# THE LINEAR HYPOTHESIS AND LARGE SAMPLE THEORY

By G. A. F. Seber[1]

*University of Manchester*

**1. Introduction.** In a paper by Wald [7] we have the beginning of a series of papers on the testing of statistical hypotheses about unknown population parameters using large sample, maximum likelihood theory. We are given $n$ independent observations $x_1, x_2, \cdots, x_n$ (these may be vectors) from a common underlying probability density function $f(x, \boldsymbol{\theta})$ where $\boldsymbol{\theta}' = (\theta_1, \theta_2, \cdots, \theta_s)$ and is unknown save for the fact that it is known to belong to $\Omega$ a subset of $s$-disional Euclidean space $R^s$, and we wish to test whether $\boldsymbol{\theta}_0$, the true value of $\boldsymbol{\theta}$, belongs to $\omega$ an $s - r$ dimensional subspace of $\Omega$.

There are two ways of specifying $\omega$: either in the form of constraint equations $h_1(\boldsymbol{\theta}) = h_2(\boldsymbol{\theta}) = \cdots = h_r(\boldsymbol{\theta}) = 0$, or in the form of freedom equations $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \cdots, \alpha_{s-r})$, or perhaps by a combination of both constraint and freedom equations. Although to any freedom equation specification there will correspond a constraint equation specification, this relationship is often difficult to derive in practice and therefore the two forms of specification are usually dealt with separately. These alternative methods of specification have lead to the formulation of three methods of testing $\omega$: the Wald test (Wald [7]), the Lagrange Multiplier test (Rao [4] and Silvey [6]) and the likelihood ratio test, all of which are described fully in Aitchison and Silvey [2] and Aitchison [1]. The choice of which method to use depends largely on the ease of computation of the test statistic and therefore to some extent on the method of specification of $\omega$.

In [1], Aitchison considers the problem of testing more than one hypothesis and introduces the notion of "separability" which is analogous to the idea of orthogonality in linear hypothesis theory (c.f. Darroch and Silvey [3]). He defines two hypotheses $\omega_1$ and $\omega_2$ to be separable with respect to $\Omega$ if and only if the critical region for the test of $\omega_i \cap \omega_j$ against $\omega_j - \omega_i \cap \omega_j$ $(i, j = (1, 2), (2, 1))$ is the same as the critical region for the test of $\omega_i$ against $\Omega - \omega_i$ and derives sufficient conditions for two hypotheses to be separable when the hypotheses are defined by constraint equations. However Darroch and Silvey [3] indicate briefly that, under reasonable assumptions, all large sample problems can be interpreted as linear problems with known variance matrix of residuals and therefore sufficient conditions for separability are given by sufficient conditions for orthogonality in the corresponding linear model.

The aim of this paper is to demonstrate in detail this relationship between large sample maximum likelihood theory and linear theory and hence find sufficient conditions for the separability of two hypotheses where the hypotheses

Received 29 July 1963.

[1] Now at London School of Economics.

are defined by freedom equations only. Thus if

$$\omega_i = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \varepsilon \Omega \text{ and } \boldsymbol{\theta} = \boldsymbol{\theta}_i(\beta_i)\}$$

where $\boldsymbol{\theta}_i$ is a function of an $s - r_i$ dimensional vector $\boldsymbol{\beta}_i$, then we shall derive sufficient conditions for $\omega_1$ and $\omega_2$ to be separable.

The main results will be stated as theorems without proofs in Sections 2 and 3 and the proofs will be given briefly in Section 4.

**2. Linear model approximation.** Let $\omega$ be defined by the following freedom equation specification,

$$\omega = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \varepsilon \Omega \text{ and } \boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\alpha})\}$$

where $\boldsymbol{\alpha}$ is the $s - r$ dimensional vector defined above. In this section we shall show that for large sample size $n$, $\omega$ may be approximated by a linear hypothesis $\omega^*$. We shall require the following notation.

Let $\Theta_\alpha$ be the $s - r \times s$ matrix of rank $s - r$ with $i, j$th element $\partial\theta_j/\partial\alpha_i$.

Let $L_n(\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta})$ be the likelihood function and let $\mathbf{D}_\theta \log L_n(\boldsymbol{\theta})$, $\mathbf{D}_\alpha \log L_n(\boldsymbol{\theta})$ be the column vectors whose $i$th elements are $\partial \log L_n(\boldsymbol{\theta})/\partial\theta_i$ and $\partial \log L_n(\boldsymbol{\theta}(\boldsymbol{\alpha}))/\partial\alpha_i$ respectively. We denote the information matrix for $\boldsymbol{\theta}$ by $B_\theta$ i.e., its $i, j$th element is

$$n^{-1}E_\theta[\partial^2 \log L(\boldsymbol{\theta})/\partial\theta_i\partial\theta_j].$$

To simplify notation we use $[\cdot]_\alpha$ to denote that the matrix in square brackets is evaluated at $\boldsymbol{\alpha}$; thus $[\Theta B_\theta \Theta']_\alpha = \Theta_\alpha B_{\theta(\alpha)} \Theta'_\alpha$.

We have two cases to consider as $B_{\theta_0}$ may be positive definite or positive semidefinite and of rank $s - r_0$ say. In this latter case $\boldsymbol{\theta}_0$ is non-identifiable and we introduce $r_0$ independent identifiability constraints

$$\mathbf{h}_0'(\boldsymbol{\theta}) = (h_{01}(\boldsymbol{\theta}) = h_{02}(\boldsymbol{\theta}) = \cdots = h_{0r_0}(\boldsymbol{\theta})) = \mathbf{0}'.$$

If $H_{0\theta_0}$ is the $s \times r_0$ matrix of rank $r_0$ with $i, j$th element $\partial h_{0j}(\boldsymbol{\theta})/\partial\theta_i$, then we assume that $[B + H_0H_0']_{\theta_0}$ is positive definite. (We shall see later that this assumption follows naturally from linear hypothesis theory).

As we shall use maximum likelihood theory in what follows we shall require certain underlying assumptions to hold and these are provided by Silvey [6]. The maximum likelihood estimates of $\boldsymbol{\theta}$ for $\boldsymbol{\theta}$ belonging to $\omega$ and $\Omega$ are denoted by $\hat{\boldsymbol{\theta}}_n$ and $\dot{\boldsymbol{\theta}}_n$ respectively. If $\boldsymbol{\theta}^*$ is the value of $\boldsymbol{\theta}$ which maximises $E_\theta[\log f(x, \boldsymbol{\theta}]$ for $\boldsymbol{\theta} \varepsilon \omega$, and if certain regularity conditions on $\Omega$, $f$ and $\boldsymbol{\theta}(\boldsymbol{\alpha})$ are satisfied (c.f. Assumptions 1–6 of [6]), then $\hat{\boldsymbol{\theta}}_n$ and $\dot{\boldsymbol{\theta}}_n$ will converge with probability one to $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}_0$ respectively. In addition Silvey points out that if $\boldsymbol{\theta}_0$ is not "near" $\omega$ then the powers of the three tests will be asymptotically near one. Thus the situation of interest is when $\boldsymbol{\theta}_0$ is near $\omega$ i.e. when $\boldsymbol{\theta}_0$ is near $\boldsymbol{\theta}^*$. (For rigour we could define what we mean by nearness by choosing $\boldsymbol{\theta}_0$ close enough to $\omega$ such that $\|(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)\| = O(n^{-\frac{1}{2}})$. In what follows we assume that Assumptions 1–10, 11B and 12 of [6] holds (suitably interpreted of course in terms of a freedom equation specification), and for convenience, the dependence on the sample size $n$ is dropped from the notation. We shall now give the linear model approximation to $\omega$.

CASE 1. Let $B_{\theta_0}$ be positive definite, i.e. there exists a non-singular $s \times s$ matrix $V_{\theta_0}$ such that $B_{\theta_0} = [VV']_{\theta_0}$.

Let $\boldsymbol{\alpha}^*$ and $\hat{\boldsymbol{\alpha}}$ be defined by $\boldsymbol{\theta}^* = \boldsymbol{\theta}(\boldsymbol{\alpha}^*)$ and $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})$. Then assuming $\boldsymbol{\theta}_0$ to be near $\boldsymbol{\theta}^*$ we have the following theorem.

THEOREM 1. *The hypothesis that $\boldsymbol{\theta}_0$ belongs to $\omega$ given that $\boldsymbol{\theta}_0$ is in $\Omega$ is asymptotically equivalent to the linear hypothesis* $\mathbf{y} = \boldsymbol{\psi} + \boldsymbol{\varepsilon}$ where $\mathbf{y}$, $\boldsymbol{\psi}$, $\boldsymbol{\varepsilon}$ are $s$ dimensional vectors, $\boldsymbol{\varepsilon}$ is a random vector distributed as $N(\mathbf{0}, I_s)$, $I_s$ being the $s \times s$ unit matrix,

$$\Omega^* = \{\boldsymbol{\psi} \mid \boldsymbol{\psi} = V'_{\theta_0}\boldsymbol{\beta}\} = R^s$$

and

$$\omega^* = \{\boldsymbol{\psi} \mid \boldsymbol{\psi} = V'_{\theta_0}\Theta_{\alpha^*}\boldsymbol{\gamma}\}.$$

CASE 2. If $B_{\theta_0}$ is positive semi-definite, then there exists an $s \times s - r_0$ matrix $W_{\theta_0}$ of rank $s - r_0$ such that $B_{\theta_0} = [WW']_{\theta_0}$. Theorem 1 is now modified as follows.

THEOREM 2. *If $[B + H_0H_0']_{\theta_0}$ is positive definite then the original hypothesis is asymptotically equivalent to the linear hypothesis* $\mathbf{y} = \boldsymbol{\phi} + \boldsymbol{\varepsilon}$ where $\mathbf{y}$, $\boldsymbol{\phi}$, $\boldsymbol{\varepsilon}$ are $s - r_0$ dimensional vectors, $\boldsymbol{\varepsilon}$ is distributed as $N(\mathbf{0}, I_{s-r_0})$,

$$\Omega^* = \{\boldsymbol{\phi} \mid \boldsymbol{\phi} = W'_{\theta_0}\boldsymbol{\beta}; H'_{0\theta_0}\boldsymbol{\beta} = \mathbf{0}\} = R^{s-r_0}$$

*and*

$$\omega^* = \{\boldsymbol{\phi} \mid \boldsymbol{\phi} = W'_{\theta_0}\Theta'_{\alpha^*}\boldsymbol{\gamma}; H'_{0\theta_0}\Theta'_{\alpha^*}\boldsymbol{\gamma} = \mathbf{0}\}.$$

REMARK. We note that the conditions $H'_{0\theta_0}\boldsymbol{\beta} = \mathbf{0}$ are necessary and sufficient for the identifiability of $\boldsymbol{\beta}$ if and only if $[W\vdots H_0]_{\theta_0}$ is of rank $s$ and the rank of $H_{0\theta_0}$ is $r_0$ (see Scheffé [5], p. 17) i.e., if and only if $[B + H_0H_0']_{\theta_0}$ is of rank $s$ and therefore positive definite. Thus the stipulation made in Section 2 that $[B + H_0H_0']_{\theta_0}$ should be positive definite follows from the identifiability of the corresponding linear model. This implies that since $\boldsymbol{\beta}$ is of greater dimension than $\boldsymbol{\phi}$, $\Omega^* = \{\boldsymbol{\phi}\} = R^{s-r_0}$ which we would expect intuitively.

**3. Sufficient conditions for separability.** Suppose we wish to test two hypotheses

$$\omega_i = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \,\varepsilon\, \Omega, \boldsymbol{\theta} = \boldsymbol{\theta}_i(\boldsymbol{\alpha}_i)\}, \qquad i = 1, 2,$$

where $\boldsymbol{\theta}_i$ is a function of an $s - r_i$ dimensional vector $\boldsymbol{\alpha}_i$. We assume that $B_{\theta_0}$ is positive semidefinite of rank $s - r_0$ as in Section 2. If $\boldsymbol{\theta}_0$ is near $\omega_1 \cap \omega_2$ then, dropping the dependence on $\boldsymbol{\theta}_0$ and $\boldsymbol{\alpha}_i^*$ from the notation and letting $\Theta_1$ and $\Theta_2$ denote the corresponding matrices of derivatives, we have from Theorem 2 that the two hypotheses are asymptotically equivalent to the linear hypotheses

$$\omega_i^* = \{\boldsymbol{\phi} \mid \boldsymbol{\phi} = W'\Theta_i'\boldsymbol{\gamma}; H_0'\Theta_i'\boldsymbol{\gamma} = \mathbf{0}\}.$$

Then $\omega_1$ and $\omega_2$ are separable if and only if $\omega_1^*$ and $\omega_2^*$ are separable. If the tests of $\omega_i^* \cap \omega_j^*$ against $\omega_j^* - \omega_i^* \cap \omega_j^*$ $(i, j = (1, 2), (2, 1))$ and $\omega_i^*$ against $\Omega^* - \omega_i^*$ not only have the same critical region but are identical, then the linear hypotheses $\omega_1^*$ and $\omega_2^*$ are said to be orthogonal. From Darroch and Silvey [3] we have that $\omega_1^*$ and $\omega_2^*$ are orthogonal with respect to $\Omega^*$ if and only if the orthogonal complements of $\omega_1^*$ and $\omega_2^*$ with respect to $\Omega^*$ are mutually perpendicular i.e. if

$$[\omega_1^*]^\perp \cap \Omega^* \perp [\omega_2^*]^\perp \cap \Omega^*.$$

Thus this condition is *sufficient* for the separability of $\omega_1$ and $\omega_2$ (Aitchison [1]) and we have the following theorem.

THEOREM 3. *The hypotheses $\omega_1$ and $\omega_2$ are separable if*

$$\text{(3.1)} \quad \Big[B - \sum_{i=1}^{2} B\Theta_i'[\Theta_i(B + H_0H_0')\Theta_i']^{-1}\Theta_i B + B\Theta_1'[\Theta_1(B + H_0H_0')\Theta_1']^{-1}\Theta_1 B\Theta_2'[\Theta_2(B + H_0H_0')\Theta_2']^{-1}\Theta_2' B\Big]_\theta = 0$$

*for every $\boldsymbol{\theta}$ in $\omega_1 \cap \omega_2$.*

In (3.1) we interpret $[\Theta_i]_\theta$ as meaning $[\Theta_i]_{\alpha_i}$ where $\boldsymbol{\theta} = \boldsymbol{\theta}_i(\boldsymbol{\alpha}_i)$.

Slightly stronger conditions for separability are as follows:

THEOREM 4. *The hypotheses $\omega_1$ and $\omega_2$ are separable if* $[(I_s - \Theta_i'[\Theta_i\Theta_i']^{-1}\Theta_i)\cdot[B + H_0H_0']^{-1}H_0]_\theta = 0$ *for either $i = 1$ or 2 and*

$$[(I_s - \Theta_1'[\Theta_1\Theta_1']^{-1}\Theta_1)[B + H_0H_0']^{-1}(I_s - \Theta_2'[\Theta_2\Theta_2']^{-1}\Theta_2)]_\theta = 0$$

*for every $\boldsymbol{\theta} \in \omega_1 \cap \omega_2$.*

COROLLARY. *If $B_{\theta_0}$ is positive definite then sufficient conditions for separability are obtained from Theorems 3 and 4 by putting $H_0 = 0$.*

We observe that the above sufficient conditions are not as simple as those obtained when $\omega_1$ and $\omega_2$ are defined by constraint equations (c.f. Aitchison [1]) and therefore as far as separability is concerned, the constraint equation specification should be used where possible. However we shall demonstrate the use of Theorem 4 in a simple example based on Example 6 of [1].

EXAMPLE. Consider the model

$$f(x_1, x_2, x_3, x_4, \boldsymbol{\theta}) = n! \prod_{i=1}^{4} \{\theta_i^{x_i}/x_i!\}$$

where $h_0(\boldsymbol{\theta}) = \sum_{i=1}^4 \theta_i - 1 = 0$. Suppose we have the following subspaces

$$\Omega: \boldsymbol{\theta}_0' = (\theta_1, \theta_2, \theta_3, \theta_4); \qquad 0 < \theta_i < 1 \text{ for } i = 1, \cdots, 4,$$

$$\omega_1: \theta_1 = (1 - \alpha_1)^2, \quad \theta_2 = \alpha_1^2, \quad \theta_3 = \beta_1, \quad \theta_4 = \gamma_1,$$

and

$$\omega_2: \theta_1 = \alpha_2, \quad \theta_2 = \alpha_2, \quad \theta_3 = \beta_2, \quad \theta_4 = \gamma_2.$$

We wish to know whether $\omega_1$ and $\omega_2$ are separable.

To apply the above theorems we first of all define $\omega_1 \cap \omega_2$. For a point $\boldsymbol{\theta}$ in $\omega_1 \cap \omega_2$ we have $(1 - \alpha_1)^2 = \alpha_2$, $\alpha_1^2 = \alpha_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$ i.e. $\alpha_1 = \frac{1}{2}$ and $\alpha_2 = \frac{1}{4}$. As $H_{0\theta_0} = [1\ 1\ 1\ 1]$, $[B + H_0H_0']_\theta^{-1} = \text{diag}\,[\theta_1, \theta_2, \theta_3, \theta_4]$ and this evaluated for $\boldsymbol{\theta}\ \varepsilon\ \omega_1 \cap \omega_2$ takes the form diag $[\frac{1}{4}, \frac{1}{4}, \theta_3, \theta_4]$. The matrices $\Theta_1$ and $\Theta_2$ for $\boldsymbol{\theta}\ \varepsilon\ \omega_1 \cap \omega_2$ are given by

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

respectively. By direct matrix calculation we find that the conditions of Theorem 4 are satisfied for $i = 2$ and $\omega_1$ and $\omega_2$ are therefore separable.

**4. Proofs of theorems.** In this section we shall give brief proofs of Theorems 1–4.

PROOF OF THEOREM 1. The maximum likelihood equation for $\dot{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ are given by

$$n^{-1}\mathbf{D}_\theta \log L(\dot{\boldsymbol{\theta}}) = \mathbf{0} \tag{4.1}$$

and

$$n^{-1}\mathbf{D}_\alpha \log L(\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})) = \mathbf{0}. \tag{4.2}$$

As $\boldsymbol{\theta}^*$ is near $\boldsymbol{\theta}_0$ and therefore near $\dot{\boldsymbol{\theta}}$ we have from (4.1) using a Taylor expansion that

$$\mathbf{m} - B_{\theta_0}n^{\frac{1}{2}}(\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + B_{\theta_0}n^{\frac{1}{2}}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*) + \mathbf{o}_p(1) = \mathbf{0}$$

where $\mathbf{m} = n^{-\frac{1}{2}}\mathbf{D}_\theta \log L(\boldsymbol{\theta}_0)$ is distributed as $N(\boldsymbol{\theta}, B_{\theta_0})$ (by assumptions 7–10 of [6]) and $\mathbf{o}_p(1)$ means "$o(1)$ with probability one for each element of the vector." As $\hat{\boldsymbol{\theta}}$ is near $\boldsymbol{\theta}^*$ and therefore also near $\dot{\boldsymbol{\theta}}$ we have a similar equation from (4.2) namely

$$\begin{aligned} \mathbf{0} &= n^{\frac{1}{2}}n^{-1}\Theta_{\hat{\alpha}}\mathbf{D}_\theta \log L(\hat{\boldsymbol{\theta}}), \\ &= n^{\frac{1}{2}}n^{-1}\Theta_{\alpha^*}\mathbf{D}_\theta \log L(\hat{\boldsymbol{\theta}}) + \mathbf{o}_p(1), \\ &= -n^{\frac{1}{2}}\Theta_{\alpha^*}B_{\theta_0}[\hat{\boldsymbol{\theta}} - \dot{\boldsymbol{\theta}}] + \mathbf{o}_p(1). \end{aligned}$$

Thus by putting $\mathbf{y} = n^{\frac{1}{2}}V_{\theta_0}'(\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$, $\boldsymbol{\phi} = n^{\frac{1}{2}}V_{\theta_0}'(\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)$ and $\hat{\boldsymbol{\phi}} = n^{\frac{1}{2}}V_{\theta_0}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ the theorem is proved.

We note that as $V_{\theta_0}'$ is non-singular its range space is just $R^s$ i.e. $\Omega^* = R^s$.

PROOF OF THEOREM 2. For $B_{\theta_0}$ positive semidefinite, the estimates $\dot{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ are given by

$$n^{-1}\mathbf{D}_\theta \log L(\dot{\boldsymbol{\theta}}) + H_{0\dot{\theta}}\dot{\boldsymbol{\lambda}} = \mathbf{0}, \qquad \mathbf{h}_0(\dot{\boldsymbol{\theta}}) = \mathbf{0} \tag{4.3}$$

$$n^{-1}\mathbf{D}_\alpha \log L(\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})) + \Theta_{\hat{\alpha}}H_{0\hat{\theta}}\hat{\boldsymbol{\lambda}} = \mathbf{0}, \qquad \mathbf{h}_0(\boldsymbol{\theta}(\hat{\boldsymbol{\alpha}})) = \mathbf{0} \tag{4.4}$$

where $\dot{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}$ are the appropriate Lagrange multipliers. As $\boldsymbol{\theta}_0$ is near $\boldsymbol{\theta}^*$ we can express the Equations (4.3) as a system of matrix equations (cf Silvey [6]), solve for $\dot{\boldsymbol{\lambda}}$ and thus prove the following lemmas.

LEMMA 1. *If* $[B + H_0H_0']_{\theta_0}$ *is positive definite then* $n^{\frac{1}{2}}\dot{\boldsymbol{\lambda}} = \mathbf{o}_p(1)$.

LEMMA 2. *If* $\Theta_\alpha$ *is of rank* $s - r$ *then* $n^{\frac{1}{2}}\hat{\boldsymbol{\lambda}} = \mathbf{o}_p(1)$.

We omit the proof of Lemma 1 as it is straightforward. Lemma 2 follows from Lemma 1 by treating the unknown parameter as $\boldsymbol{\alpha}$ instead of $\boldsymbol{\theta}$ and replacing $B_{\theta_0}$ by $\Theta_{\alpha^*}B_{\theta_0}\Theta_{\alpha^*}'$ and $H_{0\theta_0}$ by $\Theta_{\alpha^*}H_{0\theta_0}$.

Theorem 2 is now proved by using Lemmas 1 and 2, expanding the Equations (4.3) and (4.4) by Taylor's theorem as in Theorem 1 and putting $\mathbf{y} = n^{\frac{1}{2}}W_{\theta_0}'(\dot{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ etc.

PROOF OF THEOREM 3.

$$\begin{aligned}\omega_i^* &= \{\boldsymbol{\phi} \mid \boldsymbol{\phi} = W'\Theta_i'\boldsymbol{\gamma}_i\,;\, H_0'\Theta_i'\boldsymbol{\gamma}_i = \mathbf{0}\}\\ &= \{\boldsymbol{\phi} \mid (I_{s-r_0} - W'\Theta_i'[\Theta_i(B + H_0H_0')\Theta_i']^{-1}\Theta_iW)\boldsymbol{\phi} = \mathbf{0}\}\\ &= \{\boldsymbol{\phi} \mid M_i\boldsymbol{\phi} = \mathbf{0}\} \text{ say.}\end{aligned}$$

Since the null space of any matrix is the orthogonal complement of the range space of its transpose,

$$[\omega_1^*]^\perp \cap \Omega^* \perp [\omega_2^*]^\perp \cap \Omega^*$$

if and only if $M_1M_2' = 0$ or $WM_1M_2'W' = 0$ (since $W$ is of full rank). As the dependence on $\boldsymbol{\theta}$ is understood in the above equation and we are concerned only with critical regions, we have proved Theorem 3.

PROOF OF THEOREM 4. It can be shown that

$$\boldsymbol{\gamma}_i = [\Theta_i\Theta_i']^{-1}\Theta_i[B + H_0H_0']^{-1}W\boldsymbol{\phi}$$

and thus

$$\omega_i^* = \{\boldsymbol{\phi} \mid N_i\boldsymbol{\phi} = \mathbf{0}\}$$

where $N_i = I_{s-r_0} - W'\Theta_i'[\Theta_i\Theta_i']^{-1}\Theta_i[B + H_0H_0']^{-1}W$. As in the proof of Theorem 3, $\omega_1^*$ and $\omega_2^*$ are orthogonal with respect to $\Omega^*$ if and only if

$$WN_1N_2'W' = 0. \tag{4.5}$$

Now if $G = [W \vdots H_0]$, then $G$ is a regular $s \times s$ matrix and $G'[GG']^{-1}G = I_s$. Thus $W'[GG']^{-1}W = I_{s-r_0}$ and $H_0'[GG']^{-1}W = 0$. Using these relations in Equation (4.5) we complete the proof of Theorem 4.

REMARK. If we apply the method of Section 2 of approximating a non-linear problem by a linear one for two hypotheses expressed in constraint equation form, namely

$$\omega_i = \{\boldsymbol{\theta} \mid h_{1i}(\boldsymbol{\theta}) = \cdots = h_{r_i i}(\boldsymbol{\theta}) = \mathbf{0};\, \mathbf{h}_0(\boldsymbol{\theta})\}$$

then it can be shown that the corresponding linear hypotheses are

$$\omega_i^* = \{\boldsymbol{\phi} \mid [H_i'[B + H_0H_0']^{-1}W]_{\theta_0}\boldsymbol{\phi} = \mathbf{0}\}$$

where $H_{i\theta}$ is the matrix with $j$, $k$th element $\partial h_{ki}/\partial\theta_j$. Thus as in the proof of Theorem 3 we have that the condition

$$[H_1'[B + H_0H_0']^{-1}B[B + H_0H_0']^{-1}H_2]_\theta = 0$$

for every $\boldsymbol{\theta} \varepsilon \omega_1 \cap \omega_2$ is sufficient for the separability of $\omega_1$ and $\omega_2$. If either

$$(4.6) \quad [H_1'[B + H_0H_0']^{-1}H_0]_\theta = 0 \quad \text{or} \quad [H_2'[B + H_0H_0']^{-1}H_0]_\theta = 0$$

for every $\boldsymbol{\theta} \varepsilon \omega_1 \cap \omega_2$ then the above condition reduces to the stronger but simpler condition

$$(4.7) \qquad [H_1'[B + H_0H_0']^{-1}H_2]_\theta = 0$$

for every $\boldsymbol{\theta} \varepsilon \omega_1 \cap \omega_2$ which is derived in Aitchison [1] by a different method.

Turning our attention once again to the freedom equation specification we see that as the range space of $\Theta_i'$ is the same as the null space of $(I_s - \Theta_i'[\Theta_i\Theta_i']^{-1}\Theta_i)$, the sufficient conditions for separability in Theorem 4 are thus completely analogous to Conditions (4.6) and (4.7).

In conclusion, I would like to thank Dr. S. D. Silvey for suggesting the above problem to me and for many useful discussions on large sample theory. My thanks also go to the referee for helpful criticisms of a previous draft of this paper.

## REFERENCES

[1] Aitchison, J. (1962). Large sample restricted parametric tests. *J. Roy. Statist. Soc. Ser. B.* **24** 234–250.

[2] Aitchison, J. and Silvey, S. D. (1960). Maximum likelihood estimation procedures and associated tests of significance. *J. Roy. Statist. Soc. Ser. B.* **22** 154–171.

[3] Darroch, J. N. and Silvey, S. D. (1963). On testing one or more hypothesis. *Ann. Math. Statist.* **34** 555–567.

[4] Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proc. Cambridge Philos. Soc.* **44** 50–57.

[5] Scheffé, H. (1959). *The Analysis of Variance.* Wiley, New York.

[6] Silvey, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

[7] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.