

MARGINAL LIKELIHOOD METHODS FOR
ESTIMATING VARIANCE PARAMETERS

O.B. Allen
University of Guelph and Cornell University

Statistical Series 1986-183

BU-904-M(Revised)

July 1986

MARGINAL LIKELIHOOD METHODS FOR ESTIMATING
VARIANCE PARAMETERS

O.B. Allen
University of Guelph and Cornell University

ABSTRACT

This paper investigates the use of marginal likelihood to estimate variance parameters. Marginal likelihood methods have been proposed as an alternative to full maximum likelihood when one wishes to estimate a parameter in the presence of other nuisance parameters which are of no interest. Marginal likelihood methods are appropriate in such a setting if:

- (i) there exists a transformation $Y_1, \dots, Y_n \rightarrow A_1, \dots, A_r, U_1, \dots, U_{n-r}$ such that the marginal distribution of A_1, \dots, A_r does not depend on the nuisance parameter β and
- (ii) the conditional distribution of U_1, \dots, U_{n-r} given A_1, \dots, A_r contains no available information about θ in the absence of knowledge of β .

We consider a response vector Y , having a normal distribution with $E(Y) = X\beta$ and $\text{Var}(Y) = \sigma^2 V$ where $V = V(\theta)$ is a function of unknown parameters, θ . We wish to estimate θ ; β and σ^2 are regarded as nuisance parameters. The marginal likelihood chosen is based on the distribution of the standardized residuals, $(Y-Xb)/s$, where $b = (X'X)^{-1}X'Y$ and $s^2 = (Y-Xb)'(Y-Xb)$. This likelihood, proposed by Fraser (1968), can be shown to be marginally sufficient as defined by Sprott (1975). The use of this likelihood is examined for the estimation of variance components.

1. INTRODUCTION

Partial likelihood methods of inference have been proposed as a means of estimating a parameter in the presence of other nuisance parameters which are of little interest. The use of partial likelihood methods goes back to at least to Bartlett (1937). There are two reasons why one may not wish to use the more usual analysis based on the full likelihood. With a large number of nuisance parameters, the numerical problems associated with maximization may be formidable. Furthermore, graphical examination of the likelihood surface in more than two dimensions is difficult. Even more importantly, however, the usual M.L. estimator of the parameters of interest may turn out to be inconsistent if the number of nuisance parameters is not fixed (Neyman and Scott, 1948). This is the situation, for example in estimating σ^2 in the usual one-way ANOVA if one fixes the number of responses per treatment, r , but lets the number of treatments, t , increase. Then the maximum likelihood estimate of σ^2 is (residual sum of squares)/ rt and the expectation of this is $(r-1)\sigma^2/r$. The bias clearly does not go to zero as $t \rightarrow \infty$. This suggests that even when the number of nuisance parameters is fixed, but large, the usual M.L. estimator may behave poorly. The marginal likelihood procedure partitions the likelihood into the product of two components - one of which is not a function of any nuisance parameters but which contains all or most of the information about the parameters of interest in the absence of knowledge about the nuisance parameters. This component of the likelihood is then used as the basis of inferences about the parameters of interest.

More formally, following Kalbfleisch and Sprott (1970), consider a transformation of the response Y_1, \dots, Y_n to $U_1, \dots, U_{n-r}, A_1, \dots, A_r$. Then the full likelihood can be written as the product of the marginal density of

A_1, \dots, A_r and the density of U_1, \dots, U_{n-r} conditional on A_1, \dots, A_r . To make inferences about θ attention can be restricted to the marginal density of A_1, \dots, A_r if (i) this density does not depend on β and (ii) the conditional density of U_1, \dots, U_{n-r} contains no available information about β in the absence of knowledge of β (even though it may depend on θ).

We shall assume throughout that Y follows a multivariate normal distribution with mean and variance given by:

$$\begin{aligned} E(Y) &= X\beta \\ \text{Var}(Y) &= \sigma^2 V. \end{aligned} \tag{1}$$

In order to avoid nonessential details, we will assume that $E(Y)$ is parameterized in such a way that X has full column rank, p . Further $V = V(\theta)$ is assumed to be a function of unknown parameters, θ . These are in fact the parameters we wish to estimate. Finally, β and σ^2 are considered nuisance parameters.

2. ONE-WAY ANOVA

As an example consider the one-way ANOVA discussed earlier under the usual assumptions. The usual unbiased ANOVA estimator of σ^2 maximizes a marginal likelihood based on the error contrasts. Let $Y'_i = (Y_{i1}, \dots, Y_{ir})$ be the vector of observations on treatment i and let A be an $r-1$ by r matrix whose rows are any set of orthonormal contrasts (i.e., $A \cdot 1 = 0$ and $AA' = I$). Then $Z_i = AY_i$ has a $(r-1)$ variate normal distribution with $E(AY_i) = 0$ and $\text{Var}(AY_i) = \sigma^2 I$. Note that this distribution does not depend on $\mu_{ij} = E(Y_{ij})$. The log of the likelihood based on Z_1, \dots, Z_t is proportional to:

$$-t(r-1) \log \sigma - (\sum_i Z'_i Z_i) / 2\sigma^2$$

and this is maximized by:

$$\hat{\sigma}^2 = (\sum_i Z'_i Z_i) / t(r-1) = (\sum Y'_i A' A Y_i) / t(r-1).$$

Now,

$$P = \begin{bmatrix} \frac{1}{\sqrt{r}} & 1 & A' \end{bmatrix}$$

is an orthogonal matrix. Hence:

$$I_r = PP' = \frac{1}{r} 11' + A'A$$

and thus

$$A'A = I - \frac{1}{r} 11'$$

and

$$\hat{\sigma}^2 = \frac{\sum \sum (Y_{ij} - \bar{Y}_{i.})^2}{t(r-1)}.$$

This same reasoning extends easily to justify the usual ANOVA estimator of σ^2 in any fixed effects linear model. Furthermore, the same approach yields REML estimators of variance components (Patterson and Thompson, 1971). In either case if $E(Y) = X\beta$ where X has column rank p , then inferences about variance components are based on the distribution of AY where A is of rank $n-p$ and $E(AY) = 0$.

3. THE MARGINAL DISTRIBUTION OF THE STANDARDIZED RESIDUALS

Our interest is in estimators of variance parameters, θ (excluding σ^2 and β). Fraser (1968) has argued that inferences about θ should be based on the marginal distribution of the normalized residuals, $d = (Y-Xb)/s$ where $b = (X'X)^{-1}X'Y$ and $s^2 = (Y-Xb)'(Y-Xb)$. (Recall that REML is based on the marginal distribution of the residuals, $Y-Xb$.)

We shall derive the marginal distribution of d . Consider the following transformation of Y .

$$\begin{aligned} b &= (X'X)^{-1}X'Y \\ s &= \{(Y-Xb)'(Y-Xb)\}^{1/2} \\ d &= (Y-Xb)/s \end{aligned} \tag{2}$$

This is a transformation from Euclidean n space onto an n dimensional

subspace of Euclidean $n+p+1$ space. The transformation is, however, one to one. The subspace is characterized by $X'd = 0$, $d'd = 1$ and $s > 0$. Geometrically, the first two of these conditions represent the intersection in Euclidean n -space, of the n -sphere centered at the origin and with radius 1 with an $(n-p)$ dimensional plane passing through the origin. This yields an $n-p-1$ dimensional surface on which the density of d sits. The Cartesian product of this surface with the half plane $s > 0$ and with the Euclidean p -space associated with b yields the n -dimensional space on which b , s , and d sit. The inverse transformation is:

$$Y = Xb + sd \quad (3)$$

Saw (1973) provides a technique for computing the Jacobian of the transformation (2). It turns out to be

$$s^{n-p-1} |X'X|^{1/2}$$

Thus the density of b , s , d is:

$$(2\pi)^{-n/2} \sigma^{-n} |V|^{-1/2} \exp\{-(Xb - X\beta + sd)' V^{-1} (Xb - X\beta + sd) / 2\sigma^2\} \cdot s^{n-p-1} |X'X|^{1/2}. \quad (4)$$

Note that if $V = I$ then the exponent reduces to

$$-[(b-\beta)'(X'X)(b-\beta) + s^2] / 2\sigma^2 \quad \text{since } X'd = 0 \quad \text{and} \quad d'd = 1$$

and (4) can be written as the product of the marginal densities of b , s and d . Hence b , s , and d are mutually independent with the distribution of d being constant. This however, no longer holds if $V \neq I$.

To obtain the marginal density of d we shall integrate (4) first with respect to b , then with respect to s . The first integral can be made to be the integral of the multivariate normal density by rewriting the exponent as:

$$-[(b-\beta + s(X'V^{-1}X)^{-1}X'V^{-1}d)'(X'V^{-1}X)(b-\beta + s(X'V^{-1}X)^{-1}X'V^{-1}d) + s^2d'V^{-1}d - s^2d'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}d] / 2\sigma^2.$$

The second integral reduces to $\Gamma(\frac{n-p}{2})$ with the change of variable $z = s^2 d' (V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}) d / 2\sigma^2$. The resulting marginal density of d is

$$(2\pi)^{-(n-p)/2} \Gamma(\frac{n-p}{2}) 2^{(n-p-1)/2} |X'X|^{1/2} |V|^{-1/2} |X'V^{-1}X|^{-1/2} \cdot \{d'V^{-1}d - d'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}d\}^{-(n-p)/2} \quad (5)$$

This agrees with Levenbach (1973) who appeals to Fraser's method of invariant differentials to derive the Jacobson.

We shall also need the conditional density of b, s given d . But this is just the ratio of (4) over (5). Thus, it is

$$f(b, s; \beta, \sigma, \theta | d) = (2\pi)^{-p/2} \sigma^{-n} s^{n-p-1} \Gamma^{-1}((n-p)/2) 2^{-(n-p-1)/2} |X'V^{-1}X|^{1/2} \cdot \{d'V^{-1}d - d'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}d\}^{(n-p)/2} \cdot \exp\{- (Xb - X\beta + sd)' V^{-1} (Xb - X\beta + sd) / 2\sigma^2\} \quad (6)$$

4. MARGINAL SUFFICIENCY

Next, we would like to establish that d is marginally sufficient for θ in the sense defined by Sprott (1975). d is marginally sufficient for θ provided that the conditional distribution of b and s given d (6) contains no information about θ in the absence of any knowledge about β or σ . This is defined to be the case if the likelihood ratio

$$r((\theta, \theta_0) : b, s, \beta, \sigma) = f(b, s; \beta, \sigma, \theta | d) / f(b, s; \beta, \sigma, \theta_0 | d) \\ = \left\{ \frac{|X'V^{-1}X|}{|X'V_0^{-1}X|} \right\}^{\frac{1}{2}} \left\{ \frac{d'V^{-1}d - d'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}d}{d'_0V_0^{-1}d_0 - d'_0V_0^{-1}X(X'_0V_0^{-1}X)^{-1}X'_0V_0^{-1}d_0} \right\}^{-(n-p)/2} \cdot \exp\{- (Xb - X\beta + sd)(V^{-1} - V_0^{-1})(Xb - X\beta + sd) / 2\sigma^2\} \quad (7)$$

can be written as a function of θ, θ_0 and u only (remembering that X and d are fixed quantities here). It is required, further, that the distribution of u (conditional on d) be independent of β and σ - and finally that, fixing b, s and θ , there is at least one value of (b, σ) for each value of u . Choosing $u' = (u_1', u_2')$ where $u_1 = (b - \beta)/\sigma$ and $u_2 = s/\sigma$, it is easily verified from (6) that the conditional density of u will not involve β or σ . Further, (7) can be written as a function of θ, θ_0 and u only. Finally, fixing b and s , there is a one-to-one relationship between u and (β', σ) . Hence, d is marginally sufficient for θ .

5. BLOCKED DESIGNS

Assume we have a block design with b blocks of sizes k_1, \dots, k_b and t treatments with r_1, \dots, r_t replicates. Thus $\sum k_j = \sum r_i$. For notational convenience we shall assume that the responses are ordered by treatments within blocks. For the most part we shall focus on designs where a treatment occurs at most once in any block. However, since it is convenient, we shall initially allow treatment to occur more than once in a block. Subsequently, we will consider as special cases, the balanced incomplete block design and the randomized complete block design. Throughout, we assume the following model where $Y_{ij\ell}$, represents the ℓ th occurrence of the i -th treatment in block j .

$$Y_{ij\ell} = \mu_i + \beta_j + \epsilon_{ij\ell} \quad (8)$$

Further, the ϵ 's and β 's are assumed normally distributed with $E(\epsilon_{ij\ell}) = E(\beta_j) = 0$, $\text{Var}(\epsilon_{ij\ell}) = \sigma^2$, $\text{Var}(\beta_j) = \sigma_\beta^2$ and finally all ϵ 's and β 's are mutually independent. The μ_i are fixed effects and all are estimable.

The model (8) is a special case of (1). Specifically $V = (+)A_j$ where $(+)$ denotes direct sum, $A_j = I_{k_j} + \gamma J_{k_j}$, $\gamma = \sigma_\beta^2/\sigma^2$ and I_{k_j} and J_{k_j} denote the

identity matrix and the matrix of ones, both of order k_j by k_j ; $\beta' = (\mu_1, \dots, \mu_t)$ and $X' = (X'_1, \dots, X'_b)$ is the design matrix appropriate for the model $Y_{ikl} = \mu_i + \epsilon_{ijl}$. Then $A_j^{-1} = I_{k_j} - \gamma(1+k_j\gamma)^{-1} J_{k_j}$ and $|A_j| = 1 + k_j\gamma$. Thus

$$|V| = \prod_{j=1}^b (1+k_j\gamma)$$

and

$$X'V^{-1}X = \sum X'_j X_j - \gamma \sum (1+k_j\gamma)^{-1} X'_j J_{k_j} X_j = (+)r_i - \gamma \Delta [(+)(1+k_j\gamma)^{-1}] \Delta' \quad (9)$$

where Δ_{ij} is the number of occurrences of treatment i in block j . if

$$d' = (d'_1, \dots, d'_b)$$

then

$$d'V^{-1}d = \sum d'_j d_j - \gamma \sum (1+k_j\gamma)^{-1} d'_j J_{k_j} d_j = 1-\gamma s^{-2} Q' [(+)(1+k_j\gamma)^{-1}] Q \quad (10)$$

where $Q = y_b - \Delta' \bar{y}_t$, y_b is the vector of unadjusted block totals and \bar{y}_t is the vector of unadjusted treatment means. Note that in this setting s^2 can be expressed as the sum of the error sum of squares and the block sum of squares adjusted for treatments. Finally,

$$X'V^{-1}d = \sum X'_j d_j - \gamma \sum (1+k_j\gamma)^{-1} X'_j J_{k_j} d_j = -\gamma s^{-1} \Delta [(+)(1+k_j\gamma)^{-1}] Q \quad (11)$$

when the block sizes are all equal to k , the term $(1+k_j\gamma)^{-1}$ can be factored out of the summation in each of (9), (10) and (11) and we define $\theta = \gamma(1+k\gamma)^{-1}$. Then $|V| = (1-k\theta)^{-b}$. The algebraic inversion of $X'V^{-1}X$ is not straightforward even with equal block sizes without further structure in the design. However, since the likelihood is one-dimensional in this case, it is relatively straightforward to find the value of γ (or θ) which maximizes the likelihood.

Expressions (9), (10) and (11) simplify further in the case that the

design is a balanced incomplete block design. Then, if λ is the number of times treatment pairs are block-mates

$$X'V^{-1}X = (r-(r-\lambda)\theta)I - \theta\lambda J$$

$$d'V^{-1}d = 1 - s^{-2}\theta Q'Q$$

$$X'V^{-1}d = \theta(t-1)^{-1}s^{-1}W$$

where $W = (t-k)y_t + (k-1)Y_{..1} - (t-1)B_t$, where y_t is the vector of treatment totals, and B_t is the total of all blocks in which the treatment occurs.

(This quantity is given in Cochran and Cox (1957), pg. 444.)

Then

$$|X'V^{-1}X| = r^t(1 - c\theta)^{t-1}(1-k\theta)$$

where

$$c = (t-k)/(t-1)$$

and

$$(X'V^{-1}X)^{-1} = r^{-1}(1-c\theta)^{-1}[I_t + (k-1)\theta(t-1)^{-1}(1-k\theta)^{-1}J_t] .$$

Then, since $1'W = 0$

$$d'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}d = r^{-1}(t-1)^{-2}\theta^2(1-c\theta)^{-1}s^{-2}W'W$$

and the likelihood (5) is proportional to:

$$(1-k\theta)^{(b-1)/2}(1-c\theta)^{-(t-1)/2}(1 - A\theta - B\theta^2(1-c\theta)^{-1})^{-t(r-1)/2} \quad (12)$$

where

$$A = Q' \cdot Q \cdot s^{-2}$$

$$B = r^{-1}(t-1)^{-2}s^{-2}W'W .$$

Taking the log of (12), differentiating with respect to θ and setting the result to zero results in

$$k(b-1)(1-k\theta)^{-1} - c(t-1)(1-c\theta)^{-1} - t(r-1)[A(1-c\theta)^2 + B\theta(2-c\theta)][1 - (A+c)\theta + (Ac-B)\theta^2]^{-1}[1-c\theta]^{-1} = 0.$$

Multiplying by $(1-k\theta)(1-c\theta)[1 - (A+c)\theta + (Ac-B)\theta^2]$ yields a cubic polynomial

in θ :

$$\begin{aligned}
 & (\Delta c - B)ctr(k-1)\theta^3 + \{(\Delta c - B)[k(b-1) - c(tr-1) - 2kt(r-1)] \\
 & \quad + (A+c)ct(r-k)\}\theta^2 + t\{(r-1)(k-1)A+2(r-1)(\Delta c - B) \\
 & \quad + c(k+1-2r)\}\theta - t(r-1)(A-1) = 0. \tag{13}
 \end{aligned}$$

6. BLOCKED DESIGNS WHICH GROUP INTO COMPLETE REPLICATES

The results above are modified slightly when the blocks group into complete replicates. We shall assume as before that we have b blocks but that they now group into r replicates. We shall initially, however, allow for missing data so that replicates may not be complete. The b blocks are of sizes k_1, \dots, k_b and the t treatments have r_1, \dots, r_t replicates. We shall regard the replicate effects as fixed resulting in the following model. Let $y_{ij\ell}$ be a response to the i -th treatment in the ℓ -th block from the j -th replicate. Then

$$y_{ij\ell} = \tau_i + \rho_j + \beta_{j\ell} + \epsilon_{ij\ell}$$

where $E(y_{ij\ell}) = \tau_i + \rho_j$. In order to make all parameters estimable we set $\rho_r = 0$. Let X be the n by $(t+r-1)$ design matrix associated with these fixed effects. The assumptions with regard to β and ϵ are unchanged.

In the sequel, when subscripting expressions which change over blocks, we shall use $m = 1, \dots, b$ rather than the double subscript j and ℓ . With this notational change V is unchanged.

Define r_b to be a b dimensional vector whose j -th element is the sum of the residuals from block j . These residuals are given by $Y - Xb$ when $b = (X'X)^{-1}X'Y$ and $s^2 = (Y - Xb)'(Y - Xb)$. Define K to be the $r-1$ by b matrix

$$K = \begin{pmatrix} r-1 \\ (+) k_j' \\ k=1 \end{pmatrix}$$

where k_j' is the row vector of block sizes for replicate j . The definition

for Δ remains unchanged. Then

$$d'V^{-1}d = 1 - \gamma s^{-2} r_b' [(+) (1+k_j \gamma)^{-1}] r_b$$

$$X'V^{-1}d = -\gamma s^{-1} \begin{pmatrix} \Delta \\ K \end{pmatrix} [(+) (1+k_k \gamma)^{-1}] r_b$$

$$X'V^{-1}X = X'X - \gamma \begin{pmatrix} \Delta \\ K \end{pmatrix} [(+) (1+k_j \gamma)^{-1}] (\Delta'K')$$

When all replicates are complete (there is no missing data), these formulae simplify considerably. r_b becomes

$$r_b = y_b - \Delta' \bar{y}_t - k(\bar{y}_r(x) 1_s - \bar{y} \dots 1_b)$$

where y_b , \bar{y}_t , \bar{y}_r , are respectively the vectors containing the block totals, the treatment means and the replicate means, $s = b/r$ is the number of blocks per replicate and (x) denotes direct product. Then

$$d'V^{-1}d = 1 = \theta s^{-2} r_b' r_b$$

$$X'V^{-1}d = \theta s^{-1} \begin{pmatrix} \Delta r_b \\ 0 \end{pmatrix}$$

$$X'V^{-1}X = \begin{pmatrix} rI - \theta \Delta \Delta' & (1-k\theta)J \\ (1-k\theta)J & t(1-k\theta)I \end{pmatrix}$$

Finally if $P = rI_t - (1-k\theta)t^{-1}(r-1)J_t - \theta \Delta \Delta'$ then

$$|X'V^{-1}X| = t(1-k\theta)|P|$$

and

$$(X'V^{-1}X)^{-1} = \begin{bmatrix} P^{-1} & -t^{-1}P^{-1}J \\ -t^{-1}JP^{-1} & t^{-1}(1-k\theta)^{-1}I - t^{-2}JP^{-1}J \end{bmatrix}$$

7. SPECIAL CASES WHICH COINCIDE WITH ANOVA ESTIMATORS

In some special cases, the value of θ which maximizes the marginal likelihood can be derived explicitly and is found to be equivalent to the estimator one would obtain by equating observed and expected mean squares in

the analysis of variance. We consider the following three special cases:

1) The one way random model with balanced data. That is, let

$$Y_{ij} = \mu + \beta_j + \epsilon_{ij} \quad i = 1, \dots, k, \quad j = 1, \dots, b$$

Thus, $E(Y_{ij}) = \mu$ and the covariance of responses in the j th group is

$$\text{Cov}(Y_j) = \sigma^2(I_k + \gamma J_k) \text{ where } \gamma = \sigma_\beta^2/\sigma^2.$$

Simplifying (9), (10) and (11), we get

$$\begin{aligned} X'V^{-1}X &= bk(1+k\gamma)^{-1} \\ d'V^{-1}d &= 1 - k\gamma(\text{BSS})/[(1+k\gamma) \cdot s^2] \\ X'V^{-1}d &= 0 \end{aligned} \tag{14}$$

where $s^2 = \text{BSS} + \text{ESS}$, BSS is the among group sum of squares and ESS is the residual or error sum of squares. The marginal likelihood (5) can then be simplified to

$$(1-k\theta)^{(b-1)/2} (1-k(\text{BSS})\theta/s^2)^{-(bk-1)/2} \tag{15}$$

where $\theta = \gamma/(1+k\gamma) = \sigma_\beta^2/(\sigma^2 + k\sigma_\beta^2)$. The value of θ which maximizes this likelihood can be shown algebraically to be $\hat{\theta} = (\text{BMS} - \text{EMS})/k \cdot \text{BMS}$ where BMS is the among groups mean square and EMS is the error mean square. But $\hat{\theta}$ can be written

$$\hat{\theta} = \tilde{\sigma}_\beta^2/(\tilde{\sigma}^2 + k\tilde{\sigma}_\beta^2)$$

where $\tilde{\sigma}_\beta^2$ and $\tilde{\sigma}^2$ are the (unbiased) ANOVA based estimators.

2) The randomized complete block design with blocks regarded as being random. Let

$$Y_{ij} = \mu_i + \beta_j + \epsilon_{ij} \quad i=1, \dots, k, \quad j=1, \dots, b \tag{16}$$

where $E(Y_{ij}) = \mu_i$ and the covariance of the responses in the j th block is

$$\text{Cov}(Y_j) = \sigma^2(I + \gamma J). \text{ The expressions (14) remain unchanged except that}$$

$$X'V^{-1}X = b(I_k - \theta J_k). \text{ The likelihood (15) remains unchanged except that the}$$

exponent of the second term becomes $-k(b-1)/2$. Thus, the θ which maximizes

this likelihood is $\hat{\theta} = (\text{BMS} - \text{EMS})/k \cdot \text{BMS}$ which is the estimator one obtains by

substituting the ANOVA estimators $\bar{\sigma}^2$ and $\bar{\sigma}_\beta^2$ into θ .

3) A balanced incomplete block design in which the number of blocks equals the number of treatments ($b = t$). In this case, $W'W = \text{tr}(t-k)(k-1) \cdot \text{BSS}(\text{adj})$ and $A = t(r-1)(t-1)^{-1} s^{-2} \text{BSS}(\text{adj})$ where $\text{BSS}(\text{adj})$ is the block SS adjusted for treatments so that, in the cubic polynomial (13), $Ac-B = 0$ and $k = r$ reducing the polynomial to $[(r-1)A-c]\theta - (A-1) = 0$. This yields

$$\hat{\theta} = \frac{(\text{tr}-2t+1)\text{BSS}(\text{adj}) - (t-1)\text{ESS}}{r(\text{tr}-2t+1)\text{BSS}(\text{adj}) - (t-r)\text{ESS}}$$

and this is the estimator of θ one obtains by substituting into θ , the ANOVA estimates of σ^2 and σ_β^2 .

The fact that estimators based on the marginal likelihood (5) coincide with ANOVA based estimators in simple cases with a good deal of data structure, is re-assuring since the ANOVA based estimators are known to have good properties in these simple settings. Corbeil and Searle (1976), have shown that restricted maximum likelihood (REML) estimators also coincide with ANOVA estimators in similar simple balanced data settings.

8. ANALYSIS OF A BIBD IN WHICH $t \neq b$

We close with a data set from a BIBD for which the value of $\delta = \sigma_\beta^2/\sigma^2$ which maximizes the marginal likelihood (5) does not coincide with the ANOVA based estimator. The data, taken from Federer (1955), page 420, is from a BIBD with $b = 15$, $t = 10$, $k = 4$ and $r = 6$. Preliminary calculations yield $\text{BSS}(\text{adj}) = 23.8745$, $s^2 = 41.755$ and $W'W = 24295.86$. Hence, $A = 2.0421$ and $B = 1.19726$. The likelihood (12) is plotted in figure 1 verses $\gamma = \sigma_\beta^2/\sigma^2$. This curve reaches a maximum at $\hat{\gamma} = 0.70$. The ANOVA estimators, however, are $\bar{\sigma}^2 = .4967$ and $\bar{\sigma}_\beta^2 = .3384$ yielding $\bar{\gamma} = .68$. Because this is a balanced incomplete block design, it is not surprising that the two estimates of γ are

close.

REFERENCES

- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. Proc. Royal Soc. A 160, 268-82.
- Corbeil, R.R. and Searle, S.R. (1976). A comparison of variance component estimators. Biometrics 32, 779-91.
- Fraser, D.A.S. (1968). The Structure of Inference. New York, Wiley.
- Kalbfleisch, J.D. and Sprott, D.A. (1970). Application of likelihood methods to models involving large numbers of parameters J.R. Statist. Soc. B 32, 175-208.
- Levenbach, Hans (1973). The estimation of heteroscedasticity from a marginal likelihood function. J. Amer. Statist. Assoc. 68, 436-39.
- Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica 16, 1-32.
- Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-54.
- Saw, J.G. (1973). Jacobians of singular transformations with applications to statistical distribution theory, Comm. in Statist. 1, 81-91.
- Sprott, D.A. (1975). Marginal and conditional sufficiency. Biometrika 62, 599-605.

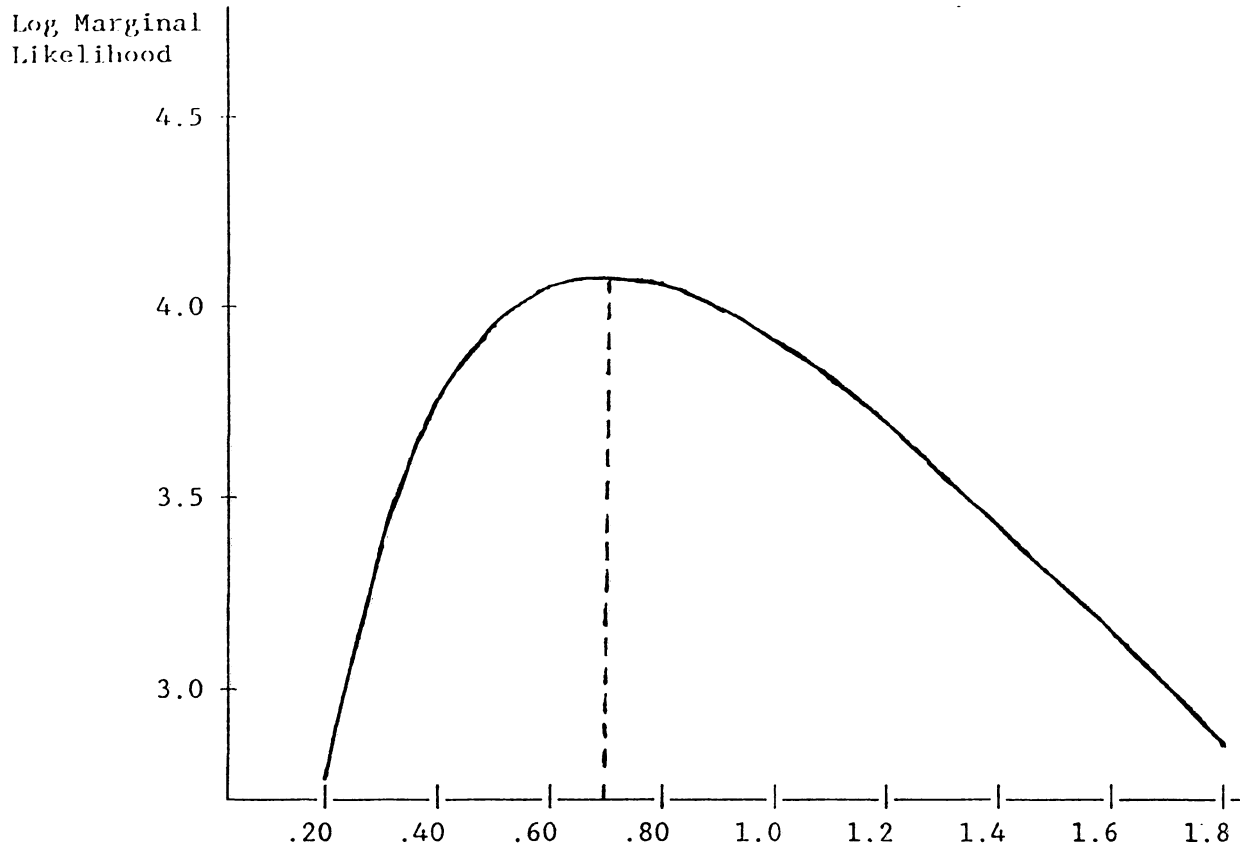


Figure 1 A plot of the marginal likelihood (12) against $\sigma_{\beta}^2/\sigma^2$ for data from Federer (1955), page 420. The ANOVA based estimate is .68 whereas .70 maximizes this likelihood.