COMPARISON OF METHODS OF ESTIMATING VARIANCE COMPONENTS IN BALANCED TWO-WAY RANDOM NESTED DESIGNS

Ebru KAYA BAŞAR [1], Mehmet Ziya FIRAT [2]

Abstract

The objective of this paper is to compare different methods of estimating variance components, such as analysis of variance (ANOVA), maximum likelihood (ML), restricted maximum likelihood (REML) and Bayesian. The ANOVA method for estimating variance components is based on equating the mean squares to their expectations. However, the problem with this method is that it gives negative estimates of variance components. This can be overcame by the use of likelihood based methods and Bayesian. In this study, four different methods of estimating variance components were compared and also demonstrated how these methods overcome the problem of negative estimates of variance components in balanced two-way random nested designs.

Keywords: Variance components, Bayesian analysis, REML, ML.

DENGELİ İKİ SEVİYELİ ŞANSA BAĞLI İÇ İÇE DÜZENLENMİŞ DENEMELERDE VARYANS UNSURLARI TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Özet

Bu çalışmanın amacı, varyansı analizi (ANOVA), en çok olabilirlik (ML), kısıtlanmış en çok olabilirlik (REML) ve Bayesyen gibi farklı varyans unsurları tahmin yöntemlerini karşılaştırmaktır. ANOVA yöntemi kareler ortalamalarının beklenen değerlerine eşitlenmesine dayanır. Ancak bu yöntemin varyans unsurlarının negatif tahminlerini vermesi bir problem teşkil eder. Olabilirlik yöntemlerine dayanan metotlar ve Bayesyen yöntemle bu problemin üstesinden gelinebilir. Bu çalışmada varyans unsurlarının dört farklı yöntemle tahminleri karşılaştırılmıştır. Ayrıca dengeli iki seviyeli şansa bağlı iç içe düzenlenmiş deneme planlarının varyans unsurlarının negatif tahmin probleminin üstesinden nasıl gelineceği gösterilmiştir.

Anahtar Kelimeler: Varyans unsurları, Bayesyen analiz, REML, ML.

1. INTRODUCTION

An understanding of variability and the nature of measurement error is a fundamental issue for the researchers. Estimating variance components in experimental design is a way to assess the amount of variation in a dependent variable which is associated with one or more random effects variables. Variance components are used extensively in developing many of the basic concepts of many fields such as animal breeding. Sources of variation in the analysis of variance context were partitioned into their expected components, which were particularly useful to the researchers [1].

A wide array of methods has been developed for estimating variance components, for example, Analysis of Variance (ANOVA) and likelihood based methods, such as, Maximum Likelihood (ML)

---

[1] Akdeniz University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, Antalya, Turkey.
E-mail: ebrukaya10@gmail.com.tr
[2] Corresponding Author: Akdeniz University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, Antalya, Turkey.
E-mail: mzfirat@akdeniz.edu.tr

and Restricted Maximum Likelihood (REML), and Bayesian methods. The ANOVA method for estimating variance components is based on equating the mean squares to their expectations. However, the problem with this method is that ANOVA estimates may lie outside the parameter space so it gives negative estimates of variance components. In light of this weakness, alternative methods were required and methods based on likelihood have become common [1,2]. The ML estimation of variance components does not take into account the loss of degrees of freedom caused by estimation of the fixed effects. This can be worrying when the model includes fixed effects with many levels. These problems can be overcome by the use of REML method which was first proposed by Thompson [3] and generalized by Patterson and Thompson [4]. REML metahod is more useful than ML method in some aspects. This approach allows for several random factors in the model and is based on maximizing with respect to the variances only the part of the likelihood function that does not depend on fixed effects. In contrast to ANOVA estimations, estimators which are based on likelihood functions are functions of every sufficient statistic and are consistent, asymptotically normal and efficient.

An alternative to the methods mentioned above is the Bayesian method for estimating the variance components. The Bayesian framework was introduced by Thomas Bayes and Bayesian estimation was first used by Laplace in 1786. Estimation of variance components using Bayesian framework is found in Hill [5] and have been applied to animal breeding by many researchers [2,6,7,8]. The likelihood corresponding to the data and the prior distribution of the parameters are the components of posterior distribution which is required in order to carry out a Bayesian analysis. Inferences about each parameter are based on the corresponding marginal posterior distributions. For many models, the joint distribution of the dispersion parameters can be derived, but numerical integration techniques are required to obtain the marginal posterior distribution of functions of interest. Gibbs sampling is a numerical integration method introduced by Geman and Geman [9] which operates by generating samples from a sequence of full conditional distributions. Bayesian method allows for the formal incorporation of prior knowledge into the process of inference. Within the Bayesian framework, it is possible to obtain the marginal posterior distributions of parameters of interest using Gibbs sampling. This, in general, is not possible in the classical approaches.

The objective of this study is to compare four different methods of estimating variance components ANOVA, ML, REML and Bayesian and also to demonstrate how these methods overcome the problem of negative estimates of variance components in balanced two-way random nested designs.

2. MODEL

In some experiments samples must be chosen in two or more steps. If we consider an experiment with two factors, let factor A has a levels and factor B b levels within each level of factor A. The levels of B are nested within levels of A and within each level of B some number n random samples are chosen. The model for this two-way random nested design is given as follows

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{ijk} \qquad i = 1,2,\ldots,a; \; j = 1,2,\ldots,b; \; k = 1,2,\ldots,n \tag{1}$$

where y_{ijk} is the observation k in level i of factor A and level j of factor B; μ is the overall mean, α_i is the effect of level i of factor A; $\beta_{j(i)}$ is the effect of level j of factor B within level i of factor A and e_{ijk} is the random error with $e_{ijk} \sim N(0, \sigma_e^2)$. Also, α_i and $\beta_{j(i)}$ are both random effects in the model and these parameters assumed to be normally distributed $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$, respectively.

3. METHODS OF ESTIMATING VARIANCE COMPONENTS
3.1. Analysis of Variance (ANOVA)

The analysis of variance (ANOVA) estimation of the variance components consists of equating mean squares to their respective expected value. The resulting equations are solved for the variance components and the solutions are the estimators of σ_α^2, $\sigma_{\beta(\alpha)}^2$ and σ_e^2. These ANOVA estimators are given as

$$\hat{\sigma}_{e,ANOVA}^2 = MS_E \,, \ \ \hat{\sigma}_{\beta(\alpha),ANOVA}^2 = \frac{MS_{B(A)} - MS_E}{n} \,, \ \ \hat{\sigma}_{\alpha,ANOVA}^2 = \frac{MS_A - MS_{B(A)}}{bn} \tag{2}$$

where

$$MS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \left(y_{ijk} - \bar{y}_{ij.}\right)^2 \Big/ ab(n-1)$$

$$MS_{B(A)} = \sum_{i=1}^a \sum_{j=1}^b n\left(\bar{y}_{ij.} - \bar{y}_{i..}\right)^2 \Big/ a(b-1)$$

$$MS_A = \sum_{i=1}^b bn\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 \Big/ (a-1)$$

However, it is known that variance estimates may lie outside their parameter space. Negative variance estimation problem is the most important issue for ANOVA method [10].

3.2. Maximum Likelihood (ML)

The conditional distribution of $\{y_{ijk}\}$ given μ, $\{\alpha_i\}$, $\{\beta_{j(i)}\}$ and σ_e^2 is

$$\left(\{y_{ijk}\}|\mu,\{\alpha_i\},\{\beta_{j(i)}\},\sigma_e^2\right) \sim N\left(\mu + \alpha_i + \beta_{j(i)}, \sigma_e^2\right)$$

Similarly, the conditional distributions of $\{\alpha_i\}$ and $\{\beta_{j(i)}\}$ are respectively,

$$\left(\{\alpha_i\}|\sigma_\alpha^2\right) \sim N\left(0,\sigma_\alpha^2\right) \text{ and } \left(\{\beta_{j(i)}\}|\sigma_{\beta(\alpha)}^2\right) \sim N\left(0,\sigma_{\beta(\alpha)}^2\right)$$

Under these assumptions of normality, the likelihood function is as follows:

$$L\left(\mu,\{\alpha_i\},\{\beta_{j(i)}\},\sigma_\alpha^2,\sigma_{\beta(\alpha)}^2,\sigma_e^2 \mid \{y_{ijk}\}\right) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}abn}\left(\sigma_\alpha^2\right)^{-\frac{1}{2}a}\left(\sigma_{\beta(\alpha)}^2\right)^{-\frac{1}{2}ab}$$

$$\times \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \left(y_{ijk} - \mu - \alpha_i - \beta_{j(i)}\right)^2}{\sigma_e^2} + \frac{\sum_{i=1}^a \alpha_i^2}{\sigma_\alpha^2} + \frac{\sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2}{\sigma_{\beta(\alpha)}^2}\right]\right\} \tag{3}$$

The log-likelihood function can be obtained from the equation (3) and then is solved for its ML estimators by equating to zero and taking the partial derivate of $\ln(L)$ with respect to μ, $\left(bn\sigma_\alpha^2 + n\sigma_{\beta(\alpha)}^2 + \sigma_e^2\right)$, $\left(n\sigma_{\beta(\alpha)}^2 + \sigma_e^2\right)$ and σ_e^2. We can demonstrate these ML solutions as follows:

$$\dot{\sigma}^2_{\alpha,ML} = \frac{\left(1-\frac{1}{a}\right)MS_A - MS_{B(A)}}{bn} \; , \; \dot{\sigma}^2_{\beta(\alpha),ML} = \frac{MS_{B(A)} - MS_E}{n} \; \text{and} \; \dot{\sigma}^2_{e,ML} = MS_E$$

Table 1. ML estimators of the variance components

Conditions	$\tilde{\sigma}^2_e$	$\tilde{\sigma}^2_{\beta(\alpha)}$	$\tilde{\sigma}^2_{\alpha}$
$\dot{\sigma}^2_{\alpha} \geq 0$, $\dot{\sigma}^2_{\beta(\alpha)} \geq 0$	MS_E	$\frac{MS_{B(A)} - MS_E}{n}$	$\frac{\left(1-\frac{1}{a}\right)MS_A - MS_{B(A)}}{bn}$
$\dot{\sigma}^2_{\alpha} \geq 0$, $\dot{\sigma}^2_{\beta(\alpha)} < 0$	$\frac{SS_E + SS_{B(A)}}{a(bn-1)}$	0	$\frac{\left(1-\frac{1}{a}\right)MS_A - \tilde{\sigma}^2_e}{bn}$
$\dot{\sigma}^2_{\alpha} < 0$, $\dot{\sigma}^2_{\beta(\alpha)} \geq 0$	MS_E	$\frac{1}{n}\left(\frac{SS_A + SS_{B(A)}}{ab} - MS_E\right)$	0
$\dot{\sigma}^2_{\alpha} < 0$, $\dot{\sigma}^2_{\beta(\alpha)} < 0$	$\frac{SS_{B(A)} + SS_A + SS_E}{abn}$	0	0

However, these are only the equations of ML solutions, not the ML estimators, which can be shown as $\tilde{\mu}$, $\tilde{\sigma}^2_{\alpha}$, $\tilde{\sigma}^2_{\beta(\alpha)}$ and $\tilde{\sigma}^2_e$ because, especially, $\dot{\sigma}^2_{\alpha}$ and $\dot{\sigma}^2_{\beta(\alpha)}$ can be negative which is contrast to the definition of variance. So, generally if $\dot{\sigma}^2_{\alpha}$ and $\dot{\sigma}^2_{\beta(\alpha)}$ are non-negative, $\tilde{\sigma}^2_{\alpha}$ and $\tilde{\sigma}^2_{\beta(\alpha)}$ will be ML estimators. The ML estimators of σ^2_{α} , $\sigma^2_{\beta(\alpha)}$ and σ^2_e under various conditions on $\dot{\sigma}^2_{\alpha}$ and $\dot{\sigma}^2_{\beta(\alpha)}$ are given in Table 1 (See [2,10,11] for details of derivations).

3.3. Restricted Maximum Likelihood (REML)

REML estimators of σ^2_{α} , $\sigma^2_{\beta(\alpha)}$ and σ^2_e can be obtained by maximizing that part of the likelihood function which is invariant to the location parameters of the model [2,11]. That is, in contrast to ML method, REML method is not interested in the fixed effects of the model. However, it uses fixed effects' degrees of freedoms. This is one of the features of REML method. In order to derive REML estimations, the likelihood function in (3) is separated into fixed and random effects of the model which is given as follows:

$$L\left(\mu,\sigma^2_{\alpha},\sigma^2_{\beta(\alpha)},\sigma^2_e \big| \{y_{ijk}\}\right) = L\left(\mu|\bar{y}\right) \times L\left(\sigma^2_{\alpha},\sigma^2_{\beta(\alpha)},\sigma^2_e \big| MS_A, MS_{B(A)}, MS_E\right) \tag{4}$$

As REML method takes only random part of the model, likelihood function, corresponding to REML, is given as follows:

$$L_{REML} = L\left(\sigma^2_{\alpha},\sigma^2_{\beta(\alpha)},\sigma^2_e \big| MS_A, MS_{B(A)}, MS_E\right)$$

$$= (2\pi)^{-\frac{1}{2}(abn-1)} \left(bn\sigma^2_{\alpha} + n\sigma^2_{\beta(\alpha)} + \sigma^2_e\right)^{-\frac{1}{2}(a-1)} \left(n\sigma^2_{\beta(\alpha)} + \sigma^2_e\right)^{-\frac{1}{2}a(b-1)} \left(\sigma^2_e\right)^{-\frac{1}{2}ab(n-1)} \tag{5}$$

$$\times (abn)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{SS_A}{bn\sigma^2_{\alpha} + n\sigma^2_{\beta(\alpha)} + \sigma^2_e} + \frac{SS_{B(A)}}{n\sigma^2_{\beta(\alpha)} + \sigma^2_e} + \frac{SS_E}{\sigma^2_e}\right]\right\}$$

Then we can obtain REML estimators by taking the log of the likelihood function in (5) and equating to zero and taking the partial derivatives of $\ln(L_{REML})$ with respect to $\left(bn\sigma_\alpha^2 + n\sigma_{\beta(\alpha)}^2 + \sigma_e^2\right)$, $\left(n\sigma_{\beta(\alpha)}^2 + \sigma_e^2\right)$ and σ_e^2, respectively. After editing of these equations, finally we obtain the REML estimators as follows:

$$\dot{\sigma}_{\alpha,REML}^2 = \frac{MS_A - MS_{B(A)}}{bn}, \ \dot{\sigma}_{\beta(\alpha),REML}^2 = \frac{MS_{B(A)} - MS_A}{bn} \text{ and } \dot{\sigma}_{e,REML}^2 = MS_E$$

However, as in the ML estimators, these solutions are not the REML estimators due to requirements of non-negativity. The REML estimators of $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_{\beta(\alpha)}^2$ and $\hat{\sigma}_e^2$ under various conditions on $\dot{\sigma}_\alpha^2$ and $\dot{\sigma}_{\beta(\alpha)}^2$ are given in Table 2 (See [2,10,11] for details of derivations).

Table 2. REML estimators of the variance components

Conditions	$\hat{\sigma}_e^2$	$\hat{\sigma}_{\beta(\alpha)}^2$	$\hat{\sigma}_\alpha^2$
$\dot{\sigma}_\alpha^2 \geq 0$, $\dot{\sigma}_{\beta(\alpha)}^2 \geq 0$	MS_E	$\dfrac{MS_{B(A)} - MS_E}{n}$	$\dfrac{MS_A - MS_{B(A)}}{bn}$
$\dot{\sigma}_\alpha^2 \geq 0$, $\dot{\sigma}_{\beta(\alpha)}^2 < 0$	$\dfrac{SS_E + SS_{B(A)}}{a(bn-1)}$	0	$\dfrac{MS_A - \tilde{\sigma}_{e,REML}^2}{bn}$
$\dot{\sigma}_\alpha^2 < 0$, $\dot{\sigma}_{\beta(\alpha)}^2 \geq 0$	MS_E	$\dfrac{1}{n}\left(\dfrac{SS_A + SS_{B(A)}}{ab-1} - MS_E\right)$	0
$\dot{\sigma}_\alpha^2 < 0$, $\dot{\sigma}_{\beta(\alpha)}^2 < 0$	$\dfrac{SS_{B(A)} + SS_A + SS_E}{abn-1}$	0	0

3.4. Bayesian Method

In order to carry out Bayesian analysis, in addition to the likelihood function we need prior distributions for each parameter of the model. For the fixed effect, a flat prior was used, so that $f(\mu) \propto$ constant, indicating no prior knowledge about this parameter. The normal distribution (N) was used as the prior distribution of the random effects, $f\left(\{\alpha_i\}\big|\sigma_\alpha^2\right) \sim N\left(0, \sigma_\alpha^2\right)$ and $f\left(\{\beta_{j(i)}\}\big|\sigma_\beta^2\right) \sim N\left(0, \sigma_\beta^2\right)$. The inverse gamma (IG) distribution is assigned for the variance components, $f\left(\sigma_i^2\right) \sim IG\left(\upsilon_i, s_i^2\right)$ The forms of the prior distributions are

$$f(\mu) \propto \text{constant}$$

$$f\left(\{\alpha_i\}\big|\sigma_\alpha^2\right) \propto \left(\sigma_\alpha^2\right)^{-\frac{1}{2}a} \exp\left(-\frac{\sum\limits_{i=1}^{a}\alpha_i^2}{2\sigma_\alpha^2}\right)$$

$$f\left(\{\beta_{j(i)}\}\big|\sigma_{\beta(\alpha)}^2\right) \propto \left(\sigma_{\beta(\alpha)}^2\right)^{-\frac{1}{2}ab} \exp\left(-\frac{\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{b}\beta_{j(i)}^2}{2\sigma_{\beta(\alpha)}^2}\right) \tag{6}$$

$$f\left(\sigma_i^2\middle|\upsilon_i,s_i^2\right)\propto\left(\sigma_i^2\right)^{-\frac{1}{2}(\upsilon_i+2)}\exp\left\{-\frac{\upsilon_i s_i^2}{2\sigma_i^2}\right\}\quad i=\alpha,\beta,e$$

By multiplying likelihood function in (3) with the prior distributions of all the parameters in (6), the joint posterior density of parameters is obtained as:

$$f(\boldsymbol{\theta}|\mathbf{y})\propto L(\boldsymbol{\theta}|\mathbf{y})f(\mu)f\left(\{\alpha_i\}\middle|\sigma_\alpha^2\right)f\left(\{\beta_{j(i)}\}\middle|\sigma_{\beta_{j(i)}}^2\right)f\left(\sigma_\alpha^2\middle|\upsilon_\alpha,s_\alpha^2\right)f\left(\sigma_{\beta(\alpha)}^2\middle|\upsilon_{\beta(\alpha)},s_{\beta(\alpha)}^2\right)f\left(\sigma_e^2\middle|\upsilon_e,s_e^2\right)$$

$$\propto\left(\sigma_\alpha^2\right)^{-\frac{1}{2}(a+\upsilon_\alpha+2)}\left(\sigma_{\beta(\alpha)}^2\right)^{-\frac{1}{2}(ab+\upsilon_{\beta(\alpha)}+2)}\left(\sigma_e^2\right)^{-\frac{1}{2}(abn+\upsilon_e+2)}$$

$$\times\exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\mu-\alpha_i-\beta_{j(i)}\right)^2+\upsilon_e s_e^2}{\sigma_e^2}+\frac{\sum_{i=1}^{a}\alpha_i^2+\upsilon_\alpha s_\alpha^2}{\sigma_\alpha^2}+\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2+\upsilon_{\beta(\alpha)}s_{\beta(\alpha)}^2}{\sigma_{\beta(\alpha)}^2}\right]\right\}$$

To implement the Gibbs sampling algorithm, we require the full conditional posterior distributions of μ, α, β, σ_α^2, $\sigma_{\beta(\alpha)}^2$ and σ_e^2. The full conditional posterior distribution of any parameter of interest can be obtained by integrating over the remaining parameters from joint posterior distribution. These are summarized as follows:

$$\left[\mu\middle|\{\alpha_i\},\{\beta_{j(i)}\},\sigma_\alpha^2,\sigma_{\beta(\alpha)}^2,\sigma_e^2,\mathbf{y}\right]\sim N\left(\left(\bar{y}_{...}-\bar{\alpha}_.-\bar{\beta}_{j(i)}\right),\sigma_e^2/abn\right)$$

$$\left[\{\alpha_i\}\middle|\mu,\{\beta_{j(i)}\},\sigma_\alpha^2,\sigma_{\beta(\alpha)}^2,\sigma_e^2,\mathbf{y}\right]\sim N\left(\left(\frac{\sigma_\alpha^2}{bn\sigma_\alpha^2+\sigma_e^2}\right)n\left(b\left(\bar{y}_{i..}-\mu\right)-b\bar{\beta}_{j(i)}\right),\frac{\sigma_\alpha^2\sigma_e^2}{bn\sigma_\alpha^2+\sigma_e^2}\right)$$

$$\left[\{\beta_{j(i)}\}\middle|\mu,\{\alpha_i\},\sigma_\alpha^2,\sigma_{\beta(\alpha)}^2,\sigma_e^2,\mathbf{y}\right]\sim N\left(\frac{n^2\sigma_{\beta(\alpha)}^2\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\mu-\alpha_i\right)}{n\sigma_{\beta(\alpha)}^2+\sigma_e^2},\frac{\sigma_e^2\sigma_{\beta(\alpha)}^2}{n\sigma_{\beta(\alpha)}^2+\sigma_e^2}\right)$$

$$\left[\sigma_\alpha^2\middle|\mu,\{\alpha_i\},\{\beta_{j(i)}\},\sigma_{\beta(\alpha)}^2,\sigma_e^2,\mathbf{y}\right]\sim IG\left(\frac{a+\upsilon_\alpha}{2},\frac{\sum_{i=1}^{a}\alpha_i^2+\upsilon_\alpha s_\alpha^2}{2}\right)$$

$$\left[\sigma_{\beta(\alpha)}^2\middle|\mu,\{\alpha_i\},\{\beta_{j(i)}\},\sigma_\alpha^2,\sigma_e^2,\mathbf{y}\right]\sim IG\left(\frac{ab+\upsilon_{\beta(\alpha)}}{2},\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2+\upsilon_{\beta(\alpha)}s_{\beta(\alpha)}^2}{2}\right)$$

$$\left[\sigma_e^2 \middle| \mu, \{\alpha_i\}, \{\beta_{j(i)}\}, \sigma_\alpha^2, \sigma_{\beta(\alpha)}^2, \mathbf{y}\right] \sim IG\left(\frac{abn + \upsilon_e}{2}, \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \mu - \alpha_i - \beta_{j(i)}\right)^2 + \upsilon_e s_e^2}{2}\right)$$

The Gibbs sampling algorithm generates random samples from the full conditional distributions of the parameters, without having to calculate the density. Gibbs sampling algorithm requires an initial starting point for the parameters. Then, one at a time, a value for each parameter of interest is sampled given the values for the other parameters and data. Once all of the parameters of interest have been sampled, the nuisance parameters are sampled given the parameters of interest and the observed data. At this point, the process is started over. The power of Gibbs sampling is that the joint distribution of the parameters will converge to the joint probability of the parameters given the observed data.

4. NUMERICAL EXAMPLE

Let's consider an experiment from animal breeding in which the levels of factor A are sires of a particular breed, and the levels of factor B are dams mated to those sires. The dams are a random sample within the sires. Milk yield was measured on offspring of those sires and dams. The offspring present random samples within the dams. If any relationship among the dams is ignored, then the dams bred by different sires are independent. Also, the offspring of different dams and sires are independent of each other. Therefore, in this example the first step is to choose sires, and the second step is to choose dams within sires. Such an experiment is typically a two-way nested design which is commonly used in variance estimation problems.

Table 3. Records for milk yields (kg)

Sire	Dam	Offspring's yield							
		Data set 1		Data set 2		Data set 3		Data set 4	
	1	4379	6560	6379	6560	6379	6560	6679	6560
	2	5560	7733	6560	6733	6560	6733	6560	6333
1	3	4637	5639	6637	6639	6637	6639	6637	6639
	4	5726	5576	6726	6576	6726	6576	6826	6576
	5	4968	4574	6968	6574	6968	6574	6968	6574
	6	5355	7057	4355	4057	6355	7057	6355	7057
	7	4605	4180	4605	4180	6605	6180	6605	6180
2	8	4393	4530	4393	4530	6393	6530	6393	6530
	9	5195	7024	4195	4024	6195	7024	6195	7024
	10	6137	4748	4137	4748	6137	6748	6137	6748
	11	6253	6666	6253	6666	6253	6666	6553	6666
	12	5553	6026	6553	6026	5553	6026	6453	6726
3	13	6268	7575	6268	6575	6268	6575	6268	6575
	14	7112	6382	6112	6382	7112	6382	6612	6782
	15	5840	7316	6840	6316	6840	7316	6840	6316
	16	6246	5595	5246	5595	6246	6595	6946	6795
	17	5400	6440	5400	5440	6400	6440	6400	6440
4	18	7301	6615	5301	5615	7301	6615	6801	6515
	19	5453	6592	5453	5592	6453	6592	6453	6592
	20	7374	6693	5374	5693	7374	6693	6374	6693

We consider 4 different data sets to determine variance components from such an experiment. For this reason, 4 sires were randomly chosen with 5 dams per sire and 2 offspring per dam. Then, the milk yield of offspring was recorded (305 days). This data set is shown in the first column of Table 3. Other data sets (data set 2, data set 3 and data set 4) were obtained by modifying the data set 1.

In the analysis of data sets 1 to 4, a normal distribution with means obtained from REML method (5932, 5707, 6582 and 6584, respectively) were used with a variance of 1000. An inverse gamma distribution with shape and scale parameters set to 0.001, IG(0.001,0.001), was chosen as the prior distribution for each of the variance components in models $\left(\sigma_\alpha^2, \sigma_{\beta(\alpha)}^2, \sigma_e^2\right)$ to reflect prior ignorance about the parameters.

Statistical analyses for ANOVA, ML and REML estimations were obtained using PROC MIXED procedure, and Bayesian analysis was conducted using the PROC MCMC procedure of SAS software [12]. A single chain of size 500000 iterations was run. The initial 10000 iterations were discarded as a burn-in, and every 50th sample was recorded to reduce the auto-correlation. In total, 9800 samples were stored for each parameter, and means of the sample values were used as an estimate of the parameters.

5. RESULTS

The results from four different methods are shown in Table 4. As can be seen from this table, the ANOVA estimates of the parameters outside the parameter space are treated as they are and data sets yield ANOVA estimates of σ_α^2 and $\sigma_{\beta(\alpha)}^2$ ranging from -16480 to 1101634. Although the data set 1 has positive ANOVA estimates for σ_α^2 and $\sigma_{\beta(\alpha)}^2$, the data sets 2, 3 and 4 are the most difficult ones and badly behaved, in that the estimates of σ_α^2 and $\sigma_{\beta(\alpha)}^2$ are negative, rendering inferences about these parameters very difficult. The ANOVA estimators of σ_α^2 and $\sigma_{\beta(\alpha)}^2$ are unbiased yet they can be negative. The ML and REML estimators are obtained by deleting those negative values and substituting zero. Therefore, these estimators are biased upwards.

The posterior means for σ_e^2, σ_α^2 and $\sigma_{\beta(\alpha)}^2$ from the analysis of data sets 1-4 in Table 4 are based on 9800 Gibbs sampler. These point estimates of variance components from Bayesian analysis are within the permissible parameter space in contrast to the estimates obtained from ANOVA. The Bayesian method is feasible computationally and appears to give much more sensible answer to the inferential problems than ANOVA and likelihood-based estimation methods.

Table 4. ANOVA, ML, REML and Bayesian estimations of variance components

Methods	σ_α^2	$\sigma_{\beta(\alpha)}^2$	σ_e^2
Data Set 1			
ANOVA	259296	97566	689557
ML	172355	97567	689557
REML	259297	97567	689557
Bayesian	251784	105408	785807
Data Set 2			
ANOVA	1101634	-12864	54107
ML	823985	0	42673
REML	1100051	0	42673
Bayesian	1672998	994	45069
Data Set 3			
ANOVA	-10245	32510	107957
ML	0	20501	107958
REML	0	24392	107981
Bayesian	3453	23362	116676
Data Set 4			
ANOVA	-1332	-16480	68712
ML	0	0	50339
REML	0	0	51629

Bayesian	623	1050	53412

It can be noted that the Bayesian method overestimates the variance components compared with the estimates of ANOVA, ML and REML. Variance components obtained by REML are only marginal with respect to fixed effects but conditionals to other nuisance parameters of the model. The Bayesian analysis allows further marginalization via Markov chain Monte Carlo methods. This approach is particularly interesting for models, as the present, with high number of variance components. In consequence, point estimates of variance components obtained in the Bayesian analysis under that priors presented some differences with the ML and REML estimates.

6. DISCUSSIONS

In this paper we have presented four different methods of estimating the variance components. Bayesian estimators depend not only on the information about the parameters contained in the data, but also on prior knowledge. This is one of the potential advantages of the Bayesian methods. Therefore, it is expected that the Bayesian method will do better than the classical procedures when the data contain little information about the parameters of interest. Moreover, the Bayesian method implicitly account for the uncertainty about the values at the parameters of interest.

The computations required to implement the Bayesian method are of the same order of magnitude as those required for the classical methods, and therefore the Bayesian method are likely to be computationally feasible whenever the classical procedures are computationally feasible.

One of the main differences between the Bayesian and maximum likelihood approaches is the way in which they deal with nuisance parameters. This is apparent from our results. The likelihood function is obtained by maximizing with respect to the nuisance parameters, whereas the conditional posterior density is obtained by a Monte Carlo numerical integration method, which is known as a Gibbs Sampler. In certain cases, the two operations may produce sharply contrasting results.

While variance components estimations of all parameters for each method are non-negative in the first data set, there are negative ANOVA estimations for the other data sets. So that, when ANOVA estimations are negative, ML and REML estimations are zero. REML estimators are quite different from ML estimators because REML method eliminates deviations which are originated from ML method. Moreover, REML estimations of first data set are almost similar with ANOVA estimations [13].

Based on the results from our four data sets, we can conclude that the estimates of ANOVA, ML and REML are accurate but the posterior point estimates from the Gibbs sampling can be overestimated depending on the nature of the data set. The differences in the results of different estimation methods occurred the most in the estimation of sire and dam variances. Data sets 2, 3 and 4 give highly negative estimates of sire and dam variance components from ANOVA. ML and REML methods truncate these estimates to zero, and the posterior point estimates from the Gibbs sampler tend to be biased. This is due to the fact that each data set provides relatively small amount of information for sire and dam variances than for the error variance.

Finally, we can conclude that the Bayesian method of estimation using the Gibbs sampling approach is suitable for estimating the variance components under balanced two-way nested design as compared to traditional methods, particularly for small sample data sets.

REFERENCES

[1] Theobald, C. M., Fırat, M. Z. and Thompson, R. Gibbs sampling, adaptive rejection sampling and robustness to prior specification for a mixed linear model. Genetic Selection Evolution, 29:57-72, 1997.

[2] Searle, R. S., Casella, G. and McCulloch, C. E. Variance Components. New York: Wiley; 1992.

[3] Thompson, Jr. W. A. The problems of negative estimates of variance components. Annals of Mathematical Statistics, 33:273-289, 1962.

[4] Patterson, H. D. and Thompson, R. Recovery of inter-block information when block sizes are unequel. Biometrika, 58:545-554, 1971.

[5] Hill, B. M. Inference about variance components in the one-way model. Journal of the American Statistical Association, 60:806-825, 1965.

[6] Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72:320-338, 1977.

[7] Gianola, D. and Fernando, R. L. Bayesian methods in animal breeding theory. Journal of Animal Science, 63:217-244, 1986.

[8] Gianola, D., Foulley, J. L. and Fernando, R. L. Prediction of breeding values when variances are not known. Genetic Selection Evolution, 18:485-498, 1986.

[9] Geman, S. and Geman, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions and Pattern Analysis and Machine Intelligence, 6:721-741, 1984.

[10] Fırat, M. Z. Hayvan ıslahında negatif varyans unsuru tahmini ve tahmin yöntemlerinin incelenmesi. Ç.Ü. Ziraat Fakültesi Dergisi, 12:169-176, 1997.

[11] Sahai, H. and Ojeda, M. M. Analysis of Variance for Random Models- Balanced Data: Vol. I: Theory, Methods, Applications and Data Analysis. Birkhauser Boston, Cambridge; 2004.

[12] Sas Institute, Inc. Sas version 9.3, Online Help, Cary, NC: Sas Institute. 2009.

[13] Fırat, M. Z. ve Bek, Y. Varyans unsurlarının tahmini için maksimum olabilirlik metotlarının karşılaştırmalı olarak incelenmesi. Ç.Ü. Ziraat Fakültesi Dergisi, 12:1-8, 1997.