ELSEVIER

Journal of Statistical Planning and Inference 103 (2002) 117–135

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A composite likelihood approach to (co)variance components estimation

Subhash Lele[a,*], Mark L. Taper[b]

[a] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada AB T6G 2G1*

[b] *Department of Ecology, Montana State University, Bozeman, MT 59717, USA*

Abstract

Estimation of variance and covariance components has importance in various substantive fields such as animal breeding and evolutionary biology among others. The most popular methods of variance components estimation are maximum likelihood (ML), restricted maximum likelihood (REML), analysis of variance and covariance (ANOVA) and minimum quadratic norm (MINQUE). All these methods are computationally intensive. This computational barrier is particularly limiting in data obtained from large animal breeding experiments involving multiple traits. The purpose of this paper is to introduce a new method, which we call maximum composite likelihood (MCL), for the estimation of variance and covariance components. This method is as generally applicable as the method of maximum likelihood: to cases where designs are balanced or unbalanced, involving mixed effects and multiple traits or designs where random effects are correlated to each other. The MCL approach, in contrast to ML/REML or ANOVA, however, does not require inversion of matrices. As a consequence the computational burden is reduced from $O(N^3)$ to $O(N^2)$ where N denotes the total sample size. Moreover, and in contrast to the ML/REML estimating functions, the estimating functions obtained for MCL, after a minor modification, are shown to possess a unique solution thus guaranteeing convergence of the numerical optimization routine. Conditions are specified that assure consistency and asymptotic normality of these estimators. These results do not depend on the assumption of a Gaussian distribution of the random effects. Simulation study indicates that there is only a small loss of statistical efficiency in using MCL as compared to REML but a substantial gain in the computational efficiency.

Keywords: Restricted maximum likelihood; Heritability; Maximum likelihood; Analysis of variance; Henderson's methods; Animal breeding; Multiple traits; Minimum norm quadratic unbiased estimation; Unbalanced designs; Genetics; Variance components

1. Introduction

Without question, analysis of (co)variance and estimation of (co)variance components has held a central position in both statistical theory and statistical practice (Searle

* Corresponding author.
E-mail address: slele@ualberta.ca (S. Lele).

0378-3758/02/$ - see front matter
PII: S0378-3758(01)00215-4

and McCulloch, 1996). Discussion of various applications and a thorough review of the existing techniques for the estimation of variance and covariance components can be found in Gianola and Hammond (1990), Searle et al. (1992) and Rao and Kleffe (1988). Estimation of the covariance components is also important for many evolutionary studies. The issues related to the applications of covariance components in evolutionary studies are discussed in Shaw (1987, 1991), Lande (1979) and Kohn and Atcheley (1988) and references therein.

There are a number of different approaches available for the estimation of (co)variance components. The most prominent methods are analysis of variance (ANOVA) (Henderson, 1953), based on equating the properly chosen sums of squares to their expectations; maximum likelihood (ML) estimation (Hartley and Rao, 1967), based on maximizing the likelihood of the observations, and restricted maximum likelihood (REML) estimation (Patterson and Thompson, 1971), based on maximizing the likelihood of the contrasts that eliminate the fixed effects. Other commonly used methods are those based on minimum norm quadratic estimation (MINQUE) (Rao, 1971; Rao and Kleffe, 1988) and dispersion-mean approach (Seely, 1970; Pukelsheim, 1976). Each of these methods has merits and demerits. For example, ANOVA estimators are usually easy to understand and are unbiased but may lead to negative estimates of a nonnegative parameter. In unbalanced data situations, ANOVA estimates may not be unique, in the sense that the same data set analyzed by different researchers can lead to different numerical estimates of the (co)variance components. See Chapter 5 of Searle et al. (1992) for a detailed discussion of the non-uniqueness problem associated with the ANOVA method for unbalanced data. MINQUE estimators are optimal only near the a priori chosen value and also suffer from non-uniqueness problems. ML and REML estimators are optimal under the model assumptions but can be extremely difficult computationally, especially for analyzing data involving multiple traits with missing data. Searle et al. (1992) is a good source for these comparisons.

This paper proposes a new method for the estimation of (co)variance components based on the idea of composite likelihood (Lindsay, 1988). This method is computationally simple, amenable to modern distributed computing environments and is statistically satisfactory in the sense that it is as generally applicable as the method of maximum likelihood. The salient features of the method of maximum composite likelihood (MCL) are: (1) No need for the inversion of large matrices, (2) Consistency and asymptotic normality of the estimators, (3) Uniqueness of the solution and (4) Model robustness. The asymptotic loss of statistical efficiency as compared to REML is small, but the gain in computational simplicity is substantial.

The outline of the paper is as follows. In Section 2 we provide a brief review of the basic ideas behind composite likelihood. In Section 3, we study some theoretical properties of the MCL estimators. In particular, we provide conditions for the uniqueness, consistency and asymptotic normality of the MCL estimators for the single trait, random effects model. Further, in Section 4, we show that the MCL estimating functions for mixed effects and multiple traits cases can be reduced to the same structural form as for the simple random effects case. Hence all the results from the random effects case

generalize to these cases. In Section 5, we compare, using simulations, the statistical efficiency of the MCL estimators with the REML estimators for one-way unbalanced designs with Gaussian random effects. For this simulation study we replicate the study design used by Swallow and Monahan (1984) with some extensions to consider sample sizes that are commonly obtained in practice. We also present efficiency comparisons under non-normal random effects with skew and kurtosis. We conclude the paper in Section 6 by discussing future directions.

2. Composite likelihood preliminaries

A composite likelihood is formed by adding together individual component log-likelihoods each of which is a valid marginal or conditional log-likelihood (Lindsay, 1988). The key utility of the composite likelihood is that the component score functions form an additive estimating function that can be used to provide consistent parameter estimates in settings where a full maximum likelihood is not feasible or is computationally difficult. Key examples of successful composite likelihood approaches include working independence generalized estimating equations (Liang and Zeger, 1986) for longitudinal data, pseudolikelihood methods for Markov random fields (Besag, 1975) and statistical inference for hierarchical spatial models (Heagerty and Lele, 1998). The term composite likelihood simply refers to the pooling of likelihood contributions in a multiplicative fashion in circumstances where the components do not necessarily represent independent replicates. As an example, consider a random vector Y_N of length N with joint probability density $f_N(.,\theta)$, that is $Y_N \sim f_N(.,\theta)$. Let $f(Y_i,\theta)$ denote the marginal density of Y_i, the ith component of Y_N. Similarly let $f(Y_i,Y_j,\theta)$ denote the joint density of any two components and let $f(Y_i|Y_j,\ j\neq i,\theta)$ denote the conditional density of the ith component given all the other components. There are a variety of different "composite likelihoods" that one can write: $\mathrm{CL}_1(\theta,Y_N)=\prod_{i=1}^{N} f(Y_i,\theta)$, which is a product of marginal distributions, $\mathrm{CL}_2(\theta,Y_N)=\prod_{i=1}^{N-1}\prod_{j>i}^{N} f(Y_i,Y_j,\theta)$, product of all bivariate marginal distributions, or, $\mathrm{CL}_3(\theta,Y_N)=\prod_{i=1}^{N} f(Y_i|Y_j,\ j\neq i;\theta)$, product of all conditional distributions. If the marginal distributions are easier to describe, as in the multivariate Gaussian distributions, one may prefer the first two types of composite likelihoods whereas if the conditional distributions are easily specified, as in the Markov random field situation, one may prefer the third type of composite likelihood. Analytical simplicity, computational convenience, and the questions and parameters of interest determine the form of the composite likelihood.

A simple example of the use of composite likelihood is in the case of spatial data. Let Y denote the vector of observations at K locations. Suppose Y is a multivariate Gaussian vector, that is $Y \sim N_K(0,\Sigma(\rho))$ where $\mathrm{cov}(Y_i,Y_j)=\rho^{d(i,j)}$ with $d(i,j)$ denoting the distance between the locations and $|\rho|<1$. The full likelihood for ρ involves inversion of a $K\times K$ matrix as well as evaluation of its determinant. This could be prohibitive if K, the number of locations, is large. Now suppose we consider all possible contrasts of the type (Y_i-Y_j). The marginal distribution of these contrasts is Gaussian

with mean 0 and variance $2(1-\rho^{d(i,j)})$. One possible version of composite likelihood can be based on the product of the marginal distribution of all possible pairwise contrasts: $\mathrm{CL}(\rho, Y) = \prod_{i=1}^{K-1} \prod_{j>i} f(Y_i - Y_j, \rho)$. Notice that this composite likelihood involves no inversion of matrices. Maximizing this composite likelihood with respect to ρ provides the maximum composite likelihood estimator of ρ. This particular example corresponds to the estimation of variogram parameters as described in Lele (1997) and Curriero and Lele (1999). Notice that the evaluation of the full likelihood involves $\mathrm{O}(K^3)$ operations for the inversion of a $K \times K$ matrix, whereas the evaluation of the composite likelihood involves $\mathrm{O}(K^2)$ operations. This idea of using the composite likelihood based on contrasts can be successfully exploited in the estimation of variance components. The computational advantage garnered by the use of composite likelihood in the variance and covariance components estimation can be substantial. This may be of particular importance in the analysis of large breeding pedigrees where K may run into the millions.

3. Composite likelihood for variance components estimation—general theory

In this section, we develop some general theoretical results for the variance components estimators based on the composite likelihood approach. In particular, we prove that these estimators are consistent and asymptotically normal. We also show that a consistent solution to the estimating function corresponding to the maximization of the composite likelihood, called the composite score function, can be obtained uniquely. This guarantees convergence of the numerical algorithm for any dataset. These theoretical developments for the composite likelihood estimator are shown to be general enough to cover all the cases where ML or REML method can be applied: designs with random effects, mixed effects, covariance components estimation that allows for relationships between individuals and between random effects and missing traits. We start with the notationally simplest situation, the simple random effects model.

3.1. Simple random effects model

Following the notation of Searle et al. (1992, p. 234), we write the linear model corresponding to the simple random effects case as:

$Y = \mu 1 + Z_1 u_1 + Z_2 u_2 + \cdots + Z_R u_R + e$ where Z_r's are design matrices, u_r's are random effects and environmental variation is denoted by e. Assuming that the random effects are Gaussian, that is, $u_r \sim N(0, \sigma_r^2 I), e \sim N(0, \sigma_0^2 I)$, and, they are independent of each other, it follows that $Y \sim N(\mu \underline{1}, \sum_{r=0}^{R} \sigma_r^2 Z_r Z_r^{\mathrm{T}})$ where the environmental variation is included in the term σ_0^2 with $Z_0 = I_{\mathrm{N}}$. Let N denote the total number of observations, the length of the vector Y. Consider a matrix A of dimension $N(N-1)/2$ by N such that each of its rows has exactly one entry of '1', one entry of '-1' with all other entries '0' and no two rows are identical. Premultiplying Y by A gives us the vector of all possible contrasts between pairs of observations. Let $W = AY$ then the distribution

of the vector W is given by $W \sim N(\underline{0}, \sum_{r=0}^{R} \sigma_r^2 A Z_r Z_r^{\mathrm{T}} A^{\mathrm{T}})$. For the sake of notational simplicity, let $C_r = A Z_r Z_r^{\mathrm{T}} A^{\mathrm{T}}$.

Consider the composite likelihood written as the product of only the marginal distributions of the components of W, namely, $\mathrm{CL}(\theta, W) = \prod_{i=1}^{L} f(W_i, \theta)$ where $\theta = (\sigma_r^2, r = 0, 1, \ldots, R)^{\mathrm{T}}$ and $L = N(N-1)/2$. This can be explicitly written as

$$\mathrm{CL}(\theta, W) = \prod_{i=1}^{L} \left\{ \frac{1}{\sqrt{2\pi}(\sum_{r=0}^{R} \sigma_r^2 C_{ii,r})^{1/2}} \exp\left(-\frac{1}{2} \frac{W_i^2}{(\sum_{r=0}^{R} \sigma_r^2 C_{ii,r})} \right) \right\}, \tag{1}$$

where $C_{ii,r}$ denotes the ith diagonal element of the matrix C_r. The set of estimating functions for the variance components are given by differentiating the log-composite likelihood with respect to each variance component and equating the result to zero:

$$\sum_{i=1}^{L} \frac{C_{ii,k}}{\sum_{r=0}^{R} \sigma_r^2 C_{ii,r}} \left(\frac{W_i^2}{\sum_{r=0}^{R} \sigma_r^2 C_{ii,r}} - 1 \right) = 0 \quad \text{for } k = 0, 1, 2, \ldots, R.$$

We call these estimating functions the 'composite score functions'.

Let $B^{\langle i \rangle} = (C_{ii,0}, C_{ii,1}, \ldots, C_{ii,R})^{\mathrm{T}}$, be a vector consisting of the ith diagonal elements of the matrices $C_r = A Z_r Z_r^{\mathrm{T}} A^{\mathrm{T}}$ and let $B = (B^{\langle 1 \rangle}, B^{\langle 2 \rangle}, \ldots, B^{\langle L \rangle})$. The matrix B is a $(R+1)$ by L matrix consisting of the diagonal elements of the C_r matrices. With this notation, the marginal distribution of W_i can be written as $W_i \sim N(0, \theta^{\mathrm{T}} B^{\langle i \rangle})$ and the composite score functions can be written as

$$\sum_{i=1}^{L} \frac{B_k^{\langle i \rangle}}{\theta^{\mathrm{T}} B^{\langle i \rangle}} \left(\frac{W_i^2}{\theta^{\mathrm{T}} B^{\langle i \rangle}} - 1 \right) = 0 \quad \text{for } k = 0, 1, 2, \ldots, R. \tag{2}$$

Strictly speaking, the MCL estimator is that value of the parameter vector that maximizes the composite likelihood function (Eq. (1)). However, we will somewhat loosely call the consistent solution to the above set of equations (Eq. (2)) the maximum composite likelihood (MCL) estimator. Notice also that this solution itself is not guaranteed to be positive in contrast to the true MCL estimator which, by definition, is positive. However we ignore this subtle difference in the rest of the paper.

The reason for the use of the notation in Eq. (2) will be clear when we deal with more complicated situations such as correlated random effects, multiple traits with missing data among others. For these general cases we will show that the estimating functions can be reduced to the above generic form. Thus, the arguments developed below for the uniqueness of the solution and asymptotic properties such as consistency hold for all situations.

3.2. Consistency and asymptotic normality of the MCL estimators

In the following, we provide conditions under which the MCL estimators are consistent and asymptotically normal. First we show that there exists at least one solution to the composite score function which is consistent, that is, the estimators of variance components converge in probability to the true values as the sample size increases. Then

we show that this consistent solution, under additional assumptions, is also asymptotically normal. Then we tackle the issue of finding the consistent solution out of all possible solutions. Towards this goal, we modify the set of composite score functions so that the resultant set has a unique, consistent solution. This solution is then used as a starting value in the Fisher scoring algorithm for solving the composite score function and obtain a one-step estimator, which is shown to be asymptotically equivalent to the MCL estimator. This program is similar to the one-step efficient estimators described, for example, in Lehmann (1985, pp. 422–423) or Serfling (1980, pp. 258–259).

3.2.1. Consistency of the MCL estimator

Throughout, without explicit mention, we assume that the variance components are unconfounded, that is to say, we assume that the design matrices are linearly independent of each other. We also assume that the parameter space is an open subset of the Euclidean space E^{R+1}, that is, none of the variance components are zero. This assumption avoids the technical difficulties involved when the parameters lie on the boundary of the parameter space. The asymptotic theory for REML estimators in such situation can possibly be adapted to show similar results for MCL estimators as well. We refer the reader to Miller (1977) or Jiang (1996) for more details on the identifiability conditions for the variance components models.

3.2.2. Assumptions

A0: The entries in the design matrices Z_r are uniformly bounded for all $r = 0, 1, \ldots, R$.
A1: Let $n_k = \#(B_k^{\langle i \rangle} \neq 0,\ i, 1, 2, \ldots, L)$. That is, n_k denotes the number of non-zero entries in the kth row of the B matrix. The number n_k corresponds to the number of contrasts that contribute to the estimating function for the kth variance component. We assume that $n_k \to \infty$ for all $k = 0, 1, 2, \ldots, R$ as $N \to \infty$.
A2: Let $C = \sum_{r=0}^{R} A Z_r Z_r^{\mathrm{T}} A^{\mathrm{T}}$. Assume that the number of non-zero elements in any row of this matrix divided by $n = \min(n_k, k = 0, 1, 2, \ldots, R)$ converges to zero as N, the sample size, converges to infinity. Notice that the matrix C is related to the covariance matrix for the contrast vector W. This assumption says that there are enough uncorrelated components in the estimating functions.
A3: Assume that matrix B is of full rank $(R + 1)$.
A4:

$$\lim_{N \to \infty} \frac{1}{\min(n_k, n_m)} \sum_{i=1}^{L} \frac{C_{ii,k} C_{ii,m}}{(\sum_{r=0}^{R} \sigma_r^2 C_{ii,r})^2} = I_{k,m}.$$

We assume that the matrix $I = [I_{k,m}]$ is positive definite.

These assumptions are both intuitive and easily satisfied in most practical situations. Assumption A0 holds, for example, when Z_r's are incidence matrices. Assumption A1 says that, for all k, the number of contrasts that contribute to the estimation of σ_k^2 becomes large as the sample size increases. Assumption A2 assures that there are enough uncorrelated pieces of information to estimate all the variance components.

Assumption A3 relates to the identifiability of the variance components. Assumption A4 says that information about each variance components goes to infinity fast enough.

Theorem 1. *Under the assumptions* A0–A4, *one of the solutions to the composite score functions is a consistent estimator of the variance components.*

The theorem can be proved by showing that with probability tending to 1, there exists a solution to the estimating functions in an arbitrarily small neighborhood of the true value. This follows from the zero unbiasedness of the estimating functions, Taylor series expansion and convergence of the empirical information matrix to a positive definite matrix. A detailed proof is given in the appendix.

3.2.3. Asymptotic normality of the MCL estimator

To prove the asymptotic normality of the consistent solution obtained using MCL we need some additional notation and assumptions.

Additional notation: Let the vector of the 'composite score functions' be denoted by

$$G(\theta)=\left[\sum_{i=1}^{L}\frac{C_{ii,k}}{\sum_{r=0}^{R}\sigma_r^2C_{ii,r}}\left(\frac{W_i^2}{\sum_{r=0}^{R}\sigma_r^2C_{ii,r}}-1\right)\right]_{k=0,1,2,\ldots,R}.$$

We need to calculate the covariance matrix for these estimating functions, namely, $E(GG')$. Notice that

$$\mathrm{Cov}\left(\frac{W_i^2}{\sum_{r=0}^{R}\sigma_r^2C_{ii,r}}-1,\frac{W_j^2}{\sum_{r=0}^{R}\sigma_r^2C_{ii,r}}-1\right)=\mathrm{Corr}^2(W_i,W_j).$$

Hence

$$E(GG')=\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{C_{ii,k1}C_{jj,k2}}{(\sum_{r=0}^{R}\sigma_r^2C_{ii,r})^2(\sum_{r=0}^{R}\sigma_r^2C_{jj,r})^2}\left(\sum_{r=0}^{R}\sigma_r^2C_{ij,r}\right)^2.$$

Let a matrix J of dimension $(R+1)$ by $(R+1)$ be defined by the entries

$$J_{k1,k2}=\lim_{N\to\infty}\frac{1}{n_{k1}n_{k2}}\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{C_{ii,k1}C_{jj,k2}}{(\sum_{r=0}^{R}\sigma_r^2C_{ii,r})^2(\sum_{r=0}^{R}\sigma_r^2C_{jj,r})^2}\left(\sum_{r=0}^{R}\sigma_r^2C_{ij,r}\right)^2,$$

where $k1,k2=0,1,2,\ldots,R$.

Additional assumptions:

A5: Assume that matrix J is positive definite.

A6: Assume that $n_k/L\to h_k$ where h_k's are uniformly bounded, for all $k=0,1,\ldots,R$.

Note: Assumption A6 is not necessary if only componentwise asymptotic normality is needed.

Theorem 2. *Let* $\hat{\theta}$ *denote the consistent estimator obtained by solving the composite score functions. Under the assumptions* A0–A6, *as* L *converges to infinity*,

the distribution of $\sqrt{L}(\hat{\theta}-\theta)$ *converges to a Normal distribution with mean zero and covariance matrix* $V(\theta)$ *where* $V(\theta)=I^{-1}JI^{-1}$.

Proof. The main step in proving the asymptotic normality of a consistent estimator is the application of the appropriate central limit theorem for the estimating functions evaluated at the true value of the parameters (See Serfling, 1980 p. 147). Consider the set of composite score functions. These are sums of dependent random variables. However, by assumption A2 the proportion of random variables that are dependent is of order o(L). Thus the central limit theorem for a (correlation) mixing process is applicable. We refer the reader to Heagerty and Lele (1998) for a detailed argument for the asymptotic normality of the maximum composite likelihood estimators in the dependent data situation. A very similar argument can be applied in the above situation. We do not provide the details here.

Assumption A2 is the key in the proofs for consistency and asymptotic normality. We would like to emphasize that both the consistency and asymptotic normality depend only on the correlation decay and not on the independence of the random variables and hence are valid for non-Gaussian random effects as well. Also note that, unlike Miller (1977) and Jiang (1996), we do not have different normalizing constants for each variance component estimator. At first sight, this may seem surprising. However, notice that assumption A6 essentially says that number of contrasts for each of the variance components is approaching infinity at the same rate. Without this assumption, one can prove componentwise asymptotic normality with normalizing constants that are different for each variance components estimator. For joint convergence A6 is a sufficient but probably not a necessary condition.

3.2.4. Finding the consistent solution

The estimating functions corresponding to the composite likelihood do not necessarily have a unique solution. Hence the real issue in practice is to identify which solution is *the* consistent solution.

Towards this end, consider the following modified set of estimating functions:

$$\sum_{i=1}^{L} B_k^{\langle i\rangle}\left(\frac{W_i^2}{\theta^{\mathrm{T}}B^{\langle i\rangle}}-1\right)=0 \quad \text{for } k=0,1,2,\ldots,R.$$

Theorem 3. *Under assumptions* A0–A4, *the above system of equations has a unique consistent solution.*

Proof. First we show the uniqueness of the solution. Consider the Hessian matrix (or equivalently the matrix of the first derivatives of the estimating functions). A typical entry in this matrix is given by

$$\frac{\mathrm{d}}{\mathrm{d}\sigma_m^2}\sum_{i=1}^{L} C_{ii,k}\left(\frac{W_i^2}{\sum_{r=0}^{R}\sigma_r^2 C_{ii,r}}-1\right)=\sum_{i=1}^{L} C_{ii,k}C_{ii,m}\left(\frac{-W_i^2}{(\sum_{r=0}^{R}\sigma_r^2 C_{ii,r})^2}\right).$$

Let $D = \mathrm{diag}(\frac{W_i^2}{(\sum_{r=0}^{R} \sigma_r^2 C_{ii,r})^2},\ i = 1, 2, \ldots, L)$. Then the Hessian matrix can be written as $H = -BDB^{\mathrm{T}}$. To prove that this is a negative definite matrix, recall that $\mathrm{rank}(-H) = \mathrm{rank}(BD^{1/2})$. Moreover $\mathrm{rank}(BD^{1/2}) \geqslant \mathrm{rank}(B) + \mathrm{rank}(D^{1/2}) - L = R + 1$ and $\mathrm{rank}(BD^{1/2}) \leqslant \min(\mathrm{rank}(B), \mathrm{rank}(D^{1/2}) = R + 1$ (Barnett, 1990, p. 94, Eq. (5.11)). Hence $\mathrm{rank}(-H) = R + 1$. The assumption that B is a positive definite matrix, D is a diagonal matrix with positive elements along with this rank result proves that H is negative definite. Hence the solution to this set of estimating functions is unique. The consistency of the solution follows along the same lines as Theorem 1 and we do not repeat the arguments.

Now we use the standard method of one-step M-estimators as described in Lehmann (1985) or Serfling (1980) to obtain estimators that are equivalent to the MCL estimators.

Theorem 4. *Let $\tilde{\theta}$ denote the solution of the system of Eq.* (2). *Let*

$$\hat{\theta} = \tilde{\theta} + I^{-1}(\tilde{\theta}) G(\tilde{\theta}).$$

Then $\sqrt{L}(\hat{\theta} - \theta) \xrightarrow{L} N(0, V(\theta))$ *as* $L \to \infty$.

Proof. Follows from Theorems 1–3 above and Theorem 3.1 (Corollary 3.1) of Lehmann (1985, pp. 422–23) or Serfling (1980, pp. 258–259). This is a one-step estimator with a consistent estimator as the starting value. Also note that this only involves inversion of a $(R + 1)$ by $(R + 1)$ matrix.

It should be noted that the composite score functions are valid estimating equations, in the sense that they are zero unbiased (Godambe and Kale, 1991), irrespective of whether the random effects are normally distributed or not. The consistency and the asymptotic normality of the resultant estimators also hold as long as the linear model is correct whether the random effects are normally distributed or not. However, we shall see in the discussion of our simulations that for small or medium sample sizes the asymptotic normality of MCL (as well as REML) may not be a good approximation.

3.2.5. Optimal estimating functions and MCL estimators

Let us now turn to efficiency issues. Consider a set of L estimating functions

$$g_i(\theta) = \frac{W_i^2}{\sum_{r=0}^{R} \sigma_r^2 C_{ii,r}} - 1 \quad \text{where } i = 1, 2, \ldots, L.$$

There are L equations in $R + 1$ unknowns. Let us denote the vector of these equations by $g(\theta) = (g_i(\theta),\ i = 1, 2, \ldots, L)$. One can combine these L equations to obtain $R + 1$ equations in $R + 1$ unknowns. The general formula for such an optimal linear combination is given in Lindsay (1988). Let us denote the optimal linear combination of the L estimating functions by $g^*(\theta) = \omega^* \cdot g$ where ω^* is a $R + 1$ by L matrix. The set of

optimal weights ω^* for the above set of estimating functions is given by

$$\omega^* = \begin{bmatrix} 1 & \rho_{12}^2 & \dots & \rho_{1L}^2 \\ \rho_{21}^2 & 1 & \dots & \rho_{2L}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{L1}^2 & \dots & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{C_{11,0}}{\sum_{r=0}^{R} \sigma_r^2 C_{11,r}} & \cdots\cdots & \dfrac{C_{11,R}}{\sum_{r=0}^{R} \sigma_r^2 C_{11,r}} \\ \vdots & \vdots\ \vdots & \vdots \\ \vdots & \vdots\ \vdots & \vdots \\ \dfrac{C_{LL,0}}{\sum_{r=0}^{R} \sigma_r^2 C_{LL,r}} & \cdots\cdots & \dfrac{C_{LL,R}}{\sum_{r=0}^{R} \sigma_r^2 C_{LL,r}} \end{bmatrix}.$$

Where $\rho_{ij} = \mathrm{corr}(W_i, W_j)$.

It is easy to see that the composite score functions correspond to ignoring the correlation terms in the first matrix similar to the working independence model for the generalized estimating equations (GEE) discussed by Liang and Zeger (1986). However the key difference between the GEE and the composite likelihood method proposed above is that GEEs utilize only the univariate marginal distributions whereas we use all bivariate marginal distributions. Utilization of the pairwise correlation information in the bivariate distributions seems to lead to high efficiency for the MCL estimators (see Section 5). This result parallels the result that GEE2 (Liang et al., 1992), in which bivariate marginals are used, improves efficiency over GEE1 in which only the univariate marginal distributions are used. The idea of using a wrong correlation matrix, to avoid matrix inversion, was also proposed by Marshall and Mardia (1985) in the spatial statistics context.

4. MCL estimators for general linear mixed models

We now show how the above framework can be used for more complicated situations than the simple random effects models.

4.1. Handling relationships

In the above development it was assumed that the realizations of any random effect are independent of each other, that is, $\mathrm{var}(u_r) = \sigma_r^2 I$. However, in many practical situations, this may not hold. For example, in a typical animal breeding experiment sires are related to each other through common parentage. A matrix that quantifies such genetic relationships is called the relationship matrix. Let us denote this matrix by Ψ. This matrix is typically known a priori or can be calculated in $O(N^2)$ operations (Henderson, 1976). We now describe how the formulation of the MCL for the estimation of variance components from simple random effects model can be generalized to such a situation. We provide some of the details here.

When the random effects are related to each other, notice that $\mathrm{var}(u_r) = \sigma_r^2 \Psi_r$. It is easy to see that in such a situation $Y \sim N(\mu \underline{1}, \sum_{r=0}^{R} \sigma_r^2 Z_r \Psi_r Z_r^{\mathrm{T}})$ and $W \sim N(\underline{0}, \sum_{r=0}^{R} \sigma_r^2 A Z_r \Psi_r Z_r^{\mathrm{T}} A^{\mathrm{T}})$. Now one can replace C_r's by $C_r^{\Psi} = A Z_r \Psi_r Z_r^{\mathrm{T}} A^{\mathrm{T}}$ in all the

formulae in Section 3 and modify the conditions on C_r's by similar conditions on C_r^{Ψ}'s. The uniqueness, consistency and asymptotic normality follow in a similar fashion.

It is also possible that two or more random effects may be related to each other. This case can be handled similarly by noting the fact that in this case: $Y \sim N(\mu\underline{1}, \sum_{r=0}^{R} \sigma_r^2 Z_r \Psi_r Z_r^{\mathrm{T}} + \sum_{r_1=0}^{R} \sum_{r_2 \neq r_1}^{R} \tau_{r_1 r_2} Z_{r_1} \Psi_{r_1 r_2} Z_{r_2}^{\mathrm{T}})$ where τ_{r_1,r_2} is the covariance between the random effects r_1 and r_2. Let us consider some of the details in this situation.

Let $C_r^{\Psi} = AZ_r \Psi_r Z_r^{\mathrm{T}} A^{\mathrm{T}}$ and $C_{r_1,r_2}^{\Psi} = AZ_{r_1} \Psi_{r_1,r_2} Z_{r_2}^{\mathrm{T}} A^{\mathrm{T}}$. Let us also modify the parameter vector to include the covariance parameters:

$\theta = (\sigma_r^2,\ r = 0, 1, 2, \ldots, R; \tau_{r_1,r_2},\ r_1 \neq r_2 = 0, 1, 2, \ldots, R)^{\mathrm{T}}$. The B matrix now has dimension $(R+2)(R+1)/2$ by L with ith column containing the ith diagonal elements of the above C matrices. With this notation, we can now write the marginal distribution of $W_i \sim N(0, \theta^{\mathrm{T}} B^{\langle i \rangle})$ and the composite score functions have the same form as equation 2. This is of the same form as in the previous situation with more C_r^{Ψ}'s to evaluate. The rest of the theoretical results follow easily.

4.2. Multiple traits models

In many biological situations, scientists measure several different traits on the same individual. In such situations we are interested in estimating the covariance as well as variance components. We now generalize the method of MCL to such situations. To avoid notational complexity, we consider two traits situation. Also for the sake of notational simplicity, we consider only the random effects case. We indicate how this formulation may be extended to more than two traits. Extensions to the mixed effects case and to random effects with relationships are straightforward.

We follow the notation of Searle et al. (1992, p. 382). The vector of observations corresponding to the first trait is denoted by Y_1 and marginally has distribution $Y_1 \sim N(\mu_1 \underline{1}, \sum_{r=0}^{R} \sigma_{r,1}^2 Z_{r,1} Z_{r,1}^{\mathrm{T}})$ and similarly the vector of observations corresponding to the second trait is denoted by Y_2 and has distribution $Y_2 \sim N(\mu_2 \underline{1}, \sum_{r=0}^{R} \sigma_{r,2}^2 Z_{r,2} Z_{r,2}^{\mathrm{T}})$.

To begin with we assume that none of the traits are missing. Note that in general, $Z_{r,1}$ and $Z_{r,2}$ will be identical to each other; however, if there are missing traits they may be different. Let $W_1 = AY_1$ and $W_2 = AY_2$ be the contrast vectors as defined in the single trait case. Notice that

$$\mathrm{var}(W_1) = \sum_{r=0}^{R} \sigma_{r,1}^2 AZ_{r,1} Z_{r,1}^{\mathrm{T}} A^{\mathrm{T}}, \qquad \mathrm{var}(W_2) = \sum_{r=0}^{R} \sigma_{r,2}^2 AZ_{r,2} Z_{r,2}^{\mathrm{T}} A^{\mathrm{T}}$$

and $\mathrm{cov}(W_1, W_2) = \sum_{r=0}^{R} \xi_{r,12} AZ_{r,1} Z_{r,2}^{\mathrm{T}} A^{\mathrm{T}}$ where $\xi_{r,12}$'s are the covariance components. Define a new vector W_3 such that $W_3 = W_1 - W_2$. It follows that

$$\mathrm{var}(W_3) = \sum_{r=0}^{R} \sigma_{r,1}^2 (AZ_{r,1} Z_{r,1}^{\mathrm{T}} A^{\mathrm{T}}) + \sum_{r=0}^{R} \sigma_{r,2}^2 (AZ_{r,2} Z_{r,2}^{\mathrm{T}} A^{\mathrm{T}}) - 2 \sum_{r=0}^{R} \xi_{r,12} (AZ_{r,1} Z_{r,1}^{\mathrm{T}} A^{\mathrm{T}}).$$

Now combine these three vectors to construct a vector $W = (W_1, W_2, W_3)^{\mathrm{T}}$. The composite likelihood in the multiple traits case can be constructed by taking the product of the marginal distributions of the components of W.

Let $C_{r,1} = AZ_{r,1}Z_{r,1}^{\mathrm{T}}A^{\mathrm{T}}$, $C_{r,2} = AZ_{r,2}Z_{r,2}^{\mathrm{T}}A^{\mathrm{T}}$ and $C_{r,12} = AZ_{r,1}Z_{r,2}^{\mathrm{T}}A^{\mathrm{T}}$. Define the new parameter vector $\theta = (\sigma_{r,i}^2,\ r = 0,1,2,\ldots,R,\ i = 1,2; \xi_{r,12},\ r = 0,1,2,\ldots,R)^{\mathrm{T}}$ of length $3R$ and the new

$$B = \begin{bmatrix} B_1 & 0 & B_1 \\ 0 & B_2 & B_2 \\ 0 & 0 & -2B_{12} \end{bmatrix},$$

where matrices B_1, B_2, B_{12} are defined by the diagonal elements of $C_{r,1}, C_{r,2}, C_{r,12}$, respectively. The dimension of the matrix B is $3R \times 3L$. With this notation, it is easy to see that the marginal distribution of $W_i \sim N(0, \theta^{\mathrm{T}}B^{\langle i \rangle})$ and the composite score functions have the same form as in Eq. (2). The theoretical results follow similarly.

The extension to more than two traits is straightforward. For example, for the three traits situation, we construct $W_i = AY_i$, $i = 1,2,3$ and $W_i - W_j$, $i < j = 2,3$ and stack them together to get W. The parameter vector and B matrix is changed accordingly. These will be of length $6R$ and dimension $6R \times 6L$, respectively.

4.2.1. Missing traits

Missing traits, that is, observations for which not all traits are available, is a substantial problem for many types of data. If the researcher believes that the traits are effectively missing at random, then there is a straightforward mechanism for calculating the MCL estimates in the presence of missing traits. We fill missing values with any missing traits code. The columns of the B matrix, corresponding to contrasts involving missing traits, are filled with zeros. As we can see from Eq. (2), any contrast involving a missing trait is thereby given a zero weight in the estimating functions. With this minor change, the rest of the estimation procedure as well as the conditions for consistency remain the same. Notice if only one trait is observed on a particular observation, it is utilized in the estimation of the variance components corresponding to that trait. This is clearly superior to the common practice of deleting all observations with any missing traits (e.g. Wilkinson et al., 1990). The relative statistical and computational efficiency of this approach to handling missing traits and the "expectation maximization" approach (Dempster et al., 1977) has yet to be investigated. However, particularly for large problems, computationally our approach seems substantially simpler than the EM algorithm.

4.3. Mixed effects models

Now consider the mixed effects models where $Y \sim N(X\beta, \sum_{r=0}^{R} \sigma_r^2 Z_r \Psi_r Z_r^{\mathrm{T}})$.

Approach 1: One can estimate the fixed effects β using the least squares approach and behave as if $Y^* = Y - \widehat{X\beta} \sim N(0, \sum_{r=0}^{R} \sigma_r^2 Z_r \Psi_r Z_r^{\mathrm{T}})$. One can then follow the MCL approach as described above behaving as if Y^* are observations obtained from a random effects model. It is clear that the estimating functions corresponding to the MCL approach in this situation will not be zero unbiased. However, the bias in the

estimating functions approaches zero as the sample size approaches infinity and the asymptotic properties remain unaffected.

Approach 2: This approach parallels that of REML where we eliminate the fixed effects by using invariance. The steps are:

Let $P_X = I - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ be the projection matrix. Consider $P_X Y \sim N(0, \sum_{r=0}^{R} \sigma_r^2 P_X Z_r \Psi_r Z_r^{\mathrm{T}} P_X^{\mathrm{T}})$. These are the observations that we will base the MCL estimators on. Now one can replace C_r's by $C_r^I = AP_X Z_r \Psi_r Z_r^{\mathrm{T}} P_X^{\mathrm{T}} A^{\mathrm{T}}$ (I for invariance) in all the above development and modify the conditions on C_r's by corresponding conditions on C_r^I's. The uniqueness, consistency and asymptotic normality follow in a similar fashion. Although, the Y^* in the first approach is identical to $P_X Y$, the distribution used in the first approach is approximate whereas the distribution used in the second approach is exact. What approach 2 acknowledges is that the transformation influences the variance of the Y^*.

The above discussion clearly illustrates the general applicability of the method of maximum composite likelihood to the problem of variance and covariance components estimation.

5. Efficiency comparisons (unbalanced designs and non-normal effects)

In deciding whether to use an alternative method to REML for the estimation of variance components, aside from computational ease, one needs to consider the relative statistical efficiencies. In this simulation study, we investigate the effects of three factors on statistical efficiency: (1) Imbalance in the data, (2) Sample size, and (3) Distributional assumptions.

(1) *Imbalance and sample size*: For simple models, it can be shown that for balanced designs MCL and REML are equivalent. This appears to be true in more complicated situations as well. However, real experiments, despite the best intentions of empiricists, rarely generate balanced designs. The degree of imbalance in designs is difficult to characterize. Following Swallow and Monahan (1974), we investigate the impact of mild and severe imbalance in small and large data sets for one-way random effects designs. We consider the model $Y_{ij} = \mu + u_i + e_{ij}$ where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, n_i$. Data sets with mild imbalance have effects with 3, 5, or 7 observations each, while data sets with severe imbalance have effects with 1, 5, or 9 observations each. For small sample size we replicate these basic patterns 6 times and for large sample size we replicate 67 times. Thus, for small sample size and mild imbalance, we have 6 effects with 3 observations, 6 effects with 5 observations, and 6 effects with 7 observations for a total of 18 effects and 90 observations. For large sample size and severe imbalance, we have 67 effects with 1 observation, 67 effects with 5 observations, and 67 effects with 9 observations for a total of 201 effects and 1005 observations.

In the notation of Eq. (1) one-way designs have two variance components, a residual or environmental variance (σ_0^2), and an effect variance (σ_1^2). In the simulations, the

Table 1
Mean absolute deviation ratios

Effect distribution	Imbalance	Data size	Effect variance					
			0.1	0.2	0.5	1.0	2.0	5.0
Normal	Mild	S	0.95	0.96	0.98	0.98	0.97	0.96
		L	0.92	0.94	0.91	0.99	0.98	0.96
	Severe	S	0.95	1.00	1.00	0.94	0.95	0.89
		L	0.89	0.96	0.96	0.96	0.93	0.88
Skewed	Mild	S	0.97	1.19	1.00	1.04	1.01	1.06
		L	0.97	0.99	1.00	0.99	0.97	0.97
	Severe	S	0.95	0.96	1.02	1.03	0.96	0.95
		L	0.99	0.97	0.97	0.91	0.89	0.85
Kurtotic	Mild	S	0.96	0.98	0.99	0.98	0.97	0.94
		L	0.94	1.00	0.97	0.96	0.95	1.00
	Severe	S	0.96	0.99	0.98	0.94	0.92	0.88
		L	0.95	1.00	0.97	0.93	0.91	1.00

environmental variance (σ_0^2) is held constant at a value of 1. The effect variance (σ_1^2) ranges over the set $\{0.1, 0.2, 0.5, 1, 2, 5\}$.

(2) *Distributional assumptions*: We also investigated the relative impact of deviations from normality in the distributions of the random effects by drawing random effects from three kinds of distributions: (1) a normal distribution, (2) a symmetric but highly peaked distribution (kurtosis = 7), and (3) a highly skewed distribution (skewness = 7). All distributions have mean zero and a specified variance. The distribution for the environmental errors is always taken to be normal.

For every combination of control factors, 1000 data sets were constructed and parameters estimated using both REML (SAS 6.12 proc VARCOMP) and our own MCL program. The measure of relative efficiency we report is the ratio of the mean absolute deviations of estimates from truth (MAD) for REML estimates to the MAD for the MCL estimates. A MAD ratio less than 1 indicates that REML is more efficient than MCL, while a MAD ratio greater than 1 indicates that MCL is more efficient than REML. Table 1 gives the MAD ratios estimated from simulations for all 72 combinations of our control factors.

We use the MAD ratio rather than the MSE ratio as our comparative measure because for some combinations of our control factors the distributions of variance component estimates are substantially skewed (Table 2) for both MCL and REML estimation methods. Under these conditions the estimates of the MSEs are quite unstable, with the deletion of a single estimate out of 1000 determining whether the MSE ratio substantially favors the MCL method or the REML method. For comparative purposes, negative MCL variance component estimates were truncated to zero before the MAD was calculated. However, very few estimates were negative even with our smallest data size and smallest true variance.

Table 2
The estimated skewness of the distribution of the estimator of σ_1^2

Effect distribution	Imbalance	Data size	Effect variance					
			0.1	0.2	0.5	1.0	2.0	5.0
Normal	Mild	S	0.44	0.54	0.55	0.77	0.80	0.90
		L	0.19	0.22	0.32	0.21	0.03	0.18
	Severe	S	0.41	0.66	0.72	0.93	0.83	0.82
		L	0.12	0.22	0.22	0.25	0.15	0.23
Skewed	Mild	S	7.50	29.98	14.62	15.29	5.11	22.74
		L	5.88	5.55	3.77	3.87	3.92	5.36
	Severe	S	8.92	20.56	13.58	13.19	7.05	5.66
		L	4.26	10.33	5.14	4.45	6.98	19.67
Kurtotic	Mild	S	0.75	0.88	1.12	1.00	0.97	0.96
		L	0.22	0.43	0.27	0.23	0.30	0.32
	Severe	S	0.80	0.99	0.95	1.30	1.30	1.19
		L	0.20	0.18	0.32	0.35	0.27	0.49

The MAD ratios in Table 1 indicate that while in general the REML method is more efficient, the difference between the two methods is quite insignificant. Another way to see this is to calculate the correlation between the MCL estimates and the REML estimates for each group of 1000 data sets. Out of all 72 correlations, the minimum was 0.87 and the average was 0.95. Put quite simply, the choice of estimation method (MCL or REML) makes very little contribution to the uncertainty of the estimate. The great bulk of the uncertainty in an estimate is due to the random processes involved in the realization of the data.

Neither method produces estimates under any combination of control factors with biases that are detectable with 1000 simulations given estimate variability. The distribution of variance component estimates for normal and kurtotic effect distributions, is indistinguishable from normal. On the other hand, when effect distributions are skewed, distributions were strongly positively skewed and strongly positively kurtotic (peaky). There was, however, no discernable difference between the distributions of the estimators obtained from MCL and REML, both were either close to normal or different from normal in a similar fashion.

Probably the most important advantage of the method of composite likelihood method for variance components estimation is computational. Theoretically it is clear that instead of inverting an $N \times N$ matrix that requires $O(N^3)$ operations (unless one exploits a particular structure of the covariance matrix), we have replaced it by the computational burden of calculating $N(N-1)/2$ contrasts. This has reduced the computational problem to $O(N^2)$, which may represent a gain of many orders of magnitude if N is large. The actual time scaling of working variance component programs validate the theoretical calculations. We plotted the change in the computational time using the MCL approach using our program and for REML using the SAS-VARCOMP program as a function of the number of observation. The increase in the computational time was quadratic for the MCL whereas it was cubic for the REML.

6. Discussion

In this paper we have discussed a new approach to the estimation of covariance components based on the concept of composite likelihood. We have shown that this approach is as generally applicable as the method of maximum likelihood and that the estimates are consistent and asymptotically normal. However, in contrast to the ML or REML approach, MCL approach is computationally substantially less intensive. It avoids the inversion of large matrices and thereby reduces the computational burden from $O(N^3)$ operations to $O(N^2)$ operations. This computational gain is achieved without substantial loss in the statistical efficiency as shown by our simulation study. Another important point to note that this computational advantage is obtained across the border for all types of designs, relationship matrices and multiple traits situations. In some special structured design matrices, computational burden for ML or REML method can be reduced substantially. However, although one may design the study that leads to special structured matrices, such a structure may be destroyed if there are missing observations, making special methods for matrix inversion inapplicable. This is not an infrequent situation in practice. In contrast, the method of MCL needs no such special structures for the design matrices. In this paper, we have shown that for the designs considered in the simulation study reported, the computational time for ML using SAS programs increases as a cubic function of the number of observations whereas the computational time for MCL increases as a quadratic function of the number of observations. Although, as pointed out by the referee, more comprehensive simulation studies will be useful to substantiate this claim. Further, the MCL estimate is unique, something that cannot be claimed by ML, REML, ANOVA, or MINQUE estimates. A final advantage to the MCL method is that a solution is guaranteed. There are no convergence problems. This property becomes important if inference is to be undertaken using bootstrap methods.

This paper does not develop procedures for estimation of fixed effects or prediction of random effects using the composite likelihood concept. These topics will be discussed elsewhere. Similarly methods for obtaining confidence intervals for the variance components and testing for the equality of covariance matrices will also be discussed in detail elsewhere.

Acknowledgements

This research was partially supported by a National Science Foundation grant (DEB-94-11770) to Gerald Wilkinson and Mark Taper, by Department of Energy cooperative agreement (DE-FC07-94ID-13317) to Daniel Goodman and National Sciences and Engineering Research Council of Canada to Subhash Lele. The authors would like to thank Derek Sonderegger for his assistance in the computer implementation of the method. The authors are grateful to Charles E. McCulloch, George Casella and Ranjini Natarajan for their encouragement and insightful comments.

Appendix A

A.1. Consistency of MCL estimators

(1) *Zero unbiasedness*: Let

$$G1(W,\theta)=\left\{\frac{1}{\theta^{\mathrm{T}}B^{\langle i\rangle}}\left(\frac{W_i^2}{\theta^{\mathrm{T}}B^{\langle i\rangle}}-1\right)\right\}_{i=1,2,\ldots,L}$$

be a vector of estimating function. Notice that each component of this estimating function has mean zero.

Let $G(W,\theta)=B^{*}G1(W,\theta)$. It is clear that these are the composite score functions in Eq. (1). These are also zero unbiased.

To show consistency, we need to show that within an $\underline{\varepsilon}$ neighborhood of the true parameter vector $\underline{\theta_0}$, with probability converging to 1 there exists a solution to the above estimating functions.

Towards that end, define a matrix

$$F_N=\begin{bmatrix}1/n_1 & 0 & \cdots & 0\\ 0 & 1/n_2 & \cdots & \vdots\\ \vdots & 0 & \cdots & \vdots\\ \vdots & \vdots & \cdots & 1/n_1\end{bmatrix}.$$

(2) We will argue that $F_NG(W,\theta)\xrightarrow{\mathrm{p}}0$ as $N\to\infty$ under A0–A4.

Clearly $E(F_NG(W,\theta))=0$. Moreover $\mathrm{Var}(F_NG(W,\theta))=F_NB\,\mathrm{Var}(G1(W,\theta))B^{\mathrm{T}}F_N\to 0$. This follows from the observation that the diagonal elements of the above covariance matrix:

$$\frac{1}{n_k^2}\left\{\sum_{i=1}^{L}\frac{2(B_k^{\langle i\rangle})^2}{(\theta^{\mathrm{T}}B^{\langle i\rangle})^2}+\sum_{i=1}^{L}\sum_{j\neq i}^{L}r_{ij}^2\frac{(B_k^{\langle i\rangle}B_k^{\langle j\rangle})}{(\theta^{\mathrm{T}}B^{\langle i\rangle})(\theta^{\mathrm{T}}B^{\langle j\rangle})}\right\}\to 0$$

as $N\to\infty$ under A0–A4.

The off-diagonal elements converge to zero by Cauchy–Schwartz inequality. The convergence in probability follows by Chebychev inequality.

(3) Now apply the Taylor series expansion to get

$$F_N(G(W,\theta+\underline{\varepsilon})-G(W,\theta))=F_NG'(W,\theta^{*})\underline{\varepsilon}$$

and $F_N(G(W,\theta-\underline{\varepsilon})-G(W,\theta))=-F_NG'(W,\theta^{*})\underline{\varepsilon}$.

If $-F_NG'(W,\theta^{*})\xrightarrow{\mathrm{p}}I$, a positive definite matrix as $N\to\infty$, then $F_N(G(W,\theta+\underline{\varepsilon})$ and $F_N(G(W,\theta-\underline{\varepsilon})$ are reflections of each other. By continuity of $G(W,\theta)$, it follows that $G(W,\theta)$ has to cross zero in the $\underline{\varepsilon}$ neighborhood of the true parameter vector $\underline{\theta_0}$, with probability converging to 1, which in turn implies that $P\{||\underline{\theta_N}-\theta||>\varepsilon\}\to 0$ as $N\to\infty$ where $\underline{\theta_N}$ is a solution of $G(W,\theta)$. The convergence $-F_NG'(W,\theta^{*})\xrightarrow{\mathrm{p}}I$

follows from application of Chebychev inequality and assumptions A0–A4 in a manner similar to Step 2.

References

Barnett, S., 1990. Matrices: Methods and Applications. Oxford University Press, NY.
Besag, J.E., 1975. Statistical analysis of non-lattice data. The Statistician 24, 179–195.
Curriero, F., Lele, S., 1999. A composite likelihood approach to semivariogram estimation. J. Agr. Biol. Environ. Statist. 4 (1), 9–28.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1–38.
Gianola, D., Hammond, K., 1990. Advances in Statistical Methods for Genetic Improvement of Livestock. Springer, New York.
Godambe, V.P., Kale, B.K., 1991. In: Godambe, V.P. (Ed.), Estimating Functions: an Overview Estimating Functions. Oxford University Press, Oxford, pp. 1–20.
Hartley, H.O., Rao, J.N.K., 1967. Maximum likelihood estimation for the mixed analysis of variance models. Biometrika 54, 93–108.
Heagerty, P., Lele, S.R., 1998. A Composite Likelihood approach to binary spatial data. J. Amer. Statist. Assoc. 93, 1099–1111.
Henderson, C.R., 1953. Estimation of variance and covariance components. Biometrics 9, 226–252.
Henderson, C.R., 1976. A simple method for computing the inverse of a numerator matrix and in predicting breeding values. Biometrics 32, 69–84.
Jiang, J., 1996. REML estimation: asymptotic behaviour and related topics. Ann. Statist 24, 255–286.
Kohn, L.A.P., Atcheley, W.R., 1988. How similar are genetic correlation structures? Data from mice and rat. Evolution 42, 467–481.
Lande, R., 1979. Quantitative genetics analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33, 402–416.
Lehmann, E.L., 1985. Theory of Point Estimation. Wiley, NY.
Lele, S., 1997. Estimating functions for semivariogram estimation. In: Basawa, I.V., Godambe, V.P., Taylor, R.L. (Eds.), Selected Proc. of the Symp. on Estimating Functions, IMS Lecture Notes, Monograph Series Vol., Springer, Berlin, pp. 360–375.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 17, 13–22.
Liang, K.Y., Zeger, S.L., Qaqish, B., 1992. Longitudinal data analysis using generalized linear models. J. Roy. Statist. Soc. Ser. B. 54 (1), 3–40.
Lindsay, B.G., 1988. Composite likelihood methods. Contemporary Math. 80, 221–239.
Marshall, R.J., Mardia, K.V., 1985. Minimum norm quadratic estimation of components of spatial covariance. Math. Geol. 17 (5), 517–525.
Miller, J.J., 1977. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. Ann. Statist. 5, 746–762.
Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.
Pukelsheim, F., 1976. Estimating variance components in linear models. J. Multivariate Analysis 6, 626–629.
Rao, C.R., 1971. Estimation of variance and covariance components-MINQUE theory. J. Multivariate Anal. 1, 257–275.
Rao, C.R., Kleffe, J., 1988. Estimation of Variance Components and Applications. North-Holland, Amsterdam.
SAS-STAT User's Guide release 6.12 (1996) SAS Inst., Inc. Carey, NC, USA.
Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York.
Searle, S.R., McCulloch, C.E., 1996. Linear models 1945–1995. In: Armitage, P., David, H.A. (Eds.), Advances in Biometry. Wiley, New York, pp. 65–88.
Seely, J., 1970. Linear spaces and unbiased estimation-application to mixed model. Ann. Math. Stat. 41, 1735–1748.
Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley, NY.

Shaw, R.G., 1987. Maximum likelihood approaches to quantitative genetics of natural populations. Evolution 41, 812–826.

Shaw, R.G., 1991. The comparison of quantitative genetic parameters between populations. Evolution 45, 143–151.

Swallow, W.H., Monahan, J.F., 1984. Monte–carlo comparison of ANOVA, MINQUE, REML and ML estimators of variance components. Technometrics 26, 47–57.

Wilkinson, G.S., Fowler, K., Partridge, L., 1990. Resistance of genetic correlation structure to directional selection in Drosophila melanogaster. Evolution 44 (8), 1990–2003.