Modified Large Sample Method for Variance Components Model
(소표본 실험에서 분산 성분 모형에 대한 붓스트랩 보정 신뢰구간)

Yonghee Lee
Department of Statistics, Ewha Womans University

제 1 절 Introduction

Linear combinations of variance components are frequently considered when parameter of interest is individual or gross variability in statistical experimental designs. Because of the lack of an exact confidence interval for a general linear combination of variance components, many researchers have suggested various methods for setting approximate confidence intervals on linear combinations of variance components. Among those methods, the Modified Large Sample (MLS) Method has its merit in terms of simplicity, flexibility, and good empirical performance. Generally, the MLS method can give an approximate confidence interval on any linear combination of variance components, i.e.

$$c_1\sigma_1^2 + c_2\sigma_2^2 + \cdots + c_p\sigma_p^2 \tag{1}$$

where $\sigma_1^2, \ldots, \sigma_p^2$ are variance components and $c_1^2, \ldots, c_p^2$ are fixed non-zero real numbers. The idea of MLS method was proposed by (4w, o) and (8, W) independently. The main idea of the MLS method was well described by (4w, o) as follows: when each of the marginally exact quantiles of estimators of variance components are known, we can incorporate this knowledge into setting a good approximate confidence interval for a linear combination of variance components by interpolating the given exact quantiles.

제 2 절 Overview

In the next section, the main idea of the original MLS method will be introduced when all coefficients are positive and estimators of variance components are independent. The following two sections will discuss the extensions of the MLS method for the general cases such that (i) signs of coefficients are unrestricted, (ii) setting confidence bound on a ratio of linear combinations of variance components.

Throughout this article we denote the distribution of a random quantity $\mathbf{x}$ by $\mathcal{L}(\mathbf{x})$. Hence, $\mathcal{L}(\mathbf{x}) = \mathcal{L}(\mathbf{y})$ means that $\mathbf{x}$ and $\mathbf{y}$ have the same distribution. We use $N(\boldsymbol{\mu}, \mathbf{V})$ to denote the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ and use χ_r^2 to denote a random variable having the central chi-square distribution with r degrees of freedom. The αth quantile of $N(0,1)$, $\mathcal{L}(\chi_r^2)$, and the central t-distribution with degrees of freedom r are denoted by z_α, $\chi^2_{\alpha,r}$, and $t_{\alpha,r}$, respectively. The transpose of the vector $\mathbf{x}$ is denoted by $\mathbf{x}'$ and the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$.

제 3 절 Main Idea of Modified Large Sample Method

Consider the problem of setting a $1-\alpha$ upper confidence bound for $\eta = c_1\sigma_1^2 + \cdots + c_p\sigma_p^2$, where $\sigma_1^2,\ldots,\sigma_p^2$ are unknown variance components and $c_1,\ldots,c_p$ are known positive constants. Let $\hat{\sigma}_i^2$ be an unbiased estimator of σ_i^2 such that $\mathcal{L}(\hat{\sigma}_i^2) = \mathcal{L}(\sigma_i^2 n_i^{-1}\chi^2_{n_i})$, $i = 1, \ldots, p$. In applications, $\hat{\sigma}_i^2$ is typically a quadratic form of an observed normal random vector (e.g., a sample variance or a mean square error). For any fixed i, an exact $1-\alpha$ upper confidence bound for σ_i^2 is

$$\frac{n_i}{\chi^2_{\alpha,n_i}}\hat{\sigma}_i^2 = \hat{\sigma}_i^2 + \sqrt{\hat{\sigma}_i^4\left(\frac{n_i}{\chi^2_{\alpha,n_i}} - 1\right)^2}. \tag{2}$$

When $p \geq 2$, however, an exact upper confidence bound for η usually does not exist.

Suppose that $\hat{\sigma}_1^2,...,\hat{\sigma}_p^2$ are independent. Then

$$\begin{aligned}\mathrm{Var}(c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2) &= c_1^2\mathrm{Var}(\hat{\sigma}_1^2) + \cdots + c_p^2\mathrm{Var}(\hat{\sigma}_p^2)\\ &= c_1^2\sigma_1^4 n_1^{-2}\mathrm{Var}(\chi^2_{n_1}) + \cdots + c_p^2\sigma_p^4 n_p^{-2}\mathrm{Var}(\chi^2_{n_p})\\ &= c_1^2\sigma_1^4 2n_1^{-1} + \cdots + c_p^2\sigma_p^4 2n_p^{-1}. \end{aligned} \tag{3}$$

An estimator of the variance in (3) is then obtained by replacing σ_i^4 in (3) by its estimator $\hat{\sigma}_i^4$. For large n_i's, these results and the Central Limit Theorem lead to the following approximate $1-\alpha$ upper confidence bound for η:

$$\begin{aligned}& c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 + z_{1-\alpha}\sqrt{c_1^2\hat{\sigma}_1^4 2n_1^{-1} + \cdots + c_p^2\hat{\sigma}_p^4 2n_p^{-1}}\\ &= c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 + \sqrt{c_1^2\hat{\sigma}_1^4 2z^2_{1-\alpha}n_1^{-1} + \cdots + c_p^2\hat{\sigma}_p^4 2z^2_{1-\alpha}n_p^{-1}}\end{aligned} \tag{4}$$

In view of (2), (8, W) proposed to replace $2z^2_{1-\alpha}n_i^{-1}$ in (4) by $\left(\frac{n_i}{\chi^2_{\alpha,n_i}} - 1\right)^2$, $i = 1,...,p$, and named this method as the modified large sample (MLS) method. The resulting upper confidence bound for η is then

$$c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 + \sqrt{c_1^2\hat{\sigma}_1^4\left(\frac{n_1}{\chi^2_{\alpha,n_1}} - 1\right)^2 + \cdots + c_p^2\hat{\sigma}_p^4\left(\frac{n_p}{\chi^2_{\alpha,n_p}} - 1\right)^2}. \tag{5}$$

Using a similar argument, we can obtain the following MLS lower confidence bound for η:

$$c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 - \sqrt{c_1^2\hat{\sigma}_1^4\left(\frac{n_1}{\chi^2_{1-\alpha,n_1}} - 1\right)^2 + \cdots + c_p^2\hat{\sigma}_p^4\left(\frac{n_p}{\chi^2_{1-\alpha,n_p}} - 1\right)^2}. \tag{6}$$

An equal-tail two sided MLS confidence interval for η can be obtained by using the upper bounds (5) and lower bound (6) as the interval limits.

The MLS confidence bound has the following properties:

Property 1. The confidence bound in (5) is still asymptotically correct, i.e., the coverage probability of the confidence bound converges to $1-\alpha$ when all n_i's increase to infinity. This is because

$$\lim_{n\to} \frac{n}{2}\left(\frac{n}{\chi^2_{\alpha,n}} - 1\right)^2 = z^2_{1-\alpha},$$

which can be proved using the Cornish-Fisher expansion.

Property 2. When $\sigma_i^2 > 0$ and $\sigma_j^2 = 0$ for all $j \neq i$, the confidence bound in (5) reduces to the confidence bound in (2) and, hence, is an exact confidence bound. This property is not enjoyed by the confidence bound in (4).

The Property 1 implies that the coverage probability of the MLS confidence bound converge to the intended confidence coefficient $1 - \alpha$ as sample sizes become large. The finite sample evaluation of coverage probability of the MLS method are investigated by using numerical integration in (8, W). The numerical evaluation in (8, W) indicates that the MLS method is comparable to the methods by (1t, a6,a) and is also conservative in sense that the finite sample coverage probabilities are larger than $1-\alpha$ for most of cases considered. The Properties 2 is the main and unique advantage of the MLS method over other competing methods. When one of variance components is relatively larger than the others, the MLS method gives an almost exact confidence bound. This property was also discussed by (4w, o). Other than these two properties of the MLS method, another nice property is its simplicity.

제 4 절 An Extension of MLS method to Difference of Variance Components

In this section, we will consider an extension of the MLS method when signs of coefficient in (1) are unrestricted. Originally, (4w, o) considered an approximate confidence bound on inter-subject (between group) variance in one-way classification model by using similar idea to (8, W). First, the Howe's idea will be introduceed and its generalization will be discussed.

Consider two mean squares $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, which are distributed independently and

$$\mathcal{L}(\hat{\sigma}_1^2) = \mathcal{L}\left(\frac{(\sigma_a^2 + \sigma_e^2)\chi_{n_1}^2}{n_1}\right), \quad \mathcal{L}(\hat{\sigma}_2^2) = \mathcal{L}\left(\frac{\sigma_e^2\chi_{n_2}^2}{n_2}\right).$$

Therefore, an unbiased estimator of σ_a^2 is given by $\hat{\sigma}_1^2 - \hat{\sigma}_2^2$. For this example, if

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \le F_{\alpha,n_1,n_2},$$

then the upper confidence bound of σ_a^2 is zero. (4w, o) considered this as an extra piece of readily available information and incorporated this information with the idea of MLS method to construct more accurate confidence bound on σ_a^2. The $1-\alpha$ MLS upper confidence bound of σ_a^2 can be defined as

$$\hat{\sigma}_1^2 - \hat{\sigma}_2^2 + \sqrt{V}$$

where

$$V = \hat{\sigma}_1^4 \left(\frac{n_1}{\chi^2_{\alpha,n_1}} - 1 \right)^2 + A\hat{\sigma}_2^4. \tag{7}$$

Now, the constant A in (7) can be determined by forcing the upper confidence bound $\hat{\sigma}_1^2 - \hat{\sigma}_2^2 + \sqrt{V}$ to be 0 when $\hat{\sigma}_1^2/\hat{\sigma}_2^2 = F_{\alpha,n_1,n_2}$. This imposing condition leads to the equation for A. By using the solution for A from the given equation, $1-\alpha$ MLS upper confidence bound of σ_a^2 is given by

$$\begin{cases} 0 & \text{if } F \le F_{\alpha,n_1,n_2} \\ \hat{\sigma}_1^2 - \hat{\sigma}_2^2 + \sqrt{V} & \text{if } F > F_{\alpha,n_1,n_2}, \end{cases} \tag{8}$$

where $F = \hat{\sigma}_1^2/\hat{\sigma}_2^2$ and

$$V = \hat{\sigma}_1^4 \left(\frac{n_1}{\chi^2_{\alpha,n_1}} - 1 \right)^2 + \hat{\sigma}_2^4 \left[(F_{\alpha,n_1,n_2} - 1)^2 - F^2_{\alpha,n_1,n_2} \left(\frac{n_1}{\chi^2_{\alpha,n_1}} - 1 \right)^2 \right].$$

Note that the confidence bound in (8) use an extra information such as $\sigma_a^2 > 0$ in the model considered.

Now, consider more general parameter $c_1\sigma_1^2 - c_2\sigma_2^2$ where c_1 and c_2 are known positive real numbers. For this general case, however, if the parameter $c_1\sigma_1^2 - c_2\sigma_2^2$ can be negative, the MLS confidence upper bound in (8) is no longer valid. For this case, (8B, G) extended Howe's confidence bound in (8) as

$$\begin{cases} c_1\hat{\sigma}_1^2 - c_2\hat{\sigma}_2^2 + \sqrt{V^*} & \text{if } F \le F_{\alpha,n_1,n_2} \\ c_1\hat{\sigma}_1^2 - c_2\hat{\sigma}_2^2 + \sqrt{V} & \text{if } F > F_{\alpha,n_1,n_2}, \end{cases} \tag{9}$$

where F and V are same as in (8) except using $c_i\hat{\sigma}_i^2$ instead of $\hat{\sigma}_i^2$ and

$$V^* = c_1^2\hat{\sigma}_1^4\left[\left(\frac{1}{F_{\alpha,n_1,n_2}}-1\right)^2-\left(\frac{1}{F_{\alpha,n_1,n_2}}\right)^2\left(\frac{n_2}{\chi^2_{\alpha,n_2}}-1\right)^2\right]+\hat{c}_2^2\sigma_2^4\left(\frac{n_2}{\chi^2_{1-\alpha,n_2}}-1\right)^2.$$

Furthermore, (90G, B) considered MLS confidence bounds for more general case such that the parameter of interest can be described as

$$\sum_{i=1}^{q} c_i\sigma_i^2 - \sum_{j=q+1}^{p} c_j\sigma_j^2$$

where $q < p$ and $c_1, \ldots, c_p$ are all positive real numbers. Also, the alternative form of the MLS upper confidence bound was proposed by (90G, B) without considering F-test in (11) and it can be described as

$$\sum_{i=1}^{q} c_i\hat{\sigma}_i^2 - \sum_{j=q+1}^{p} c_j\hat{\sigma}_j^2 + \sqrt{W} \tag{10}$$

where

$$W = \sum_{i=1}^{q} c_i^2\hat{\sigma}_i^4 L_i^2 + \sum_{j=q+1}^{p} c_j^2\sigma_j^2 L_j^2 + \sum_{i=1}^{q}\sum_{j=q+1}^{p} c_i c_j \hat{\sigma}_i^2\hat{\sigma}_j^2 L_{ij}$$

and L_i, L_j, L_{ij} are appropriately chosen to make the MLS upper confidence bound exact for some special cases. Note that we can simplify the confidence bound in (8) by ignoring the cross-product terms $\sum_{i=1}^{q}\sum_{j=q+1}^{p} c_i c_j \hat{\sigma}_i^2\hat{\sigma}_j^2 L_{ij}$. Moreover, if we consider Property 1 in the previous section as the imposed condition for special cases (i.e., an exact confidence bound when only one variance component is positive and the others are zero), the following simplified $1-\alpha$ MLS confidence upper bound can be obtained:

$$\sum_{i=1}^{q} c_i\hat{\sigma}_i^2 - \sum_{j=q+1}^{p} c_j\hat{\sigma}_j^2 + \sqrt{W^*} \tag{11}$$

where

$$W^* = \sum_{i=1}^{q} c_i^2\hat{\sigma}_i^4 L_i^2 + \sum_{j=q+1}^{p} c_j^2\sigma_j^2 L_j^2$$

and

$$L_i^2 = \left(\frac{n_i}{\chi^2_{\alpha,n_i}}-1\right)^2, i = 1, \ldots, q; \quad L_j^2 = \left(\frac{n_j}{\chi^2_{1-\alpha,n_j}}-1\right)^2, j = q+1, \ldots, p.$$

제 5 절 An Extension of MLS method to Ratio of Variance Components

In many experimental designs with variance components models, a certain ratio of variance components is the parameter of interest. For example, under one-way random effect model, a ratio of between variance and total variance is often considered as important parameter which is called by heritability measure in genetics. The MLS confidence bound on the ratio of variance components have been extensively developed by (9B, G1; G, B; 9, G95; B, G). In this section, the MLS confidence bound by (95B, G) will be introduced.

Consider the problem of setting a confidence upper bound on the ratio

$$\rho = \frac{\sigma_2^2 + \sigma_3^2}{\sigma_1^2 + \sigma_2^2}$$

The $1-\alpha$ upper bound on the ratio ρ can be defined by U such that $P(\rho < U)$ is close to $1-\alpha$. Note that $P[\rho < U] = P[(\sigma_2^2+\sigma_3^2)/(\sigma_1^2+\sigma_2^2) < U] = P(\gamma_U < 0)$ such that

$$\gamma_U = -U\sigma_1^2 + (1-U)\sigma_2^2 + \sigma_3^2. \qquad (12)$$

Since γ_U in (12) is a linear combination of variance components, the MLS method discussed in the previous section can be applied to find $1-\alpha$ MLS confidence upper bound on γ_U such as $P(\gamma_U < U^*) \approx 1-\alpha$. In the other hand, we also desire $P(\gamma_U < 0) \approx 1-\alpha$. Therefore, we have the quadratic equation $U^* = 0$ in terms of U and the solution for U can be considered as $1-\alpha$ MLS confidence upper bound on the ratio ρ.

But, there is one difficulty in constructing confidence bound on γ_U in (12) since the sign of U in (12) cannot be determined prior to setting the bound. For example, if $\sigma_3^2 > \sigma_1^2$, then $\rho > 1$ which, in turn, implies $U > 1$. Therefore, we can determine the sign of $(1-U)$ which is the coefficient of σ_2^2 in (12). On the other hand, if $\sigma_3^2 > \sigma_1^2$, then $\rho < 1$. For this case, (95B, G) recommend to restrict the range of U by $0 < U < 1$. For either case, as discussed in the previous section, we can apply the MLS method by (8B, G) to construct

$1-\alpha$ MLS confidence upper bound on γ_U and its form is given in (11). (95B, G) proposed the procedure to find MLS confidence bound on ρ such that (i) Assume $\sigma_3^2 > \sigma_1^2$ (ii) compute an upper bound U (iii) if $U > 1$, then stop (iv) if $U < 1$, then assume $\sigma_3^2 \leq \sigma_1^2$ and find U again.

(95B, G) considered constructing MLS confidence bound on more general parameter such that

$$\rho = \frac{\sum_{i=1}^{q} c_i\sigma_i^2 - \sum_{j=q+1}^{p} c_j\sigma_j^2}{\sum_{k=1}^{r} c_k\sigma_k^2}$$

The explicit form of confidence interval and more details can be found in (95B, G).

참고 문헌

Richard K. Burdick and Franklin A. Graybill. *Confidence intervals on variance components*, volume 127. Marcel Dekker Inc., New York, 1992.

Rongde Gui, Franklin A. Graybill, Richard K. Burdick, and Naitee Ting. Confidence intervals on ratios of linear combinations for non-disjoint sets of expected mean squares. *Journal of Statistical Planning and Inference*, 48:215–227, 1995.

Franklin A. Graybill and Chih-Ming Wang. Confidence intervals on nonnegative linear combinations of variances. *Journal of the American Statistical Association*, 75:869–873, 1980.

W. G. Howe. Approximate confidence limits on the mean of $x+y$ where x and y are two tabled independent random variables. *Journal of the American Statistical Association*, 69:789–794, 1974.

Tai-Fang C. Lu, Franklin A. Graybill, and Richard K. Burdick. Confidence intervals on a difference of expected mean squares. *Journal of Statistical Planning and Inference*, 18:35–43, 1988.

Tai-Fang C. Lu, Franklin A. Graybill, and Richard K. Burdick. Confidence intervals on the ratio of expected mean squares $(\theta_1 - d\theta_2)/\theta_3$. *Journal of Statistical Planning and Inference*, 21(2):179–190, 1989.

F. E. Satterthwaite. Symthesis of variance. *Psychometrika*, 6:309–316, 1941.

F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2:110–114, 1946.

Naitee Ting, Richard K. Burdick, Franklin A. Graybill, S. Jeyaratnam, and Tai-Fang C. Lu. Confidence intervals on linear combinations of variance components that are unrestricted in sign. *Journal of Statistical Computation and Simulation*, 35:135–143, 1990.

Naitee Ting, Richard K. Burdick, and Franklin A. Graybill. Confidence intervals on ratios of positive linear combinations of variance components. *Statistics & Probability Letters*, 11(6):523–528, 1991.