. \quad (2.4)$$

It should be noted that while the covariance matrix $\boldsymbol{\Sigma}_1$ associated with the random effects $\mathbf{b}_i$, $i = 1, ..., m$, is unknown, it is not explicitly included in the prior as a parameter. The prior in (2.4) appears to be the same as would be the case for a regression model that contains only fixed effects. However, $\mathbf{B}$ is viewed as resulting from a random selection process.

3. Posterior distributions

Our interest will be focused on estimation of

$$\mathbf{H}_0 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0, \quad \mathbf{H}_1 = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_1, \quad \text{where } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1. \quad (3.1)$$

The largest eigenvalue of $\mathbf{H}_1$ is called the generalized heritability.

The mean of the posterior distribution of $\mathbf{H}_1$ is to be found as the "natural" Bayesian estimator if a quadratic loss function is appropriate. For our problem we will use the approximate posterior density function for the covariance matrices $(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1)$ of the form (see Jelenkowska, 1995)

$$p(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1 \mid \mathbf{Y}) \propto p(\boldsymbol{\Sigma}_0 \mid \mathbf{Y}) \cdot p(\boldsymbol{\Sigma}_1 \mid \mathbf{Y}),$$

where

$$p(\boldsymbol{\Sigma}_0 \mid \mathbf{Y}) \propto |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}(n-m+p+1)} \exp\left\{-\frac{1}{2} tr(\mathbf{S}\boldsymbol{\Sigma}_0^{-1})\right\},$$

$$\mathbf{S} = \mathbf{Y}'\mathbf{R}\mathbf{Y} - \widehat{\mathbf{B}}'\mathbf{U}'\mathbf{R}\mathbf{U}\widehat{\mathbf{B}}, \quad \mathbf{R} = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\mathbf{J}_n', \quad (3.2)$$

$$\widehat{\mathbf{B}} = (\mathbf{U}'\mathbf{R}\mathbf{U})^{-}\mathbf{U}'\mathbf{R}\mathbf{Y},$$

and

$$p(\Sigma_1 \mid \mathbf{Y}) \propto |\Sigma_1|^{-\frac{1}{2}(m+p+1)} \exp\left\{-\frac{1}{2}tr(\mathbf{S}_1\Sigma_1^{-1})\right\}, \tag{3.3}$$

where

$$\mathbf{S}_1 = \frac{m-p-2}{m}\left[\widehat{\mathbf{B}}'\widehat{\mathbf{B}} + \frac{\mathbf{S}}{n-m-p-1}\right], \; m > p+2, \; n > m+p+1.$$

The functions (3.2) and (3.3) are the density functions of inverted Wishart distribution.

Thus the approximate joint posterior distribution of (Σ_0, Σ_1) is

$$p(\Sigma_0, \Sigma_1 \mid \mathbf{Y}) \propto |\Sigma_0|^{-\frac{n-m+p+1}{2}} |\Sigma_1|^{-\frac{m+p+1}{2}} \exp\left\{-\frac{1}{2}tr(\mathbf{S}\Sigma_0^{-1} + \mathbf{S}_1\Sigma_1^{-1})\right\}. \tag{3.4}$$

From (3.1) we have

$$\Sigma_1^{-1} = \mathbf{H}_1^{-1}\Sigma^{-1} \quad \text{and} \quad \Sigma_0^{-1} = (\mathbf{I} - \mathbf{H}_1)^{-1}\Sigma^{-1}.$$

The jacobian is

$$J(\Sigma_0, \Sigma_1 \to \Sigma, \mathbf{H}_1) = |\Sigma|^{\frac{1}{2}(p+1)}.$$

The joint posterior distribution of $(\Sigma, \mathbf{H}_1)$ is

$$p(\Sigma, \mathbf{H}_1 \mid \mathbf{Y}) \propto |\mathbf{I} - \mathbf{H}_1|^{-\frac{1}{2}(n-m+p+1)} |\mathbf{H}_1|^{-\frac{1}{2}(m+p+1)} |\Sigma|^{-\frac{1}{2}(n+p+1)}. \tag{3.5}$$

Integrating (3.5) with respect to Σ, we get the marginal posterior density of $(\mathbf{H}_1 \mid \mathbf{Y})$ as

$$p(\mathbf{H}_1 \mid \mathbf{Y}) \propto |\mathbf{S}|^{-\frac{n}{2}} \left|\mathbf{S}_1^{-1}\mathbf{S}\right|^{\frac{1}{2}(m-p-1)} |\mathbf{H}_1|^{(p+1)} \left|\mathbf{S}_1\mathbf{H}_1^{-1}(\mathbf{I} - \mathbf{H}_1)\mathbf{S}^{-1}\right|^{\frac{1}{2}(m-p-1)}. \tag{3.6}$$

Let

$$\mathbf{W}_1 = \mathbf{S}_1\mathbf{H}_1^{-1}(\mathbf{I} - \mathbf{H}_1)\mathbf{S}^{-1}$$

be a transformation from $\mathbf{H}_1$ to $\mathbf{W}_1$ with jacobian

$$J(\mathbf{H}_1 \to \mathbf{W}_1) = \left|\mathbf{S}_1^{-1}\mathbf{S}\right|^p \left|\mathbf{H}_1^{-1}\right|^{p+1}.$$

This yields the marginal posterior distribution of $\mathbf{W}_1$

$$p(\mathbf{W}_1 \mid \mathbf{Y}) \propto |\mathbf{W}_1|^{\frac{1}{2}(m-p-1)} |\mathbf{I} + \mathbf{W}_1|^{-\frac{n}{2}}. \tag{3.7}$$

This proves the following.

THEOREM 1. *The approximate posterior distribution of* $\mathbf{W}_1$ *is the multivariate inverted beta distribution given by* (3.7).

4. Bayesian Estimators for $(\mathbf{H}_0, \mathbf{H}_1)$

The posterior mean for $\mathbf{W}_1$ can now be written as follows:

$$\mathbf{E}(\mathbf{W}_1 \mid \mathbf{Y}) = \mathbf{S}_1\mathbf{E}(\mathbf{H}_1^{-1}\mathbf{H}_0 \mid \mathbf{Y})\mathbf{S}^{-1}.$$

From the inverted multivariate beta distribution, we have

$$\mathbf{E}(\mathbf{W}_1 \mid \mathbf{Y}) = \frac{m-p+1}{n-m-2}\mathbf{I}_p. \tag{4.1}$$

Note that

$$\mathbf{H}_1^{-1}\mathbf{H}_0 = \mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_0 = \mathbf{R}_1^{-1}$$

and recall that

$$\mathbf{H}_1 = \mathbf{\Sigma}_1(\mathbf{\Sigma}_0 + \mathbf{\Sigma}_1)^{-1} = (\mathbf{R}_1^{-1} + \mathbf{I})^{-1}.$$

Then,

$$\hat{\mathbf{R}}_1^{-1} = E(\mathbf{H}_1^{-1}\mathbf{H}_0) = \frac{m-p+1}{n-m-2}\mathbf{S}_1^{-1}\mathbf{S},$$

and, finally, the estimators for $\mathbf{H}_0$, $\mathbf{H}_1$ are given by

$$\hat{\mathbf{H}}_0 = \hat{\mathbf{R}}^{-1}, \qquad \hat{\mathbf{H}}_1 = \hat{\mathbf{R}}_1\hat{\mathbf{R}}^{-1}, \tag{4.2}$$

where

$$\hat{\mathbf{R}}_1 = \frac{n-m-2}{m-p+1}\mathbf{S}^{-1}\mathbf{S}_1 \quad \text{and} \quad \hat{\mathbf{R}} = \mathbf{I} + \hat{\mathbf{R}}_1$$

The result is summarized in

THEOREM 2. *The bayesian estimators of* $\mathbf{H}_1$ *and* $\mathbf{H}_0$ *are given approximately in* (4.2).

5. Numerical example

The model used to generate the data is a one-way layout (2.1) consisting of one random effect (parental lines). The dependent variables are ear length, grain weight per ear and 1000-grain weight ($p = 3$). We also have $n_1 = n_2 = \ldots = n_7 = 3$, $n = 21$ and $m = 7$. The covariance matrix among parental lines is denoted by $\mathbf{\Sigma}_1$ and the covariance matrix for the experimental errors is $\mathbf{\Sigma}_0$. The data were taken from Kaczmarek and Krajewski (1994), page 315.

Using the posterior mean as a point estimate of the correlation matrix, we obtained

$$\hat{\mathbf{H}}_1 = \begin{bmatrix} 0.2855 & 0.0225 & 1.0547 \\ -0.2253 & 0.1027 & -5.3095 \\ 0.0692 & 0.0081 & 0.6656 \end{bmatrix}.$$

The eigenvalues λ_i of $\hat{\mathbf{H}}_1$ denote in genetics the heritabilities of the characters and will be used as multivariate measures of the intraclass correlations. They are

$$\lambda_1 = 0.7238, \quad \lambda_2 = 0.1102, \quad \lambda_3 = 0.2198.$$

Another possible approach to the estimation problem is to use the eigenvalues λ_i, $i = 1, 2, 3$, of $\hat{\Sigma}_1(\hat{\Sigma}_0 + \hat{\Sigma}_1)^{-1}$ ($\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ are maximum likelihood estimators for Σ_0 and Σ_1) as multivariate measures of the intraclass correlation (see Konishi et al., 1991).

Then,

$$\hat{\Sigma}_1(\hat{\Sigma}_0 + \hat{\Sigma}_1)^{-1} = \begin{bmatrix} 0.1745 & -0.1377 & 0.0421 \\ 0.0137 & 0.0627 & 0.0049 \\ 0.6445 & -3.2447 & 0.4067 \end{bmatrix}$$

and

$$\lambda_1 = 0.4423, \quad \lambda_2 = 0.0673, \quad \lambda_3 = 0.1343.$$

6. Discussion

Extensive simulation studies are needed to compare the estimation method introduced in Section 4 with the method of Konishi et al. (1991) in terms of the mean square error, to answer the question: which one is closer to the true value?

In the univariate situation Palmer and Broemeling (1990) proposed the median of a conditional posterior density as a Bayes estimator of the intraclass correlation coefficient. Using the simulation study, they showed that the Bayes estimator has a smaller mean square error then the maximum likelihood estimator. The estimators proposed in Section 4 may be considered as the multivariate generalization of these univariate estimators.

The main aim of this paper was to describe a method of estimation of the degree of resemblance between family members with respect to more than one characteristic from a Bayesian point of view. In the fields of research such as genetics, biology, etc. we frequently have a sample of families and then take measurements of several characteristics, e.g. physical or chemical variables. Our unified estimators are based on the posterior mean.

References

Box, G.E., Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. *Reading Massachusetts,* Addison Wesley.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review* **54**,1, 67-82.

Jelenkowska, T.H. (1995). A Bayesian analysis of the multivariate mixed linear model. *Comm. in Statist., Theory and Methods* **24**, 3183-3196.

Jelenkowska, T.H. (1996). Bayesian estimation of the intraclass correlation coefficients in mixed linear model. *Applications of Mathematics* (in print).

Kaczmarek, Z., Krajewski, P. (1994). Multivariate evaluation of parental lines on the basis of experiment with line × tester hybrides. *XXIV Biometrical Colloquium,* 309-320.

Konishi, Khatri, C.G., Rao, C.R. (1991). Inferences on multivariate measures of interclass and intraclass correlations in familial data. *J.R. Statist. Soc. B* **53**, 649-659.

Narain, P. (1990). Statistical Genetics. *Wiley Eastern Limited, New Delhi, India.*

Olkin, J., Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Annals of Mathem. Statistics* **35**, 261-269.

Palmer, J.L., Broemeling, L.D. (1990). A comparison of Bayes and maximum likelihood estimation of the intraclass correlation coefficient. *Comm. in Stat., Theory and Methods* **19**, 953-975.

Srivastava, M.S. (1993). Estimation of the intraclass correlation coefficient. *Ann. Hum. Genet.* **57**, 159-165.

Received 4 April 1996; revised 15 April 1998

Estymacja bayesowska wewnątrzklasowych współczynników korelacji w wielowymiarowym modelu klasyfikacji pojedynczej

Streszczenie

W pracy została przedstawiona estymacja macierzy wewnątrzklasowych współczynników korelacji z bayesowskiego punktu widzenia. W celu oszacowania stopnia podobieństwa pomiędzy członkami rodziny z uwzględnieniem więcej niż jednej cechy została wprowadzona wielowymiarowa miara współczynników korelacji. Jako ujednolicony estymator tej miary została zaproponowana średnia rozkładu a posteriori odwrotnego wielowymiarowego rozkładu beta. Zaproponowana metoda została przedstawiona na przykładzie modelu klasyfikacji pojedynczej. Rozważania teoretyczne zostały zilustrowane na przykładzie liczbowym.

Słowa kluczowe: estymacja Bayesowska, model MANOVA, model klasyfikacji pojedynczej, odwrotny wielowymiarowy rozkład beta, współczynniki korelacji.