論文の内容の要旨

Multivariate Linear Mixed Models with Application to Small Area Estimation（多変量線形混合モデルと小地域推定への応用）

Tsubasa Ito
伊藤　翼

Linear mixed models and model-based predictors in small area estimation have been studied extensively and actively in recent years due to the growing demand for reliable small area estimates. In small area estimation, direct design-based estimates for small area means have large standard errors due to small sample sizes from small areas. In order to improve accuracy, the linear mixed models are considered which consist of fixed effects based on common parametes and random effects depending on areas, and the resulting empirical best linear unbiased predictors (EBLUP) provide more reliable estimates by 'borrowing strength' from neighboring areas. This is because EBLUP shrinks the sample mean of the small area towards a stable quantity costructed by pooling all the data through the fixed effects, and the shrinkage arises from the random eects.

When multivariate data with correlations are observed from small areas for estimating multidimensional characteristics, like poverty and unemployment indicators, it is desirable to incorporate all the data to the one multivariate model and construct multivariate model-based predictors rather than to handle the data separately.

In this thesis, the consideration is the multivariate linear mixed models, especially the models used in small area estimation, that is, the Fay-Herriot model for analyzing area-level data and the nested error regression (NER) model for analyzing unit-level data, and the following three problems are mainly discussed.

At first, the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown is considered. As a specific estimator of the covariance matrix, Prasad-Rao type estimators with closed forms and use the modified versions which are restricted over the space of nonnegative definite matrices is employed. The empirical best linear unbiased predictors are provided based on the Prasad-Rao type estimators, and second-order approximation of their mean squared error matrices (MSEM) and their second-order unbiased estimators of the MSEM are derived with closed expressions.

Confidence regions for multivariate small area means are also constructed. Naive confidence regions can be constructed easily by using the Bayes estimators of small area means and their MSEM. However, the coverage probability of the naive methods cannot be guaranteed to be greater than or equal to the nominal confidence coefficient. To overcome this problem, a confidence region based on the Mahalanobis distance centerd around EBLUP is considered here, and the asymptotic expansion of the characteristic function of this distance is used to approximate the coverage probability based on the chi-square distributions. Then, the correction term is obtained in a closed form, and the confidence region that is second order correct is provided. Concerning the estimation of the covariance matrix, the Prasad-Rao type estimator with non-negative definite modification can be given in a closed form by the moment method. When the covariance matrix is estimated with the zero matrix or a singular matrix close to the zero matrix, however, the correction term becomes instable in the confidence region. This fact is well known in the univariate confidence interval. Thus, a new method for obtaining a positive-definite and sencon-order unbiased estimator of the covariance matrix is suggested. Moreover, the extension of our results to construction of corrected confidence regions for the difference of two small area mean vectors is considered. Another approach to construction of corrected confidence regions is the bootstrap method which needs heavy burden in computation. Because the corrected confidence region suggested here is provided in closed forms, it is easy to implement, which is a merit of the proposed method.

Next, the multivariate Fay-Herriot model without assuming multivariate normal distributions for random effects and sampling errors is considered. The differences between the method discussed above and the existing research are as follows: (1) The existing research assumes that the error terms have univariate normal distributions with known variances, while no distributional assumptions are imposed on the random effects. The present method does not assume the normality for the error terms, but assumes that the second and fourth moments are known. (2) The present method handles the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown without nomality assumption, while the existing research treated the univariate Fay-Herriot model with unknown variance of random effects.

A consistent and nonnegative-definite estimator of the covariance matrix of the random effects is suggested, and the EBLUP for a vector of small-area characteristics is provided. A second-order approximation of the mean squared error matrix (MSEM) of the EBLUP is derived and the second-order unbiased estimator of the MSEM is obtained. This MSEM estimator is achieved only under the moment assumptions for random effects, and then, our estimator of MSEM is useful because the normality assumption is very resrtrictive and the specification of the underlying distributions for random effects and sampling errors are difficulut in practice. However, in the multivariate problem, when deriving a second-order approximation of the MSEM of EBLUP and a second-order unbiased estimator of the MSEM, we cannot use the standard technique of approximation via the Taylor's expansion. Then, the results for the multivariate version are not obvious and we must consider them separately from the univariate problem.

Lastly, multivariate nested error regression models (MNER) is considered. The MNER model has the two components of covariance: 'between' component and 'within' component. Here, an exact unbiased estimator for 'within' component is used, and for 'between' component, a nonnegative definite and consistent estimator which is a second-order unbiased estimator is suggested. Substituting these estimators for the components of covariance into the Bayes estimator and estimating fixed effects by the generalized least squares estimator, one gets the empirical Bayes estimator or EBLUP. A second-order approximation of the MSE matrix of the EBLUP is derived analytically and a closed form expression of a second-order unbiased estimator is provided. Another topic is the confidence interval problem. One difficulty with traditional confidence intervals is that the coverage probabilities do not have second-order accuracy. It is also numerically confirmed that the coverage probabilities are smaller than the nominal confidence coefficient. Here, the confidence interval for the liner combination of a small area mean and some vector is constructed in the closed-form whose coverage probability is identical to the nominal confidence coefficient up to second order.

Finite sample performances of all the results obtained above are investigated by numerical simulation studies. In both of Fay-Herriot model and nested error regression model, the multivariate EBLUP improves the prediction risk of the direct estimator significantly for all simulation setup. Similarly, the multivariate EBLUP improves the univariate EBLUP when the correlation of random effects are considerably large. However, the multivariate EBLUP is worse than the univariate EBLUP when the correlation of random effects are small. This is because the low accuracy in estimation of the covariance matrix has more adverse influence on prediction than the benefit from incorporating the small correlation into the estimation.

A second-order approximations of the MSEM of the EBLUP and a second-order unbiased estimators are also investegated by comparing these values with the simulated true MSEM of the EBLUP, and the approximations seem to be well done.

Lastly, the proposed confidence regions and intervals are investigated numerically. From the simulation results, it can be seen that the proposed methods achieve the coverage probability equal to the nominal confidence coefficient up to second order while the coverage probability of naive methods is considerably smaller than the nominal confidence coefficient.