PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS OF VARIANCE COMPONENTS IN THE ONE-WAY CLASSIFICATION, BALANCED DATA

Hongjian Yu, Shayle R. Searle, and Charles E. McCulloch

Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, N. Y. 14853

PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS OF VARIANCE COMPONENTS IN THE ONE-WAY CLASSIFICATION, BALANCED DATA

Hongjian Yu, Shayle R. Searle, and Charles E. McCulloch

Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, N. Y. 14853

Key Words and Phrases: variance components; expected values; variances; asymptotic results; probability of negative solution to the ML equation; robustness.

Category: Variance components

ABSTRACT

We studied properties of maximum likelihood estimators (MLEs) of the variance components obtained from balanced data of the one-way classification. Exact and asymptotic expected values and variances of these MLEs were derived under the usual normality assumptions. Numerical studies illustrate these expected values and variances, and also illustrate the probability of obtaining a negative solution to the maximum likelihood (ML) equation for the between-class variance component. Simulations were used to study the robustness of the ML estimators under non-normal distributions.

1. Introduction

The variance components model for the one-way classification is based on

$$y_{ij} = \mu + \alpha_i + e_{ij}, \tag{1.1}$$

where y_{ij} is the j-th observation in the i-th class, α_i is the random effect of class i having mean zero and variance σ_α^2; μ is the overall mean and e_{ij} is the random residual error with mean zero and variance σ_e^2. The covariance of every α_i with every other α_i and with every e_{ij} is taken to be zero. The data are balanced when the number of observations within each class (n_i in class i) is the same, that is, $n_i = n$ for $i = 1, \cdots, c$, where c is the number of classes. The variance of the observation y_{ij} is then

$$\text{Var}(y_{ij}) = \sigma_\alpha^2 + \sigma_e^2, \tag{1.2}$$

with σ_α^2 and σ_e^2 being components of variance due to α-effects and the error term, respectively.

The normality assumption for the maximum likelihood (ML) method to estimate σ_α^2 and σ_e^2 means is, for balanced data,

$$\alpha \sim \mathcal{N}\left(0, \sigma_\alpha^2 I_c\right) \quad \text{and} \quad e \sim \mathcal{N}\left(0, \sigma_e^2 I_{cn}\right), \tag{1.3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_c)'$, $\mathbf{e} = (e_{11}, \cdots, e_{1n}, \cdots, e_{cn})'$, and $\mathbf{I}_c$ and $\mathbf{I}_c$ are identity matrices with order c and cn, respectively. So

$$\mathbf{y} \sim \mathcal{N}\left(\mu \mathbf{1}_{cn}, \mathbf{I}_c \otimes (\sigma_\alpha^2 \mathbf{J}_n + \sigma_e^2 \mathbf{I}_n)\right),$$

where $\mathbf{J}_n$ is a square matrix with order n and all elements being 1, and $\otimes$ is the Kronecker product operator.

The log likelihood function is

$$L = -\tfrac{1}{2}c n \log(2\pi) - \tfrac{1}{2}c(n-1)\log\sigma_e^2 - \tfrac{1}{2}c\log\left(\sigma_e^2 + \sigma_\alpha^2\right)$$

$$-\frac{\sum_i \sum_j (y_{ij} - \mu)^2}{2\sigma_e^2} + \frac{1}{2\sigma_e^2} \sum_i \frac{\sigma_\alpha^2}{\sigma_e^2 + n\sigma_\alpha^2}(y_i. - n\mu)^2 \,,$$

where $y_i. = \sum_j y_{ij}$. Defining

$$SSA = n \sum_{i=1}^c (\bar{y}_i. - \bar{y}..)^2 \qquad MSA = SSA/(c-1) \,,$$

$$SSE = \sum_{i=1}^c \sum_{j=1}^n (y_{ij} - \bar{y}_i.)^2 \qquad MSE = SSE/c(n-1) \,,$$

$$SST_m = \sum_{i=1}^c \sum_{j=1}^n (y_{ij} - \bar{y}..)^2 = SSA + SSE \,,$$

where $\bar{y}_i. = \sum_{j=1}^n y_{ij}/n$ and $\bar{y}.. = \sum_{i=1}^c \sum_{j=1}^n y_{ij}/cn$, and equating to zero the partial derivatives of L with respect to μ, σ_α^2 and σ_e^2, the solutions to the ML equations are (Searle 1971, p. 418):

$$\dot{\sigma}_e^2 = MSE \quad \text{and} \quad \dot{\sigma}_\alpha^2 = \frac{(1-1/c)MSA - MSE}{n} \,.$$

But these are not always the MLEs because when MSE is greater than $(1-1/c)MSA$, using $\dot{\sigma}_\alpha^2$ as the MLE of σ_α^2 would yield a negative estimate of σ_α^2 which, by the nature of a MLE of a non-negative parameter, must never be negative.

Crump (1947) derived equations for both balanced and unbalanced data and, at least for balanced data gave partial consideration to the need for replacing a negative solution to the equations by a zero. But it was Herbach (1959) who dealt with this problem fully, for balanced data. If $(\frac{c-1}{c})MSA \geq MSE$, the MLEs are $\hat{\sigma}_\alpha^2 = \left((\frac{c-1}{c})MSA - MSE\right)/n$ and $\hat{\sigma}_e^2 = MSE$. If $(\frac{c-1}{c})MSA < MSE$, then $\hat{\sigma}_\alpha^2 = 0$ and $\hat{\sigma}_e^2 = SST_m/cn$. That is

$$\hat{\sigma}_\alpha^2 = \left\{[(1-1/c)MSA - MSE]/n\right\}\mathfrak{I}[(1-1/c)MSA \geq MSE] \,,$$

$$\hat{\sigma}_e^2 = \left\{MSE\right\}\mathfrak{I}[(1-1/c)MSA \geq MSE] + \left\{SST_m/cn\right\}\mathfrak{I}[(1-1/c)MSA < MSE] \,,$$

(1.4)

where $\mathfrak{I}$ is the indicator function.

Anderson and Bancroft (1952, Sec. 22.2) and Thompson (1962) introduced the idea of maximizing the part of the likelihood which is invariant to the location parameters of the

model: to the fixed effects, or in (1.1) the overall mean μ. This procedure is now known as restricted maximum likelihood (REML). But in this paper we focus on the MLEs defined in (1.4).

It is easily shown and well known that MSE and MSA are independent under the normality assumption and we then have

$$\frac{(c-1)MSA}{n\sigma_\alpha^2 + \sigma_e^2} \sim \chi_{c-1}^2 \quad \text{and} \quad \frac{c(n-1)MSE}{\sigma_e^2} \sim \chi_{c(n-1)}^2 \ . \tag{1.5}$$

So, letting X = MSA, Y = MSE and $\lambda = n\sigma_\alpha^2 + \sigma_e^2$, and since MSE and MSA are independent, the joint density of MSA and MSE is

$$f_{X,Y}(x,y) = \frac{c-1}{\lambda} \frac{1}{\Gamma\left(\frac{c-1}{2}\right)} \left(\frac{1}{2}\right)^{\frac{c-1}{2}} \left(\frac{c-1}{\lambda}x\right)^{\frac{c-1}{2}-1} \exp\left(-\frac{1}{2}\frac{c-1}{\lambda}x\right)$$

$$\times \frac{c(n-1)}{\sigma_e^2} \frac{1}{\Gamma\left(\frac{c(n-1)}{2}\right)} \left(\frac{1}{2}\right)^{\frac{c(n-1)}{2}} \left(\frac{c(n-1)}{\sigma_e^2}y\right)^{\frac{c(n-1)}{2}-1} \exp\left(-\frac{1}{2}\frac{c(n-1)}{\sigma_e^2}y\right) . \tag{1.6}$$

In Section 2 we use (1.6) to derive $E(\hat{\sigma}_e^2)$, $Var(\hat{\sigma}_e^2)$, $E(\hat{\sigma}_\alpha^2)$ and $Var(\hat{\sigma}_\alpha^2)$. The final forms of the expectations and variances involve integrals, and numerical computation of these forms has been done for a small set of parameter values to study the behavior of expected values and variances of the estimators. This is done in Section 3, which discusses the effects of sample size c and n, and of the ratio of $\rho = \sigma_\alpha^2/\sigma_e^2$ on the expected values and variances. It also discusses the probability of obtaining for the between-class variance component a negative solution to the ML equation and how the MLEs behave under some non-normal distributions such as uniform, lognormal and "Student"-t distributions.

2. Derivation of the Expectations and Variances of the MLEs

We present the derivation of the expected values and variances of the estimators $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_e^2$ of equation (1.4). Throughout this section we let X = MSA and Y = MSE and define:

U : a random variable having distribution χ_{c-1}^2

V : a random variable having distribution χ_{c+1}^2

W : a random variable having distribution χ_{c+3}^2

$\rho = \sigma_\alpha^2/\sigma_e^2$, the ratio of σ_α^2 to σ_e^2

$\lambda = n\sigma_\alpha^2 + \sigma_e^2$

b = c−1

m = n−1 .

Since $SST_m = (c-1)MSA + c(N-1)MSE$, for $\hat{\sigma}_e^2$ defined in (1.4) we have

$$E(\hat{\sigma}_e^2) = \int_0^\infty \int_0^{(1-1/c)x} y f_{X,Y} \, dy \, dx + \int_0^\infty \int_{(1-1/c)x}^\infty \frac{(c-1)x + c(n-1)y}{cn} f_{X,Y} \, dy \, dx$$

$$= A_1 + A_2 ,$$
(2.1)

so defining A_1 and A_2. Then because of the independence of X and Y

$$A_1 = \int_0^\infty f_X \int_0^{bx/c} \left(\frac{\Gamma\left(\frac{cm}{2}+1\right)}{\Gamma\left(\frac{cm}{2}\right)} \frac{2}{cm/\sigma_e^2} \right) \frac{1}{\Gamma\left(\frac{cm}{2}+1\right)} \left(\frac{1}{2}\right)^{\frac{cm}{2}+1} \left(\frac{cm}{\sigma_e^2} y\right)^{\frac{cm}{2}} \exp\left(\frac{cm}{2\sigma_e^2} y\right) d\left(\frac{cm}{\sigma_e^2} y\right) dx .$$

Let $Q = \frac{cm}{\sigma_e^2} Y$ and note that

$$Q = \frac{bm}{\sigma_e^2} X \quad \text{when} \quad Y = bX/c .$$

Therefore, since

$$\left(\frac{\Gamma\left(\frac{cm}{2}+1\right)}{\Gamma\left(\frac{cm}{2}\right)} \frac{2}{cm/\sigma_e^2} \right) = \left(\frac{\Gamma\left(\frac{cm}{2}\right)\frac{cm}{2}}{\Gamma\left(\frac{cm}{2}\right)} \frac{2\sigma_e^2}{cm} \right) = \sigma_e^2 ,$$

$$A_1 = \sigma_e^2 \int_0^\infty f_X \int_0^{bmx/\sigma_e^2} \left\{ \frac{1}{\Gamma\left(\frac{cm+2}{2}\right)} \left(\frac{1}{2}\right)^{\frac{cm+2}{2}} q^{\frac{cm+2}{2}-1} \exp\left(-\frac{1}{2}q\right) \right\} dq \, dx .$$

The term in the braces is the density of a chi-square distribution with $cm+2$ degrees of freedom. Hence the integral we have here with respect to q, namely,

$$\int_0^{bmx/\sigma_e^2} \frac{1}{\Gamma\left(\frac{cm+2}{2}\right)} \left(\frac{1}{2}\right)^{\frac{cm+2}{2}} q^{\frac{cm+2}{2}-1} \exp\left(-\frac{1}{2}q\right) dq = \Pr\left(Q < bmx/\sigma_e^2\right),$$
(2.2)

is the probability of a random variable, say Q, being less than $\frac{bm}{\sigma_e^2} X$, where $Q \sim \chi^2_{cm+2}$. For this we introduce a new notation:

$$P_k(w) = \Pr(T < w) \quad \text{where} \quad T \sim \chi^2_k .$$

The integral (2.2) can now be written as $P_{cm+2}\left(\frac{bm}{\sigma_e^2} x\right)$, whereupon

$$A_1 = \int_0^\infty \sigma_e^2 P_{cm+2}\left(\frac{bm}{\sigma_e^2} x\right) f_X \, dx .$$

Hence from (1.6),

$$A_1 = \sigma_e^2 \int_0^\infty P_{cm+2}\left(\frac{bm}{\sigma_e^2} x\right) \frac{b}{\lambda} \frac{1}{\Gamma\left(\frac{1}{2}b\right)} \left(\frac{1}{2}\right)^{\frac{b}{2}} \left(\frac{b}{\lambda} x\right)^{\frac{b}{2}-1} \exp\left(-\frac{b}{2\lambda} x\right) dx .$$

Let $U = \frac{b}{\lambda}X$, then $X = \frac{\lambda}{b}U$,

$$A_1 = \sigma_e^2 \int_0^\infty P_{cm+2}\left(\frac{\lambda m}{\sigma_e^2}u\right)\left\{\frac{1}{\Gamma(\frac{1}{2}b)}\left(\frac{1}{2}\right)^{\frac{b}{2}}\left(\frac{b}{\lambda}x\right)^{\frac{b}{2}-1}\exp(-u/2)\right\}du .$$

Since $b = c-1$, the expression in the braces is the density function of χ^2_{c-1}, and so

$$A_1 = \sigma_e^2 E\left[P_{cm+2}\left(\lambda m U/\sigma_e^2\right)\right],$$

where E stands for expectation and $U \sim \chi^2_{c-1}$, as defined at the beginning of this section.

Similarly, from (2.1)

$$A_2 = \int_0^\infty \int_{(1-1/c)x}^\infty \frac{(c-1)x + c(n-1)y}{cn}f_{X,Y}\,dydx$$

$$= \frac{b}{cn}\int_0^\infty \int_{bx/c}^\infty xf_{X,Y}\,dydx + \frac{cm}{cn}\int_0^\infty \int_{bx/c}^\infty xf_{X,Y}\,dyd$$

$$= \frac{b}{cn}A_{21} + \frac{m}{n}A_{22} ,$$

so defining A_{21} and A_{22}. Using the same technique as above we derive

$$A_{21} = \lambda\left\{1 - E\left[P_{cm}\left(\lambda m V/\sigma_e^2\right)\right]\right\} \quad \text{and} \quad A_{22} = \sigma_e^2\left\{1 - E\left[P_{cm+2}\left(\lambda m U/\sigma_e^2\right)\right]\right\} .$$

Substituting A_{21} and A_{22} into A_2, and then A_1 and A_2 into (2.1) gives

$$E(\hat{\sigma}_e^2) = \sigma_e^2 E\left[P_{cm+2}\left(\lambda m U/\sigma_e^2\right)\right] + \lambda\frac{b}{cn}\left\{1 - E\left[P_{cm}\left(\lambda m V/\sigma_e^2\right)\right]\right\}$$

$$+ \frac{m}{n}\sigma_e^2\left\{1 - E\left[P_{cm+2}\left(\lambda m U/\sigma_e^2\right)\right]\right\} .$$

Reverting to c and n, rather than $b = c-1$ and $m = n-1$, gives

$$E(\hat{\sigma}_e^2) = \sigma_e^2 E\left\{P_{c(n-1)+2}\left[\lambda(n-1)U/\sigma_e^2\right]\right\} + \left(n\sigma_\alpha^2 + \sigma_e^2\right)\frac{c-1}{cn}$$

$$- \left(n\sigma_\alpha^2 + \sigma_e^2\right)\frac{c-1}{cn}E\left\{P_{c(n-1)}\left[\lambda(n-1)V/\sigma_e^2\right]\right\}$$

$$+ \frac{(n-1)}{n}\sigma_e^2 - \frac{(n-1)}{n}\sigma_e^2 E\left\{P_{c(n-1)+2}\left[\lambda(n-1)U/\sigma_e^2\right]\right\}$$

$$= \sigma_e^2(1-1/cn) + \sigma_\alpha^2(1-1/c) + \frac{\sigma_e^2}{n}E\left\{P_{c(n-1)+2}\left[\lambda(n-1)U/\sigma_e^2\right]\right\}$$

$$- \frac{c-1}{cn}\left(n\sigma_\alpha^2 + \sigma_e^2\right)E\left\{P_{c(n-1)}\left[\lambda(n-1)V/\sigma_e^2\right]\right\} .$$

A little more algebra gives

$$E(\hat{\sigma}_e^2) = \sigma_e^2\Bigg((1-1/cn) + \rho(1-1/c) + \tfrac{1}{n}E\Big\{P_{c(n-1)+2}[(n\rho+1)(n-1)U]\Big\}$$
$$-\tfrac{c-1}{cn}(n\rho+1)E\Big\{P_{c(n-1)}[(n\rho+1)(n-1)V]\Big\}\Bigg).$$

For $\text{Var}(\hat{\sigma}_e^2)$ we have

$$\text{Var}\left(\hat{\sigma}_e^2\right) = E\left[\left(\hat{\sigma}_e^2 - E\hat{\sigma}_\alpha^2\right)^2\right] = E\left[\left(\hat{\sigma}_e^2\right)^2\right] - \left(E\left[\hat{\sigma}_e^2\right]\right)^2$$

and

$$E\left[(\hat{\sigma}_e^2)^2\right] = E\left(\Big\{\text{MSE}\,\mathfrak{I}[(1-1/c)\text{MSA} \geq \text{MSE}] + (\text{SST}_m/cn)\,\mathfrak{I}[(1-1/c)\text{MSA} < \text{MSE}]\Big\}^2\right)$$
$$= E\left(\Big\{\text{MSE}^2\,\mathfrak{I}[(1-1/c)\text{MSA} \geq \text{MSE}] + (\text{SST}_m/cn)\,\mathfrak{I}[(1-1/c)\text{MSA} < \text{MSE}]\Big\}\right).$$

Using the same technique as used in deriving $E\left(\hat{\sigma}_e^2\right)$, we get

$$E\left[(\hat{\sigma}_e^2)^2\right] = \sigma_e^4\Bigg\{\left(1+\frac{2}{c(n-1)}\right)E\Big\{P_{c(n-1)+4}[(n\rho+1)(n-1)U]\Big\}$$
$$+\frac{1}{c^2n^2}\Big((c^2-1)(n\rho+1)^2\Big(1-E\Big\{P_{c(n-1)}[(n\rho+1)(n-1)W]\Big\}\Big)$$
$$+ 2c(c-1)(n-1)(n\rho+1)\Big(1-E\Big\{P_{c(n-1)+2}[(n\rho+1)(n-1)V]\Big\}\Big)$$
$$+ c^2(n-1)^2\Big(1+\frac{2}{c(n-1)}\Big)\Big(1-E\Big\{P_{c(n-1)+4}[(n\rho+1)(n-1)U]\Big\}\Big)\Bigg\}.$$

Similarly, for $\hat{\sigma}_\alpha^2$ we get

$$E(\hat{\sigma}_e^2) = \frac{\sigma_\alpha^2}{\rho}\Bigg\{\frac{c-1}{cn}(n\rho+1)E\Big\{P_{c(n-1)}[(n\rho+1)(n-1)V]\Big\} - \tfrac{1}{n}E\Big\{P_{c(n-1)+2}[(n\rho+1)(n-1)U]\Big\}\Bigg\}$$

and

$$E\left[(\hat{\sigma}_\alpha^2)^2\right] = \frac{\sigma_d^4}{\rho^2}\Bigg\{\frac{(n\rho+1)^2(c^2-1)}{c^2n^2}E\Big\{P_{c(n-1)}[(n\rho+1)(n-1)W]\Big\}$$
$$+\frac{1}{n^2}\left(1+\frac{2}{c(n-1)}\right)E\Big\{P_{c(n-1)+4}[(n\rho+1)(n-1)U]\Big\}$$
$$-\frac{2(c-1)}{cn^2}(n\rho+1)E\Big\{P_{c(n-1)+2}[(n\rho+1)(n-1)V]\Big\}\Bigg\}.$$

These two expressions can be utilized in $\text{Var}\left(\hat{\sigma}_\alpha^2\right) = E\left[(\hat{\sigma}_\alpha^2)^2\right] - \left[E(\hat{\sigma}_\alpha^2)\right]^2$.

3. Properties of the ML Estimators

3.1. Numerical studies of expected values and variances

From the derivations of Section 2, note that $E(\hat{\sigma}_e^2)$ and $Var(\hat{\sigma}_e^2)$ are multiples of σ_e^2 and σ_e^4, respectively; $E(\hat{\sigma}_\alpha^2)$ and $Var(\hat{\sigma}_\alpha^2)$ are multiples of σ_α^2 and σ_α^4, respectively. These multiples depend only on c, n and ρ. This means that in numerical studies of these expectations and variances, all we need to focus on are these multiplying factors. So without loss of generality we can take $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = \rho$ when we study $E(\hat{\sigma}_e^2)$ and $Var(\hat{\sigma}_e^2)$, and take $\sigma_e^2 = 1/\rho$ and $\sigma_\alpha^2 = 1$ when we study $E(\hat{\sigma}_\alpha^2)$ and $Var(\hat{\sigma}_\alpha^2)$. Other values of σ_e^2 and σ_α^2 change only the scales of the expectations and variances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$. Figures 1 through 4 illustrate the behaviors of the means and variances of the MLEs. For every figure the coordinates for c and n start from (2, 2) and not from (0, 0), because from (1.5) we see that the model (1.1) is meaningful only when both c and n are greater than 1.

With $\sigma_e^2 = 1$ we see from Figure 1 that $\hat{\sigma}_e^2$ is nearly unbiased for most values of c and n except when both c and n are small. In this case a downward bias occurs, so that $\hat{\sigma}_e^2$ underestimates σ_e^2. As c and/or n increases the bias disappears. The magnitude of the bias also decreases as ρ increases. The variance of $\hat{\sigma}_e^2$ is seen in Figure 2 to be largest when both c and n are small and it decreases as c and/or n increase. The ratio ρ also plays a role here: $Var(\hat{\sigma}_e^2)$ increases as ρ increases, given that c and n are small. This is because for a fixed σ_e^2, increasing ρ means increasing σ_α^2 and this contributes to larger $Var(\hat{\sigma}_e^2)$. But the effect of ρ on $Var(\hat{\sigma}_e^2)$ is much less than that of c and n, and this effect fades as c or n, or both, increase.

In Figure 3 we see, with $\sigma_\alpha^2 = 1$, that $E(\hat{\sigma}_\alpha^2)$ is not greatly affected by ρ except when ρ is very small. When ρ and n are small there is a large bias. This is because for fixed σ_α^2, small ρ means large σ_e^2 and this large σ_e^2 causes a bias in estimating σ_α^2. In $\lambda = n\sigma_\alpha^2 + \sigma_e^2$, the main term of $E(\hat{\sigma}_\alpha^2)$, when n and ρ are small, $\sigma_e^2 = \sigma_\alpha^2/\rho$ becomes the dominant term. This effect is reduced when n increases. From Figure 3, this seems the only place where n plays an important role in affecting $E(\hat{\sigma}_\alpha^2)$. In contrast, the number of groups, c, has greater impact on $E(\hat{\sigma}_\alpha^2)$ than either ρ or n. When c is small, bias occurs and $\hat{\sigma}_\alpha^2$ underestimates σ_α^2. The bias fades as c increases. In Figure 4, $Var(\hat{\sigma}_\alpha^2)$ decreases as c increases in somewhat the same manner for all ρ. In particular, the sample size n has little influence on $Var(\hat{\sigma}_\alpha^2)$ for $\rho > 1$. When n and ρ are small, for the same reason as when discussing $E(\hat{\sigma}_\alpha^2)$, σ_e^2 becomes the dominant term in λ and causes a large variance of $\hat{\sigma}_\alpha^2$. This effect fades as n increases.

Generally speaking, from Figures 1 through 4 we see that c and n both have considerable effect on the values of $E(\hat{\sigma}_e^2)$ and $Var(\hat{\sigma}_e^2)$. But for $E(\hat{\sigma}_\alpha^2)$ and $Var(\hat{\sigma}_\alpha^2)$, the same size c plays the

major role, and n has little effect when ρ is very small. The role of ρ can be explained intuitively: since $\rho = \sigma_\alpha^2/\sigma_e^2$, small ρ will make estimating σ_α^2 difficult because we are estimating a small σ_α^2 with is camouflaged by a large σ_e^2.

3.2. Asymptotic results

Asymptotic expressions for $E(\hat{\sigma}_e^2)$, $Var(\hat{\sigma}_e^2)$, $E(\hat{\sigma}_\alpha^2)$ and $Var(\hat{\sigma}_\alpha^2)$ depend on integrals of the form

$$E\left\{P_{c(n-1)}[(n\rho+1)(n-1)U]\right\} = \int_0^\infty P[Q \le (n\rho+1)(n-1)x]f_U(x)dx , \qquad (3.1)$$

which is the expected value of the probability that $Q \sim \chi_{c(n-1)}^2$ and so is less than $(n\rho+1) \times (n-1)U$, where $U \sim \chi_{c-1}^2$ and U and Q are independent. The upper bound of the integral (3.1) is 1. Defining h such that for given $\epsilon > 0$, $P[Q \le (n\rho+1)(n-1)x] > 1-\epsilon$ if $x \ge h$, so we have

$$EP = \int_0^h P[Q \le (n\rho+1)(n-1)x]f_U(x)dx + \int_h^\infty (1-\epsilon)f_U(x)dx$$

$$< E\left\{P_{c(n-1)}[(n\rho+1)(n-1)U]\right\} \le 1 . \qquad (3.2)$$

Notice $\int_0^h P[Q \le (n\rho+1)(n-1)x]f_U(x)dx < \int_0^h f_U(x)dx$ and if $\int_0^h f_U(x)dx$ is so small that it can be neglected, then the expression in (3.2) is approximately $\int_0^\infty (1-6-1)$, the probability

infinity, so h approaches 0. Then we have, from (3.2)

$$\lim_{n\to\infty} EP = \int_0^\infty (1-\epsilon)f_U(x)dx < \lim_{n\to\infty} \int_0^\infty P[Q \le (n\rho+1)(n-1)x]f_U(x)dx \le 1 ,$$

and EP approaches 1 as $\epsilon \to 0$. So the integral (3.1) approaches 1 as n goes to infinity.

Likewise for a given h, we have

$$\lim_{c\to\infty} \int_0^h f_U(x)dx = 0 ,$$

and for large n the rate of convergence will be faster than for small n.

So generally, if one of c and n is fixed, if the other increases, then (3.1) will approach 1. This gives the asymptotic results shown in Table I. The discussion and figures in Section 3.1 show that in most cases the expectations and variances of $\hat{\sigma}_e^2$ approaches their asymptotic values very rapidly. The expected value of $\hat{\sigma}_\alpha^2$ also approaches its asymptotic value value rapidly; for example, for c = 2 and n = 9, when $\rho = 0.5$, the ratio of the asymptotic value to the exact expectation is 1.086; when $\rho = 1$, the ratio is 1.061 and when $\rho = 5$, the ratio is 1.017. And the ratio monotonically decreases to one as we fix one of the c or n and increase the other. For the

variance of $\hat{\sigma}_\alpha^2$, the process is different. With c fixed, the approach to the asymptote is much slower than for the expected value of $\hat{\sigma}_\alpha^2$. For example, for c = 9 and $\rho = 1$, when n = 100, the ratio of asymptotic variance to the exact variance is 0.832; when n = 150, the ratio is 0.882 and when n = 300, the ratio is 0.937. And it is slower for smaller ρ. For fixed n, both exact and asymptotic variances of $\hat{\sigma}_\alpha^2$ would descend to zero but progress to the asymptote is faster. So the ratio of asymptotic to exact decreases as c increases.

TABLE I

Asymptotic form of means and variances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$

As n → ∞, c fixed	As c → ∞, n fixed
$E(\hat{\sigma}_e^2) = \sigma_e^2$	$E(\hat{\sigma}_e^2) = \sigma_e^2$
$\mathrm{Var}(\hat{\sigma}_e^2) = 0$	$\mathrm{Var}(\hat{\sigma}_e^2) = 0$
$E(\hat{\sigma}_\alpha^2) = \sigma_\alpha^2(1-1/c)$	$E(\hat{\sigma}_\alpha^2) = \sigma_\alpha^2$
$\mathrm{Var}(\hat{\sigma}_\alpha^2) = \frac{2}{c}\sigma_\alpha^4(1-1/c)$	$\mathrm{Var}(\hat{\sigma}_\alpha^2) = 0$

3.3. Probability of negative solutions for σ_α^2 to ML equations

The unusual characteristic of the ML equations is the possibility of a negative value for the solution $\dot{\sigma}_\alpha^2 = [(1-1/c)\mathrm{MSA}-\mathrm{MSE}]/n$. To avoid this, the MLEs of σ_e^2 and σ_α^2 are defined as in (1.4). But the possibility of a negative $\dot{\sigma}_\alpha^2$ still exists and this causes other problems. For example, it is well known that MSE is an unbiased estimator of σ_e^2. But, using (1.4) leads to bias occurring in estimating σ_e^2.

The probability of having a negative $\dot{\sigma}_\alpha^2$ can be expressed as

$$P = \Pr\left\{\dot{\sigma}_\alpha^2 < 0\right\} = \Pr\{(1-1/c)\mathrm{MSA} < \mathrm{MSE}\} = \Pr\left\{\frac{(1-1/c)\mathrm{MSA}}{\mathrm{MSE}} < 1\right\}.$$

From (1.5) define $F = \dfrac{\mathrm{MSA}/\left(\sigma_e^2 + n\sigma_\alpha^2\right)}{\mathrm{MSE}/\sigma_e^2} \sim F_{c-1,\,c(n-1)}$, the F-distribution with degrees of freedom c-1 for the numerator and c(n-1) for the denominator. Then

$$P = \Pr\left\{\frac{(1-1/c)\left(\sigma_e^2 + n\sigma_\alpha^2\right)}{\sigma_e^2}F < 1\right\} = \Pr\left\{F^* > (1-1/c)(1+n\rho)\right\},$$

where $F^* \sim F_{c(n-1),\ c-1}$.

Table II shows values of P for a set of values of c, n and ρ (the symbol 0^* indicates P < 0.001). In each cell of Table II, the three numbers are the Ps corre-sponding to ρ equal to 0.1, 1 and 3. We see that for certain combinations of c and n, the smaller that ρ is, the larger

the chance of having a negative $\dot{\sigma}_\alpha^2$. Indeed for small c and n there is always an appreciable probability of negative $\dot{\sigma}_\alpha^2$, unless ρ is extremely large. This partly explains why properties such as bias and large variance can happen when c and n are small; and why for the same c and n, the smaller the value of ρ is, the larger the variance and bias can be.

TABLE II

Probability of negative $\dot{\sigma}_\alpha^2$

C

	4	9	16	25	36	49	64	81	100
2	0.723	0.680	0.613	0.547	0.488	0.438	0.396	0.360	0.330
	0.450	0.339	0.266	0.217	0.183	0.158	0.139	0.124	0.112
	0.292	0.201	0.159	0.128	0.107	0.092	0.081	0.072	0.065
4	0.554	0.442	0.325	0.233	0.167	0.122	0.090	0.068	0.052
	0.152	0.061	0.029	0.015	0.009	0.006	0.004	0.003	0.002
	0.143	0.014	0.006	0.003	0.002	0.001	0*	0*	0*
9	0.395	0.219	0.100	0.042	0.018	0.008	0.004	0.002	0.001
	0.017	0.001	0*	0*	0*	0*	0*	0*	0*
	0*	0*	0*	0*	0*	0*	0*	0*	0*
16	0.289	0.100	0.023	0.001	0.001	0*	0*	0*	0*
	0.001	0*	0*	0*	0*	0*	0*	0*	0*
	0*	0*	0*	0*	0*	0*	0*	0*	0*

n

3.4. Estimation for non-normal distributions

Simulations for distributions other than normal were conducted to study the robustness of the MLEs derived under the assumption of normality. These distributions are the uniform, 'Student'-t with three degrees of freedom (t_3), and lognormal distribution. All these distributions have been modified to have $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = \rho$ when used for studying $\dot{\sigma}_e^2$, and to have $\sigma_\alpha^2 = 1$ and $\sigma_e^2 = 1/\rho$ when used for studying $\dot{\sigma}_\alpha^2$.

When the distribution is light-tailed, such as the uniform, the estimators defined in (1.4) work very well. For example, when either α or e is uniform and the other is normal, or both α and e are uniform, simulation results show that the patterns of the expected values and variances of both $\dot{\sigma}_e^2$, and $\dot{\sigma}_\alpha^2$ look almost identical to those when both α and e are normal. Figure 5 is for both α and e having a uniform distribution with $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = 1$; thus $\rho = 1$ and so Figure 5 can be compared with the $\rho = 1$ case of Figures 1 through 4.

But when the distribution is heavier-tailed than the normal, the results change dramatically. For example, in the model (1.1) when e is normal and α is t_3, simulation shows

that $\hat{\sigma}_e^2$ behaves quite similarly to when both e and α are normal; but $\hat{\sigma}_\alpha^2$ behaves very differently: the simulated $\text{Var}(\hat{\sigma}_\alpha^2)$ of $\hat{\sigma}_\alpha^2$ becomes very large for any combination of c and n. If we switch the distributions to have α being normal and e being t_3, then the behaviors of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$ switch also: $\hat{\sigma}_\alpha^2$ has expectation and variance similar to when both e and α are normal; but the variance of $\hat{\sigma}_e^2$ is very large for any combination of c and n. When both α and e are t_3, neither of them give acceptable estimates because of large variances. Similar results occur when t_3 is replaced with the lognormal distribution.

These results led to the investigation of the behavior of the samples of the simulations. Since large variance may indicate the existence of some extreme observations, we used percentiles instead of mean and variance to describe the sample distribution. With 2,000 simulated observations for each combination of c and n, Table III gives the estimated 75th and 95th percentile of the distributions of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$ when both α and e have $t_3/\sqrt{3}$ distributions (so $\sigma_e^2 = \sigma_\alpha^2 = 1$ and $\rho = 1$). The table shows that the estimators behave well most of the time: 95 percent of the estimates are close to the true parameter values. And $\hat{\sigma}_e^2$ is a better estimator for σ_e^2 (which is one) than $\hat{\sigma}_\alpha^2$ is for σ_α^2 (which is one also). This hints that there is a small number of extreme observations that cause great variation of the estimators.

The same of 2,000 observations of $\hat{\sigma}_e^2$ at the combination of c = 2 and n = 9 yielded the greatest variation. The median of these 2,000 observed $\hat{\sigma}_e^2$ values is 0.66 while the mean is 1.8 and the standard deviation is 39.42, compared with having α and e both N(0, 1), whereupon $\text{Var}(\hat{\sigma}_e^2)$ is only about 1. One $\hat{\sigma}_e^2$ was as large as 1,800. Extreme observations like this pull the mean far from the median and contribute large variation to the simulation results. Simulations show that this can happen at any combination of c and n for both $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$. This is because both the t_3 and lognormal are heavier-tailed distributions than the normal, and the chance of extreme observations is then greater than that when using normal distributions, and so the variances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$ under these distributions are much larger than under normals. To get a better view of the distribution of the rest of the observed $\hat{\sigma}_e^2$, Figure 6 shows the distribution of the sample with that extreme one, which is about 1800, deleted. We see that the skewness is clearly evident.

TABLE IIIa

75th and 95th percentile (in parentheses) of $\hat{\sigma}_e^2$ for 2,000 simulated

observations with $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = 1$

C

		2	4	9	16	25	36
	2	0.715	0.895	1.033	1.067	1.078	1.115
		(2.246)	(2.562)	(2.319)	(2.197)	(2.044)	(2.060)
	4	0.917	0.995	1.047	1.065	1.067	1.046
		(2.403)	(2.199)	(1.997)	(1.934)	(1.894)	(1.681)
	9	1.013	1.066	1.054	1.054	1.072	1.048
n		(2.379)	(2.120)	(1.746)	(1.722)	(1.680)	(1.522)
	16	1.033	1.039	1.044	1.067	1.052	1.061
		(2.051)	(1.948)	(1.763)	(1.641)	(1.546)	(1.498)
	25	1.065	1.047	1.054	1.058	1.054	1.043
		(1.981)	(1.739)	(1.618)	(1.568)	(1.478)	(1.428)
	36	1.061	1.045	1.073	1.042	1.045	1.042
		(1.900)	(1.720)	(1.638)	(1.504)	(1.447)	(1.357)

TABLE IIIb

75th and 95th percentile (in parentheses) of $\hat{\sigma}_\alpha^2$ for 2,000 simulated

observations with $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = 1$

C

		2	4	9	16	25	36
	2	0.379	0.796	0.979	1.043	1.097	1.068
		(1.993)	(2.919)	(2.444)	(2.520)	(2.264)	(2.077)
	4	0.392	0.759	0.924	1.020	1.018	1.064
		(1.941)	(2.537)	(2.225)	(2.638)	(2.204)	(2.076)
	9	0.376	0.728	0.914	0.981	0.991	1.032
n		(1.803)	(2.389)	(2.343)	(2.219)	(2.135)	(2.063)
	16	0.403	0.710	0.876	1.012	1.006	1.025
		(1.855)	(2.180)	(2.261)	(2.204)	(2.006)	(2.062)
	25	0.354	0.731	0.920	0.997	1.029	1.027
		(1.460)	(2.165)	(2.263)	(2.246)	(2.172)	(2.168)
	36	0.363	0.725	0.867	0.983	1.043	1.018
		(1.646)	(2.280)	(2.436)	(2.283)	(2.411)	(2.029)

3.4. Conclusions

From the numerical study we see that under normality $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$ give nearly unbiased estimates for most combinations of c and n. For some small values of c and/or n there is large variance for the MLEs of the variance components in model (1.1) and also a bias. This is because for small c and/or n there is substantial probability of getting negative $\hat{\sigma}_\alpha^2$. The magnitude of the probability depends on the ratio ρ which is usually unknown. The best way to reduce the probability of negative $\hat{\sigma}_\alpha^2$ is to have both c and n as large as possible.

Under the distributions other than normal for α and/or e, the estimators of (1.4) usually give estimates which are similar to those when both α and e are normal. But we should be aware that for heavier-tailed distributions like "Student-t" and lognormal there is an appreciable probability of very extreme estimates.

Another frequently used method of estimation is restricted likelihood (REML), which differs from ML in that account is taken of the fixed factors. REML estimators are adjusted so that, for balanced data, the solutions to the REML equations are the unbiased, ANOVA estimators. The REML estimator of σ_α^2 is biased upward since it is required to be nonnegative, the consequence of which is that when the ANOVA estimator is negative the REML estimator is zero, otherwise the two estimators are equal. Thus the REML estimator is never less than the unbiased ANOVA estimator. This will exacerbate the problem seen in Figure 3 for $\rho = 0.1$ and small n. On the other hand, the REML estimators will be asymptotically correct as *either* c or n approaches ∞ (contrast ML in Table I) and the slight downward bias seen in Figure 3 for $\rho = 0.5$, 1 and 10 for small c will be alleviated.

Finally, this work is based on the assumption of having balanced data. In real life this is not always achieved, which makes studying the MLEs for the model (1.1) much more difficult, because there is then no closed-form expression for the MLEs. More importantly, MSA no longer has a chi-square distribution, so the probability of negative $\hat{\sigma}_\alpha^2$ is no longer based on the F-distribution, and we lose a very important property for conducting the research.

BIBLIOGRAPHY

Anderson, R. L. and Bancroft, T. A. (1952). *Statistical Theory in Research*, McGraw-Hill Book Co., Inc., New York.

Crump, S. L. (1947). "The estimation of components of variance in multiple classifications," Ph.D. Thesis, Iowa State University, Ames, IA, 78 pp.

Herbach, L. H. (1959). "Properties of model H-type analysis of variance tests. A: Optimum nature of the F-test for model H in the balanced case," *Annals of Mathematical Statistics*, **30**: 583-588.

Searle, S. R. (1971). *Linear Models*, John Wiley & Sons, New York.

Thompson, W. A. (1962). "The problem of negative estimates of variance components," *Annals of Mathematical Statistics*, **33**: 273-289.

List of figure captions

FIG. 1 The expected value of $\hat{\sigma}_e^2$ for various values of c, n and $\rho = \sigma_\alpha^2/\sigma_e^2$ with $\sigma_e^2 = 1$.

FIG. 2 The variance of $\hat{\sigma}_e^2$ for various values of c, n and $\rho = \sigma_\alpha^2/\sigma_e^2$ with $\sigma_e^2 = 1$.

FIG. 3 The expected value of $\hat{\sigma}_\alpha^2$ for various values of c, n and $\rho = \sigma_\alpha^2/\sigma_e^2$ with $\sigma_\alpha^2 = 1$.

FIG. 4 The variance of $\hat{\sigma}_\alpha^2$ for various values of c, n and $\rho = \sigma_\alpha^2/\sigma_e^2$ with $\sigma_e^2 = 1$.

FIG. 5 The expected values and variances of $\hat{\sigma}_e^2$ and $\hat{\sigma}_\alpha^2$ by simulation where e and α are both uniformly distributed with $\sigma_e^2 = 1$ and $\sigma_\alpha^2 = 1$.

FIG. 6 The histogram of a simulation sample of 2,000 values of $\hat{\sigma}_e^2$ for a = 2, n = 9 and $\rho = 1$. One of the observations in this sample is approximately 1,800. In order to show the distribution of the sample, this observation was deleted from the histogram.