THE PROBLEM OF NEGATIVE ESTIMATES OF VARIANCE COMPONENTS

By W. A. Thompson, Jr.

University of Delaware[1]

0. Summary. The usefulness of variance component techniques is frequently limited by the occurrence of negative estimates of essentially positive parameters. This paper uses a restricted maximum likelihood principle to remove this objectionable characteristic for certain experimental models. Section 2 discusses certain necessary results from the theory of non-linear programming. Section 3 derives specific formulae for estimating the variance components of the random one-way and two-way classification models. The problem of determining the precision of instruments in the two instrument case is dealt with in section 4, and a surprising though not unreasonable answer is obtained.

The remaining sections provide an algorithm for solving the problem of negative estimates of variance components for all random effects models whose expected mean square column may be thought of as forming a mathematical tree in a certain sense. The algorithm is as follows: Consider the minimum mean square in the entire array; if this mean square is the root of the tree then equate it to its expectation. If the minimum mean square is not the root then pool it with its predecessor. In either case the problem is reduced to an identical one having one less variable, and hence in a finite number of steps the process will yield estimates of the variance components. These estimates are non-negative and have a maximum likelihood property.

1. Introduction. In order to discuss the philosophy of this paper consider a series of J astronomical observations of the same quantity taken on each of K nights [13]. There are then KJ observations in all, which will be denoted by y_{kj}, $k = 1, \cdots, K$; $j = 1, \cdots, J$. If μ is the "true" value of the observation being made and a_k is the error peculiar to the kth night, then it may be appropriate to assume that the conditional random variables y_{kj} given a_k are independent and normally distributed with common variance σ^2 and with conditional expectations equal to $\mu + a_k$. Further, if the a_k's are independent and normal with means 0 and variances σ_a^2, then the unconditional distribution of the y_{kj}'s is normal. Each observation has expectation μ, variance $\sigma^2 + \sigma_a^2$, and the covariance of y_{kj} and y_{il} is σ_a^2 if $i = k$ and is 0 otherwise. This is, of course, equivalent to the usual balanced one-way variance component model: $y_{kj} = \mu + a_k + e_{kj}$ where μ is an unknown constant and the a's and e's are independent normal variates with zero means and variances σ_a^2, σ^2 respectively. However, the above conditional probability argument constitutes, to our way of thinking, a derivation of the usual model from more reasonable assumptions.

Received October 26, 1959; revised June 30, 1961.

[1] A portion of this research was carried out while the author was a summer employee at White Sands Missile Range.

Denoting the overall mean of the observations by $\bar{y}_{..}$ and the mean of the observation taken on the kth night by $\bar{y}_{k.}$, then it may be demonstrated [14, p. 226] that $\bar{y}_{..}$, $s_a^2 = J\sum_k (\bar{y}_{k\cdot} - \bar{y}_{..})^2/(K-1)$, and $s^2 = \sum_{j,k} (y_{kj} - \bar{y}_{k\cdot})^2/(KJ - K)$ are a set of independent sufficient statistics. s_a^2 and s^2 are, of course, the mean squares due to the night effect and error respectively. If $h(\alpha, p)$ denotes the density function of a random variable which is α times a mean-square deviate with p degrees of freedom, i.e.,

$$h(\alpha, p) = [\Gamma(p/2)(2\alpha/p)^{p/2}]^{-1}x^{\frac{1}{2}p-1} \exp(-px/2\alpha), \quad \alpha > 0, x \geqq 0, \tag{1.1}$$

then s^2 and s_a^2 have the densities $h[\sigma^2, K(J-1)]$ and $h(\sigma^2 + J\sigma_a^2, K-1)$ respectively.

The traditional estimators of σ^2 and σ_a^2, which are s^2 and $(s_a^2 - s^2)/J$ respectively, are obtained from Table 1 by equating the mean-squares to their expectations and solving. Clearly the traditional estimate of σ_a^2 may be negative; should this occur, we do not believe that any such statistical analysis would become useful until it is decided what to do with the negative estimate. This,

TABLE 1

Source	d.f.	Mean Square	Expected Mean Square
nights	$K-1$	s_a^2	$\sigma^2 + J\sigma_a^2$
error	$K(J-1)$	s^2	σ^2

then, is an example of what we mean by "the problem of negative estimates of variance components". Two possible explanations of a negative estimate present themselves: (1) the assumed model may be incorrect and (2) statistical noise may have obscured the underlying physical situation. Anscombe [2] and Nelder [10] have done valuable work which adopts the first explanation; in this paper we explore the second point of view. That is, we take the assumed model to be correct and ask how our estimation procedures should be changed when negative estimates occur.

Herbach [7] has used the maximum likelihood principle to obtain variance component estimators which are non-negative. For the balanced one-way classification that is being discussed in this section, the full maximum likelihood estimators are $\hat{u} = \bar{y}_{..}$,

$$\hat{\sigma}^2 = s^2, \qquad J\hat{\sigma}_a^2 = (1 - K^{-1})s_a^2 - s^2, \qquad \text{when } (1 - K^{-1})s_a^2 \geqq s^2,$$

$$\hat{\sigma}_a^2 = 0, \qquad \hat{\sigma}^2 = [(K-1)s_a^2 + K(J-1)s^2]/KJ, \qquad \text{when } (1 - K^{-1})s_a^2 < s^2.$$

In the present paper we do not use these true maximum likelihood estimators. Instead, we use a method of restricted maximum likelihood which, for the one-way classification, yields estimators only slightly different from those obtained by full maximum likelihood. We illustrate the method by continuing our discussion of the balanced one-way classification. By sufficiency it is immaterial whether we maximize the joint likelihood of the original JK observations or the joint likelihood of $\bar{y}..$, s^2 and s_a^2. For the restricted maximum likelihood method we go further and maximize the joint likelihood of s^2 and s_a^2.

For estimating scale parameters in general, the *restricted maximum likelihood* method, used in this paper, consists of maximizing the joint likelihood of that portion of the set of sufficient statistics which is location invariant. This procedure is utilized in similar problems by Anderson and Bancroft [1, p. 320], for example, and has two arguments in its favor. First, it gives the accepted estimates in similar cases where it is well established what estimates "should" be used. For example, if $x_1, \cdots, x_n$ are independent and identically distributed normal deviates from a population whose mean and variance are both unknown then full maximum likelihood would yield the variance estimator $\Sigma(x_i - \bar{x})^2/n$. We believe that very few statisticians would use this estimator in a practical situation; they would prefer $\Sigma(x_i - \bar{x})^2/(n - 1)$, the estimator given by restricted maximum likelihood.

Second, the intention here is not to question the traditional estimates, but to investigate the way in which the traditional estimates should be altered if the problem of negative estimates arises. Consideration of the likelihood function of s^2 and s_a^2 yields the traditional estimates, while working with all JK observations does not. Similar considerations apply in the other models considered in this paper. A tilde placed over a parameter will denote the restricted maximum likelihood estimator of the parameter. Thus $\tilde{\sigma}^2$ is the restricted maximum likelihood estimator of the parameter σ^2.

The log-likelihood function of s_a^2 and s^2 is

$$L(\omega_1, \omega_2) = -\tfrac{1}{2}(c + f_1 \log \omega_1 + f_2 \log \omega_2 + f_1 w_1/\omega_1 + f_2 w_2/\omega_2), \tag{1.2}$$

where $f_1 = K(J - 1)$, $f_2 = K - 1$, $\omega_1 = \sigma^2$ and $\omega_2 = \sigma^2 + J\sigma_a^2$; and c does not depend on σ^2 or σ_a^2. w_1 and w_2 denote the values which may be assumed by the random variables s^2 and s_a^2 respectively. The notation introduced at this point anticipates a terminology which will be used throughout the paper.

From a conceptual viewpoint, the solution of the problem of negative estimates of variance components, at least in so far as restricted maximum likelihood is concerned, is that L should be maximized *subject to the constraints* $\sigma^2 \geqq 0$ and $\sigma_a^2 \geqq 0$. Because of the invariance of maximum likelihood estimates, it is enough to maximize L with respect to ω_1 and ω_2 subject to the constraints $\omega_2 \geqq \omega_1 > 0$. Note that

$$\partial L/\partial \omega_i = \tfrac{1}{2} f_i (w_i - \omega_i)/\omega_i^2 ; \qquad i = 1, 2; \tag{1.3}$$

and hence if $0 \leqq w_1 \leqq w_2$, then the restricted maximum likelihood estimates of ω_1 and ω_2 are respectively w_1 and w_2.

The other possibility is that $0 \leqq w_2 < w_1$ in which case $(\tilde{\omega}_1, \tilde{\omega}_2)$, the vector of restricted maximum likelihood estimates, lies on the boundary of the admissible region. It is shown in the next section that $\tilde{\omega}_1$ cannot equal zero and hence $\tilde{\omega}_1 = \tilde{\omega}_2$ in this case. Throughout the region $\omega_1 = \omega_2 = \omega$, the relation $\partial L/\partial\omega = \frac{1}{2}(f_1 + f_2)[(f_1 w_1 + f_2 w_2)/(f_1 + f_2) - \omega]/\omega^2$ is valid and $\tilde{\omega}_1 = \tilde{\omega}_2 = (f_1 w_1 + f_2 w_2)/(f_1 + f_2)$.

Now reinterpreting, if $s_a^2 \geqq s^2$ then $\tilde{\sigma}^2 = s^2$ and $\tilde{\sigma}_a^2 = (s_a^2 - s^2)/J$. If, however, $s_a^2 < s^2$ then $\tilde{\sigma}_a^2 = 0$ and $\tilde{\sigma}^2 = [(K-1)s_a^2 + K(J-1)s^2]/(KJ-1)$. Referring to Table 1, it is likely that this is what one would have done without the detailed mathematical analysis. We have, however, relieved our consciences a little and at the same time laid the foundation for the analysis of more complex situations.

2. A maximization problem. In the following, ω and $\tilde{\omega}$ are p dimensional column vectors while A is a non-singular p by p matrix. The transpose of the matrix M is denoted by M'. $g(\omega)$ is a function of p variables, $G_i(\omega) = \partial g/\partial \omega_i$ and $G(\omega)$ is the column vector with elements $G_1(\omega), \cdots, G_p(\omega)$. A vector is said to be $\geqq 0$ or $\leqq 0$ if and only if each of its elements satisfies these inequalities. Denote the ith rows of A^{-1} and A' by b_i and a_i respectively.

THEOREM 1. *If $g(\omega)$ is a differentiable function at $\tilde{\omega}$, then a necessary set of conditions that it has a relative maximum at $\tilde{\omega}$ subject to the constraints $A^{-1}\omega \geqq 0$ is:*

$$(2.1)\quad \text{(a) } A^{-1}\tilde{\omega} \geqq 0, \qquad \text{(b) } A'G(\tilde{\omega}) \leqq 0 \quad \text{and} \quad \text{(c) } \tilde{\omega}'G(\tilde{\omega}) = 0.$$

Furthermore the conditions (2.1) *are equivalent to, for each $i(=1, \cdots, p)$, either*

$$(2.2)\quad \begin{array}{c} b_i\tilde{\omega} = 0 \quad \text{and} \quad a_i G(\tilde{\omega}) \leqq 0 \\ \text{or} \\ b_i\tilde{\omega} > 0 \quad \text{and} \quad a_i G(\tilde{\omega}) = 0. \end{array}$$

The set of all points ω satisfying $A^{-1}\omega \geqq 0$ will be referred to as the admissible region. The previous theorem can be proved from the results of Kuhn and Tucker [8] or more simply by considering the directional derivative at $\tilde{\omega}$ in the direction of an arbitrary point in the admissible region. In this way one first establishes that (c) must hold and then that $\omega'G(\tilde{\omega}) \leqq 0$ for all admissible ω. The Farkas theorem (a statement may be found in [8]) is then used to establish (b).

In order to study conditions under which there will be a unique solution to the relations (2.1) or (2.2) it is necessary to remind the reader of a property of strictly concave functions and their tangent planes. The equation of the plane tangent to $g(\omega)$ at $\omega = \tilde{\omega}$ is $(\omega - \tilde{\omega})'G(\tilde{\omega}) + g(\tilde{\omega}) = t(\omega)$, say. If $g(\omega)$ is a strictly concave function having gradient vector $G(\omega)$, then $t(\omega) \geqq g(\omega)$, equality holding only at $\omega = \tilde{\omega}$.

THEOREM 2. *If $g(\omega)$ is a strictly concave function having gradient vector $G(\omega)$ and if $\tilde{\omega}$ satisfies* (2.1), *then $g(\tilde{\omega}) > g(\omega)$ whenever $A^{-1}\omega \geqq 0$ and $\omega \neq \tilde{\omega}$.*

PROOF. Assume for purposes of contradiction that $g(\bar{\omega}) \geqq g(\tilde{\omega})$, $A^{-1}\bar{\omega} \geqq 0$ and $\bar{\omega} \neq \tilde{\omega}$. Then $t(\bar{\omega}) > g(\bar{\omega}) \geqq g(\tilde{\omega})$, $(\bar{\omega} - \tilde{\omega})'G(\tilde{\omega}) + g(\tilde{\omega}) > g(\tilde{\omega})$ and $\bar{\omega}'G(\tilde{\omega}) > \tilde{\omega}'G(\tilde{\omega}) = 0$. But $\bar{\omega}'G(\tilde{\omega}) = (A^{-1}\bar{\omega})'A'G(\tilde{\omega}) \leqq 0$ a contradiction.

COROLLARY. *If $g(\omega)$ is a strictly concave function having gradient vector $G(\omega)$, then the solution to* (2.1) *or* (2.2) *is unique.*

3. Multiple classifications.

General Theory. The analysis of the one-way classification has already been treated in the introduction. The problem treated in this section can be described as follows: $w_1, \cdots, w_p$ are p independently distributed variables. Using the notation of equation (1.1), the probability density function of w_i is $h(\omega_i, f_i)$. w, σ^2 and ω are the column vectors $(w_1, \cdots, w_p)'$, $(\sigma_1^2, \cdots, \sigma_p^2)'$ and $(\omega_1, \cdots, \omega_p)'$ respectively.

$$E(w) = \omega = A_{p\times p}\sigma^2.$$

$\sigma^2 = A^{-1}\omega$ if the inverse exists. In order to find $\tilde{\sigma}^2$ (the vector of restricted maximum likelihood estimates of the components of σ^2) it is sufficient to maximize twice the log of the density function of $w_1, \cdots, w_p$, which is

$$g(\omega) = \text{const} - \sum_{i=1}^{p} f_i(\log \omega_i + w_i/\omega_i), \tag{3.1}$$

with respect to $\omega_1, \cdots, \omega_p$ and subject to the set of linear constraints $A^{-1}\omega \geqq 0$.

For the function (3.1),

$$G_i = \partial g/\partial \omega_i = f_i(w_i - \omega_i)/\omega_i^2, \tag{3.2}$$

and (2.1c) becomes

$$\sum f_i = \sum f_i w_i/\tilde{\omega}_i. \tag{3.3}$$

Note that $g(\omega) = \text{const} - \sum f_i \log \omega_i - \sum f_i$ on the curve (3.3).

One final point needs to be emphasized before proceeding to special cases. From (1.1) it is clear that (1.2) and (3.1) are valid only if $w_i \geqq 0$ and $\omega_i > 0$. If $w_i > 0$ then the value of the likelihood function at $\omega_i = 0$ is $-\infty$ and in looking for a maximum the possibility $\omega_i = 0$ need not be considered. If $w_i = 0$, then $\tilde{\omega}_i = 0$.

The two way classification. For a development of this model see [9, p. 345], Table 2 gives the expected mean squares.

Here σ_1^2 is the interaction component, $\sigma_2^2 = r_1\sigma_b^2$, $\sigma_3^2 = r_2\sigma_a^2$, σ_a^2 and σ_b^2 are the row and column components, and r_1 and r_2 are the number of rows and columns respectively.

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

TABLE 2

Source	Mean Square	Expected Mean Square
Rows	w_3	$\omega_3 = \sigma_1^2 + \sigma_3^2$
Columns	w_2	$\omega_2 = \sigma_1^2 + \sigma_2^2$
Deviations	w_1	$\omega_1 = \sigma_1^2$

The equations (2.1) become

(i) $\tilde{\omega}_1 \geqq 0$;

(i′) $0 \leqq f_1(\tilde{\omega}_1 - w_1)/\tilde{\omega}_1^2 + f_2(\tilde{\omega}_2 - w_2)/\tilde{\omega}_2^2 + f_3(\tilde{\omega}_3 - w_3)/\tilde{\omega}_3^2$

(ii) $\tilde{\omega}_1 \leqq \tilde{\omega}_2$; (ii′) $\tilde{\omega}_2 \geqq w_2$

(iii) $\tilde{\omega}_1 \leqq \tilde{\omega}_3$; (iii′) $\tilde{\omega}_3 \geqq w_3$.

According to Section 2, necessary conditions for maximum likelihood estimates are that equality must hold in at least one equation of each pair; further, if $w_1 > 0$ a maximum likelihood estimate cannot occur at $\omega_1 = 0$. There are thus four possibilities according to whether equality does or does not hold in (ii) and (iii).

$w_{i \cdots k}$ is the mean square obtained by pooling $w_i, \cdots, w_k$, i.e.,

$$w_{i \cdots k} = (f_i w_i + \cdots + f_k w_k)/(f_i + \cdots + f_k). \tag{3.4}$$

The four possible cases with their solutions and implications for the w's are listed.

(a) $\tilde{\omega}_1 < \tilde{\omega}_2$, $\tilde{\omega}_1 < \tilde{\omega}_3$; $\tilde{\omega}_1 = w_1$, $\tilde{\omega}_2 = w_2$, $\tilde{\omega}_3 = w_3$; $w_1 < w_2$, $w_1 < w_3$.
(b) $\tilde{\omega}_1 = \tilde{\omega}_2$, $\tilde{\omega}_1 < \tilde{\omega}_3$; $\tilde{\omega}_1 = \tilde{\omega}_2 = w_{12}$, $\tilde{\omega}_3 = w_3$; $w_1 \geqq w_2$, $w_2 \leqq w_{12} < w_3$.
(c) $\tilde{\omega}_1 = \tilde{\omega}_3$, $\tilde{\omega}_1 < \tilde{\omega}_2$; $\tilde{\omega}_1 = \tilde{\omega}_3 = w_{13}$, $\tilde{\omega}_2 = w_2$; $w_1 \geqq w_3$, $w_3 \leqq w_{13} < w_2$.
(d) $\tilde{\omega}_1 = \tilde{\omega}_2$, $\tilde{\omega}_1 = \tilde{\omega}_3$; $\tilde{\omega}_1 = \tilde{\omega}_2 = \tilde{\omega}_3 = w_{123}$; $w_{123} \geqq w_2$, $w_{123} \geqq w_3$; $w_{13} \geqq w_2$, $w_{12} \geqq w_3$ and either $w_1 \geqq w_2$ or $w_1 \geqq w_3$.

Table 3 gives the appropriate estimates under the various conditions. Since (a), (b), (c) and (d) are mutually exclusive, the maximum likelihood estimates are unique in this case.

The two factor experiment with m observations per cell. The table of expected mean squares appears here as Table 4, see [9, p. 346] for the background.

Equations (2.1) give

(i′) $0 = f_1(\tilde{\omega}_1 - w_1)/\tilde{\omega}_1^2 + f_2(\tilde{\omega}_2 - w_2)/\tilde{\omega}_2^2 + f_3(\tilde{\omega}_3 - w_3)/\tilde{\omega}_3^2 + f_4(\tilde{\omega}_4 - w_4)/\tilde{\omega}_4^2$

(ii) $\tilde{\omega}_1 \leqq \tilde{\omega}_2$ (ii′) $\tilde{\omega}_1 \leqq w_1$

TABLE 3

Conditions	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$	$\tilde{\sigma}_3^2$
$w_1 < w_2$, $w_1 < w_3$	w_1	$w_2 - w_1$	$w_3 - w_1$
$w_1 \geqq w_2$, $w_2 \leqq w_{12} < w_3$	w_{12}	0	$w_3 - w_{12}$
$w_1 \geqq w_3$, $w_3 \leqq w_{13} < w_2$	w_{13}	$w_2 - w_{13}$	0
$w_{13} \geqq w_2$, $w_{12} \geqq w_3$ and either $w_1 \geqq w_2$ or $w_1 \geqq w_3$	w_{123}	0	0

TABLE 4

Source	Mean Square	Expected Mean Square
Rows	w_4	$\omega_4 = \sigma_1^2 + \sigma_2^2 + \sigma_4^2$
Columns	w_3	$\omega_3 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$
Interactions	w_2	$\omega_2 = \sigma_1^2 + \sigma_2^2$
Deviations	w_1	$\omega_1 = \sigma_1^2$

TABLE 5

Conditions	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$	$\tilde{\sigma}_3^2$	$\tilde{\sigma}_4^2$
$w_1 < w_2 < w_3$, $w_2 < w_4$	w_1	$w_2 - w_1$	$w_3 - w_2$	$w_4 - w_2$
$w_1 < w_{24} < w_3$, $w_2 \geqq w_4$	w_1	$w_{24} - w_1$	$w_3 - w_{24}$	0
$w_1 < w_{23} < w_4$, $w_2 \geqq w_3$	w_1	$w_{23} - w_1$	0	$w_4 - w_{23}$
$w_1 \geqq w_2$, $w_{12} < w_3$, $w_{12} < w_4$	w_{12}	0	$w_3 - w_{12}$	$w_4 - w_{12}$
$w_{123} < w_4$, $w_3 \leqq w_{12}$, $w_{23} \leqq w_1$	w_{123}	0	0	$w_4 - w_{123}$
$w_{124} < w_3$, $w_4 \leqq w_{12}$, $w_{24} \leqq w_1$	w_{124}	0	$w_3 - w_{124}$	0
$w_1 < w_{234}$, $w_3 \leqq w_{24}$, $w_4 \leqq w_{23}$	w_1	$w_{234} - w_1$	0	0
$w_3 \leqq w_{124}$, $w_4 \leqq w_{123}$, $w_1 \geqq w_{234}$	w_{1234}	0	0	0

(iii) $\tilde{\omega}_2 \leqq \tilde{\omega}_3$ (iii′) $\tilde{\omega}_3 \geqq w_3$

(iv) $\tilde{\omega}_2 \leqq \tilde{\omega}_4$ (iv′) $\tilde{\omega}_4 \geqq w_4$.

Hence in this case, there are eight possibilities depending on whether equality does or does not hold in (ii) (iii) and (iv). Table 5 gives the various solutions and again they are mutually exclusive showing that the maximum likelihood estimates are unique. The notation is that of (3.4).

4. Determination of the precision of instruments. The succeeding portions of this paper may be read without reference to this section. The problem of determining the precision of instruments has been studied by Simon [15] and Grubbs [6], and by Russell and Bradley [12]. Russell and Bradley have a slightly different practical application in mind, and their model assumes that all effects are fixed except those which are called the instrument errors in this paper. The Simon, Grubbs technique results in estimates of the variance components which are frequently negative. In this section the model of Grubbs is assumed but estimators are developed which are non-negative. We treat only the "two instrument" case.

It is instructive to "derive" the Grubbs model from the point of view of conditional probability. With this in mind, consider a random sample $a_1, \cdots, a_n$ of size n selected from a normal distribution with mean μ and variance σ^2. Each of the items in this sample is then measured by two instruments. y_{ij} is the measurement of the ith item according to the jth instrument. y_{ij} given a_i is normal with mean $a_i + \beta_j$ and variance σ_j^2. The instrument errors are assumed to be independent. β_j is the bias of the jth instrument. Thus y_{i1} and y_{i2}, $i = 1$,

$\cdots, n$, are a sample of size n from a bivariate normal distribution with $E(y_{ij}) = \mu + \beta_j$. The variance covariance matrix of y_{i1} and y_{i2} is

$$\Sigma = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma_2^2 + \sigma^2 \end{pmatrix}.$$

and $|\Sigma| = \sigma_1^2\sigma^2 + \sigma_1^2\sigma_2^2 + \sigma^2\sigma_2^2$. If $|\Sigma|$ is not zero then σ^{ij} denotes the element in the ith row and jth column of Σ^{-1}.

Once again we apply the restricted maximum likelihood method. The location invariant portion of the set of sufficient statistics consists of the sample variances and covariance:

$$s_{11} = \Sigma(y_{i1} - \bar{y}_{\cdot 1})^2/(n-1), \qquad s_{22} = \Sigma(y_{i2} - \bar{y}_{\cdot 2})^2/(n-1), \qquad \text{and}$$
$$s_{12} = \Sigma(y_{i1} - \bar{y}_{\cdot 1})(y_{i2} - \bar{y}_{\cdot 2})/(n-1).$$

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}.$$

The determination of $f(s_{11}, s_{22}, s_{12})$, the joint probability density function, is complicated by the fact that the observations may follow the singular as well as the non-singular normal distribution. Using the notation of (1.1)

$$\begin{aligned} f(s_{11}, s_{22}, s_{12}) & \\ &= h(\sigma^2, n-1), \quad s_{11} = s_{12} = s_{22}, \quad \sigma_1^2 = \sigma_2^2 = 0, \quad \sigma^2 > 0, \\ &= h(\sigma_1^2, n-1), \quad s_{12} = s_{22} = 0, \quad \sigma^2 = \sigma_2^2 = 0, \quad \sigma_1^2 > 0, \\ &= h(\sigma_2^2, n-1), \quad s_{11} = s_{12} = 0, \quad \sigma^2 = \sigma_1^2 = 0, \quad \sigma_2^2 > 0, \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

for the singular cases. For the non-singular case [4, p. 397]

$$\begin{aligned} f(s_{11}, s_{22}, s_{12}) & \\ &= \text{constant } |S|^{\frac{1}{2}(n-4)} |\Sigma|^{-\frac{1}{2}(n-1)} \exp\left[-\tfrac{1}{2}(n-1)(\sigma^{11}s_{11} + \sigma^{22}s_{22} + 2\sigma^{12}s_{12})\right] \end{aligned}$$

in the domain where S is positive definite and is zero elsewhere. Hence if $L = L(\sigma^2, \sigma_1^2, \sigma_2^2) = \log f(s_{11}, s_{22}, s_{12})$, then

$$\begin{aligned} 2L/(n-1) &= c - \log \sigma^2 - s_{11}/\sigma^2, \quad \sigma_1^2 = \sigma_2^2 = 0, \quad \sigma^2 > 0, \\ &\qquad s_{11} = s_{22} = s_{12}, \\ &= c - \log \sigma_2^2 - s_{22}/\sigma_2^2, \quad \sigma_1^2 = \sigma^2 = 0, \quad \sigma_2^2 > 0, \\ &\qquad s_{12} = s_{11} = 0, \\ &= c - \log \sigma_1^2 - s_{11}/\sigma_1^2, \quad \sigma_2^2 = \sigma^2 = 0, \quad \sigma_1^2 > 0, \qquad (4.1) \\ &\qquad s_{12} = s_{22} = 0, \\ &= c - \log |\Sigma| - (\sigma^{11}s_{11} + \sigma^{22} + 2\sigma^{12}s_{12}), \quad |\Sigma| > 0, \end{aligned}$$

S positive definite,

$= -\infty$ otherwise.

Here c is a generic symbol denoting a quantity independent of σ_1^2, σ_2^2 and σ^2.

Consider the case where S is positive definite; the region $|\Sigma| = 0$ need not be investigated for maxima since $L = -\infty$ throughout this region. With a view to using Theorem 1 write

$$\omega_1 = \sigma^{11} = (\sigma_2^2 + \sigma^2)/|\Sigma|, \qquad \omega_2 = \sigma^{22} = (\sigma_1^2 + \sigma^2)/|\Sigma|,$$
$$\omega_3 = \sigma^{12} = \sigma^{21} = -\sigma^2/|\Sigma|.$$

The restrictions on these parameters are $\omega_1 + \omega_3 \geqq 0$, $\omega_2 + \omega_3 \geqq 0$, $-\omega_3 \geqq 0$.

$$A^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \qquad A' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix},$$

$$G_1 = n/2(s_{11} - \sigma_1^2 - \sigma^2), \qquad G_2 = n/2(s_{22} - \sigma_2^2 - \sigma^2),$$
$$G_3 = n/2(s_{12} - \sigma^2).$$

The inequalities (2.1) become

(i) $\tilde{\sigma}_2^2 \geqq 0$	(i′) $s_{11} \geqq \tilde{\sigma}_1^2 + \tilde{\sigma}^2$
(ii) $\tilde{\sigma}_1^2 \geqq 0$	(ii′) $s_{22} \geqq \tilde{\sigma}_2^2 + \tilde{\sigma}^2$
(iii) $\tilde{\sigma}^2 \geqq 0$	(iii′) $s_{11} + s_{22} - 2s_{12} \geqq \tilde{\sigma}_1^2 + \tilde{\sigma}_2^2$

At least one inequality of each pair must be an equality. It is clear from (4.1) what the variance component estimates should be if S is not positive definite. Table 6 lists the various possibilities. Two consequences of the positive semi-definite nature of S are worthy of mention. First, Table 6 covers all possibilities and second, $s_{11} + s_{22} - 2s_{12}$ is non-negative. This follows from the relation $\frac{1}{4}(s_{11} + s_{22})^2 \geqq s_{11}s_{22} \geqq s_{12}^2$. The top line of the table agrees with the estimates of Grubbs [6]. The further entries show how Grubbs' estimates should be modified under various circumstances.

TABLE 6

Conditions	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$	$\tilde{\sigma}^2$
$s_{11} > s_{12}$, $s_{22} > s_{12} > 0$	$s_{11} - s_{12}$	$s_{22} - s_{12}$	s_{12}
$s_{22} > s_{12} \geqq s_{11}$	0	$s_{11} + s_{22} - 2s_{12}$	s_{11}
$s_{11} > s_{12} \geqq s_{22}$	$s_{11} + s_{22} - 2s_{12}$	0	s_{22}
$s_{12} \leqq 0$	s_{11}	s_{22}	0
$s_{11} = s_{22} = s_{12}$	0	0	s_{11}
$s_{12} = s_{11} = 0$	0	s_{22}	0
$s_{12} = s_{22} = 0$	s_{11}	0	0

5. Trees and mathematical programming. The remaining mathematical portions of this paper develop some theorems in non-linear programming which yield rules for pooling the mean squares when negative estimates of variance components occur. "A linear graph is defined informally as a collection of points, and a collection of lines or pairs of points which describes the connections of points; $\cdots$.

A cycle of a graph is a collection of lines of the form: $t_1t_2, t_2t_3, \cdots, t_{k-1}t_1$, where t_it_j designates the line joining points t_i and t_j, all points in the collection save t_1 are distinct. A graph is connected if every pair of points is joined by a path, that is, a collection of lines of the form $t_1t_2, t_2t_3, \cdots, t_{k-1}t_k$, with all points t_1 to t_k distinct $\cdots$.

A connected linear graph without cycles (or lines in parallel or "slings") is a tree. This mathematical object has a closer affinity to a family tree than to the growing varieties. A tree with one point, the root, distinguished from all other points by this very fact, is called a rooted tree" [11, pp. 109–110].

The root of a rooted tree is denoted by r, and $\overline{ts}$ is the (unique) path connecting the points t and s. It will be convenient to write $t = s$ if t and s are two labels for the same point. If t and s are distinct points, but $\overline{tr}$ is coincident with s then s will be said to be $< t$ or $t > s$. Note that not all points are ordered by this relationship. The predecessor of t denoted by t^{-1} is the (unique) point on $\overline{rt}$ for which

(a) $t^{-1} < t$

(b) $s \leqq t^{-1}$ for all $s < t$.

In a perfectly natural manner, $(t^{-1})^{-1} = t^{-2}$ and $t^{-(k+1)} = (t^{-k})^{-1}$ are defined inductively.

Let T denote a rooted tree with points $t_1, \cdots, t_p$. Several kinds of subsets of T will be found useful.

Let U be an arbitrary subset of T and define

$$(5.1) \qquad B(t) = (s \mid s \geqq t), \qquad \bar{B}(t) = (s \mid s \text{ is not in } B(t)),$$

$$(5.2) \qquad S(t, U) = (u \in U \mid t < u \text{ but for no } \bar{u} \text{ in } U \text{ is } t < \bar{u} < u),$$

$$(5.3) \qquad E(t, U) = B(t) \bigcap_{u \in S(t,U)} \bar{B}(u).$$

B, S and E stand for branch, successor and equivalence respectively. Note that $E(t, U)$ depends on the set U under discussion, but $B(t)$ is independent of U. Also if t is an element of $E(s, U)$ then $E(t, U) = B(t) \cap E(s, U)$.

Let the ω vector of Section 2, be given by $\omega = [\omega(t_1), \cdots, \omega(t_p)]'$. Thus $\omega(t)$ is a scalar function defined over the points of T, $\omega(t_i)$ corresponding to the point t_i of T. The problem which is now to be solved is that of maximizing a differentiable function

$$(5.4) \qquad g(\omega) = g[\omega(t_1), \cdots \omega(t_p)]$$

subject to the constraints

$$(5.5) \qquad \omega(t) \geqq \omega(t^{-1}),$$

where $\omega(r^{-1})$ is defined to be equal to zero in order to simplify the notation.

Further let $G(\omega, t)$ be the partial derivative of g with respect to $\omega(t)$ evaluated at the p dimensional vector point, ω, then the gradient vector $G(\omega)$ of Section 2 is given by $G(\omega) = [G(\omega, t_1), \cdots, G(\omega, t_p)]'$.

THEOREM 3. *$g(\omega)$ has a relative maximum at $\omega = \tilde{\omega} \equiv [\tilde{\omega}(t_1), \cdots, \tilde{\omega}(t_p)]'$ subject to the constraints* (5.5) *only if for some subset of T, call it Q, $\tilde{\omega}$ satisfies*

$$\text{(a)} \qquad \tilde{\omega}(t) > \tilde{\omega}(t^{-1}), \qquad \sum_{s \varepsilon B(t,Q)} G(\tilde{\omega}, s) = 0, \qquad t \varepsilon Q;$$

$$\text{(5.6)} \quad \text{(b)} \qquad \sum_{s \varepsilon B(t,Q)} G(\tilde{\omega}, s) \leqq 0 \qquad t \not\varepsilon Q;$$

$$\text{(c)} \qquad \tilde{\omega}(t) = \tilde{\omega}(q) \quad \textit{whenever } q \varepsilon Q \textit{ and } t \varepsilon E(q, Q).$$

$\tilde{\omega}$ will be called a solution corresponding to the solution set Q.

PROOF. The matrix of coefficients in (5.5) is the matrix A^{-1} of Theorem 1. Hence the relations (2.2) become

$$\text{(5.7)} \qquad \begin{array}{lll} \text{either} & \tilde{\omega}(t) = \tilde{\omega}(t^{-1}) \text{ and } \sum_{B(t)} G(\tilde{\omega}, s) \leqq 0 \\ \text{or} & \tilde{\omega}(t) > \tilde{\omega}(t^{-1}) \text{ and } \sum_{B(t)} G(\tilde{\omega}, s) = 0 \end{array}$$

for each $t \varepsilon T$. Thus to prove Theorem 3 we need only show that the conditions (5.6) and (5.7) are equivalent. First assume that (5.7) holds and define $Q = \{t \mid \tilde{\omega}(t) > \tilde{\omega}(t^{-1})\}$.

$$B(t) = [B(t) \bigcap_{q\varepsilon S(t,Q)} \bar{B}(q)] \bigcup_{q\varepsilon S(t,Q)} B(q) = E(t, Q) \bigcup_{q\varepsilon S(t,Q)} B(q).$$

The sets on the right are disjoint and hence

$$\sum_{s\varepsilon B(t)} G(\tilde{\omega}, s) = \sum_{s\varepsilon B(t,Q)} G(\tilde{\omega}, s) + \sum_{q\varepsilon S(t,Q)} \sum_{B(q)} G(\tilde{\omega}, s).$$

From this last relationship it is clear that (5.6) (a) and (b) hold. If $t = q$ then $\tilde{\omega}(t) = \tilde{\omega}(q)$ trivially; if $t \varepsilon E(q, Q)$ but $t \neq q$ then according to (5.3) and (5.7) there is a chain $t, t^{-1}, \cdots, t^{-k} = q$ such that $\tilde{\omega}(t) = \tilde{\omega}(t^{-1}) = \cdots = \tilde{\omega}(t^{-k}) = \tilde{\omega}(q)$ and hence (5.6) (c) must hold. Now assume that (5.6) is valid.

$$B(t) = E(t, Q) \bigcup_{\substack{q\varepsilon Q \\ q>t}} E(q, Q).$$

$$\sum_{s\varepsilon B(t)} G(\tilde{\omega}, s) = \sum_{s\varepsilon B(t,Q)} G(\tilde{\omega}, s) + \sum_{\substack{q\varepsilon Q \\ q>t}} \sum_{s\varepsilon B(q,Q)} G(\tilde{\omega}, s),$$

and (5.7) is easily established.

6. A quadratic programming problem. This section specializes the results of 5 by choosing for the function $g(\omega)$ of (5.4) the special form

$$\text{(6.1)} \qquad g(\omega) = g(w, \omega) = -\tfrac{1}{2} \sum_{j=1}^{p} f(t_j)[w(t_j) - \omega(t_j)]^2$$

where $f(t_j)$ and $w(t_j)$ are known positive constants associated with the point

t_j of T. The objective here, then, is to maximize (6.1) subject to the constraints (5.5).

In order to make the statement of the next theorem as tractable as possible it is desirable to introduce some terminology concerning finite weighted averages. With this in mind let D and E be subsets of T and define

$$f(D) = \sum_{t \varepsilon D} f(t), \tag{6.2}$$

$$w(D) = \sum_{t \varepsilon D} f(t) w(t) / f(D). \tag{6.3}$$

It is clear from this definition that if D and E are disjoint, then

$$w(D \cup E) = [f(D)w(D) + f(E)w(E)]/[f(D) + f(E)]. \tag{6.4}$$

THEOREM 4. *If $g(\omega)$ has the special form* (6.1), *then $\tilde{\omega}$ is a relative maximum point subject to the constraints* (5.5) *only if there exists a set Q such that*

$$\begin{array}{ll} \text{(a)} & \tilde{\omega}(t) = w[E(q, Q)] \quad \textit{whenever } q \varepsilon Q \textit{ and } t \varepsilon E(q, Q), \\ \text{(b)} & w[E(q, Q)] < w[E(\bar{q}, Q)] \quad \textit{whenever } q \varepsilon Q, \bar{q} \varepsilon Q \textit{ and } q < \bar{q}, \\ \text{(c)} & w[E(t, Q)] \leqq w[E(q, Q)] \quad \textit{whenever } q \varepsilon Q, \textit{ and } t \varepsilon E(q, Q). \end{array} \tag{6.5}$$

PROOF. For brevity write $E(q)$ for $E(q, Q)$ and $G(t)$ for $G(\omega, t)$.

$$G(t) = f(t)(w(t) - \omega(t)), \qquad t \varepsilon T$$

First assume that (5.6) holds.

$$0 = \sum_{E(q)} G(s) = \sum_{E(q)} f(s)(w(s) - \tilde{\omega}(q)).$$

Solving for $\tilde{\omega}(q)$

$$\tilde{\omega}(q) = \sum_{E(q)} f(s) w(s) / \sum_{E(q)} f(s) = w[E(q)]$$

and (6.5) (a) is necessary. To prove (6.5) (b) one merely needs to notice that there exists a chain $\bar{q}, \bar{q}^{-1}, \cdots \bar{q}^{-k} = q$ such that $w[E(\bar{q})] = \tilde{\omega}(\bar{q}) > \tilde{\omega}(\bar{q}^{-1}) \geqq \cdots \geqq \tilde{\omega}(\bar{q}^{-k}) = \tilde{\omega}(q) = w[E(q)]$. The proof that (6.5) (c) is necessary proceeds as follows. $0 \geqq \sum_{s \varepsilon E(t)} G(s) = \sum_{E(t)} f(s)\{w(s) - w[E(q)]\}$ and hence

$$w[E(q)] \geqq \sum_{E(t)} f(s) w(s) / \sum_{E(t)} f(s) = w[E(t)].$$

That the conditions (5.6) are true whenever (6.5) is valid, may be seen by noting that the steps are reversible in the above argument.

THEOREM 5. *If $g(\omega)$ has the special form* (6.1), *then for given $w(t_1), \cdots, w(t_p)$ the relations* (6.5) *have a unique solution, $\tilde{\omega}(t_1), \cdots, \tilde{\omega}(t_p)$. This solution is the absolute maximum point of* (6.1) *subject to the constraints* (5.5).

PROOF. By calculating the second partial derivatives, (6.1) may be shown to be a strictly concave function. The theorem then follows from Theorem 4 and the corollary to Theorem 2.

The remainder of this section establishes an algorithm for computing the solution $\tilde{\omega}(t_1), \cdots, \tilde{\omega}(t_p)$ of Theorem 5.

LEMMA 1. *If Q is the solution set defined in Theorem* 4, then $w[E(q, Q)] \leqq w(q)$ *whenever $q \varepsilon Q$.*

PROOF.

$$E(q) = \{q\} \bigcup_{\substack{t\varepsilon S(q,T) \\ t\notin Q}} E(t)$$

where the role of Q is suppressed.

$$w[E(q)] = \{f(q)w(q) + \sum_{\substack{t\varepsilon S(q,T) \\ t\notin Q}} f[E(t)]w[E(t)]\}/f[E(q)]$$

$$\leqq \{f(q)w(q) + w[E(q)][f[E(q)] - f(q)]\}/f[E(q)].$$

Hence the lemma.

THEOREM 6. *Let $w(t^*) \leqq w(t)$ for all $t \varepsilon T$. If $t^* = r$, then the solution, $\tilde{\omega}$, of Theorems 4 and 5 lies on the plane $\omega(r) = w(r)$. If $t^* > r$, then $\tilde{\omega}$ lies on the plane $\omega(t^*) = \omega(t^{*-1})$.*

PROOF. First consider the case $t^* = r$, r is always $\varepsilon\ Q$ so that $\tilde{\omega}(r) = w[E(r, Q)] \geqq w(r)$. If $\tilde{\omega}(r) \neq w(r)$ then $w[E(r, Q)] > w(r)$. This however, contradicts Lemma 1. Next consider the case $t^* > r$. Assume for purposes of contradiction, that $w(t^*) \leqq w(t)$ for all $t \varepsilon T$, but $\tilde{\omega}(t^*) \neq \tilde{\omega}(t^{*-1})$. Thus $\tilde{\omega}(t^*) > \tilde{\omega}(t^{*-1})$ or $t^* \varepsilon Q$. Since $t^* \neq r$ there exists $t \varepsilon T$ such that $t^* \varepsilon S(t, Q)$ and $w[E(t)] < w[E(t^*)]$. But according to Lemma 1, $w[E(t^*)] \leqq w(t^*)$. Therefore $w[E(t)] < w(t^*)$ contradicting the minimum property of $w(t^*)$.

THEOREM 7. *If g has the special form* (6.1), *H is a hyperplane containing the arbitrary set K, and if $\omega_h \varepsilon H$ is such that $g(w, \omega_h) \geqq g(w, \omega)$ for all $\omega \varepsilon H$, then $g(w, \omega_0) \geqq g(w, \omega)$ for all $\omega \varepsilon K$, if and only if $g(\omega_h, \omega_0) \geqq g(\omega_h, \omega)$, for all $\omega \varepsilon K$.*

PROOF. The truth of the theorem will be apparent if it is established that $g(w, \omega) = g(w, \omega_h) + g(\omega_h, \omega)$ for all $\omega \varepsilon H$. In this proof $f(t_j)$, $w(t_j)$ and $\omega(t_j)$ are written as f_j, w_j and ω_j respectively. If the equation of H is

$$\sum_{i=1}^{p} h_i\omega_i = d, \tag{6.6}$$

then ω_h is the value of ω which minimizes

$$\sum_{i=1}^{p} f_i(w_i - \omega_i)^2$$

subject to the constraint $\sum_{i=1}^{p} h_i\omega_i = d$. By the method of Lagrange multipliers it can be established that

$$f_j(w_j - \omega_{hj}) = \lambda h_j \tag{6.7}$$

where λ is a constant of proportionality independent of j. Hence if $\omega \varepsilon H$ then

$$\sum f_i(w_i - \omega_{hi})(\omega_{hi} - \omega_i) = \lambda \sum h_i(\omega_{hi} - \omega_i) = \lambda(d - d) = 0,$$
$$\sum f_i(w_i - \omega_i)^2 = \sum f_i(w_i - \omega_{hi})^2 + \sum f_i(\omega_{hi} - \omega_i)^2$$

and finally $g(w, \omega) = g(w, \omega_h) + g(\omega_h, \omega)$.

Before Theorem 7 can be used in a particular case, it will be necessary to determine an expression for ω_h. This may be done by carrying on from equations (6.6) and (6.7).

$$\omega_{hj} = w_j - \lambda h_j/f_j. \tag{6.8}$$
$$\sum_{i=1}^{p} h_i\omega_{hi} = \sum_{i=1}^{p} h_iw_i - \lambda \sum_{i=1}^{p} h_i^2/f_i = d$$

since $\omega_h \varepsilon H$, and hence

$$\lambda = \left(\sum_{i=1}^{p} h_iw_i - d\right)/\left(\sum_{i=1}^{p} h_i^2/f_i\right). \tag{6.9}$$

In order to discuss the way in which Theorems 6 and 7 may be used to establish an algorithm for computing $\tilde{\omega}$, first consider the case $t^* = r$. Let ω_0 be the maximum of (6.1) subject to $\omega(t) \geqq \omega(t^{-1})$ for all t such that $t^{-1} > r$. Writing

$$g(w, \omega) = -\tfrac{1}{2}f(r)[w(r) - \omega(r)]^2 - \tfrac{1}{2} \sum_{t\varepsilon S(r,T)} \sum_{B(t)} f(t)[w(t) - \omega(t)]^2,$$

it is clear that $\omega_0(r) = w(r)$ and for $t > r$, $\omega_0(t)$ is to be found by several applications of Theorem 4. For $t > r$, $\omega_0(t)$ must be a weighted average of the $w(t)$'s (again $t > r$) and hence since $w(r) \leqq w(t)$ for all $t \varepsilon T$ (i.e., $t^* = r$) then $w(r) = \omega_0(r) \leqq \omega_0(t)$ and the vector ω_0 satisfies the relations (5.5). In this case, ω_0 maximizes (6.1) subject to (5.5) and $\omega_0 = \tilde{\omega}$.

Next consider the case $t^* > r$, according to Theorem 6, $\tilde{\omega}$ lies on the plane, H say, whose equation is $\omega(t^*) = \omega(t^{*-1})$. Defining $K = [\omega \mid \omega(t^*) = \omega(t^{*-1})$ and $\omega(t) \leqq \omega(t^{-1})]$ and using Theorem 7, maximizing (6.1) in the admissible region (5.5) is the same as maximizing $g(\omega_h, \omega)$ with respect to $\omega \varepsilon K$. According to equations (6.6), (6.8) and (6.9), $\omega_h(t^*) = \omega_h(t^{*-1}) = w(t^*, t^{*-1})$ and $\omega_h(t) = w(t)$ for $t \neq t^*$ and $t \neq t^{*-1}$. Hence

$$g(\omega_h, \omega) = -\tfrac{1}{2} \sum_{\substack{t \neq t^* \text{ and} \\ t \neq t^{*-1}}} f(t)[w(t) - \omega(t)]^2 - \tfrac{1}{2}[f(t^*) + f(t^{*-1})][w(t^*, t^{*-1}) - \omega(t^*)]^2. \tag{6.10}$$

Maximizing (6.10) over K is the same problem as the one with which we started but in one less variable.

Summarizing, $\tilde{\omega}$ may be found by applying the following algorithm: Let $T_1 = T$ and $\omega(t, T_1) = w(t)$. Define T_{i+1} inductively from T_i as follows. Denote the predecessor of t in T_i by $P(t, T_i)$ and let $\omega(t^*, T_i)$ be the least $\omega(t, T_i)$ such that $\omega(t, T_i) < \omega[P(t, T_i), T_i]$. We may express this by saying that t^* is the minimum violator in T_i. t^* depends on i, but we suppress this dependence in the interest of simplifying the notation. Writing $P(t^*, T_i) = t'$ say, then the line t^*t' is deleted from T_i and a point will be said to be $>(<)t^*$ in T_{i+1} if it was $>(<)$ t^* or t' in T_i. All other points, lines and orderings are unchanged and the resulting linear graph is T_{i+1}. T_{i+1} is a tree or perhaps several disjoint trees. $f(t^*, T_{i+1}) = f(t^*, T_i) + f(t', T_i)$ and $\omega(t^*, T_{i+1})$ is defined to be

$$[f(t^*, T_i)\omega(t^*, T_i) + f(t', T_i)\omega(t', T_i)]/f(t^*, T_{i+1}).$$

All other f's and ω's in T_{i+1} are unchanged. Record for later use that $\tilde{\omega}(t^*) = \tilde{\omega}(t')$. If T_i has no violators and hence no minimum violator, then the induction is stopped at the ith stage and $\tilde{\omega}(t) = \omega(t, T_i)$ for all t in T_i. This process would seem to lend itself best to explanation by example and it is intended to write a less mathematical companion article in another journal which will provide several examples.

7. Trees and multiple classifications. Again let T denote a rooted tree with p points. Let $\sigma^2(t)$ be a function defined over T. $\omega(t)$ is then defined by

$$\omega(t) = \sum_{s \leq t} \sigma^2(s). \tag{7.1}$$

Solving these equations one obtains

$$\begin{aligned} \sigma^2(r) &= \omega(r) \\ \sigma^2(t) &= \omega(t) - \omega(t^{-1}), \qquad t \neq r. \end{aligned} \tag{7.2}$$

As in Section 3, if $w(t_1), \cdots, w(t_p)$ are independently distributed mean squares, and if the probability density of $w(t_i)$ is $h[\omega(t_i), f(t_i)]$, where the notation of (1.1) has been used, then twice the log of the density function is given by

$$g(\omega) = \text{const.} - \sum_{i=1}^{p} f(t_i)[\log \omega(t_i) + w(t_i)/\omega(t_i)]. \tag{7.3}$$

Thus to find maximum likelihood estimates of the σ^2's defined in (7.2), one must maximize (7.3) with respect to the constraints (5.5) to find $\tilde{\omega}(t)$ and then substitute in the equations (7.2) to find $\tilde{\sigma}^2(t)$. The theory of Section 5 applies and, in particular, Theorem 3 may be used to prove

THEOREM 8. *The problem of maximizing* (7.3) *subject to* (5.5) *is equivalent to the problem of maximizing* (6.1) *subject to* (5.5).

PROOF. Since (7.3) assumes an unrestricted maximum and the admissible region is closed then (7.3) assumes a restricted maximum. Further the necessary conditions (5.6) are the same for the two different problems and hence the conditions (6.5) of Theorem 4 are necessary that either function have a restricted maximum. But according to Theorem 5 the solution of (6.5) is unique and hence yields the restricted maximum of both (6.1) and (7.3).

8. Acknowledgments, and remarks of a general nature. The author's interest in the problems of this paper, was aroused by discussions with E. Vernon Lewis, Kenneth Horowitz, Paul Cox, and Frank E. Grubbs. Suggestions made by Robert M. Lauer, Russell Remage, Christian C. Braunschweiger, Willard E. Baxter, James R. Moore and the referees have been particularly beneficial.

A problem similar to that of Section 7 is considered by Van Eeden [5] and Brunk [3]. Their problem is more general in that they consider maximum likelihood estimation with respect to lattice type constraints. Hence, the algorithm which is developed by Van Eeden and Brunk would apply here as well; however, they do not consider the problem of negative estimates of variance components explicitly nor do they specialize their results to tree type constraints. For tree type constraints the algorithm of this paper is superior to that of Van Eeden and Brunk in that (in Brunk's notation) the order of choosing the sets A_i is specified thus reducing the maximum possible number of steps from $2^N - 1$ to $N - 1$. The algorithm of this paper is of value in the Van Eeden, Brunk problem when their constraints are of the tree type but is not extendable to the more general case of lattice constraints.

In some cases, where the pooling is always in the same direction, it is clear that the restricted maximum likelihood estimates are biased but for other components, perhaps in the same experiment, the pooling sometimes increases the estimate and sometimes decreases it so that the presence of any bias is difficult to determine. Some designs, for example the three factor experiment [9, p. 346], do not yield expected mean square columns which form a tree. If the various mean squares are independent we may use general quadratic programming techniques to maximize (6.1) subject to the appropriate constraints. The resulting estimates will no longer be restricted maximum likelihood, but will have a corresponding least square property. Further, in the case of tree type constraints this procedure reduces to the method of this paper and hence may be considered as an extension of it.

REFERENCES

[1] R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*, McGraw Hill, New York, 1952.

[2] F. J. Anscombe, Contributions to the discussion on D. G. Champernowne's, "Sampling theory applied to autoregressive sequences," *J. Roy. Stat. Soc.*, Ser. B, Vol. 10 (1948), p. 239.

[3] H. D. Brunk, "On the estimation of parameters restricted by inequalities," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 437–455.

[4] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[5] Constance Van Eeden, *Testing and Estimating Ordered Parameters of Probability Distributions*, University of Amsterdam, Amsterdam, Holland, 1958.

[6] Frank E. Grubbs, "On estimating precision of measuring instruments and product variability," *J. Amer. Stat. Assn.*, Vol. 43 (1948), pp. 243–264.

[7] Leon H. Herbach, "Properties of model II—type analysis of variance tests," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 939–959.

[8] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 481–492.

[9] ALEXANDER MCFARLANE MOOD, *Introduction to the Theory of Statistics*, McGraw Hill, New York, 1950.
[10] J. A. NELDER, "The interpretation of negative components of variance," *Biometrika*, Vol. 41 (1954).
[11] JOHN RIORDAN, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.
[12] THOMAS S. RUSSELL AND RALPH ALLAN BRADLEY, "One-way variances in a two-way classification," *Biometrika*, Vol. 45 (1958), pp. 111–129.
[13] HENRY SCHEFFÉ, "Alternative models for the analysis of variance," *Ann. Math.Stat.*, Vol. 27 (1956), pp. 251–271.
[14] HENRY SCHEFFÉ, *The Analysis of Variance*, Wiley, New York, 1959.
[15] L. E. SIMON, "On the relation of instrumentation to quality control," *Instruments*, Vol. 19 (1946), pp. 654–656.