THE MULTIVARIATE ONE-WAY CLASSIFICATION MODEL WITH RANDOM EFFECTS

BY

JAMES ROBERT SCHOTT

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1981

To Susan

and

My Parents

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to Dr. John Saw for suggesting this topic to me and for constantly providing guidance and assistance. He has made this project not only a very rewarding educational experience but also an enjoyable one. I also wish to thank Dr. Andre Khuri, Dr. Mark Yang, and Dr. Dennis Wackerly for their willingness to provide help when called upon.

Finally, a special thanks goes to Mrs. Edna Larrick who has turned a somewhat unreadable rough draft into a nicely typed manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Abstract of Dissertation Presented to the Graduate Council of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

THE MULTIVARIATE ONE-WAY CLASSIFICATION MODEL WITH RANDOM EFFECTS

By

James Robert Schott

August 1981

Chairman: Dr. John G. Saw
Major Department: Statistics

A well-known model in univariate statistical analysis is the one-way random effects model. In this paper we investigate the multivariate generalization of this model, that is, the multivariate one-way random effects model.

Two specific situations, regarding the structure of the variance-covariance matrix of the random error vectors, are considered. In the first and most general case, it is only assumed that this variance-covariance matrix is symmetric and positive definite. In the second case, it is assumed, in addition, that the variance-covariance matrix is a scalar multiple of the identity matrix.

Maximum likelihood estimates are obtained and the likelihood ratio test for a hypothesis test on the rank of the variance-covariance matrix of the random effect vectors is derived. Properties of the likelihood ratio test are investigated for the general case, while for the second case an alternative test is developed and its properties are investigated. In each case a sequential procedure for determining the rank of the variance-covariance matrix of the random effect vectors is presented.

CHAPTER 1

INTRODUCTION

1.1 The Random Effects Model, Scalar Case

Suppose a physician is considering administering some particular blood test to his patients as a part of their physical examination. He suspects that the test results vary with the presence and severity of a particular pathological condition. In order to examine variability in the results of the blood test, the physician chooses to administer the blood test n times to each of g patients. This results in the observations x_{ij}: $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$.

A suitable model to explain the different values of x_{ij}: $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$ would be

$$x_{ij} = \mu + \alpha_i + z_{ij}. \qquad (1.1.1)$$

Here μ is an overall mean, α_i is an effect due to the i^{th} patient, and z_{ij} represents a random error due to the measuring process. We assume that z_{ij}: $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$ are independent and have a normal distribution with mean zero and variance σ_z^2.

If the physician is interested in using the blood test as a diagnostic tool, he will certainly be interested to know whether a major source of variation in the results of the blood test is due to variation between the patients. Since the physician will administer the test to an unlimited number of patients in the future, we should properly regard the g patients involved as a sample from the entire population of patients. The patient effects, α_i: $i = 1,2,...,g$, now have the role of random variables, and (1.1.1) is a random effects model. Again we assume that α_i: $i = 1,2,...,g$ are independent and have a normal distribution with mean zero and variance σ_α^2. Thus, from our model (1.1.1) we deduce that x_{ij} has a normal distribution with mean μ and variance $\sigma_\alpha^2 + \sigma_z^2$.

The variation in the results of the blood test is governed by $\sigma_\alpha^2 + \sigma_z^2$. The portion of this attributable to the patients is, of course, $\sigma_\alpha^2/(\sigma_\alpha^2+\sigma_z^2)$, and the physician would like to know whether this or, correspondingly, $\sigma_\alpha^2/\sigma_z^2$ is sizeable. If $\sigma_\alpha^2/\sigma_z^2$ is sufficiently large, he would choose to investigate the possible use of this test as a means of detecting the pathological condition; otherwise he would find the blood test essentially useless as a diagnostic tool. Hence, the physician might be interested in testing the hypothesis H_o: $\sigma_\alpha^2 = 0$ against the hypothesis H_1: $\sigma_\alpha^2 > 0$.

In order to derive the likelihood ratio test for testing the hypothesis H_o against H_1, we first need to obtain the likelihood function of $(\mu, \sigma_z^2, \sigma_\alpha^2)$. This is most easily done by making a transformation. Let C be an orthogonal matrix, with the element in the i^{th} row and the j^{th} column denoted by c_{ij}, such that $c_{1j} = 1/\sqrt{n} : j = 1,2,\ldots,n$. Since C is orthogonal,

$$\sum_{j=1}^{n} c_{kj} = \sqrt{n} \sum_{j=1}^{n} c_{1j}\, c_{kj} = 0 \qquad \text{for } k = 2,3,\ldots,n. \tag{1.1.2}$$

Consider the orthogonal transformation

$$(y_{i1}, y_{i2}, \ldots, y_{in})' = C(x_{i1}, x_{i2}, \ldots, x_{in})'. \tag{1.1.3}$$

Upon replacing x_{ij} by the right side of (1.1.1) and using (1.1.2), we observe that

$$y_{i1} = \sum_{k=1}^{n} x_{ik}/\sqrt{n} = \sqrt{n}\, \bar{x}_{i.},$$

$$y_{ij} = \sum_{k=1}^{n} c_{jk}\, z_{ik} \qquad \text{for } j = 2,3,\ldots,n,$$

where $\bar{x}_{i.} = \sum_{k=1}^{n} x_{ik}/n$. Thus,

$$\mathrm{Cov}(y_{ij}, y_{ik}) = 0 \qquad \text{for } j \neq k,$$

$$V(y_{ij}) = \sigma_z^2 \qquad \text{for } j = 2,3,\ldots,n,$$

and $\{\bar{x}_{i.}, y_{i2}, y_{i3}, \ldots, y_{in}\}_{i=1}^{g}$ is a set of gn mutually independent random variables, where $\bar{x}_{i.}$ has a normal distribution with mean μ and variance $\sigma_z^2/n + \sigma_\alpha^2$, and y_{ij}

has a normal distribution with mean zero and variance σ_z^2. Note also from (1.1.3) that $\sum_{j=1}^{n} (x_{ij}-\bar{x}_{i.})^2 = \sum_{j=2}^{n} y_{ij}^2$ and denote this quantity by u_i. We can now write the joint density function of $y_{i2}, y_{i3}, \ldots, y_{in}$ as

$$f(y_{i2}, y_{i3}, \ldots, y_{in}; \sigma_z^2) = \prod_{j=2}^{n} (2\pi\sigma_z^2)^{-\frac{1}{2}} \exp[-y_{ij}^2/2\sigma_z^2]$$

$$= (2\pi\sigma_z^2)^{-\frac{1}{2}(n-1)} \exp[-\sum_{j=2}^{n} y_{ij}^2/2\sigma_z^2]$$

$$= g(\sum_{j=2}^{n} y_{ij}^2; \sigma_z^2) = g(u_i; \sigma_z^2),$$

so that from the set $\{\bar{x}_{i.}, y_{i2}, y_{i3}, \ldots, y_{in}\}$, $(\bar{x}_{i.}, u_i)$ is sufficient for (μ, σ_z^2). Thus, we may assume that we have, independently, u_i and $\bar{x}_{i.}$ for $i = 1,2,\ldots,n$, where u_i/σ_z^2 has a chi-square distribution with $\nu = n-1$ degrees of freedom, and $\bar{x}_{i.}$ has a normal distribution with mean μ and variance $\sigma_z^2/n + \sigma_\alpha^2$. Note that with

$$\bar{x}_{..} = \sum_{i=1}^{g} \bar{x}_{i.}/g,$$

$$\sum_{i=1}^{g} (\bar{x}_{i.}-\mu)^2 = \sum_{i=1}^{g} (\bar{x}_{i.}-\bar{x}_{..})^2 + g(\bar{x}_{..}-\mu)^2.$$

Then putting $\sigma^2 = \sigma_z^2/n+\sigma_\alpha^2$, we can write the joint density function of $\bar{x}_{1.}, \bar{x}_{2.}, \ldots, \bar{x}_{g.}$ as

$$f(\bar{x}_{1.},\bar{x}_{2.},\ldots,\bar{x}_{g.};\mu,\sigma^2) = \prod_{i=1}^{g} (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-(\bar{x}_{i.}-\mu)^2/2\sigma^2]$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}g} \exp[-\sum_{i=1}^{g} (\bar{x}_{i.}-\mu)^2/2\sigma^2]$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}g} \exp[-(\sum_{i=1}^{g} (\bar{x}_{i.}-\bar{x}_{..})^2 + g(\bar{x}_{..}-\mu)^2)/2\sigma^2]$$

$$= g(\bar{x}_{..},v;\mu,\sigma^2),$$

where $v = n\sum_{i=1}^{g}(\bar{x}_{i.}-\bar{x}_{..})^2$. Hence, from the set $\{\bar{x}_{1.},\bar{x}_{2.},\ldots,\bar{x}_{g.}\}$, $(\bar{x}_{..},v)$ is sufficient for $(\mu,\sigma_z^2/n + \sigma_\alpha^2)$.

Also, if we let c denote a constant, we can write the joint density function of $u_1,u_2,\ldots,u_g$ as

$$f(u_1,u_2,\ldots,u_g;\sigma_z^2) = \prod_{i=1}^{g} c \exp(-u_i/2\sigma_z^2)u_i^{\frac{1}{2}\nu-1}/(\sigma_z^2)^{\frac{1}{2}\nu}$$

$$= (\sigma_z^2)^{-\frac{1}{2}g\nu} \exp(-\sum_{i=1}^{g} u_i/2\sigma_z^2) \prod_{i=1}^{g} c\, u_i^{\frac{1}{2}\nu-1}$$

$$= g(u;\sigma_z^2)h(u_1,u_2,\ldots,u_g),$$

where $u = \sum_{i=1}^{g} u_i$. Thus, from the set $\{u_1,u_2,\ldots,u_g\}$, u is sufficient for σ_z^2.

We may now assume that we have, independently, $\bar{x}_{..}$, u, and v, where $\bar{x}_{..}$ has a normal distribution with mean μ and variance $(\sigma_z^2+n\sigma_\alpha^2)/gn$; u/σ_z^2 has a chi-square distribution with $e = g(n-1)$ degrees of freedom, and $v/(\sigma_z^2+n\sigma_\alpha^2)$ has a chi-square distribution with $h = g-1$ degrees of freedom. The likelihood function of $(\mu,\sigma_z^2,\sigma_\alpha^2)$ can be expressed as

$$f(\bar{x}_{..},u,v) = \frac{\exp[-(\bar{x}_{..}-\mu)^2 gn/2(\sigma_z^2+n\sigma_\alpha^2)]}{(2\pi(\sigma_z^2+n\sigma_\alpha^2)/gn)^{\frac{1}{2}}} \frac{u^{\frac{1}{2}e-1}\exp[-u/2\sigma_z^2]}{(2\sigma_z^2)^{\frac{1}{2}e}\Gamma(\frac{1}{2}e)}$$

$$\times \frac{v^{\frac{1}{2}h-1}\exp[-v/2(\sigma_z^2+n\sigma_\alpha^2)]}{(2(\sigma_z^2+n\sigma_\alpha^2))^{\frac{1}{2}h}\Gamma(\frac{1}{2}h)} .$$

Let the set $\omega = \{(\sigma_\alpha^2,\sigma_z^2): \sigma_\alpha^2 = 0\}$ and the set $\Omega = \{(\sigma_\alpha^2,\sigma_z^2) : \sigma_\alpha^2 \geq 0\}$. We seek the maximum likelihood estimators, $\hat{\mu}_\omega$ and $\hat{\sigma}^2_{z\omega}$, of μ and σ_z^2 when the parameters are restricted to lie within ω, and the maximum likelihood estimators, $\hat{\mu}_\Omega$, $\hat{\sigma}^2_{z\Omega}$, and $\hat{\sigma}^2_{\alpha\Omega}$, of μ, σ_z^2, and σ_α^2 when the parameters are restricted to lie within Ω.
In ω $\sigma_\alpha^2 = 0$ and

$$f(\bar{x}_{..},u,v) = \frac{\exp[-(\bar{x}_{..}-\mu)^2 gn/2\sigma_z^2]}{(2\pi\sigma_z^2/gn)^{\frac{1}{2}}} \frac{u^{\frac{1}{2}e-1}\exp[-u/2\sigma_z^2]}{(2\sigma_z^2)^{\frac{1}{2}e}\Gamma(\frac{1}{2}e)}$$

$$\times \frac{v^{\frac{1}{2}h-1}\exp[-v/2\sigma_z^2]}{(2\sigma_z^2)^{\frac{1}{2}h}\Gamma(\frac{1}{2}h)} ,$$

so that the logarithm of the likelihood function, omitting a function of the observations, is

$$-\frac{(\bar{x}_{..}-\mu)^2 gn}{2\sigma_z^2} - \frac{u+v}{2\sigma_z^2} - \frac{(e+h+1)}{2}\ln\sigma_z^2 . \qquad (1.1.4)$$

Differentiating (1.1.4) with respect to μ and σ_z^2, we obtain, respectively, the equations

$$\frac{(\bar{x}_{..}-\mu)gn}{\sigma_z^2} = 0, \qquad \frac{(\bar{x}_{..}-\mu)^2 gn+u+v}{2(\sigma_z^2)^2} - \frac{e+h+1}{2\sigma_z^2} = 0,$$

which yield the maximum likelihood solutions

$$\hat{\mu}_{\omega} = \bar{x}_{..},$$

$$\hat{\sigma}^2_{z\omega} = (u+v)/(e+h+1).$$

In Ω the logarithm of the likelihood function, omitting a function of the observations, is

$$-\frac{(\bar{x}_{..}-\mu)^2 gn}{2(\sigma_z^2+n\sigma_\alpha^2)} - \frac{u}{2\sigma_z^2} - \frac{e}{2}\ell n\,\sigma_z^2 - \frac{v}{2(\sigma_z^2+n\sigma_\alpha^2)} - \frac{(h+1)}{2}\ell n(\sigma_z^2+n\sigma_\alpha^2). \qquad (1.1.5)$$

Differentiation of (1.1.5) with respect to μ, σ_α^2, and σ_z^2 yields, respectively,

$$\frac{(\bar{x}_{..}-\mu)gn}{(\sigma_z^2+n\sigma_\alpha^2)} = 0,$$

$$\frac{n((\bar{x}_{..}-\mu)^2 gn+v)}{2(\sigma_z^2+n\sigma_\alpha^2)^2} - \frac{n(h+1)}{2(\sigma_z^2+n\sigma_\alpha^2)} = 0,$$

$$\frac{(\bar{x}_{..}-\mu)^2 gn+v}{2(\sigma_z^2+n\sigma_\alpha^2)^2} + \frac{u}{2(\sigma_z^2)^2} - \frac{e}{2\sigma_z^2} - \frac{(h+1)}{2(\sigma_z^2+n\sigma_\alpha^2)} = 0.$$

Solving these equations for μ, σ_z^2, and σ_α^2, we obtain the maximal solution of the likelihood function in Ω, $(\tilde{\mu},\tilde{\sigma}_z^2, \tilde{\sigma}_\alpha^2)$, given by

$$\tilde{\mu} = \bar{x}_{..},$$

$$\tilde{\sigma}_z^2 = u/e = u_*,$$

$$\tilde{\sigma}_\alpha^2 = (v/(h+1)-u/e)/n = (v_*-u_*)/n,$$

where $u_* = u/e$ and $v_* = v/(h+1)$. Since we insist that $\hat{\sigma}^2_{\alpha\Omega}$ be greater than or equal to zero, the solution above is the maximum likelihood solution only if $v_* - u_* \geq 0$. Suppose, however, that $v_* < u_*$. Clearly (1.1.5) is still maximized when $\mu = \bar{x}_{..}$, so that we need to minimize

$$\frac{u}{\sigma_z^2} + e\, \ell n\, \sigma_z^2 + \frac{v}{(\sigma_z^2 + n\sigma_\alpha^2)} + (h+1)\ \ell n(\sigma_z^2 + n\sigma_\alpha^2)$$

subject to the contraints $\sigma_z^2 > 0$ and $\sigma_\alpha^2 \geq 0$. Equivalently, we consider the problem of minimizing

$$\psi(x,t) = u/x + e\ \ell n\ x + v/t + (h+1)\ \ell n\ t$$

subject to the constraint $t \geq x > 0$. For fixed x $\psi(x,t)$ is concave upward in t with its absolute minimum at $t = v_*$. For each x $\psi(x,t)$ is, therefore, minimized with respect to $t \geq x$ when

$$t = \begin{cases} v_* & \text{if } v_* \geq x, \\ x & \text{if } v_* < x. \end{cases}$$

Thus, $\psi(x,t)$ is minimized over $\{(t,x): t \geq x > 0\}$ by setting

$$\begin{aligned} t = v_* \text{ and } x = u_* \qquad & \text{if } v_* \geq u_*, \\ t = x = (u+v)/(e+h+1) \qquad & \text{if } v_* < u_*. \end{aligned}$$

Hence, for the maximum likelihood estimators when the parameters are restricted to be within Ω, we obtain

$$\hat{\mu}_\Omega = \bar{x}_{..},$$

$$\hat{\sigma}^2_{z\Omega} = u_*,$$

$$\hat{\sigma}^2_{\alpha\Omega} = (v_* - u_*)/n,$$

if $v_* \geq u_*$, and

$$\hat{\mu}_\Omega = \bar{x}_{..},$$

$$\hat{\sigma}^2_{z\Omega} = (u+v)/(e+h+1),$$

$$\hat{\sigma}^2_{\alpha\Omega} = 0,$$

if $v_* < u_*$.

Substituting the maximum likelihood estimates into the likelihood function, we see that in ω

$$\max_\omega f(\bar{x}_{..},u,v) = \frac{u^{\frac{1}{2}e-1} v^{\frac{1}{2}h-1} \exp[-(e+h+1)/2]}{\Gamma(\frac{1}{2}e)\Gamma(\frac{1}{2}h)\left(\frac{u+v}{e+h+1}\right)^{\frac{1}{2}(e+h+1)} (\pi/gn)^{\frac{1}{2}}\, 2^{\frac{1}{2}(e+h+1)}},$$

and in Ω

$$\max_\Omega f(\bar{x}_{..},u,v) = \begin{cases} \dfrac{u^{\frac{1}{2}e-1} v^{\frac{1}{2}h-1} \exp[-(e+h+1)/2]}{\Gamma(\frac{1}{2}e)\Gamma(\frac{1}{2}h)\left(\frac{u}{e}\right)^{\frac{1}{2}e}\left(\frac{v}{h+1}\right)^{\frac{1}{2}(h+1)} (\pi/gn)^{\frac{1}{2}}\, 2^{\frac{1}{2}(e+h+1)}} & \text{if } v_* \geq u_*, \\ \max_\omega f(\bar{x}_{..},u,v) & \text{if } v_* < u_*. \end{cases}$$

The likelihood ratio, λ, is

$$\lambda = \frac{\max_\omega f(\bar{x}_{..},u,v)}{\max_\Omega f(\bar{x}_{..},u,v)} = \begin{cases} \dfrac{(u/e)^{\frac{1}{2}e}(v/(h+1))^{\frac{1}{2}(h+1)}}{[(u+v)/(e+h+1)]^{\frac{1}{2}(e+h+1)}} & \text{if } v_* \geq u_*, \\ 1 & \text{if } v_* < u_*. \end{cases}$$

Now putting $w = u/(u+v)$ and noting that

$v_* \geq u_*$, if and only if (iff)

$\frac{v}{h+1} \geq \frac{u}{e}$, iff

$\frac{u+v}{h+1} \geq u\left(\frac{1}{e} + \frac{1}{h+1}\right)$, iff

$\frac{e}{e+h+1} \geq \frac{u}{u+v} = w,$

we can rewrite the likelihood ratio

$$\lambda = \begin{cases} \dfrac{(e+h+1)^{\frac{1}{2}(e+h+1)}}{e^{\frac{1}{2}e}\,(h+1)^{\frac{1}{2}(h+1)}}\, w^{\frac{1}{2}e}\,(1-w)^{\frac{1}{2}(h+1)} & \text{if } w \le \dfrac{e}{e+h+1}, \\ 1 & \text{if } w > \dfrac{e}{e+h+1}. \end{cases}$$

Since λ is an increasing function of w, and H_o is rejected for small values of λ, it follows that H_o is rejected for small values of w or large values of 1/w. Now

$$\frac{1}{w} = \frac{u+v}{u} = 1 + \frac{v}{u} = 1 + \frac{h}{e}\left(\frac{v/h}{u/e}\right),$$

so the likelihood ratio test rejects H_o for ev/hu large.

Recall that u/σ_z^2 has a chi-square distribution with e degrees of freedom, and $v/(\sigma_z^2+n\sigma_\alpha^2)$ has a chi-square distribution with h degrees of freedom, independent of u. Hence, the quantity $\sigma_z^2\, ev/(\sigma_z^2+n\sigma_\alpha^2)hu$ has an F distribution with h and e degrees of freedom. If we let $F(h,e,\alpha)$ denote the constant for which $P(F(h,e) \ge F(h,e,\alpha)) = \alpha$ where $F(h,e)$ has an F distribution with h and e degrees of freedom, then we will reject H_o if $ev/hu \ge F(h,e,\alpha)$. The power function of this test is a function of $\theta = \sigma_\alpha^2/\sigma_z^2$ and is given by

$$\beta(\theta) = P(F(h,e) \ge F(h,e,\alpha)/(1+n\theta)).$$

Although the analysis which we have just outlined is, by now, quite standard to any graduate level course in design and analysis, we have reproduced it since it motivates the more general problem to be described in the next section.

Indeed the situation we wish to consider contains the one-way random effects model as a special case to which we can return on occasion to check our work.

1.2 The Multivariate Random Effects Model

Suppose a physician is considering administering a battery of m distinct types of blood tests to his patients as a part of their physical examination. He believes that, based on the results of these tests, he may be able to detect any one of several particular pathological conditions. In order to examine variability in the results of the blood tests, the physician chooses to administer the battery of blood tests n times to each of g patients. This results in the observations $\underline{x}_{ij}$(mxl): i = 1,2,...,g; j = 1,2,...,n.

A suitable model to explain the different values of $\underline{x}_{ij}$: i = 1,2,...,g; j = 1,2,...,n would be

$$\underline{x}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{z}_{ij}. \tag{1.2.1}$$

Here $\underline{\mu}$(mxl) is an overall mean, $\underline{\alpha}_i$(mxl) is an effect due to the i^{th} patient, and $\underline{z}_{ij}$(mxl) represents a vector of random errors due to the measuring process. We assume that $\underline{z}_{ij}$: i = 1,2,...,g; j = 1,2,...,n are independent and have an m-variate normal distribution with mean $\underline{0}$ and variance-covariance matrix Σ.

Since the physician will administer the tests to an unlimited number of patients in the future, we should properly regard the g patients involved as a sample from the entire population of patients. The patient effects, $\underline{\alpha}_i$: i = 1,2,...,g, now have the role of random vectors, and (1.2.1) is a multivariate random effects model. We will assume that $\underline{\alpha}_i$: i = 1,2,...,g are independent and have an m-variate normal distribution with mean $\underline{0}$ and variance-covariance matrix Δ. Hence, from our model (1.2.1) we see that $\underline{x}_{ij}$ has an m-variate normal distribution with mean $\underline{\mu}$ and variance-covariance matrix $\Delta + \Sigma$.

While there are m different blood tests, it is believed that there are some groups of tests for which the tests within a group vary quite strongly together. In other words, the data from some of the tests are highly correlated. For this reason the number of sources of variation between the patients, which we will denote by p, may be less than the number of tests, m. That is, the rank of the variance-covariance matrix Δ is p where $p \leq m$. Since Δ is symmetric, nonnegative definite, and of rank p, there exists a matrix L(mxp) such that $\Delta = LL'$. Clearly L is not unique since if $\Delta = LL'$ and P(pxp) is such that $PP' = I$, then $\Delta = L_*L_*'$ where $L_* = LP$. This enables us to rewrite (1.2.1) as

$$\underline{x}_{ij} = \underline{\mu} + L\underline{f}_i + \underline{z}_{ij}, \qquad (1.2.2)$$

where $\underline{f}_i$(px1): i = 1,2,...,g are independently distributed, having a p-variate normal distribution with mean $\underline{0}$ and variance-covariance matrix equal to the identity matrix.

If the physician is interested in using the blood tests as a diagnostic tool, he will certainly be interested in determining the value p, since the p sources of variation may correspond to p different pathological disorders. So of particular interest to the physician is a test of the hypothesis $H_o^{(s)}$: the rank of the matrix $LL' \leq s-1$ against the hypothesis $H_1^{(s)}$: the rank of the matrix $LL' = s$. With such a test procedure he could develop a sequential test procedure for determining the rank of LL'. He would first test $H_o^{(m)}$ against $H_1^{(m)}$, and if he rejects $H_o^{(m)}$, he would stop and take the rank of LL' to be m; otherwise, he would proceed to test $H_o^{(m-1)}$ against $H_1^{(m-1)}$. The procedure continues until either some hypothesis $H_o^{(s)}$ is rejected, in which case he then takes the rank of LL' to be s, or the hypothesis $H_o^{(1)}$ is accepted, in which case he would conclude that there is no significant variation between patients.

In this paper we investigate the multivariate one-way classification model with random effects, given by (1.2.2). Two specific cases, regarding the structure of the variance-covariance matrix Σ, will be considered. In the first and most general case we will assume no more than that Σ is symmetric and positive definite. In the second case we will assume that the vector of random errors, $\underline{z}_{ij}$, is such that its components are independent and have the same variance. That is, we assume that Σ is equal to some constant multiple of the identity matrix. In each case we develop a test procedure for testing the hypothesis $H_o^{(s)}$: the rank of $LL' \leq s-1$ against the hypothesis $H_1^{(s)}$: the rank of $LL' = s$. In addition, we investigate some of the properties of these test procedures and present a numerical example to illustrate the use of these procedures.

1.3 Notation

The following notation will be used whenever convenient:

Notation	Interpretation
$(A)_{i.}$	row i of the matrix A
$(A)_{.j}$	column j of the matrix A
$(A)_{ij}$	the element in row i and column j of the matrix A
a_{ij}	the element in row i and column j of the matrix A
A^{-1}	the inverse of the matrix A
A'	the transpose of the matrix A
$\|A\|$	the determinant of the matrix A
tr A	the trace of the matrix A
dg(A)	the diagonal matrix which has as its diagonal elements the diagonal elements of A
$\text{diag}(a_1,a_2,\ldots,a_m)$	the diagonal matrix which has $a_1, a_2,\ldots,a_m$ as its diagonal elements
$ch_i(A)$	the i^{th} largest latent root of the matrix A

Notation	Interpretation
rank (A)	the rank of the matrix A
I_m	the m x m identity matrix
I	the identity matrix (used when the order of the matrix is obvious)
(0)	the matrix which has all of its elements equal to zero
$\underline{x}$	a vector
x_i	the i^{th} element of the vector $\underline{x}$
$\underline{0}$	the vector which has all of its elements equal to zero
E(x)	the expected value of x
V(x)	the variance of x
Cov(x,y)	the covariance of x and y
P(A)	the probability of event A
P(A|B)	the probability of event A given event B
$\Gamma(x)$	the gamma function
$x_n \xrightarrow{d} x$	x_n converges to x in distribution
$a_n \longrightarrow a$	convergence of a sequence of constants
exp(x)	Euler's constant, "e," raised to the x power
ε	is contained in
$\sim$	is distributed as
$N(\mu,\sigma^2)$	the normal distribution with mean μ and variance σ^2

Notation	Interpretation
$N_m(\underline{\mu},\Sigma)$	the m-variate normal distribution with mean $\underline{\mu}$ and variance-covariance matrix Σ
χ^2_ν	the central chi-square distribution with ν degrees of freedom
$F^{\nu_1}_{\nu_2}$	the central F distribution with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom
$W_m(\Sigma,\nu,0)$	the central Wishart distribution with variance-covariance matrix Σ and degrees of freedom ν
Jones [1973]	the reference authored by Jones and published in 1973
Jones [1973:1]	page 1 of the reference authored by Jones and published in 1973

CHAPTER 2

MAXIMIZATION OF THE LIKELIHOOD FUNCTION FOR GENERAL Σ

2.1 The Likelihood Function

Suppose the vectors $\underline{x}_{ij}$(mx1): $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$ can be modeled by

$$\underline{x}_{ij} = \underline{\mu} + L\underline{f}_i + \underline{z}_{ij}, \tag{2.1.1}$$

wherein $\underline{\mu}$(mx1) is a fixed but unknown vector, L(mxp) is a fixed but unknown matrix, $\underline{f}_i \sim N_p(\underline{0},I)$: $i = 1, 2,\ldots,g$, and $\underline{z}_{ij} \sim N_m(0,\Sigma)$: $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$. We assume that the set of random vectors $\{\underline{f}_1,\underline{f}_2,\ldots,\underline{f}_g,\underline{z}_{11},\ldots,\underline{z}_{gn}\}$ are mutually independent. Thus, $\underline{x}_{ij} \sim N_m(\underline{\mu},V)$ with $V = LL' + \Sigma$. However, for any orthogonal matrix P(pxp), $V = LL' + \Sigma = LP(LP)' + \Sigma$ so that L is not unique whereas LL' is unique. The purpose of this section is to derive the likelihood function for $\underline{\mu}$, LL', and Σ. Although $\underline{x}_{ij}$ and $\underline{x}_{k\ell}$ are independent for all (j,ℓ) when $i \neq k$, $\underline{x}_{ij}$ and $\underline{x}_{i\ell}$ are not independent even when $j \neq \ell$, since $\mathrm{Cov}(\underline{x}_{ij},\underline{x}_{i\ell}) = LL'(j\neq\ell)$. Thus, the likelihood function is not simply the product of the density functions of the $\underline{x}_{ij}$'s. A transformation of the $\underline{x}_{ij}$'s will expedite the derivation of the likelihood function.

Consider the Helmert transformation (see, for example, Kendall and Stuart [1963:250]) given below:

$$\underline{x}_{i1} = \underline{\bar{x}}_{i.} + (2\cdot 1)^{-\frac{1}{2}}\underline{y}_{i1} + (3\cdot 2)^{-\frac{1}{2}}\underline{y}_{i2} + \cdots + (n(n-1))^{-\frac{1}{2}}\underline{y}_{i\nu},$$

$$\underline{x}_{i2} = \underline{\bar{x}}_{i.} - (2\cdot 1)^{-\frac{1}{2}}\underline{y}_{i1} + (3\cdot 2)^{-\frac{1}{2}}\underline{y}_{i2} + \cdots + (n(n-1))^{-\frac{1}{2}}\underline{y}_{i\nu},$$

$$\underline{x}_{i3} = \underline{\bar{x}}_{i.} \qquad - 2(3\cdot 2)^{-\frac{1}{2}}\underline{y}_{i2} + \cdots + (n(n-1))^{-\frac{1}{2}}\underline{y}_{i\nu},$$

$$\vdots$$

$$\underline{x}_{in} = \underline{\bar{x}}_{i.} \qquad - (n-1)(n(n-1))^{-\frac{1}{2}}\underline{y}_{i\nu},$$

where $\nu = n-1$. It will be helpful to note that the above equations imply the following:

$$\underline{\bar{x}}_{i.} = n^{-1}\underline{x}_{i1} + n^{-1}\underline{x}_{i2} + \cdots + n^{-1}\underline{x}_{in},$$

$$\underline{y}_{i1} = 2^{-\frac{1}{2}}\underline{x}_{i1} - 2^{-\frac{1}{2}}\underline{x}_{i2},$$

$$\underline{y}_{i2} = (3\cdot 2)^{-\frac{1}{2}}\underline{x}_{i1} + (3\cdot 2)^{-\frac{1}{2}}\underline{x}_{i2} - 2(3\cdot 2)^{-\frac{1}{2}}\underline{x}_{i3},$$

$$\vdots$$

$$\underline{y}_{i\nu} = (n(n-1))^{-\frac{1}{2}}\underline{x}_{i1} + \cdots + (n(n-1))^{-\frac{1}{2}}\underline{x}_{i,n-1} - (n-1)(n(n-1))^{-\frac{1}{2}}\underline{x}_{in}.$$

In matrix formulation we have

$$(\underline{x}_{i1},\ldots,\underline{x}_{in})' = H(\underline{\bar{x}}_{i.},\underline{y}_{i1},\ldots,\underline{y}_{i\nu})',$$

and we note that, while not an orthogonal matrix, the columns of H are orthogonal. The matrix H fails to be orthogonal since $H'H = \text{diag}(n,1,1,\ldots,1)$. Observe that, upon replacing $\underline{x}_{ij}$ by the right side of (2.1.1), we have

$$\underline{\bar{x}}_{i.} = \underline{\mu} + L\underline{f}_i + \underline{\bar{z}}_{i.},$$

$$\underline{y}_{i1} = (2)^{-\frac{1}{2}}(\underline{z}_{i1}-\underline{z}_{i2}),$$

$$\underline{y}_{i2} = (3\cdot 2)^{-\frac{1}{2}}(\underline{z}_{i1}+\underline{z}_{i2}-2\underline{z}_{i3}),$$

$$\vdots$$

$$\underline{y}_{i\nu} = \{n(n-1)\}^{-\frac{1}{2}}(\underline{z}_{i1}+\underline{z}_{i2}+\cdots+\underline{z}_{i,n-1}-(n-1)\underline{z}_{in}).$$

Thus

$$E(\underline{\bar{x}}_{i.}\underline{y}'_{ij}) = (0),$$

$$E(\underline{y}_{ij}\underline{y}'_{iq}) = (0) \quad \text{if} \quad j \neq q,$$

$$E(\underline{y}_{ij}\underline{y}'_{ij}) = \Sigma.$$

Hence, it follows that $\{\underline{\bar{x}}_{i.},\underline{y}_{i1},\ldots,\underline{y}_{i\nu}\}_{i=1}^{g}$ are a set of gn mutually independent vectors with $\underline{\bar{x}}_{i.} \sim N_m(\underline{\mu},(1/n)\Sigma+LL')$: $i = 1,2,\ldots,g$ and $\underline{y}_{ij} \sim N_m(\underline{0},\Sigma)$: $i = 1,2,\ldots,g$; $j = 1,2,\ldots,n$. Note also that $\sum_{j=1}^{n}(\underline{x}_{ij}-\underline{\bar{x}}_{i.})(\underline{x}_{ij}-\underline{\bar{x}}_{i.})' = \sum_{j=1}^{\nu}\underline{y}_{ij}\underline{y}'_{ij}$ and denote this matrix by E_i. We can now write the joint density function of $\underline{y}_{i1},\ldots,\underline{y}_{i\nu}$ as

$$\begin{aligned} f(\underline{y}_{i1},\ldots,\underline{y}_{i\nu};\Sigma) &= \prod_{j=1}^{\nu}|2\pi\Sigma|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\underline{y}'_{ij}\Sigma^{-1}\underline{y}_{ij}] \\ &= |2\pi\Sigma|^{-\frac{1}{2}\nu}\exp[-\tfrac{1}{2}\sum_{j=1}^{\nu}(\underline{y}'_{ij}\Sigma^{-1}\underline{y}_{ij})] \\ &= |2\pi\Sigma|^{-\frac{1}{2}\nu}\exp[-\tfrac{1}{2}\sum_{j=1}^{\nu}\mathrm{tr}(\underline{y}'_{ij}\Sigma^{-1}\underline{y}_{ij})] \end{aligned}$$

$$= |2\pi\Sigma|^{-\frac{1}{2}\nu} \exp[-\tfrac{1}{2} \sum_{j=1}^{\nu} \mathrm{tr}(\Sigma^{-1} \underline{y}_{ij} \underline{y}_{ij}')]$$

$$= |2\pi\Sigma|^{-\frac{1}{2}\nu} \exp[-\tfrac{1}{2} \mathrm{tr}\ \Sigma^{-1} E_i]$$

$$= g(E_i;\Sigma)$$

so that from the set $\{\underline{\bar{x}}_{i.},\underline{y}_{i1},\underline{y}_{i2},\ldots,\underline{y}_{i\nu}\}$, $(E_i,\underline{\bar{x}}_{i.})$ is sufficient.

Thus, we may assume that we have, independently,

$$\left.\begin{array}{l} E_i \sim W_m(\Sigma,\ \nu,\ 0) \\ \\ \underline{\bar{x}}_{i.} \sim N_m(\underline{\mu},\ \frac{1}{n}\Sigma + LL') \end{array}\right\} 1 \le i \le g.$$

Note that

$$\sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\mu})(\underline{\bar{x}}_{i.}-\underline{\mu})' = \sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})'+g(\underline{\bar{x}}_{..}-\underline{\mu})(\underline{\bar{x}}_{..}-\underline{\mu})',$$

where $\underline{\bar{x}}_{..} = \sum_{i=1}^{g} \underline{\bar{x}}_{i.}/g$. Then putting $W = (1/n)\Sigma + LL'$,

we can write the joint density function of $\underline{\bar{x}}_{1.},\underline{\bar{x}}_{2.},\ldots,\underline{\bar{x}}_{g.}$ as

$$f(\underline{\bar{x}}_{1.},\underline{\bar{x}}_{2.},\ldots,\underline{\bar{x}}_{g.};\underline{\mu},W) = \prod_{i=1}^{g} |2\pi W|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}(\underline{\bar{x}}_{i.}-\underline{\mu})'W^{-1}(\underline{\bar{x}}_{i.}-\underline{\mu})]$$

$$= |2\pi W|^{-\frac{1}{2}g}\exp[-\tfrac{1}{2} \sum_{i=1}^{g} \mathrm{tr}((\underline{\bar{x}}_{i.}-\underline{\mu})' W^{-1}(\underline{\bar{x}}_{i.}-\underline{\mu}))]$$

$$= |2\pi W|^{-\frac{1}{2}g}\exp[-\tfrac{1}{2} \sum_{i=1}^{g} \mathrm{tr}(W^{-1}(\underline{\bar{x}}_{i.}-\underline{\mu})(\underline{\bar{x}}_{i.}-\underline{\mu})'))]$$

$$= |2\pi W|^{-\frac{1}{2}g}\exp[-\tfrac{1}{2} \mathrm{tr}(W^{-1} \sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\mu})(\underline{\bar{x}}_{i.}-\underline{\mu})'))]$$

$$= |2\pi W|^{-\frac{1}{2}g} \exp[-\tfrac{1}{2}g \text{ tr } W^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})(\underline{\bar{x}}_{..}-\underline{\mu})'$$

$$-\tfrac{1}{2} \text{ tr}(W^{-1} \sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})'))]$$

$$= |2\pi W|^{-\frac{1}{2}g} \exp[-\tfrac{1}{2}g \text{ tr } W^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})(\underline{\bar{x}}_{..}-\underline{\mu})']$$

$$\times \exp[-\tfrac{1}{2} \text{ tr } W^{-1} \sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})']$$

$$= g(\underline{\bar{x}}_{..}, H; \underline{\mu}, W),$$

where $H = n \sum_{i=1}^{g} (\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})'$. Hence, from the set $\{\underline{\bar{x}}_{1.},\ldots,\underline{\bar{x}}_{g.}\}$, $(\underline{\bar{x}}_{..}, H)$ is sufficient for $(\underline{\mu}, (1/n)\Sigma + LL')$.

Also if we let c denote a constant, we can write the joint density function of $E_1,\ldots,E_g$ as

$$f(E_1,\ldots,E_g;\Sigma) = c \prod_{i=1}^{g} |E_i|^{\frac{1}{2}(\nu-m-1)} \exp[-\tfrac{1}{2} \text{ tr}(\Sigma^{-1}E_i)]$$

$$= c \exp[-\tfrac{1}{2} \text{ tr}(\Sigma^{-1} \sum_{i=1}^{g} E_i)] \prod_{i=1}^{g} |E_i|^{\frac{1}{2}(\nu-m-1)}$$

$$= g_1(E;\Sigma) g_2(E_1,E_2,\ldots,E_g),$$

where $E = \sum_{i=1}^{g} E_i$. Thus, from the set $\{E_1,\ldots,E_g\}$, E is sufficient for Σ.

Then we may assume that we have, independently,

$$\underline{\bar{x}}_{..} \sim N_m(\underline{\mu}, \tfrac{1}{gn}(\Sigma+nLL')),$$

$$E \sim W_m(\Sigma, e, 0),$$

$$H \sim W_m(\Sigma+nLL', h, 0),$$

where $e = g(n-1)$ and $h = g-1$. The problem is to estimate $\underline{\mu}$, Σ, and LL' or, equivalently, to estimate $\underline{\mu}$, Σ, and M where $M = nLL'$. Recall that L is not uniquely defined so that if $\hat{L}L'$ is an estimate of LL', then any $\hat{L}$, such that $\hat{L}\hat{L}' = \hat{L}L'$, is an estimate of L. The likelihood function of $(\underline{\mu},\Sigma,M)$ can be expressed as

$$f(\underline{\bar{x}}_{..},E,H) = \frac{K_m(I,e)K_m(I,h)}{|\frac{2\pi}{gn}(\Sigma+M)|^{\frac{1}{2}}|\Sigma+M|^{\frac{1}{2}h}|\Sigma|^{\frac{1}{2}e}} |H|^{\frac{1}{2}(h-m-1)}|E|^{\frac{1}{2}(e-m-1)}$$

$$\times \exp[-\tfrac{1}{2}(\underline{\bar{x}}_{..}-\underline{\mu})'(\tfrac{1}{gn}(\Sigma+M))^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1}E)-\tfrac{1}{2}\mathrm{tr}(\Sigma+M)^{-1}H],$$

where $K_m^{-1}(I,\nu) = 2^{\frac{1}{2}m\nu}\,\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\tfrac{1}{2}(\nu-j+1))$.

The logarithm of the likelihood function, omitting a function of the observations, is

$$-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}E-\tfrac{1}{2}e\,\ln|\Sigma|\;-\tfrac{1}{2}\mathrm{tr}(\Sigma+M)^{-1}H-\tfrac{1}{2}h\,\ln|\Sigma+M|$$
$$-\;\tfrac{1}{2}\,\ln|\Sigma+M|\;-\;\tfrac{1}{2}(\underline{\bar{x}}_{..}-\underline{\mu})'((1/gn)(\Sigma+M))^{-1}(\underline{\bar{x}}_{..}-\underline{\mu}).$$

We seek the solution, $(\underline{\tilde{\mu}},\tilde{\Sigma},\tilde{M})$, which maximizes the equation above, or equivalently, the solution which minimizes

$$\mathrm{tr}\,\Sigma^{-1}E+e\,\ln|\Sigma|\;+\;\mathrm{tr}(\Sigma+M)^{-1}H+(h+1)\ln|\Sigma+M|$$
$$+\;(\underline{\bar{x}}_{..}-\underline{\mu})'((1/gn)(\Sigma+M))^{-1}(\underline{\bar{x}}_{..}-\underline{\mu}). \qquad (2.1.2)$$

Before we can minimize the above equation, we need some results on differentiation. Let $W(m\times m)$, $X(m\times m)$, and $Y(m\times m)$ be symmetric matrices, and let $\underline{z}(m\times 1)$ and $\underline{a}(m\times 1)$ be vectors. The proof of the first result can be found in Graybill [1969:267].

<u>Lemma 2.1.1</u>:

$$\frac{\partial \ell n|X|}{\partial X} = 2X^{-1} - dg(X^{-1}).$$

<u>Lemma 2.1.2</u>:

$$\frac{\partial \ell n|X+Y|}{\partial X} = 2(X+Y)^{-1} - dg((X+Y)^{-1}).$$

<u>Proof</u>: Let V = X + Y. Then

$$\frac{\partial \ell n|X+Y|}{\partial x_{ij}} = \frac{\partial \ell n|V|}{\partial x_{ij}} = \sum_{1\le p\le q\le m}\sum \frac{\partial \ell n|V|}{\partial v_{pq}} \frac{\partial v_{pq}}{\partial x_{ij}} = \frac{\partial \ell n|V|}{\partial v_{ij}},$$

so $\frac{\partial \ell n|X+Y|}{\partial X} = \frac{\partial \ell n|V|}{\partial V} = 2V^{-1}-dg(V^{-1}) = 2(X+Y)^{-1}-dg((X+Y)^{-1}).$

<u>Lemma 2.1.3</u>:

$$\frac{\partial tr(X+Y)^{-1}W}{\partial X} = -2(X+Y)^{-1}W(X+Y)^{-1}+dg((X+Y)^{-1}W(X+Y)^{-1}).$$

<u>Proof</u>: Let V = X + Y. Note that

$$\frac{\partial V^{-1}V}{\partial x_{ij}} = (0) = \left(\frac{\partial V^{-1}}{\partial x_{ij}}\right)V + V^{-1}\left(\frac{\partial V}{\partial x_{ij}}\right)$$

so that $\frac{\partial V^{-1}}{\partial x_{ij}} = - V^{-1}\left(\frac{\partial V}{\partial x_{ij}}\right)V^{-1}.$

Then $\frac{\partial tr(X+Y)^{-1}W}{\partial x_{ij}} = \frac{\partial trV^{-1}W}{\partial x_{ij}} = tr\left(\frac{\partial V^{-1}}{\partial x_{ij}}\right)W$

$$= - tr\ V^{-1}\left(\frac{\partial V}{\partial x_{ij}}\right)V^{-1}W$$

$$= - \sum_{p=1}^{m}\sum_{q=1}^{m} (V^{-1}WV^{-1})_{qp} \frac{\partial v_{pq}}{\partial x_{ij}}$$

$$= \begin{cases} -(V^{-1}WV^{-1})_{ji} - (V^{-1}WV^{-1})_{ij} & \text{if } i \neq j, \\ -(V^{-1}WV^{-1})_{ii} & \text{if } i = j. \end{cases}$$

$$= \begin{cases} -\ 2(V^{-1}WV^{-1})_{ij} & \text{if } i \neq j, \\ -(V^{-1}WV^{-1})_{ii} & \text{if } i = j. \end{cases}$$

Hence, $\frac{\partial tr(X+Y)^{-1}W}{\partial X} = -2V^{-1}WV^{-1} + dg(V^{-1}WV^{-1})$

$$= -2(X+Y)^{-1}W(X+Y)^{-1} + dg((X+Y)^{-1}W(X+Y)^{-1}).$$

Lemma 2.1.4:

$$\frac{\partial(\underline{z}-\underline{a})'W(\underline{z}-\underline{a})}{\partial \underline{z}} = 2W(\underline{z}-\underline{a}).$$

Proof: $\frac{\partial(\underline{z}-\underline{a})'W(\underline{z}-\underline{a})}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\sum_{p=1}^{m}\sum_{q=1}^{m}(z_p-a_p)(z_q-a_q)w_{pq}\right)$

$$= \sum_{q=1}^{m}(z_q-a_q)w_{iq} + \sum_{p=1}^{m}(z_p-a_q)w_{pi}$$

$$= 2\sum_{q=1}^{m}(z_q-a_q)w_{iq} = 2(W)_{i.}(\underline{z}-\underline{a})$$

so that $\frac{\partial(\underline{z}-\underline{a})'W(\underline{z}-\underline{a})}{\partial \underline{z}} = 2W(\underline{z}-\underline{a})$.

If we ignore the constraints that Σ is positive definite and M is nonnegative definite and seek the stationary values of (2.1.2) over all possible $(\underline{\mu},\Sigma,M)$, we find, upon taking the partial derivatives of (2.1.2) with respect to Σ, M, and $\underline{\mu}$ and setting them equal to zero, that

$$e\Sigma^{-1}+(h+1)(\Sigma+M)^{-1} - \Sigma^{-1}E\Sigma^{-1}-(\Sigma+M)^{-1}H(\Sigma+M)^{-1}$$

$$- gn(\Sigma+M)^{-1}(\bar{\underline{x}}_{..}-\underline{\mu})(\bar{\underline{x}}_{..}-\underline{\mu})'(\Sigma+M)^{-1} = (0),$$

$$(h+1)(\Sigma+M)^{-1}-(\Sigma+M)^{-1}H(\Sigma+M)^{-1}-gn(\Sigma+M)^{-1}(\bar{\underline{x}}_{..}-\underline{\mu})(\bar{\underline{x}}_{..}-\underline{\mu})'(\Sigma+M)^{-1}$$

$$= (0),$$

$$gn(\Sigma+M)^{-1}(\bar{\underline{x}}_{..}-\underline{\mu}) = \underline{0},$$

for which the solutions are

$$\tilde{\underline{\mu}} = \bar{\underline{x}}_{..},$$

$$\tilde{\Sigma} = (1/e)E,$$

$$\tilde{M} = (1/(h+1))H-(1/e)E.$$

Since M is a nonnegative definite matrix, its maximum likelihood estimate must also be nonnegative definite, so the solutions above are the maximum likelihood estimates only if $(1/(h+1))H-(1/e)E$ is nonnegative definite. We find that, while the solutions for $\underline{\mu}$ and Σ are the natural unbiased estimates, the solution for M is not. That is,

$$E(\tilde{M}) = (1/(h+1))(hM-\Sigma).$$

Hence, we see that $E(\tilde{M})$ is also not necessarily nonnegative definite.

Suppose that instead of using the likelihood function of $(\underline{\mu},\Sigma,M)$ we use the marginal likelihood function of (Σ,M). Justification for this follows from the fact that (E,H) is "marginally sufficient" for (Σ,M) or, in other words, (E,H) is "sufficient for (Σ,M) in the absence of knowledge of $\underline{\mu}$."

For a detailed description of the principle of marginal sufficiency see Barnard [1963]. There is ample precedent for the use of this principle in multivariate theory. For example, Bartlett's test has two forms, one involving the sample size and the other involving the degrees of freedom. The marginal likelihood function of (Σ, M) can be written

$$f(E,H) = \frac{K_m(I,e)K_m(I,h)}{|\Sigma+M|^{\frac{1}{2}h}|\Sigma|^{\frac{1}{2}e}} |H|^{\frac{1}{2}(h-m-1)} \quad |E|^{\frac{1}{2}(e-m-1)}$$

$$\times \exp[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}E - \tfrac{1}{2} \operatorname{tr}(\Sigma+M)^{-1}H],$$

where

$$K_m^{-1}(I,\nu) = 2^{\frac{1}{2}m\nu} \pi^{\frac{1}{4}m(m-1)} \prod_{j=1}^{m} \Gamma(\tfrac{1}{2}(\nu-j+1)).$$

The logarithm of the likelihood, omitting a function of the observations, is

$$-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}E - \tfrac{1}{2}e \,\ell n|\Sigma| - \tfrac{1}{2} \operatorname{tr}(\Sigma+M)^{-1}H - \tfrac{1}{2}h \,\ell n|\Sigma+M|.$$

We seek the solution, $(\tilde{\Sigma}_*, \tilde{M}_*)$, which maximizes the above equation, or equivalently, the solution which minimizes

$$\operatorname{tr} \Sigma^{-1}E + e \,\ell n|\Sigma| + \operatorname{tr}(\Sigma+M)^{-1}H + h \,\ell n|\Sigma+M|. \qquad (2.1.3)$$

Again if we ignore the constraints that Σ is positive definite and M is nonnegative definite and seek the stationary values of (2.1.3) over all possible (Σ, M), we find, upon taking the partial derivatives of (2.1.3) with respect to Σ and M and setting them equal to zero, that

$$e\Sigma^{-1}+h(\Sigma+M)^{-1}-\Sigma^{-1}E\Sigma^{-1}-(\Sigma+M)^{-1}H(\Sigma+M)^{-1} = (0),$$

$$h(\Sigma+M)^{-1}-(\Sigma+M)^{-1}H(\Sigma+M)^{-1} = (0),$$

for which the solutions are

$$\tilde{\Sigma}_* = (1/e)E,$$

$$\tilde{M}_* = (1/h)H-(1/e)E.$$

We see that these solutions are the natural unbiased estimates of Σ and M, and thus $E(\tilde{M}_*) = M$ is clearly nonnegative definite. For this reason, we choose to continue our work with the marginal likelihood function of (Σ,M). Note that since M is nonnegative definite, the solutions above are the maximum likelihood estimates only if $(1/h)H-(1/e)E$ is also nonnegative definite. In the next two sections we will derive maximum likelihood estimates for Σ and M which are valid for all possible (E,H).

2.2 Some Lemmas

Consider the function

$$\phi(A,B;D,e,h) = e[\text{tr } A^{-1}+\ell n|A|] + h[\text{tr } B^{-1}D + \ell n|B|],$$

where A, B, and D are $m \times m$ matrices. We assume that D is diagonal with distinct, descending, positive diagonal elements; that is, $D = \text{diag}(d_1,d_2,\ldots,d_m)$ with $d_1 > d_2 > \ldots > d_m > 0$. We are interested in minimizing $\phi(A,B;D,e,h)$ subject to $(A,B) \in C_s = \{(A,B): A \in P_m, B \in P_m, B - A \in \bigcup_{j=0}^{s} P_j\}$, where P_j is the set of all symmetric, nonnegative definite matrices of rank j. In this section it will be shown that the required absolute minimum occurs when both A and B are diagonal.

The proof of this result relies mainly on a lemma regarding the stationary points of the function $g(P) = \text{tr } PB^{-1}P'D$ where $P(m\times m)$ is orthogonal.

Lemma 2.2.1: Consider $g(P) = \text{tr } PXP'D$ where $P(m\times m)$ is such that $PP' = I$, and $X(m\times m)$ and $D(m\times m)$ are both symmetric and positive definite. It is assumed that D is diagonal with distinct, descending, positive diagonal elements. Then the stationary points of $g(P)$ occur when PXP' is diagonal. Further, the absolute maximum of $g(P)$ is

$$\max_{P:PP'=I} g(P) = \sum_{i=1}^{m} d_i ch_i(X),$$

and the absolute minimum of $g(P)$ is

$$\min_{P:PP'=I} g(P) = \sum_{i=1}^{m} d_{m+1-i} ch_i(X).$$

Proof: Using the method of Lagrange multipliers, we look at

$$L(P,\Lambda) = \text{tr } PXP'D + \text{tr } \Lambda(PP'-I),$$

where $\Lambda = \Lambda'$. Let Δ_{ij} be the matrix that has 1 in row i, column j, and 0's elsewhere. Then

$$\begin{aligned}
\frac{\partial L}{\partial p_{ij}} &= \text{tr}(\Delta_{ij}XP'D+PX\Delta_{ji}D) + \text{tr}\Lambda(\Delta_{ij}P'+P\Delta_{ji}) \\
&= \text{tr}(DPX\Delta_{ji}+PX\Delta_{ji}D) + \text{tr}(P\Delta_{ji}\Lambda+\Lambda P\Delta_{ji}) \\
&= 2\text{tr}(\Delta_{ji}DPX) + 2\text{tr}(\Delta_{ji}\Lambda P) \\
&= 2(DPX)_{ij} + 2(\Lambda P)_{ij},
\end{aligned}$$

$$\frac{\partial L}{\partial \lambda_{ij}} = \mathrm{tr}(\Delta_{ij}+\Delta_{ji})(PP'-I) = 2(PP'-I)_{ij} \qquad \text{if } i \neq j,$$

$$\frac{\partial L}{\partial \lambda_{ii}} = \mathrm{tr}\ \Delta_{ii}(PP'-I) = (PP'-I)_{ii}.$$

Thus, the stationary values of g(P) occur at the solutions to

$$\left.\begin{aligned} 2DPX + 2\Lambda P &= 0, \\ PP' &= I. \end{aligned}\right\} \qquad (2.2.1)$$

From (2.2.1) it follows that

$$\Lambda = -DPXP',$$

so that $\Lambda = \Lambda'$ implies that

$$DPXP' = PXP'D,$$

or

$$DY = YD, \qquad (2.2.2)$$

where $Y = PXP'$.

Examining the $(i,j)^{th}$ term on each side of (2.2.2), we see that we must have $d_i y_{ij} = y_{ij} d_j$. Since $d_i \neq d_j$: $i \neq j$, it follows that $y_{ij} = 0$: $i \neq j$. Thus, $Y = PXP'$ is diagonal. It is clear then that the stationary values of tr PXP'D are given by the set of values

$$\sum_{i=1}^{m} d_{t(i)}\ ch_i(X),$$

where $\{t(1),t(2),\ldots,t(m)\}$ is a permutation of $\{1,2,\ldots,m\}$, the set being formed over all such permutations. Further, the absolute maximum of tr PXP'D is, clearly,

$$\max_{P:PP'=I} \mathrm{tr}\ PXP'D = \sum_{i=1}^{m} d_i\ ch_i(X),$$

and the absolute minimum is, clearly,

$$\min_{P:PP'=I} \operatorname{tr} PXP'D = \sum_{i=1}^{m} d_{m+1-i}\, ch_i(X).$$

We will also need the following results, the first of which can be found in Bellman [1970:117].

Lemma 2.2.2: Let $X(m\times m)$ and $Y(m\times m)$ be symmetric matrices with Y nonnegative definite. Then

$$ch_i(X+Y) \geq ch_i(X) \qquad \text{for } i = 1,2,\ldots,m.$$

If Y is positive definite, then

$$ch_i(X+Y) > ch_i(X) \qquad \text{for } i = 1,2,\ldots,m.$$

Lemma 2.2.3: The function $\phi(A,B;D,e,h)$ has an absolute minimum over the set of solutions $C_s = \{(A,B): A \in P_m,\ B \in P_m,\ B - A \in \bigcup_{j=0}^{s} P_j\}$.

Proof: Since B is positive definite, it follows that B^{-1} is also positive definite, so that the diagonal elements of B^{-1} are positive. Then we find that

$$\operatorname{tr} B^{-1}D = \sum_{i=1}^{m} (B^{-1})_{ii}\, d_i \geq d_m \sum_{i=1}^{m} (B^{-1})_{ii} = d_m \operatorname{tr} B^{-1}$$

$$= d_m \sum_{i=1}^{m} ch_i(B^{-1}) = d_m \sum_{i=1}^{m} (ch_i(B))^{-1},$$

since $ch_i(B^{-1}) = (ch_{m+1-i}(B))^{-1}$. Hence, using the fact that for any matrix $X(m\times m)$, $\operatorname{tr} X = \sum_{i=1}^{m} ch_i(X)$ and $|X| = \prod_{i=1}^{m} ch_i(X)$, we see that

$$\phi(A,B;D,e,h) \geq e \sum_{i=1}^{m} ((ch_i(A))^{-1} + \ell n(ch_i(A)))$$

$$+ h \sum_{i=1}^{m} (d_m(ch_i(B))^{-1} + \ell n(ch_i(B))). \qquad (2.2.3)$$

From Lemma 2.2.2 we know that $ch_i(B-A) < ch_i(B)$, since A is positive definite. Then C_s can be written

$C_s = \{(A,B): ch_i(A) > 0: i = 1,2,...,m; ch_i(B) > 0: i = 1,2,...,m; 0 \leq ch_i(B-A) < ch_i(B): i = 1,2,...,s; ch_i(B-A) = 0: i = s+1,...,m; A = A', B = B'\}$. The closure, $\overline{C}_s$, of C_s is $\{(A,B): ch_i(A) \geq 0: i = 1,2,...,m; ch_i(B) \geq 0: i = 1,2,...,m; 0 \leq ch_i(B-A) \leq ch_i(B): i = 1,2,...,s; ch_i(B-A) = 0: i = s+1,...,m; A = A', B = B'\}$. Since $\phi(A,B;D,e,h) \geq 0$, it has an absolute minimum over $\overline{C}_s$, since $\overline{C}_s$ is closed. Note that from Lemma 2.2.2 if $ch_i(B-A) = ch_i(B)$ for some i, then it must be true that $ch_m(A) = 0$, since A must then be positive semidefinite. Thus, for every $(A,B) \in \overline{C}_s - C_s$ it must be true that $ch_m(A) = 0$ or $ch_m(B) = 0$ or both. It then follows from (2.2.3) that $\phi(A,B;D,e,h) = \infty$ whenever $(A,B) \in \overline{C}_s - C_s$. Hence, $\phi(A,B;D,e,h)$ has an absolute minimum over C_s.

Lemma 2.2.4: Suppose the function f(x), minimized over $x \in S$, achieves a minimum at x = a. Let the set S_1 be such that for any $x \in S-S_1$, there exists an $x_1 \in S_1$ such that $f(x_1) < f(x)$. Similarly, let the set S_2 be such that for any $x \in S-S_2$, there exists an $x_2 \in S_2$ such that $f(x_2) < f(x)$. Then it follows that $a \in S_1 \cap S_2$.

Proof: Suppose $a \notin S_1 \cap S_2$. Then either $a \notin S_1$ or $a \notin S_2$ or both. However, if $a \notin S_1$, then $a \in S-S_1$, and there exists no $x_1 \in S_1$ such that $f(x_1) < f(a)$, since f is minimized at a. This then is a contradiction, so it must be true that $a \in S_1$. Similarly, if $a \notin S_2$, then $a \in S-S_2$, and there exists no $x_2 \in S_2$ such that $f(x_2) < f(a)$. This also is a contradiction, so it must be true that $a \in S_2$. Hence, it follows that $a \in S_1 \cap S_2$.

In Lemma 2.2.3 it was seen that the function $\phi(A,B;D,e,h)$ has an absolute minimum over the set C_s. We will now show that this absolute minimum will occur only when both A and B are diagonal.

Lemma 2.2.5: The absolute minimum of $\phi(A,B;D,e,h)$ subject to $(A,B) \in C_s$ occurs when both A and B are diagonal. We offer two proofs.

Proof 1: Define the sets S_1 and S_2 as follows:

$$S_1 = \{(A,B) \in C_s: \text{ A is diagonal}\},$$

$$S_2 = \{(A,B) \in C_s: \text{ B is diagonal}\}.$$

We want to show that if $\phi(A,B;D,e,h)$ achieves a minimum at $(A,B) = (A_*,B_*)$, then $(A_*,B_*) \in S_1 \cap S_2$. Now with $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\tilde{B} = D^{-\frac{1}{2}}BD^{-\frac{1}{2}}$,

$$\begin{aligned}\phi(A,B;D,e,h) &= e[\mathrm{tr}A^{-1}+\ell n|A|] + h[\mathrm{tr}B^{-1}D+\ell n|B|] \\ &= e[\mathrm{tr}\tilde{A}^{-1}D^{-1}+\ell n|\tilde{A}|] + h[\mathrm{tr}\tilde{B}^{-1}+\ell n|\tilde{B}|] + (e+h)\ell n|D| \\ &= \phi(\tilde{B},\tilde{A};D^{-1},h,e) + (e+h)\ell n|D|.\end{aligned}$$

Note that since $D^{-\frac{1}{2}}$ is positive definite, $(A,B) \in C_s$ if and only if $(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}, D^{-\frac{1}{2}}BD^{-\frac{1}{2}}) = (\tilde{A},\tilde{B}) \in C_s$. Thus, minimizing $\phi(A,B;D,e,h)$ subject to $(A,B) \in C_s$ is equivalent to minimizing $\phi(\tilde{B},\tilde{A};D^{-1},h,e)$ subject to $(\tilde{A},\tilde{B}) \in C_s$. Moreover, if $(\tilde{A}_*,\tilde{B}_*)$ minimizes $\phi(\tilde{B},\tilde{A};D^{-1},h,e)$, then $(D^{\frac{1}{2}}\tilde{A}_*D^{\frac{1}{2}}, D^{\frac{1}{2}}\tilde{B}_*D^{\frac{1}{2}})$ minimizes $\phi(A,B;D,e,h)$. Now arbitrarily fix $(\tilde{A},\tilde{B}) \in C_s$ and consider $\phi(P\tilde{B}P', P\tilde{A}P'; D^{-1},h,e)$ for all orthogonal P. Clearly the terms $\ell n|P\tilde{A}P'|$, tr $P\tilde{B}^{-1}P'$, and $\ell n|P\tilde{B}P'|$ are constant for all orthogonal P, so that $\phi(P\tilde{B}P', P\tilde{A}P'; D^{-1},h,e)$ is minimized with respect to P when tr $P\tilde{A}^{-1}P'D^{-1}$ is minimized. It follows from Lemma 2.2.1 that all the stationary points, and thus the absolute minimum, occur when $P\tilde{A}P'$ is diagonal. Hence, for any $(\tilde{A},\tilde{B}) \in C_s - S_1$ there exists an $(\tilde{A}_1,\tilde{B}_1) \in S_1$ such that

$$\phi(\tilde{B}_1,\tilde{A}_1;D^{-1},h,e) < \phi(\tilde{B},\tilde{A};D^{-1},h,e).$$

But since $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, we know that $\tilde{A}$ is diagonal if and only if A is diagonal. So we find that for any $(A,B) \in C_s - S_1$, there exists an $(A_1,B_1) \in S_1$ such that

$$\phi(A_1,B_1;D,e,h) < \phi(A,B;D,e,h).$$

In a similar manner now arbitrarily fix $(A,B) \in C_s$ and consider $\phi(PAP',PBP';D,e,h)$ for all orthogonal P. Clearly this is minimized with respect to P when tr $PB^{-1}P'D$ is minimized, since the terms tr $PA^{-1}P'$, $\ell n|PAP'|$, and $\ell n|PBP'|$ are constant for all orthogonal P. So from Lemma 2.2.1 it follows that all the stationary points, and therefore the absolute minimum, of $\phi(PAP',PBP';D,e,h)$ occur when PBP' is diagonal.

This implies that for any $(A,B) \in \mathcal{C}_s - \mathcal{S}_2$, there exists an $(A_2,B_2) \in \mathcal{S}_2$ such that

$$\phi(A_2,B_2;D,e,h) < \phi(A,B;D,e,h).$$

The result now follows from Lemma 2.2.4. Furthermore, from Lemma 2.2.1 we see that if (A_*,B_*) minimizes $\phi(A,B;D,e,h)$, then the diagonal elements of $D^{-\frac{1}{2}}A_*D^{-\frac{1}{2}}$ are increasing and the diagonal elements of B_* are decreasing.

The second proof of Lemma 2.2.5 utilizes the concept of "majorization" (see Marshall and Olkin [1974]).

Definition 2.2.6: Let $\underline{x}$ and $\underline{y}$ be real $m\times 1$ vectors with i^{th} element x_i and y_i, respectively, and i^{th} largest element $x_{(i)}$ and $y_{(i)}$, respectively. We say that $\underline{x}$ majorizes $\underline{y}$ and write $\underline{x} \overset{m}{>} \underline{y}$, if

$$\sum_{i=1}^{s} x_{(i)} \geq \sum_{i=1}^{s} y_{(i)} \qquad \text{for} \quad s = 1,2,\ldots,m,$$

with equality when $s = m$.

We will need some results which, while well known to workers in the area of majorization, may not be readily accessible to others. We prove the results here for the benefit of the uninitiated reader.

Lemma 2.2.7: If S $(m\times m)$ is doubly stochastic, then $\underline{x} \overset{m}{>} S\underline{x} = \underline{y}$.

Proof: Since S is doubly stochastic, it follows that $s_{ij} \geq 0$ for all (i,j), and

$$\sum_{j=1}^{m} s_{ij} = 1 \qquad \text{for} \quad i = 1,2,\ldots,m,$$

$$\sum_{i=1}^{m} s_{ij} = 1 \qquad \text{for} \quad j = 1,2,\ldots,m.$$

Thus, for $1 \le t \le m$ there exists $k_1, k_2, \ldots, k_t$ such that

$$\sum_{i=1}^{t} y_{(i)} = \sum_{j=1}^{m} (s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j})x_j.$$

Clearly when $t < m$,

$$s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j} \le \sum_{i=1}^{m} s_{ij} = 1 \qquad \text{for } j = 1,2,\ldots,m,$$

and

$$\sum_{j=1}^{m} (s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j}) = t.$$

Then when $t < m$,

$$\sum_{i=1}^{t} y_{(i)} = \sum_{j=1}^{m} (s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j})x_j$$

$$\le \sum_{i=1}^{t} x_{(i)}.$$

If $t = m$, then

$$s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j} = \sum_{i=1}^{t} s_{ij} = 1,$$

so that

$$\sum_{i=1}^{t} y_{(i)} = \sum_{j=1}^{t} (s_{k_1 j}+s_{k_2 j}+\ldots+s_{k_t j})x_j$$

$$= \sum_{j=1}^{t} x_j = \sum_{i=1}^{t} x_{(i)}.$$

Lemma 2.2.8: If $\underline{x} \overset{m}{>} \underline{y}$ and $\sigma_{(1)} \geq \sigma_{(2)} \geq \ldots \geq \sigma_{(m)} \geq 0$, then

$$\sum_{i=1}^{m} x_{(i)}\sigma_{(i)} \geq \sum_{i=1}^{m} y_i\sigma_{(i)}.$$

Proof: Put $d_i = x_{(i)} - y_i$. Then

$$\begin{aligned}\sum_{i=1}^{m} (x_{(i)}-y_i)\sigma_{(i)} &= \sum_{i=1}^{m} d_i\sigma_{(i)} \\ &= d_1(\sigma_{(1)}-\sigma_{(2)}) + \\ &\quad (d_1+d_2)(\sigma_{(2)}-\sigma_{(3)}) + \\ &\quad (d_1+d_2+d_3)(\sigma_{(3)}-\sigma_{(4)}) + \\ &\quad \vdots \\ &\quad (d_1+d_2+\ldots+d_{m-1})(\sigma_{(m-1)}-\sigma_{(m)}) + \\ &\quad (d_1+d_2+\ldots+d_m)\sigma_{(m)}.\end{aligned}$$

The last term is zero, since

$$\sum_{i=1}^{m} d_i = \sum_{i=1}^{m} (x_{(i)}-y_i) = \sum_{i=1}^{m} x_{(i)} - \sum_{i=1}^{m} y_{(i)} = 0.$$

The partial sums are nonnegative, since

$$\sum_{i=1}^{t} d_i = \sum_{i=1}^{t} x_{(i)} - \sum_{i=1}^{t} y_i \geq \sum_{i=1}^{t} x_{(i)} - \sum_{i=1}^{t} y_{(i)} \geq 0.$$

Further, the differences $\sigma_{(1)}-\sigma_{(2)}, \sigma_{(2)}-\sigma_{(3)}, \ldots, \sigma_{(m-1)}-\sigma_{(m)}$ are nonnegative. Hence, the result follows.

Lemma 2.2.9, Corollary: If $\underline{x}$ is an ordered vector, that is, $x_1 \geq x_2 \geq \ldots \geq x_m$, S is doubly stochastic, and $\underline{\sigma}$ is also an ordered vector, then $\underline{x}'\underline{\sigma} \geq (S\underline{x})'\underline{\sigma}$.

Lemma 2.2.10: If $\underline{x} \overset{m}{>} \underline{y}$ and $\sigma_{(1)} \geq \sigma_{(2)} \geq \ldots \geq \sigma_{(m)} \geq 0$, then

$$\sum_{i=1}^{m} x_{(i)} \sigma_{(m+1-i)} \leq \sum_{i=1}^{m} y_i \sigma_{(m+1-i)}.$$

Proof: The proof is similar to that of Lemma 2.2.8. Letting $d_i = x_{(i)} - y_i$, we have

$$\begin{aligned}\sum_{i=1}^{m} (x_{(i)} - y_i) \sigma_{(m+1-i)} &= \sum_{i=1}^{m} d_i \sigma_{(m+1-i)} \\ &= d_1(\sigma_{(m)} - \sigma_{(m-1)}) + \\ &\quad (d_1+d_2)(\sigma_{(m-1)} - \sigma_{(m-2)}) + \\ &\quad (d_1+d_2+d_3)(\sigma_{(m-2)} - \sigma_{(m-3)}) + \\ &\quad \vdots \\ &\quad (d_1+d_2+\ldots+d_{m-1})(\sigma_{(2)} - \sigma_{(1)}) + \\ &\quad (d_1+d_2+\ldots+d_m)\sigma_{(1)}.\end{aligned}$$

We have seen that the partial sums, $\sum_{i=1}^{t} d_i : t=1,\ldots,m-1$, are nonnegative and $\sum_{i=1}^{m} d_i$ is zero, so that the last term is zero. Further, the differences $\sigma_{(m)} - \sigma_{(m-1)}, \sigma_{(m-1)} - \sigma_{(m-2)}, \ldots,$ $\sigma_{(2)} - \sigma_{(1)}$ are negative or zero. Hence, the result follows.

Lemma 2.2.11, Corollary: If $\underline{x}$ is an ordered vector, and $\underline{y} = S\underline{x}$ with S doubly stochastic, then

$$\sum_{i=1}^{m} x_i \sigma_{(m+1-i)} \leq \sum_{i=1}^{m} y_i \sigma_{(m+1-i)}.$$

Furthermore, if $\sigma_{(1)} > \sigma_{(2)} > \ldots > \sigma_{(m)}$, then there is equality only if $\underline{y} = \underline{x}$.

We are now ready for the second proof of Lemma 2.2.5. Recall that we need to show that the absolute minimum of $\phi(A,B;D,e,h)$ subject to $(A,B) \in C_s$ occurs when both A and B are diagonal.

Proof 2 (Lemma 2.2.5): Let S_1 and S_2 be defined as before; that is,

$$S_1 = \{(A,B) \in C_s : A \text{ is diagonal}\},$$

$$S_2 = \{(A,B) \in C_s : B \text{ is diagonal}\},$$

and recall that we need to show that if $\phi(A,B;D,e,h)$ is minimized at (A_*,B_*), then $(A_*,B_*) \in S_1 \cap S_2$. Let $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m > 0$ be the latent roots of B^{-1}, and $PB^{-1}P' = \text{diag}(\beta_m,\beta_{m-1},\ldots,\beta_1)$. Then

$$\{\phi(PAP',PBP';D,e,h) - \phi(A,B;D,e,h)\}/h$$

$$= \text{tr } PB^{-1}P'D - \text{tr } B^{-1}D$$

$$= \sum_{j=1}^{m} \beta_{m+1-j}d_j - \sum_{j=1}^{m}\left(\sum_{i=1}^{m} \beta_{m+1-i}p_{ij}^2\right)d_j$$

$$= \sum_{j=1}^{m} \beta_j d_{m+1-j} - \sum_{j=1}^{m} y_j d_{m+1-j}, \qquad (2.2.4)$$

where $\underline{y} = P_2\underline{\beta}$, $\underline{\beta}' = (\beta_1,\beta_2,\ldots,\beta_m)$, and P_2 is the matrix with $(i,j)^{th}$ element $p_{m+1-j,m+1-i}^2$. Since $PP' = P'P = I$, we see that P_2 is doubly stochastic. Also $\underline{d} = (d_1,d_2,\ldots,d_m)$ and $\underline{\beta}$ are ordered vectors, so by Lemma 2.2.11, equation (2.2.4) is not positive. Furthermore, $d_1 > d_2 > \cdots > d_m$ so that

$$\phi(PAP',PBP';D,e,h) \leq \phi(A,B;D,e,h),$$

with equality holding only when $B^{-1} = \text{diag}(\beta_m, \beta_{m-1}, \ldots, \beta_1)$. Therefore, for any $(A,B) \in \mathcal{C}_s - \mathcal{S}_2$ there exists an $(A_2,B_2) \in \mathcal{S}_2$ such that

$$\phi(A_2,B_2;D,e,h) < \phi(A,B;D,e,h).$$

Now with $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and $\tilde{B} = D^{-\frac{1}{2}}BD^{-\frac{1}{2}}$

$$\phi(A,B;D,e,h) = e[\text{tr}\tilde{A}^{-1}D^{-1} + \ell n|\tilde{A}|] + h[\text{tr}\tilde{B}^{-1} + \ell n|\tilde{B}|] + (e+h)\ell n|D| \qquad (2.2.5)$$

$$= \phi(\tilde{B},\tilde{A};D^{-1},h,e) + (e+h)\ell n|D|.$$

Let $\tilde{\alpha}_1 \geq \tilde{\alpha}_2 \geq \ldots \geq \tilde{\alpha}_m > 0$ be the latent roots of $\tilde{A}^{-1}$ and $Q\tilde{A}^{-1}Q' = \text{diag}(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_m)$. Then by an argument identical to the previous one we find that

$$\phi(Q\tilde{B}Q', Q\tilde{A}Q'; D^{-1},h,e) \leq \phi(\tilde{B},\tilde{A};D^{-1},h,e),$$

with equality holding only when $\tilde{A}^{-1} = \text{diag}(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_m)$. From (2.2.5) it follows that

$$\phi(D^{\frac{1}{2}}Q\tilde{A}Q'D^{\frac{1}{2}}, D^{\frac{1}{2}}Q\tilde{B}Q'D^{\frac{1}{2}};D,e,h) \leq \phi(D^{\frac{1}{2}}\tilde{A}D^{\frac{1}{2}}, D^{\frac{1}{2}}\tilde{B}D^{\frac{1}{2}};D,e,h),$$

with equality holding only when $\tilde{A}^{-1} = \text{diag}(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_m)$. Note that $D^{\frac{1}{2}}Q\tilde{A}Q'D^{\frac{1}{2}} = \text{diag}(d_1\tilde{\alpha}_1^{-1}, \ldots, d_m\tilde{\alpha}_m^{-1})$. Thus, for any $(A,B) \in \mathcal{C}_s - \mathcal{S}_1$ there exists an $(A_1,B_1) \in \mathcal{S}_1$ such that

$$\phi(A_1,B_1;D,e,h) < \phi(A,B;D,e,h).$$

The result now follows from Lemma 2.2.4.

<u>Lemma 2.2.12, Corollary</u>: Let R be some restriction on the latent roots of A or B or both, and let $\mathcal{C}_s^R$ be the subset of $\mathcal{C}_s$ such that $(A,B) \in \mathcal{C}_s^R$ implies that R is satisfied. Since $(A,B) \in \mathcal{C}_s^R$ if and only if $(PAP',PBP') \in \mathcal{C}_s^R$ for any orthogonal P, it follows that the minimal value of

$\phi(A,B;D,e,h)$ over $(A,B) \varepsilon \mathcal{C}_s^R$ occurs when A and B are diagonal. For example, if the latent roots of A were known to be proportional to a given set, then the minimal value of $\phi(A,B;D,e,h)$ over $(A,B) \varepsilon \mathcal{C}_s^R$ occurs when A is a diagonal matrix with diagonal elements proportional to this set.

2.3 The Maximum Likelihood Estimates

In this section we seek the maximum likelihood estimates of Σ and M subject to the constraints $\Sigma \varepsilon \mathcal{P}_m$ and $M \varepsilon \bigcup_{j=0}^{s} \mathcal{P}_j$. Recall that the likelihood function of (Σ,M) is

$$f(E,H) = \frac{K_m(I,e)K_m(I,h)}{|\Sigma+M|^{\frac{1}{2}h}|\Sigma|^{\frac{1}{2}e}} |H|^{\frac{1}{2}(h-m-1)}|E|^{\frac{1}{2}(e-m-1)}$$

$$\times \exp[-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}E-\tfrac{1}{2}\mathrm{tr}(\Sigma+M)^{-1}H].$$

The logarithm of the likelihood function, omitting a function of the observations, is

$$-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}E-\tfrac{1}{2}e\ell n|\Sigma| - \tfrac{1}{2}\mathrm{tr}(\Sigma+M)^{-1}H-\tfrac{1}{2}h\ell n|\Sigma+M|.$$

We seek the solution, $(\hat{\Sigma},\hat{M})$, which maximizes the above equation, or equivalently, the solution which minimizes

$$\mathrm{tr}\Sigma^{-1}E+e\ell n|\Sigma| + \mathrm{tr}(\Sigma+M)^{-1}H+h\ell n|\Sigma+M| \tag{2.3.1}$$

subject to $\Sigma \varepsilon \mathcal{P}_m$ and $M \varepsilon \bigcup_{j=0}^{s} \mathcal{P}_j$.

Let $E_* = (1/e)E$ and $H_* = (1/h)H$. Note that since E_* and H_* are both symmetric matrices, and $E_* \varepsilon \mathcal{P}_m$ and $H_* \varepsilon \bigcup_{j=0}^{m} \mathcal{P}_j$, there exists a nonsingular matrix $K(m\times m)$ such that $KE_*K' = I$ and $KH_*K' = D$, where $D = \mathrm{diag}(d_1,d_2,\ldots,d_m)$, and $d_1 > d_2 > \ldots > d_m > 0$ are the latent roots of $H_*E_*^{-1}$.

Then with $\tilde{\Sigma} = K\Sigma K'$ and $\tilde{M} = KMK'$, (2.3.1) can be rewritten

$$\begin{aligned} &\text{etr}K'^{-1}\Sigma^{-1}K^{-1}I + e\ell n|\Sigma| + h\text{tr}K'^{-1}(\Sigma+M)^{-1}K^{-1}D + h\ell n|\Sigma+M| \\ &\quad = e[\text{tr}\tilde{\Sigma}^{-1} + \ell n|\tilde{\Sigma}|] + h[\text{tr}(\tilde{\Sigma}+\tilde{M})^{-1}D + \ell n|\tilde{\Sigma}+\tilde{M}|] \\ &\qquad\qquad - (e+h)\ell n|K|^2 \\ &\quad = \phi(\tilde{\Sigma},\tilde{\Sigma}+\tilde{M};D,e,h) - (e+h)\ell n|K|^2. \end{aligned}$$

Thus, the problem has been reduced to that of minimizing $\phi(\tilde{\Sigma},\tilde{\Sigma}+\tilde{M};D,e,h)$ subject to $\tilde{\Sigma} \in P_m$ and $\tilde{M} \in \bigcup_{j=0}^{s} P_j$ or, equivalently, $(\tilde{\Sigma},\tilde{\Sigma}+\tilde{M}) \in C_s$. But from Lemma 2.2.5 it is known that the minimal solution to $\phi(\tilde{\Sigma},\tilde{\Sigma}+\tilde{M};D,e,h)$ is such that $\tilde{\Sigma}$ and $\tilde{\Sigma}+\tilde{M}$ are diagonal, and in addition, it is known that the diagonal elements of $D^{-\frac{1}{2}}\tilde{\Sigma}D^{-\frac{1}{2}}$ are increasing while the diagonal elements of $\tilde{\Sigma}+\tilde{M}$ are decreasing.

Consider the function

$$g(x,y) = e(\frac{1}{x} + \ell nx) + h(\frac{d}{y} + \ell ny), \qquad (2.3.2)$$

where $d > 0$. Differentiating (2.3.2) with respect to x and y, we get the equations

$$-\frac{1}{x^2} + \frac{1}{x} = 0,$$

$$-\frac{d}{y^2} + \frac{1}{y} = 0,$$

which yield the minimal solution $x_0 = 1$ and $y_0 = d$. If instead we wanted to minimize (2.3.2) subject to $x = y$, (2.3.2) would reduce to

$$g(x) = e(\frac{1}{x} + \ell nx) + h(\frac{d}{x} + \ell nx). \qquad (2.3.3)$$

Then

$$\frac{dg(x)}{dx} = e(-\frac{1}{x^2} + \frac{1}{x}) + h(-\frac{d}{x^2} + \frac{1}{x}) = 0,$$

so that $x_1 = y_1 = \frac{e+dh}{e+h}$ minimizes (2.3.3).

Now let

$$\begin{aligned} f(d) &= g(x_1,y_1) - g(x_0,y_0) \\ &= e(\frac{e+h}{e+dh} + \ell n(\frac{e+dh}{e+h})) + h(\frac{d(e+h)}{e+dh} + \ell n(\frac{e+dh}{e+h})) \\ &\qquad - (e+h+h\ell nd) \\ &= e(1 - \frac{(d-1)h}{e+dh}) + h(1 + \frac{(d-1)e}{e+dh}) + (e+h)\ell n(\frac{e+dh}{e+h}) \\ &\qquad - (e+h+h\ell nd) \\ &= (e+h)\ell n(\frac{e+dh}{e+h}) - h\ell nd. \end{aligned}$$

Differentiating $f(d)$ with respect to d and noting that $e \geq 1$, $h \geq 1$, we find that when $d > 1$

$$\frac{df(d)}{dd} = \frac{h(e+h)}{e+dh} - \frac{h}{d} = \frac{dh(e+h)-h(e+dh)}{(e+dh)d} = \frac{eh(d-1)}{(e+dh)d} > 0.$$

In other words, the difference $g(x_1,y_1) - g(x_0,y_0)$ is an increasing function of d when $d > 1$.

Now with $X = \text{diag}(x_1,x_2,\ldots,x_m)$ and $Y = \text{diag}(y_1,y_2,\ldots,y_m)$ consider minimizing

$$\phi(X,Y;D,e,h) = e\sum_{i=1}^{m}(\frac{1}{x_i} + \ell n\ x_i) + h\sum_{i=1}^{m}(\frac{d_i}{y_i} + \ell n\ y_i) \qquad (2.3.4)$$

subject to $(X,Y) \in \mathcal{C}_s$, which in this case implies that $y_i \geq x_i > 0$ for all i, and $x_i = y_i$ for at least $m - s$ of the i's. Suppose that $d_1 > d_2 > \ldots > d_r > 1 > d_{r+1} > \ldots > d_m > 0$.

Using the fact that $f(d)$ is increasing in d for $d > 1$, it then follows that the minimal solution to (2.3.4) is (X_s, Y_s), where if $r \geq s$,

$$\begin{cases} x_{si} = y_{si} = (e+d_i h)/(e+h) & \text{for} \quad s+1 \leq i \leq m, \\ x_{si} = 1,\ y_{si} = d_i & \text{for} \quad 1 \leq i \leq s, \end{cases}$$

and if $r < s$,

$$\begin{cases} x_{si} = y_{si} = (e+d_i h)/(e+h) & \text{for} \quad r+1 \leq i \leq m, \\ x_{si} = 1,\ y_{si} = d_i & \text{for} \quad 1 \leq i \leq r. \end{cases}$$

Thus, $\phi(\tilde{\Sigma}, \tilde{\Sigma}+\tilde{M}; D, e, h)$ is minimized subject to $(\tilde{\Sigma}, \tilde{\Sigma}+\tilde{M}) \in \mathcal{C}_s$ at

$$\tilde{\Sigma} = X_s,$$

$$\tilde{M} = Y_s - X_s,$$

so that the maximum likelihood estimates of Σ and M are $\hat{\Sigma}$ and $\hat{M}$, where

$$\hat{\Sigma} = K^{-1} X_s K'^{-1},$$

$$\hat{M} = K^{-1} (Y_s - X_s) K'^{-1}.$$

We now present an example to illustrate the computation involved in deriving the maximum likelihood estimates. Consider model (2.1.1) in which we take $m = 4$, $g = 21$, $n = 6$, $\Sigma = I$, and $M = \text{diag}(99,24,0,0)$. Hence, $e = g(n-1) = 105$ and $h = g-1 = 20$. Generating a matrix E from the distribution $W_4(I,105,0)$ and a matrix H from the distribution $W_4(I+M,20,0)$, we obtain

$$E = \begin{bmatrix} 69.1329 & 4.07476 & -5.12762 & -9.94924 \\ & 127.055 & -3.77638 & 20.4629 \\ & & 116.342 & 8.12511 \\ & & & 100.186 \end{bmatrix},$$

$$H = \begin{bmatrix} 1845.85 & 63.5986 & -16.5227 & -1.43363 \\ & 688.962 & 1.14908 & -8.61601 \\ & & 20.1453 & -.0100181 \\ & & & 12.2617 \end{bmatrix}.$$

With $E_* = (1/105)E$ and $H_* = (1/20)H$ we need to find a non-singular matrix K such that $KE_*K' = I$ and $KH_*K' = D$, where D is a diagonal matrix. Let $D_1 = \text{diag}(ch_1(E_*),\ldots,ch_4(E_*))$, and let P be the orthogonal matrix for which the i^{th} column is the characteristic vector of E_* corresponding to $ch_i(E_*)$, then, since E_* is symmetric, $P'E_*P = D_1$. Similarly, let $D = \text{diag}(ch_1(D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}}),\ldots,ch_4(D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}}))$, and let Q be the orthogonal matrix for which the i^{th} column is the characteristic vector of $D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}}$ corresponding to $ch_i(D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}})$, then, since $D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}}$ is symmetric, $Q'D_1^{-\frac{1}{2}}P'H_*PD_1^{-\frac{1}{2}}Q = D$. Thus, we may take $K = Q'D_1^{-\frac{1}{2}}P'$. Using the above decomposition for K, we find that, for our example,

$$K = \begin{bmatrix} 1.24522 & -.0464049 & .0380069 & .130133 \\ .0181884 & -.925978 & -.0477042 & .213476 \\ .00831611 & -.00767896 & .940375 & -.237376 \\ -.000914637 & -.0158712 & -.153048 & -.994866 \end{bmatrix},$$

and $D = \text{diag}(142.729, 29.6669, .91847, .625404)$. Note that $d_2 > 1$ and $d_3 < 1$, so that $r = 2$. Simple calculation yields

$X_0 = Y_0 = \text{diag}(23.6766, 5.5867, .986955, .940065)$,

$X_1 = \text{diag}(1, 5.5867, .986955, .940065)$,

Y_1 = diag(142.729, 5.5867, .986955, .940065),

$X_2 = X_3 = X_4 =$ diag(1, 1, .986955, .940065),

$Y_2 = Y_3 = Y_4 =$ diag(142.729, 29.6669, .986955, .940065).

Hence, if we let $\hat{\Sigma}_i$ and $\hat{M}_i$ be the maximum likelihood estimates of Σ and M, respectively, subject to the constraints $\Sigma \varepsilon P_4$ and $M \varepsilon \bigcup_{j=0}^{i} P_j$, we find that

$$\hat{\Sigma}_0 = \begin{bmatrix} 15.3199 & .541388 & -.173203 & -.0910632 \\ & 6.52813 & -.0210184 & .0947756 \\ & & 1.0919 & .064921 \\ & & & .899582 \end{bmatrix},$$

$$\hat{M}_0 = (0),$$

$$\hat{\Sigma}_1 = \begin{bmatrix} .665836 & .246703 & -.046356 & -.0924036 \\ & 6.5222 & -.0184676 & .0947486 \\ & & 1.0908 & .0649326 \\ & & & .899582 \end{bmatrix},$$

$$\hat{M}_1 = \begin{bmatrix} 91.5881 & 1.84178 & -.792796 & .00837764 \\ & .0370371 & -.0159426 & .000168469 \\ & & .00686252 & -.000072518 \\ & & & .000000766 \end{bmatrix},$$

$$\hat{\Sigma}_2 = \hat{\Sigma}_3 = \hat{\Sigma}_4 = \begin{bmatrix} .6578 & .040037 & -.0471092 & -.0889834 \\ & 1.20736 & -.0378372 & .182707 \\ & & 1.09073 & .0652532 \\ & & & .898126 \end{bmatrix},$$

$$\hat{M}_2=\hat{M}_3=\hat{M}_4=\begin{bmatrix} 91.6383 & 3.13344 & -.788088 & -.0129987 \\ & 33.2548 & .105117 & -.549569 \\ & & .00730371 & -.002076 \\ & & & .00909865 \end{bmatrix}.$$

Further commentary on these data will be made in Sections 3.6, 4.2, and 5.4.

2.4 The Likelihood Ratio Test

Recall that $C_s = \{(A,B): A\varepsilon P_m, B\varepsilon P_m, B-A\varepsilon \bigcup_{j=0}^{s} P_j\}$,

and suppose we know that $(\Sigma,\Sigma+M)\ \varepsilon\ \Omega = C_s$. We wish to test, say, the null hypothesis that $(\Sigma,\Sigma+M)\varepsilon\ \omega = C_{s-1} \subset C_s$. The alternative hypothesis is then that $(\Sigma,\Sigma+M)\varepsilon\ \Omega - \omega =$ $C_s - C_{s-1}$. Thus, we are testing the hypothesis

$$H_0^{(s)}: \text{rank } (M) \leq s-1$$

against the hypothesis

$$H_1^{(s)}: \text{rank } (M) = s.$$

We adopt the likelihood approach and compare $\max_{\omega} f(E,H)$ with $\max_{\Omega} f(E,H)$. Specifically, we look at

$\max_{\omega} f(E,H)/\max_{\Omega} f(E,H) = \lambda\ \varepsilon\ (0,1]$.

With the matrices $X_s = \text{diag}(x_{s1},x_{s2},\ldots,x_{sm})$ and $Y_s = \text{diag}(y_{s1},y_{s2},\ldots,y_{sm})$ given by

$$\begin{cases} x_{si} = y_{si} = (e+d_i h)/(e+h) & \text{for } s+1 \leq i \leq m, \\ x_{si} = 1, y_{si} = d_i & \text{for } 1 \leq i \leq s, \end{cases}$$

if $r \geq s$, and

$$\begin{cases} x_{si} = y_{si} = (e+d_i h)/(e+h) & \text{for } r+1 \leq i \leq m, \\ x_{si} = 1,\ y_{si} = d_i & \text{for } 1 \leq i \leq r, \end{cases}$$

if $r < s$, the maximum likelihood estimators, $\hat{\Sigma}_\Omega$, of Σ and, $\hat{M}_\Omega$, of M when the parameters are restricted to lie within Ω, are given by

$$\hat{\Sigma}_\Omega = K^{-1} X_s K'^{-1},$$

$$\hat{M}_\Omega = K^{-1}(Y_s - X_s)K'^{-1},$$

where K is a nonsingular matrix. Similarly, the maximum likelihood estimators, $\hat{\Sigma}_\omega$, of Σ and, $\hat{M}_\omega$, of M where the parameters are restricted to lie within ω, are given by

$$\hat{\Sigma}_\omega = K^{-1} X_{s-1} K'^{-1},$$

$$\hat{M}_\omega = K^{-1}(Y_{s-1} - X_{s-1})K'^{-1}.$$

It should be noted that if $r < s$, then $X_s = X_{s-1}$ and $Y_s = Y_{s-1}$, and if $r \geq s$, $x_{si} = x_{s-1,i}$ and $y_{si} = y_{s-1,i}$ only for $i \neq s$.

The likelihood ratio, λ, is

$$\lambda = \frac{\max_\omega f(E.H)}{\max_\Omega f(E,H)}$$

$$= \frac{\exp[-\tfrac{1}{2}\mathrm{tr}\hat{\Sigma}_\omega^{-1}E - \tfrac{1}{2}\mathrm{tr}(\hat{\Sigma}_\omega + \hat{M}_\omega)^{-1}H]}{\exp[-\tfrac{1}{2}\mathrm{tr}\hat{\Sigma}_\Omega^{-1}E - \tfrac{1}{2}\mathrm{tr}(\hat{\Sigma}_\Omega + \hat{M}_\Omega)^{-1}H]} \; \frac{|\hat{\Sigma}_\Omega + \hat{M}_\Omega|^{\frac{1}{2}h} |\hat{\Sigma}_\Omega|^{\frac{1}{2}e}}{|\hat{\Sigma}_\omega + \hat{M}_\omega|^{\frac{1}{2}h} |\hat{\Sigma}_\omega|^{\frac{1}{2}e}}$$

$$= \frac{\exp[-\frac{1}{2}e\,tr X_{s-1}^{-1} - \frac{1}{2}h\,tr Y_{s-1}^{-1} D]}{\exp[-\frac{1}{2}e\,tr X_s^{-1} - \frac{1}{2}h\,tr Y_s^{-1} D]} \quad \frac{|Y_s|^{\frac{1}{2}h} |X_s|^{\frac{1}{2}e}}{|Y_{s-1}|^{\frac{1}{2}h} |X_{s-1}|^{\frac{1}{2}e}}$$

$$= \frac{|Y_s|^{\frac{1}{2}h} |X_s|^{\frac{1}{2}e}}{|Y_{s-1}|^{\frac{1}{2}h} |X_{s-1}|^{\frac{1}{2}e}} ,$$

since, if $r < s$,

$$e\,tr(X_{s-1}^{-1} - X_s^{-1}) + h\,tr(Y_{s-1}^{-1} - Y_s^{-1})D \tag{2.4.1}$$

$$= e\,tr(X_s^{-1} - X_s^{-1}) + h\,tr(Y_s^{-1} - Y_s^{-1})D$$

$$= 0,$$

and, if $r \geq s$, (2.4.1) becomes

$$e(x_{s-1,s}^{-1} - x_{ss}^{-1}) + h(y_{s-1,s}^{-1} - y_{ss}^{-1})d_s$$

$$= \frac{e(e+h)}{e+d_s h} - e + \frac{d_s h(e+h)}{e+d_s h} - h$$

$$= \frac{(e+d_s h)(e+h)}{e+d_s h} - (e+h) = 0.$$

So we have

$$\lambda = \begin{cases} d_s^{\frac{1}{2}h} \left(\dfrac{e+d_s h}{e+h} \right)^{-\frac{1}{2}(e+h)} & \text{if } r \geq s, \\ 1 & \text{if } r < s. \end{cases}$$

Since $d_1 > d_2 > \ldots > d_r > 1 > d_{r+1} > \ldots > d_m > 0$, clearly, $r \geq s$ if and only if $d_s > 1$. Hence, we can write λ as

$$\lambda = \begin{cases} d_s^{\frac{1}{2}h} \left(\dfrac{e+d_s h}{e+h} \right)^{-\frac{1}{2}(e+h)} & \text{if } d_s > 1, \\ 1 & \text{if } 0 < d_s \leq 1. \end{cases}$$

Now upon taking the derivative of λ with respect to d_s over the range $d_s > 1$, we get

$$\frac{d\lambda}{dd_s} = \tfrac{1}{2}h\ d_s^{\frac{1}{2}h-1}\left(\frac{e+d_sh}{e+h}\right)^{-\frac{1}{2}(e+h)-1}\left[\left(\frac{e+d_sh}{e+h}\right) - d_s\right]$$

$$= \tfrac{1}{2}h\ d_s^{\frac{1}{2}h-1}\left(\frac{e+d_sh}{e+h}\right)^{-\frac{1}{2}(e+h)-1}\left[\frac{e-d_se}{e+h}\right],$$

which is negative for $d_s > 1$. Thus, λ is a decreasing function of d_s over the range $d_s > 1$. In addition,

$$d_s^{\frac{1}{2}h}\left(\frac{e+d_sh}{e+h}\right)^{-\frac{1}{2}(e+h)} \leq 1 \qquad \text{for } d_s \geq 1,$$

with equality when $d_s = 1$, so that λ is a decreasing function of d_s.

The likelihood ratio test rejects $H_0^{(s)}$ for small values of λ. Since λ is a decreasing function of d_s, the likelihood ratio test rejects $H_0^{(s)}$ for large values of d_s. Now recall that with $H_* = (1/h)H$ and $E_* = (1/e)E$, there exists a nonsingular matrix K such that $KH_*K' = D$ and $KE_*K' = I$. It follows then that $d_i : i = 1,2,\ldots,m$ are the solutions to

$$|H_*-dE_*| = |KH_*K'-dKE_*K'| = |D-dI| = 0,$$

so we observe that d_s is the s^{th} largest solution to

$$|H_*-dE_*| = 0. \tag{2.4.2}$$

With $\phi_i = d_ih/e : i = 1,2,\ldots,m$, (2.4.2) can be written

$$|H - \frac{dh}{e}E| = 0,$$

or

$$|H - \phi E| = 0. \tag{2.4.3}$$

Hence, we would reject $H_0^{(s)}$ for large values of $\phi_s = d_s h/e$, where ϕ_s is the s^{th} largest solution to (2.4.3). It is of particular importance to recall that $H \sim W_m(\Sigma+M,h,0)$ and $E \sim W_m(\Sigma,e,0)$, independently.

We have seen that the likelihood ratio test rejects $H_0^{(s)}$ when $\phi_s > c$ for some constant c. Now we want to choose for the constant c some number, which we will denote by $c(\alpha,m,s)$ to indicate its dependence upon α, m, and s, such that $P(\phi_s > c(\alpha,m,s)|(\Sigma,M)) \leq \alpha$ for all $(\Sigma,\Sigma+M) \varepsilon \mathcal{C}_{s-1}$. For $c(\alpha,m,s)$ we propose the α level critical value for the largest root, θ_1, from amongst the m-s+1 roots of $|W_1-\theta W_2| = 0$, where $W_1 \sim W_{m-s+1}(I,h-s+1,0)$ and $W_2 \sim W_{m-s+1}(I,e,0)$, independently. That is, we take $c(\alpha,m,s)$ such that $P(\theta_1 > c(\alpha,m,s)) = \alpha$. Justification for this choice of $c(\alpha,m,s)$ will be given in the next chapter.

CHAPTER 3

PROPERTIES OF THE s^{th} LARGEST ROOT TEST

3.1 Introduction

In this chapter we investigate some properties of the s^{th} largest root test presented in the previous chapter. It would be desirable to show that this test is the uniformly most powerful test, but we were unable to do so for general m. However, in Section 3.2 we show that for $m = 1$ the test is uniformly most powerful. Also, in Sections 3.3 and 3.4 it is shown that the s^{th} largest root test is an invariant test of $H_0^{(s)}$ against $H_1^{(s)}$ and is the test obtained by the union-intersection principle (see Roy [1953]). Finally, in the last two sections we discuss an important monotonicity property of the roots $\phi_i : i = 1,2,\ldots,m$, and then use this property in deriving the asymptotic distribution of ϕ_s.

3.2 The Uniformly Most Powerful Test for m = 1

For $m = 1$ the problem reduces to that of the univariate random effects model discussed in Section 1.1. Recall that we have $\bar{x}_{..} \sim N(\mu,(\sigma_z^2+n\sigma_\alpha^2)/gn)$, $u \sim \sigma_z^2\chi_e^2$, and $v \sim (\sigma_z^2+n\sigma_\alpha^2)\chi_h^2$, independently, where μ,σ_z^2, and σ_α^2 are all unknown, and we wish to test the hypothesis $H_0:\sigma_\alpha^2 = 0$ against $H_1: \sigma_\alpha^2 > 0$.

Suppose that for some set of points, γ', in the space of $(\bar{x}_{..},u,v)$, we reject H_0 whenever the experimental $(\bar{x}_{..},u,v)$ belongs to γ'. Let $\beta(\gamma';\mu,\sigma_z^2,\sigma_\alpha^2) = P[(\bar{x}_{..},u,v)\varepsilon\gamma'|\mu,\sigma_z^2,\sigma_\alpha^2]$ and require that γ' be such that $\beta(\gamma';\mu,\sigma_z^2,0) = \alpha_0$. Let $x = u + v$ and $y = v/u$, so that $u = x/(y+1)$ and $v = xy/(y+1)$. Then the Jacobian, $||J||$, of $\bar{x}_{..}$, u, and v with respect to $\bar{x}_{..}$, x, and y is $||J|| = ||\partial(\bar{x}_{..},u,v)/\partial(\bar{x}_{..},x,y)|| = x/(y+1)^2$. Then

$$\beta(\gamma';\mu,\sigma_z^2,\sigma_\alpha^2) = \iiint_{\gamma'} g_1(u;\sigma_z^2)g_2(v;\sigma_z^2,\sigma_\alpha^2)g_3(\bar{x}_{..};\mu,\sigma_z^2,\sigma_\alpha^2)du\ dv\ d\bar{x}_{..}$$

$$= \iiint_{\gamma} f(x,y;\sigma_z^2,\sigma_\alpha^2)f_0(\bar{x}_{..};\mu,\sigma_z^2,\sigma_\alpha^2)dx\ dy\ d\bar{x}_{..},$$

where $\gamma = \{(\bar{x}_{..},x,y): (\bar{x}_{..},u,v)\ \varepsilon\ \gamma'\}$ and where, independently,

$$u \sim \sigma_z^2\chi_e^2,$$

$$v \sim (\sigma_z^2+n\sigma_\alpha^2)\chi_h^2,$$

$$\bar{x}_{..} \sim N(\mu,(\sigma_z^2+n\sigma_\alpha^2)/gn),$$

so that

$$f(x,y;\sigma_z^2,\sigma_\alpha^2) = f(x,y) = \frac{x^{\frac{1}{2}(e+h)-1}\ \exp[-x/2(\sigma_z^2+n\sigma_\alpha^2)]}{(2(\sigma_z^2+n\sigma_\alpha^2))^{\frac{1}{2}(e+h)}\Gamma(\frac{1}{2}(e+h))}$$

$$\times\frac{(1+n\sigma_\alpha^2/\sigma_z^2)^{\frac{1}{2}e}\ y^{\frac{1}{2}h-1}(y+1)^{-\frac{1}{2}(e+h)}\exp[-xn\sigma_\alpha^2/2\sigma_z^2(\sigma_z^2+n\sigma_\alpha^2)(y+1)]}{B(\frac{1}{2}e,\frac{1}{2}h)},$$

and (x,y) is independent of $\bar{x}_{..}$. We note that when $\sigma_\alpha^2 = 0$, x and y are independent; that is,

$$f(x,y;\sigma_z^2,0) = f_1(x;\sigma_z^2)f_2(y) = f_1(x)f_2(y)$$

where

$$x \sim \sigma_z^2 \chi_{e+h}^2 ,$$

$$y \sim (h/e) F_e^h .$$

Letting $\gamma(x,\bar{x}_{..}) = \{y: (\bar{x}_{..},x,y) \in \gamma\}$, we can write

$$\beta(\gamma';\mu,\sigma_z^2,0) = \int_{-\infty}^{\infty} \int_0^{\infty} \int_{\gamma(x,\bar{x}_{..})} f(x,y) f_0(\bar{x}_{..}) dy\ dx\ d\bar{x}_{..}$$

$$= \int_{-\infty}^{\infty} \int_0^{\infty} f_1(x) f_0(\bar{x}_{..}) \int_{\gamma(x,\bar{x}_{..})} f_2(y) dy\ dx\ d\bar{x}_{..} .$$

Putting

$$h(x,\bar{x}_{..};\sigma_z^2) = h(x,\bar{x}_{..}) = \int_{\gamma(x,\bar{x}_{..})} f_2(y) dy,$$

we see that

$$\beta(\gamma';\mu,\sigma_z^2,0) = \int_{-\infty}^{\infty} \int_0^{\infty} f_1(x) f_0(\bar{x}_{..}) h(x,\bar{x}_{..}) dx\ d\bar{x}_{..} .$$

When $\sigma_\alpha^2 = 0$, $(x,\bar{x}_{..})$ is sufficient for (σ_z^2,μ). Further, $\{f_1(x) f_0(\bar{x}_{..}): -\infty < \mu < \infty,\ \sigma_z^2 > 0\}$ is a complete family (see, for example, Lehmann [1959:130]). Thus, since $\beta(\gamma';\mu,\sigma_z^2,0) = \alpha_0$, we must have $h(x,\bar{x}_{..}) = \alpha_0$.

Now let $\gamma_*' = \{(\bar{x}_{..},u,v): y = v/u > c\}$ where c is some constant. Then with $\underset{\sim}{q} = (q_1,q_2)'$ where $q_1 \sim (\sigma_z^2+n\sigma_\alpha^2)\chi_{e+h}^2$ and $q_2 \sim N(\mu,(\sigma_z^2+n\sigma_\alpha^2)/gn)$, independently,

$$[\beta(\gamma_*';\mu,\sigma_z^2,\sigma_\alpha^2) - \beta(\gamma';\mu,\sigma_z^2,\sigma_\alpha^2)](1+n\sigma_\alpha^2/\sigma_z^2)^{-\frac{1}{2}e}$$

$$= E(d(\underset{\sim}{q};\sigma_z^2,\sigma_\alpha^2)),$$

where the expectation is with respect to the distribution of $\underset{\sim}{q}$.

Here

$$d(\underset{\sim}{q};\sigma_z^2,\sigma_\alpha^2) = \int_{\gamma_*()} f_2(y)Q(y,q_1)dy - \int_{\gamma()} f_2(y)Q(y,q_1)dy,$$

where $\gamma_*() = \gamma_*(q_1,q_2), \gamma() = \gamma(q_1,q_2)$, and

$$Q(y,q_1) = \exp[-q_1 n\sigma_\alpha^2/2\sigma_z^2(\sigma_z^2+n\sigma_\alpha^2)(y+1)].$$

Therefore,

$$d(\underset{\sim}{q};\sigma_z^2,\sigma_\alpha^2) = \int_{\gamma_*()-\gamma()} f_2(y)Q(y,q_1)dy - \int_{\gamma()-\gamma_*()} f_2(y)Q(y,q_1)dy.$$

Since

$$Q(y,q_1) \geq Q(c,q_1) \qquad \text{when} \quad y \in \gamma_*()-\gamma(),$$

$$Q(y,q_1) \leq Q(c,q_1) \qquad \text{when} \quad y \in \gamma()-\gamma_*(),$$

we find that

$$d(\underset{\sim}{q};\sigma_z^2,\sigma_\alpha^2) \geq Q(c,q_1)[\int_{\gamma_*()-\gamma()} f_2(y)dy - \int_{\gamma()-\gamma_*()} f_2(y)dy]$$

$$= Q(c,q_1)[\int_{\gamma_*()} f_2(y)dy - \int_{\gamma()} f_2(y)dy]$$

$$= Q(c,q_1)[\alpha_0-\alpha_0] = 0.$$

Thus,

$$E(d(\underset{\sim}{q};\sigma_z^2,\sigma_\alpha^2)) \geq 0,$$

so that

$$\beta(\gamma_*';\mu,\sigma_z^2,\sigma_\alpha^2) \geq \beta(\gamma';\mu,\sigma_z^2,\sigma_\alpha^2).$$

Therefore, amongst all critical regions of size α_0 the critical region which rejects H_0 when $v/u > c$ is uniformly most powerful in a test of H_0: $\sigma_\alpha^2 = 0$ against H_1: $\sigma_\alpha^2 > 0$. That is, the critical region $\phi > c$, where ϕ is the only root of $(v-\phi u) = 0$, is uniformly most powerful.

3.3 An Invariance Property

Consider the group of transformations $G = \{g_K: K(m\times m)$ is nonsingular$\}$, where $g_K(E,H) = (KEK',KHK')$. Since $E \sim W_m(\Sigma,e,0)$ and $H \sim W_m(\Sigma+M,h,0)$, it follows that $KEK' \sim W_m(K\Sigma K',e,0)$, $KHK' \sim W_m(K\Sigma K'+KMK',h,0)$, and rank (KMK') = rank (M). Thus, the problem of testing the hypothesis $H_0^{(s)}$: rank $(M) \le s-1$ against $H_1^{(s)}$: rank $(M) = s$ is invariant under the group G.

We will need the following definition.

Definition 3.3.1: Let X be a space and G, a group of transformations on X. A function T(x) on X is said to be a maximal invariant with respect to G if

a) $T(g(x)) = T(x)$ for all $x \in X$ and $g \in G$;

b) $T(x_1) = T(x_2)$ implies $x_1 = g(x_2)$ for some $g \in G$.

We will also need the following well-known result (see, for example, Lehmann [1959:216]).

Lemma 3.3.2: Let X be a space, let G be a group of transformations on X, and let T(x) be a maximal invariant with respect to G. A function f(x) is invariant with respect to G if and only if f(x) is a function of T(x).

Now consider the roots, $\phi_1 > \phi_2 > \dots > \phi_m$, of $|H-\phi E| = 0$ and the roots, $\theta_1 > \theta_2 > \dots > \theta_m$, of $|KHK'-\theta KEK'| = 0$, where K is nonsingular. Clearly

$$|KHK'-\theta KEK'| = 0$$

implies

$$|K|\ |H-\theta E|\ |K'| = 0$$

so that

$$|H-\theta E| = 0,$$

and hence, $\theta_i = \phi_i$: $i = 1,2,\ldots,m$. Suppose now that $\theta_i = \phi_i$: $i = 1,2,\ldots,m$ are the roots of $|H_1-\theta E_1| = 0$ and $|H_2-\phi E_2| = 0$, respectively, where E_1, E_2, H_1, and H_2 are all positive definite, symmetric matrices. Then there exist nonsingular matrices K_1 and K_2 such that

$$E_1 = K_1K_1', \qquad H_1 = K_1\Phi K_1',$$
$$E_2 = K_2K_2', \qquad H_2 = K_2\Phi K_2',$$

where $\Phi = \text{diag}(\phi_1,\phi_2,\ldots,\phi_m)$. It then follows that

$$\begin{aligned} g_{K_2K_1^{-1}}(E_1,H_1) &= (K_2K_1^{-1}E_1K_1'^{-1}K_2', K_2K_1^{-1}H_1K_1'^{-1}K_2') \\ &= (K_2K_1^{-1}K_1K_1'K_1'^{-1}K_2', K_2K_1^{-1}K_1\Phi K_1'K_1'^{-1}K_2') \\ &= (K_2K_2', K_2\Phi K_2') \\ &= (E_2,H_2), \end{aligned}$$

where $g_{K_2K_1^{-1}} \varepsilon G$ since, clearly, $K_2K_1^{-1}$ is nonsingular. So by Definition 3.3.1 $\{\phi: |H-\phi E| = 0\}$ is the maximal invariant with respect to G. The s^{th} largest root, ϕ_s, is clearly a function of $(\phi_1,\phi_2,\ldots,\phi_m)$, and hence, by Lemma 3.3.2 the test statistic ϕ_s is an invariant test statistic for testing the hypothesis $H_0^{(s)}$ against the hypothesis $H_1^{(s)}$.

3.4 <u>The Union-Intersection Principle</u>

Suppose that in testing $H_0^{(m)}$: rank $(M) \le m-1$ against $H_1^{(m)}$: rank $(M) = m$, we adopt the rule

R(m:m): reject $H_0^{(m)}$ if $\phi_m > c(\alpha,m,m)$.

Here $\phi_1 > \phi_2 > \ldots > \phi_m > 0$ are the roots of $|H-\phi E| = 0$,

$E \sim W_m(\Sigma,e,0)$ and $H \sim W_m(\Sigma+M,h,0)$, independently, and $c(\alpha,m,m)$ is chosen such that $P(\phi_m > c(\alpha,m,m)|H_0^{(m)}) \leq \alpha$. Consider now testing $H_0^{(s)}$: rank $(M) \leq s-1$ against $H_1^{(s)}$: rank $(M) = s$. The hypothesis $H_0^{(s)}$ is true if and only if the hypothesis ${}_FH_0^{(s)}$: rank $(FMF') \leq s-1$ is true for all $F \in S(m,s)$, where $S(m,s)$ is the class of all $(s\times m)$ matrices of rank s. Similarly, the hypothesis $H_0^{(s)}$ is false if and only if the hypothesis ${}_FH_0^{(s)}$ is false, and the hypothesis ${}_FH_1^{(s)}$: rank $(FMF') = s$ is true, for at least one, and in fact all $F \in S(m,s)$. Hence, we could think of $H_0^{(s)}$ as $\bigcap_{F\in S(m,s)} {}_FH_0^{(s)}$ and $H_1^{(s)}$ as $\bigcup_{F\in S(m,s)} {}_FH_1^{(s)}$ and reject $H_0^{(s)}$ if $(E,H) \in \gamma = \bigcup_{F\in S(m,s)} \gamma(F)$, where $\gamma(F)$ is the rejection region appropriate to a test of the hypothesis ${}_FH_0^{(s)}$. The sizes of $\gamma(F)$: $F \in S(m,s)$ should be such as to produce a desired overall error of the first kind of the desired size. This procedure is known as the union-intersection procedure.

Note that we will reject $H_0^{(s)}$: rank $(M) \leq s-1$ if for some $F \in S(m,s)$, we reject ${}_FH_0^{(s)}$: rank $(FMF') \leq s-1$. Let $\phi_{1F} > \phi_{2F} > \cdots > \phi_{sF} > 0$ be the roots of $|FHF' - \phi FEF'| = 0$, where, clearly, $FEF' \sim W_s(F\Sigma F',e,0)$ and $FHF' \sim W_s(F\Sigma F' + FMF',h,0)$, independently. Then by the rule $R(s{:}s)$ we reject ${}_FH_0^{(s)}$ if $\phi_{sF} > c(\alpha',s,s)$, where α' is chosen to give the desired overall error of the first kind of the desired size. Hence, we will reject $H_0^{(s)}$ if for some $F \in S(m,s)$, $\phi_{sF} > c(\alpha',s,s)$, or equivalently, if $\max_{F\in S(m,s)} \phi_{sF} > c(\alpha',s,s)$.

We need the following results, the first two of which can be found in Bellman [1970:115].

Lemma 3.4.1: Let $A(m\times m)$ be a symmetric matrix. Then the smallest latent root of A may be defined as follows:

$$ch_m(A) = \min_{\underline{u}'\underline{u}=1} \underline{u}'A\underline{u},$$

where $\underline{u}$ is a $(m\times 1)$ vector.

The next result is well known as the Poincaré separation theorem.

Lemma 3.4.2: Let $A(m\times m)$ be a symmetric matrix. Then for any matrix $F(s\times m)$ such that $FF' = I$

$$ch_j(A) \geq ch_j(FAF') \geq ch_{m-s+j}(A)$$

for $j = 1,2,\ldots,s$.

We need Lemma 3.4.2 to prove the following lemma.

Lemma 3.4.3: Let $A(m\times m)$ be a symmetric matrix. Then

$$\max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'FAF'\underline{u} = ch_s(A), \tag{3.4.1}$$

where F is a $s \times m$ matrix, and $\underline{u}$ is a $m \times 1$ vector.

Proof: Since A is symmetric, there exists an orthogonal matrix $P(m\times m)$ such that $P'AP = \Lambda = \mathrm{diag}(ch_1(A), ch_2(A),\ldots,ch_m(A))$, and hence, for any F such that $FF' = I$

$$\min_{\underline{u}'\underline{u}=1} \underline{u}'FAF'\underline{u} = \min_{\underline{u}'\underline{u}=1} \underline{u}'\tilde{F}\Lambda\tilde{F}'\underline{u},$$

where $\tilde{F} = FP$ and $\tilde{F}\tilde{F}' = FPP'F' = FF' = I$. Then we can rewrite (3.4.1) as

$$\max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'F\Lambda F'\underline{u}.$$

Let $F_*(s\times m)$ be the matrix with $(F_*)_{ii} = 1$ for all i, and $(F_*)_{ij} = 0$ for all $i \neq j$. Then

$$\max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'F\Lambda F'\underline{u} \geq \min_{\underline{u}'\underline{u}=1} \underline{u}'F_*\Lambda F_*'\underline{u} = ch_s(A).$$

Now by Lemma 3.4.2, for any F such that $FF' = I$, we know that

$$\min_{\underline{u}'\underline{u}=1} \underline{u}'F\Lambda F'\underline{u} \leq ch_s(A),$$

so that

$$\max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'F\Lambda F'\underline{u} \leq ch_s(A).$$

Therefore, it follows that

$$\max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'F\Lambda F'\underline{u} = ch_s(A).$$

We have seen that the union-intersection principle leads to the rule which rejects $H_0^{(s)}$: rank $(M) \leq s - 1$ in favor of $H_1^{(s)}$: rank $(M) = s$ if $\max_{F\varepsilon S(m,s)} \phi_{sF} > c(\alpha;s,s)$. Note that with $T(m\times m)$ and $\tilde{F}(s\times m)$ such that $TT' = E$ and $\tilde{F} = FT$, then for fixed $F \varepsilon S(m,s)$

$$|FHF' - \phi FEF'| = 0$$

implies

$$|FTT^{-1}HT'^{-1}T'F' - \phi FTT^{-1}ET'^{-1}T'F'| = 0,$$

or

$$|\tilde{F}T^{-1}HT'^{-1}\tilde{F}' - \phi\tilde{F}\tilde{F}'| = 0. \tag{3.4.2}$$

Since $\tilde{F}$ is of rank s, so also is $\tilde{F}\tilde{F}'(s\times s)$, and thus, there exists a nonsingular matrix $S(s\times s)$ such that $S\tilde{F}\tilde{F}'S' = I$.

So with $\hat{F} = S\tilde{F}$ we find that (3.4.2) implies

$$|\hat{F}T^{-1}HT'^{-1}\hat{F}' - \phi I| = 0,$$

and clearly, $\hat{F}\hat{F}' = S\tilde{F}\tilde{F}'S' = I$. Hence, it follows that

$$\max_{F\varepsilon S(m,s)} \phi_{sF} = \max_{F:FF'=I} \min_{\phi}\{\phi: |FT^{-1}HT'^{-1}F'-\phi I| = 0\}$$

$$= \max_{F:FF'=I} \min_{\underline{u}'\underline{u}=1} \underline{u}'FT^{-1}HT'^{-1}F'\underline{u},$$

with the final equality due to Lemma 3.4.1. Now using Lemma 3.4.3 and the fact that the latent roots of $T^{-1}HT'^{-1}$ are the roots of $|H-\phi E| = 0$, we observe that $\max\limits_{F\varepsilon S(m,s)} \phi_{sF} = \phi_s$, and thus, the union-intersection principle leads to the rule which rejects $H_0^{(s)}$ if $\phi_s > c(\alpha', s, s)$.

3.5 A Monotonicity Property of the Power Function

The test procedure developed in the previous sections depends on the latent roots, $\phi_1, \phi_2, \ldots, \phi_m$, of the random matrix HE^{-1}. The distribution of these roots (see James [1964]), and hence the power function of our test procedure, depends upon the latent roots of the corresponding population matrix $(\Sigma+M)\Sigma^{-1}$ as parameters. Let $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_m \geq 1$ be the latent roots of $(\Sigma+M)\Sigma^{-1}$, and note that with T defined such that $\Sigma = TT'$

$$|(\Sigma+M)\Sigma^{-1} - \delta I| = 0$$

implies

$$|M - (\delta-1)\Sigma| = 0,$$

so that $$|T^{-1}MT'^{-1} - (\delta-1)I| = 0.$$

Since Σ is nonsingular, T is also nonsingular, and so the rank of $T^{-1}MT'^{-1}$ is the same as the rank of M. Hence, M has rank of at most s-1 if and only if $\delta_s = 1$, and testing the hypothesis $H_0^{(s)}$: rank (M) $\leq$ s - 1 against $H_1^{(s)}$: rank(M) = s is equivalent to testing the hypothesis $H_0^{(s)}: \delta_s = 1$ against $H_1^{(s)}: \delta_s > 1$. A desirable property of the test statistic ϕ_s would be that it stochastically increases in δ_s, and thus, that the power function increases monotonically in δ_s. In this section we not only show that ϕ_s stochastically increases in δ_s, but also that it stochastically increases in each δ_i: i = 1,2,...,m. This more general result will be utilized in the following section.

We will first prove the result for the largest latent root, ϕ_1. That is, we will show that ϕ_1 stochastically increases in δ_i: i = 1,2,...,m.

Lemma 3.5.1: The test with the acceptance region

$$\phi_1 = ch_1(HE^{-1}) \leq c$$

has power function which is monotonically increasing in each population root δ_i.

The proof of Lemma 3.5.1 involves the following three results, the first of which is due to Anderson [1955].

Lemma 3.5.2: Let $\underline{y} \sim N_m(\underline{0},\Sigma_1)$ and $\underline{u} \sim N_m(\underline{0},\Sigma_2)$, where $\Sigma_2 - \Sigma_1$ is nonnegative definite. If ω is a convex set, symmetric about the origin, then $P(\underline{y}\varepsilon\omega) \geq P(\underline{u}\varepsilon\omega)$.

Lemma 3.5.3: Let the random vectors $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ and the matrix U be mutually independent, the distribution of $\underline{y}_i$ being $N_m(\underline{0},\Sigma)$: $i = 1,2,\ldots,n$. Let the set ω, in the space of $\{\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n,U\}$, be convex and symmetric in each $\underline{y}_i$ given the other $\underline{y}_i$'s and U. Denote by $P_{\Sigma_i}(\omega)$ the probability of the set ω when $\Sigma = \Sigma_i$. Then whenever $\Sigma_2 - \Sigma_1$ is nonnegative definite, $P_{\Sigma_1}(\omega) \geq P_{\Sigma_2}(\omega)$.

Proof: Since Σ_1 and Σ_2 are symmetric and $\Sigma_1 \in \mathcal{P}_m$ and $\Sigma_2 \in \bigcup_{j=0}^{m} \mathcal{P}_j$, it follows that there exists a nonsingular matrix K such that $K\Sigma_1K' = I$ and $K\Sigma_2K' = \Delta = \text{diag}(\delta_1,\delta_2,\ldots,\delta_m)$. Since it is assumed that $\Sigma_2 - \Sigma_1 \in \bigcup_{j=0}^{m} \mathcal{P}_j$, we know that

$\delta_i \geq 1$: $i = 1,2,\ldots,m$. Then $\underline{y}_i^* = K\underline{y}_i \sim N_m(\underline{0},I)$ if $\Sigma = \Sigma_1$, and $\underline{y}_i^* = K\underline{y}_i \sim N_m(\underline{0},\Delta)$ if $\Sigma = \Sigma_2$. Let $\omega^* = \{\underline{y}_1^*,\underline{y}_2^*,\ldots,\underline{y}_n^*,U: (\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n,U) \in \omega\}$, then $P_{\Sigma_1}(\omega) = P_I(\omega^*)$ and $P_{\Sigma_2}(\omega) = P_\Delta(\omega^*)$. So without loss of generality we can take $\Sigma_1 = I$ and $\Sigma_2 = \Delta$. Let

$$\Delta_i = \text{diag}(\theta_1,\theta_2,\ldots,\theta_{i-1},1,\theta_{i+1},\ldots,\theta_m),$$

$$\Delta_i^* = \text{diag}(\theta_1,\theta_2,\ldots,\theta_{i-1},\delta_i,\theta_{i+1},\ldots,\theta_m),$$

$$R_i = \{\underline{y}_i: (\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n,U) \in \omega;\ \underline{y}_j: j\neq i \text{ and U fixed}\},$$

where $\theta_j \in \{1,\delta_j\}$: $j \neq i$. Then from Lemma 3.5.2 it follows that

$$P_{\Delta_i}(R_i|\underline{y}_j: j\neq i,U) \geq P_{\Delta_i^*}(R_i|\underline{y}_j: j\neq i,U). \tag{3.5.1}$$

Multiplying both sides of the inequality (3.5.1) by the joint density of the temporarily fixed variables and integrating with respect to them, we obtain

$$P_{\Delta_i}(\omega) \geq P_{\Delta_i^*}(\omega).$$

Then by induction we have

$$P_I(\omega) \geq P_\Delta(\omega),$$

or equivalently,

$$P_{\Sigma_1}(\omega) \geq P_{\Sigma_2}(\omega).$$

Finally, the third result we need is due to Das Gupta, Anderson, and Mudholkar [1964].

Lemma 3.5.4: For any symmetric matrix $B(m\times m)$ the region $\omega = \{A(m\times n)\colon ch_1(AA'B) \leq c\}$ is convex in A.

Proof (Lemma 3.5.1): Recall that $H \sim W_m(\Sigma+M,h,0)$ and $E \sim W_m(\Sigma,e,0)$. Since the problem is invariant under transformations $g_K(E,H) = (KEK',KHK')$, we may assume, without loss of generality, that $\Sigma + M = \Delta = \text{diag}(\delta_1,\delta_2,\ldots,\delta_m)$ and $\Sigma = I$. Then we can write $H = YY'$, where $Y = (\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_h)$ and $\underline{y}_i \sim N_m(\underline{0},\Delta)\colon i = 1,2,\ldots,h$, independently. So the acceptance region can be written as $\{Y\colon ch_1(YY'E^{-1}) \leq c\}$. From Lemma 3.5.4 it follows that the acceptance region is convex in Y, and clearly we see that the acceptance region is also symmetric in each of the column vectors of Y. Note that the vectors $\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_h$ and E are mutually independent, and the distribution of $\underline{y}_i$ is $N_m(\underline{0},\Delta)$. The result now follows from Lemma 3.5.3.

The main result of this section follows from a result due to Anderson and Das Gupta [1964].

<u>Lemma 3.5.5</u>: Suppose $V \sim W_m(\Sigma_1, v, 0)$ and $U \sim W_m(\Sigma_2, u, 0)$, independently. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ be the latent roots of UV^{-1}, and let ω be a set in the space of $\lambda_1, \lambda_2, \ldots, \lambda_m$ such that when a point $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is in ω, so is every point $(\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)$ for which $\lambda_i^* \leq \lambda_i$: $i = 1,2,\ldots,m$. Then the probability of the set ω depends on Σ_1 and Σ_2 only through the latent roots of $\Sigma_2\Sigma_1^{-1}$ and is a monotonically decreasing function of each of the latent roots of $\Sigma_2\Sigma_1^{-1}$.

Clearly, the set $\omega = \{(\phi_1, \phi_2, \ldots, \phi_m): \phi_s \leq c\}$ satisfies the conditions of Lemma 3.5.5, so it follows that the probability of the set ω is monotonically decreasing in each of the latent roots $\delta_1, \delta_2, \ldots, \delta_m$ of $(\Sigma+M)\Sigma^{-1}$. In other words, the power function of the s^{th} largest root test is a monotonically increasing function of δ_i: $i = 1,2,\ldots,m$.

We now know that as $\delta_s \to \infty$, $P(\phi_s > c)$ increases monotonically. We will show that actually, as $\delta_s \to \infty$, $P(\phi_s > c) \to 1$, and hence, for sufficiently large values of δ_s the probability of rejecting $H_0^{(s)}$: $\delta_s = 1$ will be arbitrarily close to one. Recall that there exists a nonsingular matrix K such that $KEK' \sim W_m(I, e, 0)$ and $KHK' \sim W_m(\Delta, h, 0)$. Let $K_1(m\times m)$ be such that

$$K_1(m\times m) = \text{diag}(\alpha k_1, \alpha k_2, \ldots, \alpha k_s, 1, \ldots, 1).$$

Note that $K_1 KHK'K_1' \sim W_m(K_1\Delta K_1', h, 0)$ and

$$K_1\Delta K_1' = \text{diag}(\alpha^2 k_1^2 \delta_1, \alpha^2 k_2^2 \delta_2, \ldots, \alpha^2 k_s^2 \delta_s, \delta_{s+1}, \ldots, \delta_m),$$

so that as $\alpha \to \infty$, $\alpha^2 k_i^2 \delta_i \to \infty$, and hence $ch_i(K_1\Delta K_1') \to \infty$, for

$i = 1,2,\ldots,s$. Thus, we need to show that $P(ch_s(K_1KHK'K_1'(KEK')^{-1}) > c) \to 1$ as $\alpha \to \infty$. The following lemma provides the necessary result.

<u>Lemma 3.5.6</u>: Let $V \sim W_m(\Sigma_1, v, 0)$ and $U \sim W_m(\Sigma_2, u, 0)$, independently, and let

$$K_1(m\times m) = \mathrm{diag}(\alpha k_1, \alpha k_2, \ldots, \alpha k_s, 1, \ldots, 1).$$

Then $P(ch_s(K_1UK_1'V^{-1}) > c) \to 1$ as $\alpha \to \infty$.

<u>Proof</u>: Let

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \qquad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where U_{11} is $s \times s$, U_{21} is $(m-s)\times s$, U_{12} is $s \times (m-s)$, and U_{22} is $(m-s)\times(m-s)$. Similarly, define V_{11}, V_{21}, V_{12}, and V_{22}. Let F_* be the $s \times m$ matrix with $(F_*)_{ii} = 1$: $i = 1,2,\ldots,s$ and $(F_*)_{ij} = 0$: $i \neq j$, and let

$$K_2(s\times s) = \mathrm{diag}(k_1, k_2, \ldots, k_s).$$

Recall from Section 3.4 that

$$\begin{aligned} ch_s(K_1UK_1'V^{-1}) &= \max_{F\varepsilon S(m,s)} \min_{\lambda}\{\lambda : |FK_1UK_1'F' - \lambda FVF'| = 0\} \\ &\geq \min_{\lambda}\{\lambda : |F_*K_1UK_1'F_*' - \lambda F_*VF_*'| = 0\} \\ &= \min_{\lambda}\{\lambda : |\alpha^2 K_2U_{11}K_2' - \lambda V_{11}| = 0\} \\ &= \alpha^2 ch_s(K_2U_{11}K_2'V_{11}^{-1}). \end{aligned}$$

Thus,

$$\begin{aligned} P(ch_s(K_1UK_1'V^{-1}) > c) &\geq P(\alpha^2 ch_s(K_2U_{11}K_2'V_{11}^{-1}) > c) \\ &= P(ch_s(K_2U_{11}K_2'V_{11}^{-1}) > c/\alpha^2), \end{aligned}$$

and $K_2U_{11}K_2'V_{11}^{-1}$ is positive definite with probability one, so that

$$\lim_{\alpha\to\infty} P(ch_s(K_1UK_1'V^{-1}) > c)$$
$$= \lim_{\alpha \to \infty} P(ch_s(K_2U_{11}K_2'V_{11}^{-1}) > c/\alpha^2)$$
$$= P(ch_s(K_2U_{11}K_2'V_{11}^{-1}) > 0) = 1.$$

3.6 The Limiting Distribution of ϕ_s

We have seen that the likelihood ratio test for testing the hypothesis $H_0^{(s)}$: rank (M) $\le s - 1$ against $H_1^{(s)}$: rank (M) $= s$ is based on the s^{th} largest root, ϕ_s. However, if ϕ_s is to be used as a test statistic, it is necessary to compute the significance level, α, where

$$\alpha = \sup_{H_0^{(s)}} P(\phi_s > c | H_0^{(s)}).$$

With $\delta_1 \ge \delta_2 \ge \dots \ge \delta_m$ as the latent roots of $(\Sigma+M)\Sigma^{-1}$ the null hypothesis can be written $H_0^{(s)}$: $\delta_s = 1$, or more precisely, $H_0^{(s)}$: $\delta_1 \ge \delta_2 \ge \dots \ge \delta_{s-1} \ge 1$, $\delta_s = \delta_{s+1} = \dots = \delta_m = 1$. We will write $\phi_{s:m}(\delta_1,\delta_2,\dots,\delta_m)$ to indicate that ϕ_s is the s^{th} largest root of m roots and depends on the population roots $\delta_1,\delta_2,\dots,\delta_m$. Then we may write α, the significance level, as

$$\alpha = \sup_{\delta_1\ge\delta_2\ge\dots\ge\delta_{s-1}\ge 1} P(\phi_{s:m}(\delta_1,\delta_2,\dots,\delta_{s-1},1,\dots,1) > c).$$

But we saw in the previous section that ϕ_s is stochastically increasing in each δ_i: $i = 1,2,\dots,m$. It then follows that

$\alpha = P(\phi_{s:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) > c)$,

where $\phi_{s:m}(\infty,\infty,\ldots,\infty,1,\ldots,1)$ denotes the random variable which has the limiting distribution of $\phi_{s:m}(\delta_1,\delta_2,\ldots,\delta_{s-1}, 1,\ldots,1)$ as $\delta_i \to \infty$: $i = 1,2,\ldots,s-1$. So the problem at hand is to determine the distribution of $\phi_{s:m}(\infty,\infty,\ldots,\infty, 1,\ldots,1)$.

Recall that $E \sim W_m(\Sigma,e,0)$, $H \sim W_m(\Sigma+M,h,0)$, and there exists a matrix K such that $K\Sigma K' = I$ and $K(\Sigma+M)K' = \Delta = \text{diag}(\delta_1,\delta_2,\ldots,\delta_m)$. If we define $\tilde{E}$ and $\tilde{H}$ as

$$\tilde{E} = \Delta^{-\frac{1}{2}}KEK'\Delta^{-\frac{1}{2}} \sim W_m(\Delta^{-1},e,0),$$

$$\tilde{H} = \Delta^{-\frac{1}{2}}KHK'\Delta^{-\frac{1}{2}} \sim W_m(I,h,0),$$

where $\Delta^{-\frac{1}{2}} = \text{diag}(\delta_1^{-\frac{1}{2}},\delta_2^{-\frac{1}{2}},\ldots,\delta_m^{-\frac{1}{2}})$, then clearly $\phi_{s:m}(\delta_1, \delta_2,\ldots,\delta_m) = ch_s(HE^{-1}) = ch_s(\tilde{H}\tilde{E}^{-1})$. Hence, if we let $\tilde{E}_n \sim W_m(\Delta_n^{-1},e,0)$, where $\Delta_n = \text{diag}(n\delta_1,n\delta_2,\ldots,n\delta_{s-1},1,\ldots,1)$, then we need to find the limiting distribution of $ch_s(\tilde{H}\tilde{E}_n^{-1})$ as $n \to \infty$. Since we can write $\tilde{E}_n = Y_nY_n'$, where $Y_n = (\underline{y}_1^{(n)}, \underline{y}_2^{(n)},\ldots,\underline{y}_e^{(n)})$ and $\underline{y}_i^{(n)} \sim N_m(\underline{0},\Delta_n^{-1})$: $i = 1,2,\ldots,e$, independently, we can restate the problem as that of determining the limiting distribution of $ch_s(\tilde{H}(Y_nY_n')^{-1})$. Consider the following elementary result.

<u>Lemma 3.6.1</u>: If $u_n \sim N(0,1/n)$, then $u_n \xrightarrow{d} u$, where u is a degenerate random variable with all of its probability at zero.

We also need the following results, the first of which is well known as the continuity theorem (see, for example, Breiman [1968:236]).

Lemma 3.6.2: Let $\underline{x}_1, \underline{x}_2, \underline{x}_3, \ldots$ be a sequence of random vectors. Then $\underline{x}_n \xrightarrow{d} \underline{x}$ if and only if

$$\lim_{n\to\infty} E[\exp(i\underline{x}_n'\underline{t})] = E[\exp(i\underline{x}'\underline{t})]$$

for all $\underline{t}$ where $i = \sqrt{-1}$.

Lemma 3.6.3: Suppose that as $n \to \infty$, $\underline{x}_j^{(n)} \xrightarrow{d} \underline{x}_j$: $j = 1,2,\ldots,m$, and suppose $\{\underline{x}_1^{(n)}, \underline{x}_2^{(n)}, \ldots, \underline{x}_m^{(n)}\}$ are mutually independent for all n. Then

$$\underline{x}^{(n)} = \begin{bmatrix} \underline{x}_1^{(n)} \\ \underline{x}_2^{(n)} \\ \vdots \\ \underline{x}_m^{(n)} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_m \end{bmatrix} = \underline{x}.$$

Proof: Note that it follows from Lemma 3.6.2 that

$$\lim_{n\to\infty} E[\exp(i\underline{x}_j^{(n)\prime}\underline{t}_j)] = E[\exp(i\underline{x}_j'\underline{t}_j)].$$

Also, because of independence,

$$E[\exp(i\underline{x}^{(n)\prime}\underline{t})] = E[\exp(i\sum_{j=1}^{m}\underline{x}_j^{(n)\prime}\underline{t}_j)]$$

$$= \prod_{j=1}^{m} E[\exp(i\underline{x}_j^{(n)\prime}\underline{t}_j)],$$

so

$$\lim_{n\to\infty} E[\exp(i\underline{x}^{(n)\prime}\underline{t})] = \prod_{j=1}^{m}\lim_{n\to\infty} E[\exp(i\underline{x}_j^{(n)\prime}\underline{t}_j)]$$

$$= \prod_{j=1}^{m} E[\exp(i\underline{x}_j'\underline{t}_j)]$$

$$= E[\exp(i \sum_{j=1}^{m} \underline{x}_j' \underline{t}_j)]$$

$$= E[\exp(i\underline{x}'\underline{t})].$$

The result now follows from Lemma 3.6.2.

From Lemma 3.6.1 and Lemma 3.6.3 we observe that $Y_n \xrightarrow{d} Y$ with

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

where the elements of $Y_1((s-1)\times e)$ are all equal to zero with probability one, and $Y_2 = (\underline{y}_{21}, \underline{y}_{22}, \ldots, \underline{y}_{2e})$ with $\underline{y}_{2i} \sim N_{m-s+1}(\underline{0}, I)$: $i = 1,2,\ldots,e$, independently.

Consider the following result, the proof of which can be found in Ostrowski [1973:334].

Lemma 3.6.4: Let $A(n\times n)$ and $B(n\times n)$ be two matrices, and suppose the latent roots of A and B are λ_i and λ_i': $i = 1,2,\ldots,n$, respectively. Put

$$N = \max_{1\le i\le n, 1\le j\le n} (|a_{ij}|, |b_{ij}|),$$

and

$$\delta = \frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij} - b_{ij}|.$$

Then to every root λ_i' of B there belongs a certain root λ_i of A such that we have

$$|\lambda_i' - \lambda_i| \le (n+2)N\delta^{1/n}.$$

Further, for a suitable ordering of λ_i and λ_i' we have

$$|\lambda_i - \lambda_i'| \le 2(n+1)^2 N\delta^{1/n}.$$

Lemma 3.6.5, Corollary: If A is an $n \times n$ matrix, then for each i $ch_i(A)$ is a continuous function of the elements of A.

Lemma 3.6.6, Corollary: Let A be an $n \times n$ matrix and B, an $n \times p$ matrix. Then the roots of the equation

$$|A-\lambda BB'| = 0 \qquad (3.6.1)$$

are continuous functions of the elements of A and B except at B such that $|BB'| = 0$.

Proof: Let λ_i: $i = 1,2,\ldots,n$ be the roots of (3.6.1). Then when $|BB'| \neq 0$, it follows that these roots are also the latent roots of $A(BB')^{-1}$. So, from Lemma 3.6.5, for each i λ_i is a continuous function of the elements of $A(BB')^{-1}$. But clearly, when $|BB'| \neq 0$, the elements of $A(BB')^{-1}$ are continuous functions of the elements of A and B. Hence, for each i λ_i is a continuous function of A and B except when $|BB'| = 0$.

We need one final result involving the limiting distribution of a function of random vectors (see Mann and Wald [1943]).

Lemma 3.6.7: Let $\underline{x}_n \xrightarrow{d} \underline{x}$, and let $g(\underline{x})$ be a Borel measurable function such that the set R of discontinuity points of $g(\underline{x})$ is closed and $P(\underline{x} \varepsilon R) = 0$. Then

$$g(\underline{x}_n) \xrightarrow{d} g(\underline{x}).$$

Now recall that we seek the limiting distribution of $ch_s(\tilde{H}(Y_nY_n')^{-1})$. In order to use Lemma 3.6.7 it is necessary to show that $ch_s(\tilde{H}(YY')^{-1})$ is continuous with probability one under the distribution of $(\tilde{H},Y)$. Now with

$$\tilde{H} = \begin{pmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{pmatrix},$$

where $\tilde{H}_{11}$ is $(s-1)\times(s-1)$, $\tilde{H}_{12}$ is $(s-1)\times(m-s+1)$, $\tilde{H}_{21}$ is $(m-s+1)\times(s-1)$, and $\tilde{H}_{22}$ is $(m-s+1)\times(m-s+1)$, the roots under the distribution of $(\tilde{H},Y)$ are the solutions to

$$\left| \begin{pmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{pmatrix} - \phi \begin{pmatrix} (0) & (0) \\ (0) & Y_2Y_2' \end{pmatrix} \right| = 0. \qquad (3.6.2)$$

Since $\tilde{H}$ is nonsingular with probability one, we may put

$$\tilde{H}^{-1} = G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

so (3.6.2) can be written

$$\left| I_m - \phi \begin{pmatrix} (0) & G_{12}Y_2Y_2' \\ (0) & G_{22}Y_2Y_2' \end{pmatrix} \right| = 0,$$

or

$$\begin{vmatrix} I_{s-1} & -\phi G_{12}Y_2Y_2' \\ (0) & I_{m-s+1} - \phi G_{22}Y_2Y_2' \end{vmatrix} = 0.$$

Thus, it must be true that

$$\left| I_{m-s+1} - \phi G_{22}Y_2Y_2' \right| = 0,$$

or

$$\left| G_{22}^{-1} - \phi Y_2Y_2' \right| = 0. \qquad (3.6.3)$$

Then we see that with probability one $ch_1(\tilde{H}(YY')^{-1})$, $ch_2(\tilde{H}(YY')^{-1}),\ldots,ch_{s-1}(\tilde{H}(YY')^{-1})$ are undefined, and

$ch_s(\tilde{H}(YY')^{-1})$ is now the largest solution to (3.6.3); that is, since YY' is of rank m-s+1 with probability one, there are only m-s+1 solutions to $|\tilde{H}-\phi YY'| = 0$. It can be shown (see, for example, Graybill [1969:165]) that $G_{22} = (\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12})^{-1}$, since $\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12}$ is nonsingular with probability one, so that (3.6.3) can be written

$$|\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12} - \phi Y_2Y_2'| = 0.$$

Clearly, Y_2Y_2' is also nonsingular with probability one, and hence by Lemma 3.6.6 $ch_s(\tilde{H}(YY')^{-1})$ is continuous with probability one under the distribution of $(\tilde{H},Y)$. The set of discontinuity points, R, is closed, since $R = \{(\tilde{H},Y): |Y_2Y_2'| = 0\}$. Note also that as is well known (see, for example, Anderson [1958:85]) $\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12} \sim W_{m-s+1}(I,h-s+1,0)$. Therefore, from Lemma 3.6.7 since $(\tilde{H},Y_n) \xrightarrow{d} (\tilde{H},Y)$, it follows that

$$\phi_{s:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) \sim \phi_{1:m-s+1}(1,1,\ldots,1),$$

where $\phi_{1:m-s+1}(1,1,\ldots,1)$ denotes the distribution of the largest root of $|W_1-\phi W_2| = 0$, where $W_1 \sim W_{m-s+1}(I,h-s+1,0)$, and $W_2 \sim W_{m-s+1}(I,e,0)$, independently.

So in testing $H_0^{(s)}$: rank (M) $\leq$ s-1 against $H_1^{(s)}$: rank (M) = s we choose as our critical value $c(\alpha,m,s)$, where $P(ch_1(W_1W_2^{-1}) > c(\alpha,m,s)) = \alpha$. By so doing we will guarantee that

$$\sup_{H_0^{(s)}} P(\phi_{s:m}(\delta_1,\delta_2,\ldots,\delta_m) > c(\alpha,m,s)|H_0^{(s)}) = \alpha.$$

Charts and tables of the distribution of the largest root, θ_1, of $|W_1-\theta(W_1+W_2)| = 0$ are available (see, for example, Morrison [1976:379], Pillai [1965,1967]). These may be used to calculate $c(\alpha,m,s)$, since $\theta_1 = \phi_1/(1+\phi_1)$, where ϕ_1 is the largest root of $|W_1-\phi W_2| = 0$.

In order to determine the rank of M, a sequential test procedure is used. To illustrate this procedure, we will return to the example presented in Section 2.3. Recall that $D = \text{diag}(142.729, 29.6669, .91847, .625404)$, $h = 20$, $e = 105$, so that $\phi_1 = 27.1865$, $\phi_2 = 5.65084$, $\phi_3 = .174947$, and $\phi_4 = .119125$. First we consider testing the hypothesis $H_0^{(4)}$: rank $(M) \le 3$ against $H_1^{(4)}$: rank $(M) = 4$. The null hypothesis, $H_0^{(4)}$, is rejected if $\phi_4 > c(.05,4,4)$, where $c(.05,4,4) = 17\, F(17,105,.05)/105$, and $F(17,105,.05)$ is the constant for which $P(F(17,105) > F(17,105,.05)) = .05$ if $F(17,105) \sim F_{105}^{17}(0)$. Thus, $c(.05,4,4)$ is approximately equal to .28, and clearly, $\phi_4 = .119125 < .28$, so that we do not reject $H_0^{(4)}$. Since $H_0^{(4)}$ is not rejected, we now consider testing the hypothesis $H_0^{(3)}$: rank $(M) \le 2$ against $H_1^{(3)}$: rank $(M) = 3$. The null hypothesis, $H_0^{(3)}$, is rejected if $\phi_3 > c(.05,4,3)$. Using the charts mentioned earlier we find that $c(.05,4,3)$ is approximately equal to .36. Since $\phi_3 = .174947 < .36$, we do not reject $H_0^{(3)}$ and so next consider testing the hypothesis $H_0^{(2)}$: rank $(M) \le 1$ against $H_1^{(2)}$: rank $(M) = 2$. We find that $c(.05,4,2)$ is approximately equal to .42, and therefore, since $\phi_2 = 5.65084 > .42$, we reject $H_0^{(2)}$ and conclude that the rank of M could very reasonably be taken as two.

The procedure above is open to objections on the grounds that the significance level for the test criterion has not been adjusted to take into account the fact that a sequence of hypotheses is being tested, with each one dependent on the previous ones not being rejected. The mathematical complications involved in controlling the overall error make such an adjustment virtually impossible to carry out.

CHAPTER 4

MAXIMIZATION OF THE LIKELIHOOD FUNCTION WHEN $\Sigma = \sigma^2 I$

4.1 The Likelihood Function

Suppose the vectors $\underline{x}_{ij}(m\times 1)$: $i = 1,2,\ldots,g; j=1,2,\ldots,n$ can be modeled by

$$\underline{x}_{ij} = \underline{\mu} + L\underline{f}_i + \underline{z}_{ij}, \tag{4.1.1}$$

wherein $\underline{\mu}(m\times 1)$ is a fixed but unknown vector, $L(m\times p)$ is a fixed but unknown matrix, $\underline{f}_i \sim N_p(\underline{0},I)$: $i = 1,2,\ldots,g$, and $\underline{z}_{ij} \sim N_m(\underline{0},\sigma^2 I)$: $i = 1,2,\ldots,g; j=1,2\ldots,n$. We assume that the set of random vectors $\{\underline{f}_1,\underline{f}_2,\ldots,\underline{f}_g,\underline{z}_{11},\ldots,\underline{z}_{gn}\}$ are mutually independent. Thus, $\underline{x}_{ij} \sim N_m(\underline{\mu},V)$ with $V = LL' + \sigma^2 I$. For any orthogonal matrix $P(p\times p)$ it follows that $V = LL' + \sigma^2 I = LP(LP)' + \sigma^2 I$, so that, while LL' is unique, L is not unique. In this section we will derive the likelihood function for $\underline{\mu}$, LL', and σ^2.

By methods identical to those presented in Section 2.1 it can be shown that $(\underline{\bar{x}}_{..},E,H)$ is sufficient for $(\underline{\mu},\sigma^2,nLL')$ where

$$\underline{\bar{x}}_{..} = \sum_{i=1}^{g}\sum_{j=1}^{n}\underline{x}_{ij}/gn \sim N_m(\underline{\mu},(1/gn)(\sigma^2 I+nLL')),$$

$$E = \sum_{i=1}^{g}\sum_{j=1}^{n}(\underline{x}_{ij}-\underline{\bar{x}}_{i.})(\underline{x}_{ij}-\underline{\bar{x}}_{i.})' \sim W_m(\sigma^2 I,e,0),$$

$$H = n\sum_{i=1}^{g}(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})' \sim W_m(\sigma^2 I+nLL',h,0),$$

and $e = g(n-1)$; $h = g-1$. In addition, if we let c denote a constant, we find that the density of E can be written

$$f(E) = c|E|^{\frac{1}{2}(e-m-1)}\exp[-\tfrac{1}{2}\mathrm{tr}(\sigma^2 I)^{-1}E]$$

$$= c|E|^{\frac{1}{2}(e-m-1)}\exp[-(\sum_{i=1}^{m} e_{ii})/2\sigma^2]$$

$$= g_1(\sum_{i=1}^{m} e_{ii};\sigma^2)g_2(E).$$

Hence, from the set $\{e_{11},e_{12},\ldots,e_{m,m-1},e_{mm}\}$ $b=\mathrm{tr}E=\sum_{i=1}^{m} e_{ii}$ is sufficient for σ^2.

We may assume, then, that we have, independently, $\underline{\bar{x}}_{..}$, b, and H where

$$\underline{\bar{x}}_{..} \sim N_m(\underline{\mu},(1/gn)(\sigma^2 I+nLL')),$$

$$b/\sigma^2 \sim \chi^2_\beta,$$

$$H \sim W_m(\sigma^2 I+nLL',h,0),$$

and $\beta = mg(n-1)$. The problem is to estimate $\underline{\mu}$, σ^2, and LL', or equivalently, to estimate $\underline{\mu}$, σ^2, and M where $M = nLL'$. We have seen that L is not uniquely defined and so if $\hat{LL}'$ is an estimate of LL', then any $\hat{L}$, such that $\hat{L}\hat{L}' = \hat{LL}'$, is an estimate of L. The likelihood function of $(\underline{\mu},\sigma^2,M)$ can be expressed as

$$f(\underline{\bar{x}}_{..},b,H) = \frac{K_m(I,h)}{|(2\pi/gn)(\sigma^2 I+M)|^{\frac{1}{2}}|\sigma^2 I+M|^{\frac{1}{2}h}(2\sigma^2)^{\frac{1}{2}\beta}\Gamma(\frac{1}{2}\beta)} b^{\frac{1}{2}\beta-1}|H|^{\frac{1}{2}(h-m-1)}$$

$$\times \exp[-\tfrac{1}{2}gn(\underline{\bar{x}}_{..}-\underline{\mu})'(\sigma^2 I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})-\tfrac{1}{2}b/\sigma^2-\tfrac{1}{2}\mathrm{tr}(\sigma^2 I+M)^{-1}H],$$

where, as before,

$$K_m^{-1}(I,h) = 2^{\frac{1}{2}mh}\,\pi^{\frac{1}{4}m(m-1)}\prod_{j=1}^{m}\Gamma(\tfrac{1}{2}(h-j+1)).$$

The logarithm of the likelihood function, omitting a function of the observations, is

$$-b/2\sigma^2-\tfrac{1}{2}\beta\,\ell n\sigma^2-\tfrac{1}{2}tr(\sigma^2I+M)^{-1}H-\tfrac{1}{2}h\,\ell n|\sigma^2I+M|$$
$$-\tfrac{1}{2}\ell n|\sigma^2I+M|-\tfrac{1}{2}gn(\underline{\bar{x}}_{..}-\underline{\mu})'(\sigma^2I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu}).$$

We seek the solution which maximizes the equation above, or equivalently, the solution which minimizes

$$b/\sigma^2+\beta\,\ell n\sigma^2+tr(\sigma^2I+M)^{-1}H+(h+1)\,\ell n|\sigma^2I+M| \tag{4.1.2}$$
$$+gn(\underline{\bar{x}}_{..}-\underline{\mu})'(\sigma^2I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu}).$$

If we ignore the constraints that σ^2 is positive and M is nonnegative definite and seek the stationary values of (4.1.2) over all possible $(\underline{\mu},\sigma^2,M)$, we find, upon taking the partial derivatives of (4.1.2) with respect to σ^2, M, and $\underline{\mu}$ and setting them equal to zero, that

$$-b/(\sigma^2)^2+\beta/\sigma^2-tr(\sigma^2I+M)^{-1}H(\sigma^2I+M)^{-1}+(h+1)tr(\sigma^2I+M)^{-1}$$
$$-tr(gn(\sigma^2I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})(\underline{\bar{x}}_{..}-\underline{\mu})'(\sigma^2I+M)^{-1}) = 0,$$

$$-(\sigma^2I+M)^{-1}H(\sigma^2I+M)^{-1}+(h+1)(\sigma^2I+M)^{-1}$$
$$-gn(\sigma^2I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu})(\underline{\bar{x}}_{..}-\underline{\mu})'(\sigma^2I+M)^{-1} = (0),$$

$$gn(\sigma^2I+M)^{-1}(\underline{\bar{x}}_{..}-\underline{\mu}) = \underline{0},$$

for which the solutions are

$$\tilde{\underline{\mu}} = \underline{\bar{x}}_{..},$$
$$\tilde{\sigma}^2 = b/\beta,$$
$$\tilde{M} = (h+1)^{-1}H - (b/\beta)I.$$

Since M is a nonnegative definite matrix, its maximum likelihood estimate must also be nonnegative definite, so the solutions above are the maximum likelihood estimates only if $(h+1)^{-1}H - (b/\beta)I$ is nonnegative definite. Although the solutions for $\underline{\mu}$ and σ^2 are the natural unbiased estimates, the solution for M is not since $E(\tilde{M}) = (h+1)^{-1}(hM-\sigma^2 I)$. In addition, we observe that $E(\tilde{M})$ is also not necessarily nonnegative definite.

Using the principle of marginal sufficiency referred to in Chapter 2, we see that (b,H) is marginally sufficient for (σ^2,M). Hence, we choose to use the marginal likelihood function of (σ^2,M) instead of the likelihood function of $(\underline{\mu},\sigma^2,M)$. The marginal likelihood function of (σ^2,M) can be expressed as

$$f(b,H) = \frac{K_m(I,h)}{|\sigma^2 I+M|^{\frac{1}{2}h}(2\sigma^2)^{\frac{1}{2}\beta}\Gamma(\frac{1}{2}\beta)} \, b^{\frac{1}{2}\beta-1}|H|^{\frac{1}{2}(h-m-1)}$$

$$\times \exp[-b/2\sigma^2-\tfrac{1}{2}\mathrm{tr}(\sigma^2 I+M)^{-1}H].$$

The logarithm of the likelihood, omitting a function of the observations, is

$$-b/2\sigma^2-\tfrac{1}{2}\beta\ln\sigma^2-\tfrac{1}{2}\mathrm{tr}(\sigma^2 I+M)^{-1}H-\tfrac{1}{2}h\ln|\sigma^2 I+M|,$$

and we seek the solution which maximizes this equation, or equivalently, the solution which minimizes

$$b/\sigma^2+\beta\ln\sigma^2+\mathrm{tr}(\sigma^2 I+M)^{-1}H+h\ln|\sigma^2 I+M|. \qquad (4.1.3)$$

Again, if we ignore the constraints that σ^2 is positive and M is nonnegative definite and seek the stationary value of (4.1.3) over all possible (σ^2,M), we find, upon taking the partial derivatives of (4.1.3) with respect to σ^2 and M and setting them equal to zero, that

$$-b/(\sigma^2)^2+\beta/\sigma^2-\mathrm{tr}(\sigma^2 I+M)^{-1}H(\sigma^2 I+M)^{-1}+h\,\mathrm{tr}(\sigma^2 I+M)^{-1} = 0,$$

and

$$-(\sigma^2 I+M)^{-1}H(\sigma^2 I+M)^{-1}+h(\sigma^2 I+M)^{-1} = (0),$$

for which the solutions are

$$\tilde{\sigma}^2_* = b/\beta,$$

$$\tilde{M}_* = (1/h)H - (b/\beta)I.$$

Note that these solutions are the natural unbiased estimates of σ^2 and M and, clearly, $E(\tilde{M}_*) = M$ is nonnegative definite. Hence, we choose to continue our work with the marginal likelihood function of (σ^2,M). Since M is nonnegative definite, the solutions above are the maximum likelihood estimates only if $(1/h)H - (b/\beta)I$ is also nonnegative definite. In the next section we will derive maximum likelihood estimates for σ^2 and M which are valid for all possible (b,H).

4.2 The Maximum Likelihood Estimates

In this section we seek the maximum likelihood estimates of σ^2 and M subject to the constraints $\sigma^2 > 0$ and $M \in \bigcup_{j=0}^{s} P_j$. Recall that, aside from a constant, the logarithm of the likelihood function of (σ^2,M) is

$$-b/2\sigma^2-\tfrac{1}{2}\beta\ln\sigma^2-\tfrac{1}{2}\mathrm{tr}(\sigma^2 I+M)^{-1}H-\tfrac{1}{2}h\ln|\sigma^2 I+M|.$$

We seek the solution which maximizes the equation above, or equivalently, the solution which minimizes

$$b/\sigma^2+\beta\ln\sigma^2+\mathrm{tr}(\sigma^2 I+M)^{-1}H+h\ln|\sigma^2 I+M|$$

subject to $\sigma^2 > 0$ and $M \varepsilon \bigcup_{j=0}^{s} P_j$. Note that this can be rewritten as

$$tr(\sigma^2 I)^{-1}(\frac{b}{m}I) + \frac{\beta}{m}\ell n|\sigma^2 I| + tr(\sigma^2 I+M)^{-1}H + h\ell n|\sigma^2 I+M| . \qquad (4.2.1)$$

Since $(b/\beta)I$ and $H_* = (1/h)H$ are both symmetric matrices, and $(b/\beta)I \varepsilon P_m$ and $H_* \varepsilon \bigcup_{j=0}^{m} P_j$, there exists a nonsingular matrix $K(m\times m)$ such that $K((b/\beta)I)K' = I$ and $KH_*K' = D$ where $D = diag(d_1, d_2, \ldots, d_m)$ and $d_1 > d_2 > \ldots > d_m > 0$ are the latent roots of $H_*((b/\beta)I)^{-1} = (\beta/b)H_*$. Then with $\tilde{\sigma}^2 = \beta\sigma^2/b$ and $\tilde{M} = KMK'$, (4.2.1) can be rewritten

$$\frac{\beta}{m} tr\, K'^{-1}(\sigma^2 I)^{-1}K^{-1} + \frac{\beta}{m}\ell n|\sigma^2 I| + htrK'^{-1}(\sigma^2 I+M)^{-1}K^{-1}D + h\ell n|\sigma^2 I+M|$$

$$= \frac{\beta}{m}[tr(\tilde{\sigma}^2 I)^{-1} + \ell n|\tilde{\sigma}^2 I|] + h[tr(\tilde{\sigma}^2 I+\tilde{M})^{-1}D + \ell n|\tilde{\sigma}^2 I+\tilde{M}|] - (\frac{\beta}{m}+h)\,\ell n|K|^2$$

$$= \phi(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}; D, \frac{\beta}{m}, h) - (\frac{\beta}{m}+h)\,\ell n|K|^2,$$

where ϕ is the function discussed in Section 2.2. Thus, the problem has been reduced to that of minimizing $\phi(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}; D, \frac{\beta}{m}, h)$ subject to $\tilde{\sigma}^2 > 0$ and $\tilde{M} \varepsilon \bigcup_{j=0}^{s} P_j$, or equivalently, $(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}) \varepsilon C_s$ since $C_s = \{(A,B): A\varepsilon P_m, B\varepsilon P_m, B-A\varepsilon \bigcup_{j=0}^{s} P_j\}$. Now for fixed $(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}) \varepsilon C_s$ consider $\phi(P(\tilde{\sigma}^2 I)P', P(\tilde{\sigma}^2 I+\tilde{M})P'; D, \frac{\beta}{m}, h) = \phi(\tilde{\sigma}^2 I, P(\tilde{\sigma}^2 I+\tilde{M})P'; D, \frac{\beta}{m}, h)$ for all orthogonal P. Note that this is minimized with respect to P when $tr\, P(\tilde{\sigma}^2 I+\tilde{M})^{-1}P'D$ is minimized. So from Lemma 2.2.1 it follows that all stationary points, and therefore the absolute minimum, of $\phi(\tilde{\sigma}^2 I, P(\tilde{\sigma}^2 I+\tilde{M})P'; D, \frac{\beta}{m}, h)$ occur when $P(\tilde{\sigma}^2 I+\tilde{M})P'$

is diagonal. Hence, in searching for the absolute minimum of $\phi(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}; D, \frac{\beta}{m}, h)$ over all $(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I+\tilde{M}) \varepsilon\, C_s$ we may assume that $\tilde{\sigma}^2 I+\tilde{M}$ is diagonal. This result also follows immediately from Lemma 2.2.12.

Now with $V = \text{diag}(v_1, v_2, \ldots, v_m)$ and $f_i(v_i) = d_i/v_i + \ell n\, v_i$, consider minimizing

$$\phi(uI, V; D, \frac{\beta}{m}, h) = \beta(\frac{1}{u} + \ell n u) + h \sum_{i=1}^{m} (\frac{d_i}{v_i} + \ell n v_i) \tag{4.2.2}$$

$$= \beta(\frac{1}{u} + \ell n u) + h \sum_{i=1}^{m} f_i(v_i),$$

subject to $(uI, V) \varepsilon\, C_s$. The constraint $(uI, V) \varepsilon\, C_s$ can be equivalently written as

$$v_i \geq u > 0 \qquad \text{for} \quad i = 1, 2, \ldots, m, \tag{4.2.3}$$

and

$$v_i = u \qquad \text{for} \quad i \varepsilon J, \tag{4.2.4}$$

where $J \subset \{1, 2, \ldots, m\}$ is a set which has at least $m - s$ elements. Now

$$\frac{df_i(v_i)}{dv_i} = (1 - d_i/v_i)/v_i,$$

so that the function f_i decreases monotonically in v_i for $v_i \varepsilon (0, d_i]$, increases monotonically in v_i for $v_i \varepsilon [d_i, \infty)$, and is minimized over all $v_i \varepsilon (0, \infty)$ when $v_i = d_i$. Thus, the unrestricted minimum of (4.2.2) occurs when $u = 1$ and $V = D$. It is evident from the structure of f_i that if the unrestricted minimum does not satisfy the constraints (4.2.3) and (4.2.4), then the restricted minimum will occur when $u = v_{i_1} = v_{i_2} = \dots = v_{i_k}$ for some set of integers $\{i_1, i_2, \dots, i_k\} \subset \{1,2,\dots,m\}$. We need to determine k, the number of integers, and also we need to know exactly which k integers from amongst the integers $1,2,\dots,m$ comprise the set $\{i_1, i_2, \dots, i_k\}$.

First, we will consider the constraint given by (4.2.3). Let the variable r be defined in the following manner. If $1 < d_m < d_{m-1} < \dots < d_1$, then let $r = 0$. If $d_m < 1 < d_{m-1} < \dots < d_1$, then let $r = 1$. If $d_m < \dots < d_{m-t+1} < 1 < d_{m-t} < \dots < d_1$, then let r, $1 \le r \le t$, be the smallest value for which

$$d_{m-r} > (\beta + h \sum_{j=m-r+1}^{m} d_j)/(\beta + rh).$$

Finally, if $d_m < d_{m-1} < \dots < d_1 < 1$, then let r, $1 \le r \le m-1$, be the smallest value for which the inequality above is satisfied. If the inequality is not satisfied for any choice of r, $1 \le r \le m - 1$, then let $r = m$. Now if $r = 0$, the minimum of (4.2.2) subject to (4.2.3) is simply the unrestricted minimum of (4.2.2), and if $r > 0$, the minimum of (4.2.2) subject to (4.2.3) is just the minimum of (4.2.2) subject to $u = v_m = v_{m-1} = \dots = v_{m-r+1}$ which occurs at

$$\begin{cases} u = v_m = \dots = v_{m-r+1} = (\beta + h \sum_{j=m-r+1}^{m} d_j)/(\beta + rh), \\ v_i = d_i \qquad \text{for } i = 1,2,\dots,m-r. \end{cases}$$

Now consider the constraint given by (4.2.4). If $r \ge m - s$, then the minimum of (4.2.2) subject to (4.2.3) and (4.2.4) is simply the minimum of (4.2.2) subject to (4.2.3). If $r < m - s$, the minimum of (4.2.2) subject to (4.2.3) and (4.2.4) is obtained by minimizing (4.2.2) subject to

$$\begin{cases} u = v_{j_1} = v_{j_2} = \dots = v_{j_{m-s}} & \text{if } r = 0, \\ u = v_m = \dots = v_{m-r+1} = v_{j_1} = \dots = v_{j_{m-s-r}} & \text{if } r > 0, \end{cases} \tag{4.2.5}$$

where $\{j_1, j_2, \dots, j_{m-s-r}\} \subset \{1,2,\dots,m-r-1,m-r\}$. We will now show that, in fact, $j_1 = m-r$, $j_2 = m-r-1, \dots, j_{m-s-r} = s+1$. Note that for $q = 1,2,\dots,m-1$ (4.2.2) is minimized subject to $u = v_m = \dots = v_{m-q+1}$ when

$$\begin{cases} u = v_m = \dots = v_{m-q+1} = (\beta+h \sum_{j=m-q+1}^{m} d_j)/(\beta+qh), \\ v_j = d_j \qquad \text{for } j = 1,2,\dots,m-q, \end{cases}$$

and has a minimal value equal to

$$(\beta+qh)\ln\left(\frac{\beta+h \sum_{j=m-q+1}^{m} d_j}{\beta+qh}\right) + (\beta+qh) + h(m-q) + h\sum_{j=1}^{m-q} \ln d_j. \tag{4.2.6}$$

Similarly, (4.2.2) is minimized subject to $u = v_m = \dots = v_{m-q+1} = v_i$, where $i \in \{1,2,\dots,m-q-1,m-q\}$, when

$$\begin{cases} u = v_m = \dots = v_{m-q+1} = v_i = (\beta+h \sum_{j=m-q+1}^{m} d_j + hd_i)/(\beta+(q+1)h), \\ v_j = d_j \qquad \text{for } j = 1,\dots,i-1,i+1,\dots,m-q, \end{cases}$$

and has a minimal value equal to

$$(\beta+(q+1)h)\ln\left(\frac{\beta+h \sum_{j=m-q+1}^{m} d_j + hd_i}{\beta+(q+1)h}\right) + (\beta+(q+1)h) + h(m-q-1) + h\sum_{\substack{j=1 \\ j\neq i}}^{m-q} \ln d_j. \tag{4.2.7}$$

Now subtracting (4.2.6) from (4.2.7), we obtain

$$(\beta+(q+1)h)\,\ell n\left(\frac{\beta+h\sum_{j=m-q+1}^{m} d_j+hd_i}{\beta+(q+1)h}\right)-(\beta+qh)\,\ell n\left(\frac{\beta+h\sum_{j=m-q+1}^{m} d_j}{\beta+qh}\right)-h\ell nd_i, \tag{4.2.8}$$

which is the increase in the minimal value of (4.2.2) due to the additional constraint, $u = v_i$. Differentiation of (4.2.8) with respect to d_i yields

$$h\left(\frac{\beta+(q+1)h}{\beta+h\sum_{j=m-q+1}^{m} d_j+hd_i}\right)-\frac{h}{d_i},$$

which is negative when $d_i < (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$ and positive when $d_i > (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$. Hence, (4.2.8) is an increasing function of d_i when $d_i > (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$, so that if $d_{m-q} > (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$, choosing $i = m-q$ will yield a smaller minimum value than any other choice of $i < m-q$. In a similar manner subtracting the unrestricted minimal value of (4.2.2) from the minimal value of (4.2.2) subject to $u = v_i$ where $i \in \{1,2,...,m\}$, we obtain

$$(\beta+h)\,\ell n\left(\frac{\beta+hd_i}{\beta+h}\right)-h\,\ell nd_i,$$

which is an increasing function of d_i for $d_i > 1$. Thus, if $d_m > 1$, choosing $i = m$ will yield a smaller minimum value than any other choice of $i < m$. Recall that we are investigating the minimum of (4.2.2) subject to (4.2.3) and (4.2.4) when $r < m - s$. If $r = 0$, then $d_m > 1$, so that $m-r = m$ is the optimal choice for j_1. Further, since $d_i > 1$ for $i = 1,2,\ldots,m$ where $r = 0$, we have

$$\frac{\beta+h\sum_{j=m-q+1}^{m} d_j}{\beta+qh} < \frac{(\beta+qh)(\sum_{j=m-q+1}^{m} d_j/q)}{(\beta+qh)} = \sum_{j=m-q+1}^{m} d_j/q < d_{m-q},$$

for $q = 1,2,\ldots,m-1$, and hence, when $r = 0$ choosing $j_1 = m$, $j_2 = m-1,\ldots,j_{m-s} = s+1$ in (4.2.5) will yield a smaller minimum than any other choice of $\{j_1,j_2,\ldots,j_{m-s}\} \subset \{1,2,\ldots,m-1,m\}$. Now from the definition of r we see that

$d_{m-r} > (\beta+h\sum_{j=m-r+1}^{m} d_j)/(\beta+rh)$ if $1 \le r \le m-1$. In addition,

for $q = 1,2,\ldots,m-2$ if $d_{m-q} > (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$, then

$$\frac{\beta+h\sum_{j=m-q}^{m} d_j}{\beta+(q+1)h} = \frac{(\beta+qh)((\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)) + hd_{m-q}}{\beta+(q+1)h}$$

$$< ((\beta+qh)d_{m-q}+hd_{m-q})/(\beta+(q+1)h)$$

$$= d_{m-q} < d_{m-q-1}.$$

Thus, $d_{m-q} > (\beta+h\sum_{j=m-q+1}^{m} d_j)/(\beta+qh)$ for $q = r,r+1,\ldots,m-1$.

It then follows that, when $1 \le r < m-s$, choosing $j_1 = m-r$, $j_2 = m-r-1,\ldots,j_{m-s-r} = s+1$ in (4.2.5) will yield a smaller minimum than any other choice of $\{j_1,j_2,\ldots,j_{m-s-r}\} \subset \{1,2,\ldots,m-r-1,m-r\}$.

We can now obtain the minimal solution to (4.2.2) subject to (4.2.3) and (4.2.4). Denoting the minimal solution by (u_s,V_s), we find that if $r \ge m-s$,

$$\begin{cases} u_s = v_{sm} = v_{s,m-1} = \dots = v_{s,m-r+1} = (\beta + h \sum_{j=m-r+1}^{m} d_j)/(\beta + rh), \\ v_{sj} = d_j \qquad \text{for } j = 1,2,\dots,m-r, \end{cases}$$

and if $r < m - s$,

$$\begin{cases} u_s = v_{sm} = v_{s,m-1} = \dots = v_{s,s+1} = (\beta + h \sum_{j=s+1}^{m} d_j)/(\beta + (m-s)h), \\ v_{sj} = d_j \qquad \text{for } j = 1,2,\dots,s. \end{cases}$$

Thus, $\phi(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I + \tilde{M}; D, \frac{\beta}{m}, h)$ is minimized subject to $(\tilde{\sigma}^2 I, \tilde{\sigma}^2 I + \tilde{M}) \in \mathcal{C}_s$ at

$$\tilde{\sigma}^2 = u_s,$$

$$\tilde{M} = V_s - u_s I.$$

The maximum likelihood estimates of σ^2 and M are, therefore, $\hat{\sigma}^2$ and $\hat{M}$ given by

$$\hat{\sigma}^2 = bu_s/\beta,$$

$$\hat{M} = K^{-1}(V_s - u_s I)K'^{-1}.$$

To illustrate the computation involved in deriving the maximum likelihood estimates, we will again consider the example presented in Section 2.3. Recall that with $m = 4$, $g = 21$, $n = 6$, $\Sigma = I$, and $M = \text{diag}(99, 24, 0, 0)$ a matrix E from the distribution $W_4(I, 105, 0)$ and a matrix H from the distribution $W_4(I+M, 20, 0)$ were generated. With $\beta = mg(n-1) = 420$, $b = \text{tr}E$, and $H_* = (1/20)H$, we need to find a nonsingular matrix K such that $K((b/\beta)I)K' = I$

and $KH_*K' = D$ where D is a diagonal matrix. Let $D_1 =$ diag$(ch_1(H_*), ch_2(H_*), \ldots, ch_m(H_*))$, and let Q be the orthogonal matrix for which the i^{th} column is the characteristic vector of H_* corresponding to $ch_i(H_*)$, then since H_* is symmetric, $P'H_*P = D_1$. Clearly, $((\beta^{\frac{1}{2}}/b^{\frac{1}{2}})P)'((b/\beta)I)((\beta^{\frac{1}{2}}/b^{\frac{1}{2}})P) = P'P = I$ and $((\beta^{\frac{1}{2}}/b^{\frac{1}{2}})P)'H_*((\beta^{\frac{1}{2}}/b^{\frac{1}{2}})P) = (\beta/b)D_1$. Hence, we find that, for our example,

$$K = \begin{bmatrix} 1.00723 & .0551967 & -.00906271 & -.00104477 \\ .0551796 & -1.00719 & -.00310948 & .0127707 \\ .00921922 & -.00261055 & 1.00874 & -.00010739 \\ -.000345628 & -.0128084 & -.000137374 & -1.0087 \end{bmatrix}$$

and D = diag(94.1065, 34.8845, 1.01721, .618312). Note that $d_4 < 1 < d_3 < d_2 < d_1$, so that r = 1. Then simple calculation yields $u_0 = 6.06506$, $V_0 = 6.06506I$, $u_1 = 2.39667$, $V_1 =$ diag(94.1065, 2.39667, 2.39667, 2.39667), $u_2 = .984153$, $V_2 =$ diag(94.1065, 34.8845, .984153, .984153), $u_3 = u_4 = .982651$, $V_3 = V_4 =$ diag(94.1065, 34.8845, 1.01721, .982651). Thus, if we let $\hat{\sigma}_i^2$ and $\hat{M}_i$ denote the maximum likelihood estimates of σ^2 and M, respectively, subject to the constraints $\sigma^2 > 0$ and $M \varepsilon \bigcup_{j=0}^{i} P_j$, we see that

$$\hat{\sigma}_0^2 = 5.95987,$$

$$\hat{M}_0 = (0),$$

$$\hat{\sigma}_1^2 = 2.3551,$$

$$\hat{M}_1 = \begin{bmatrix} 89.8421 & 4.92338 & -.808366 & -.0931905 \\ & .269803 & -.0442987 & -.00510687 \\ & & .00727337 & .000838493 \\ & & & .0000966637 \end{bmatrix},$$

$$\hat{\sigma}_2^2 = .967085,$$

$$\hat{M}_2 = \begin{bmatrix} 91.3255 & 3.17993 & -.826433 & -.0715582 \\ & 33.4811 & .0575388 & -.426237 \\ & & .00770191 & -.000448497 \\ & & & .0054369 \end{bmatrix},$$

$$\hat{\sigma}_3^2 = \hat{\sigma}_4^2 = .965608,$$

$$\hat{M}_3 = \hat{M}_4 = \begin{bmatrix} 91.327 & 3.17993 & -.826136 & -.0715587 \\ & 33.4825 & .0574547 & -.426256 \\ & & .0416617 & -.000452157 \\ & & & .00543713 \end{bmatrix}.$$

4.3 The Likelihood Ratio Test

Recall that $C_s = \{(A,B): A \varepsilon P_m,\ B \varepsilon P_m,\ B-A \varepsilon \bigcup_{j=0}^{s} P_j\}$, and suppose we know that $(\sigma^2 I, \sigma^2 I+M) \varepsilon \Omega = C_s$. We wish to test, say, the null hypothesis that $(\sigma^2 I, \sigma^2 I+M) \varepsilon \omega = C_{s-1} \subset C_s$. The alternative hypothesis, then, is that $(\sigma^2 I, \sigma^2 I+M) \varepsilon \Omega - \omega = C_s - C_{s-1}$. Thus, we are testing the hypothesis

$$H_0^{(s)}: \text{rank}(M) \leq s - 1$$

against the hypothesis

$$H_1^{(s)}: \text{rank}(M) = s.$$

We adopt the likelihood approach and look at

$$\max_{\omega} f(b,H)/\max_{\Omega} f(b,H) = \lambda \varepsilon (0,1].$$

With u_s and the matrix $V_s = \text{diag}(v_{s1}, v_{s2}, \ldots, v_{sm})$ given by

$$\begin{cases} u_s = v_{sm} = \ldots = v_{s,m-r+1} = (\beta + h \sum_{j=m-r+1}^{m} d_j)/(\beta + rh), \\ v_{sj} = d_j \qquad \text{for } j = 1,2,\ldots,m-r, \end{cases}$$

if $r \geq m - s$, and

$$\begin{cases} u_s = v_{sm} = \ldots = v_{s,s+1} = (\beta + h \sum_{j=s+1}^{m} d_j)/(\beta + (m-s)h), \\ v_{sj} = d_j \qquad \text{for } j = 1,2,\ldots,s, \end{cases}$$

if $r < m - s$, the maximum likelihood estimators $\hat{\sigma}^2_\Omega$, of σ^2, and $\hat{M}_\Omega$, of M, when the parameters are restricted to lie within Ω, are given by

$$\begin{cases} \hat{\sigma}^2_\Omega = bu_s/\beta, \\ \hat{M}_\Omega = K^{-1}(V_s - u_s I)K'^{-1}, \end{cases}$$

where K is a nonsingular matrix. Similarly, the maximum likelihood estimators $\hat{\sigma}^2_\omega$, of σ^2, and $\hat{M}_\omega$, of M, when the parameters are restricted to lie within ω, are given by

$$\begin{cases} \hat{\sigma}^2_\omega = bu_{s-1}/\beta, \\ \hat{M}_\omega = K^{-1}(V_{s-1} - u_{s-1} I)K'^{-1}. \end{cases}$$

It should be noted that if $r \geq m - s + 1$, then $V_s = V_{s-1}$ and $u_s = u_{s-1}$.

The likelihood ratio, λ, is

$$\lambda = \max_{\omega} f(b,H)/\max_{\Omega} f(b,H)$$

$$= \frac{\exp[-b/2\hat{\sigma}_{\omega}^{2}-\tfrac{1}{2}\mathrm{tr}(\hat{\sigma}_{\omega}^{2}I+\hat{M}_{\omega})^{-1}H]}{\exp[-b/2\hat{\sigma}_{\Omega}^{2}-\tfrac{1}{2}\mathrm{tr}(\hat{\sigma}_{\Omega}^{2}I+\hat{M}_{\Omega})^{-1}H]} \; \frac{(\hat{\sigma}_{\Omega}^{2})^{\frac{1}{2}\beta}|\hat{\sigma}_{\Omega}^{2}I+\hat{M}_{\Omega}|^{\frac{1}{2}h}}{(\hat{\sigma}_{\omega}^{2})^{\frac{1}{2}\beta}|\hat{\sigma}_{\omega}^{2}I+\hat{M}_{\omega}|^{\frac{1}{2}h}}$$

$$= \frac{\exp[-\beta/2u_{s-1}-\tfrac{1}{2}h\mathrm{tr}V_{s-1}^{-1}D]}{\exp[-\beta/2u_{s}-\tfrac{1}{2}h\mathrm{tr}V_{s}^{-1}D]} \; \frac{|V_s|^{\frac{1}{2}h}\,u_s^{\frac{1}{2}\beta}}{|V_{s-1}|^{\frac{1}{2}h}\,u_{s-1}^{\frac{1}{2}\beta}}$$

$$= \frac{|V_s|^{\frac{1}{2}h}\,u_s^{\frac{1}{2}\beta}}{|V_{s-1}|^{\frac{1}{2}h}\,u_{s-1}^{\frac{1}{2}\beta}}$$

since, if $r \geq m - s + 1$

$$\beta(u_{s-1}^{-1}-u_{s}^{-1}) + h\,\mathrm{tr}(V_{s-1}^{-1}-V_{s}^{-1})D \qquad (4.3.1)$$

$$= \beta(u_{s}^{-1}-u_{s}^{-1}) + h\,\mathrm{tr}(V_{s}^{-1}-V_{s}^{-1})D$$

$$= 0,$$

and if $r < m - s + 1$, (4.3.1) becomes

$$\beta\left(\frac{\beta+(m-s+1)h}{\beta+h\sum_{j=s}^{m}d_j} - \frac{\beta+(m-s)h}{\beta+h\sum_{j=s+1}^{m}d_j}\right)+h\left((s-1) + \left(\frac{\beta+(m-s+1)h}{\beta+h\sum_{j=s}^{m}d_j}\right)\sum_{j=s}^{m}d_j - s - \left(\frac{\beta+(m-s)h}{\beta+h\sum_{j=s+1}^{m}d_j}\right)\sum_{j=s+1}^{m}d_j\right)$$

$$= \left(\frac{\beta+(m-s+1)h}{\beta+h\sum_{j=s}^{m}d_j}\right)(\beta+h\sum_{j=s}^{m}d_j) - \left(\frac{\beta+(m-s)h}{\beta+h\sum_{j=s+1}^{m}d_j}\right)(\beta+h\sum_{j=s+1}^{m}d_j) - h$$

$$= \beta + (m-s+1)h - (\beta+(m-s)h)-h = 0.$$

So we have

$$\lambda = \begin{cases} d_s^{\frac{1}{2}h}\left(\dfrac{\beta+(m-s+1)h}{\beta+h\sum_{j=s}^{m} d_j}\right)^{\frac{1}{2}(\beta+h(m-s+1))}\left(\dfrac{\beta+h\sum_{j=s+1}^{m} d_j}{\beta+(m-s)h}\right)^{\frac{1}{2}(\beta+h(m-s))} & \text{if } r < m-s+1, \\ 1 & \text{if } r \geq m-s+1. \end{cases}$$

Putting $t_s = hd_s/(\beta+h\sum_{j=s}^{m} d_j)$, we can rewrite λ as

$$\lambda = \begin{cases} \left(\dfrac{\beta+(m-s+1)h}{h}\right)^{\frac{1}{2}h}\left(\dfrac{\beta+(m-s+1)h}{\beta+(m-s)h}\right)^{\frac{1}{2}(\beta+h(m-s))} t_s^{\frac{1}{2}h}(1-t_s)^{\frac{1}{2}(\beta+h(m-s))} & \text{if } r < m-s+1, \\ 1 & \text{if } r \geq m-s+1. \end{cases}$$

We will now show that $r < m-s+1$ if and only if $t_s > h/(\beta+(m-s+1)h)$. First consider the case in which $s = m$. Then $r < m-s+1 = 1$ if and only if $d_m > 1$, and $t_m = hd_m/(\beta+hd_m) = h/(\beta/d_m+h) > h/(\beta+h)$ if $d_m > 1$, and

$$t_m = h/(\beta/d_m+h) < h/(\beta+h)$$

if $d_m < 1$. Consider now the case in which $1 \leq s \leq m-1$. Again we want to show that $r < m-s+1$ if and only if $t_s > h/(\beta+(m-s+1)h)$. If $r = 0$, clearly

$$d_{m-i} > (\beta+h\sum_{j=m-i+1}^{m} d_j)/(\beta+ih),$$

for $i = 1,2,\ldots,m-1$. Also, if $0 < r < m-s+1$, then

$$d_{m-r} > (\beta+h \sum_{j=m-r+1}^{m} d_j)/(\beta+rh),$$

and we have seen that this implies that

$$d_{m-q} > (\beta+h \sum_{j=m-q+1}^{m} d_j)/(\beta+qh),$$

for $q = r, r+1, \ldots, m-1$ and, more specifically, for $q = m-s$. Hence, if $r < m-s+1$,

$$d_s > (\beta+h \sum_{j=s+1}^{m} d_j)/(\beta+(m-s)h),$$

which implies

$$\beta + h \sum_{j=s}^{m} d_j < d_s(\beta+(m-s)h)+hd_s,$$

so that

$$\beta + h \sum_{j=s}^{m} d_j < hd_s(\frac{\beta}{h} + m-s+1),$$

and thus

$$t_s = \frac{hd_s}{\beta+h \sum_{j=s}^{m} d_j} > \frac{1}{\beta/h + m-s+1} = \frac{h}{\beta+(m-s+1)h}.$$

Also, if $r \geq m-s+1$, then it must be true that

$$d_{m-(m-s)} = d_s \leq (\beta+h \sum_{j=s+1}^{m} d_j)/(\beta+(m-s)h)$$

which implies that

$$t_s = \frac{hd_s}{\beta+h \sum_{j=s}^{m} d_j} \leq \frac{h}{\beta+(m-s+1)h} .$$

It follows that the likelihood ratio, λ, can be written as

$$\lambda = \begin{cases} \left(\frac{\beta+(m-s+1)h}{h}\right)^{\frac{1}{2}h}\left(\frac{\beta+(m-s+1)h}{\beta+(m-s)h}\right)^{\frac{1}{2}(\beta+h(m-s))} t_s^{\frac{1}{2}h}(1-t_s)^{\frac{1}{2}(\beta+h(m-s))} & \text{if } t_s > \frac{h}{\beta+(m-s+1)h}, \\ 1 & \text{if } t_s \le \frac{h}{\beta+(m-s+1)h}. \end{cases}$$

Consider the function $g(t_s) = t_s^{\frac{1}{2}h}(1-t_s)^{\frac{1}{2}(\beta+h(m-s))}$. The derivative of $g(t_s)$ with respect to t_s is

$$t_s^{\frac{1}{2}h-1}(1-t_s)^{\frac{1}{2}(\beta+h(m-s))-1}[\tfrac{1}{2}h(1-t_s)-\tfrac{1}{2}(\beta+h(m-s))t_s],$$

which is negative for $t_s \varepsilon (h/(\beta+(m-s+1)h),1)$. Thus, λ is a decreasing function of t_s when $t_s \varepsilon (h/(\beta+(m-s+1)h),1)$. In addition,

$$\left(\frac{\beta+(m-s+1)h}{h}\right)^{\frac{1}{2}h}\left(\frac{\beta+(m-s+1)h}{\beta+(m-s)h}\right)^{\frac{1}{2}(\beta+h(m-s))} t_s^{\frac{1}{2}h}(1-t_s)^{\frac{1}{2}(\beta+h(m-s))} \le 1,$$

for $t_s \varepsilon (h/(\beta+(m-s+1)h),1)$, with equality when $t_s = h/(\beta+(m-s+1)h)$, so that λ is a decreasing function of t_s. Since the likelihood ratio test rejects $H_0^{(s)}$ for small values of λ, it equivalently rejects $H_0^{(s)}$ for large values of t_s. However, the distribution of t_s is intractable, and so use of t_s in a test of $H_0^{(s)}$ versus $H_1^{(s)}$ is not practical. In the following chapter we present an alternative test statistic for testing $H_0^{(s)}$ versus $H_1^{(s)}$.

CHAPTER 5

AN ALTERNATIVE TEST WHEN $\Sigma = \sigma^2 I$ AND ITS PROPERTIES

5.1 Introduction

We have seen that the likelihood ratio test rejects $H_0^{(s)}$ for sufficiently large values of $hd_s/(\beta+h\sum_{i=s}^{m} d_i)$, where $d_1>d_2>\ldots>d_m$ are the solutions to $|H_*-d(b/\beta)I| = 0$. Let $\psi_1>\psi_2>\ldots>\psi_m$ be the solutions to $|H-\psi bI| = 0$, that is, $\psi_i = hd_i/\beta$ for $i = 1,2,\ldots,m$. Then the likelihood ratio test rejects $H_0^{(s)}$ for sufficiently large values of $\psi_s/(1+\sum_{i=s}^{m} \psi_i)$. This quantity is an increasing function of ψ_s, so that it would be reasonable to reject $H_0^{(s)}$ for sufficiently large values of ψ_s. However, the complexity of the null distribution of ψ_s makes the use of ψ_s in a test of $H_0^{(s)}$ versus $H_1^{(s)}$ impractical. Therefore, in this chapter we present an alternative test statistic for testing $H_0^{(s)}$ against $H_1^{(s)}$ and consider the test which rejects $H_0^{(s)}$ when $\sum_{i=s}^{m} \psi_i$ is sufficiently large. In the remainder of this chapter we investigate some properties of this new test. In Section 5.2 it is shown that the test based on $\sum_{i=s}^{m} \psi_i$ is an invariant test of $H_0^{(s)}$ against $H_1^{(s)}$. In the last two sections we discuss an important monotonicity property of the roots ψ_i: $i = 1,2,\ldots,m$ and use this property in deriving the asymptotic distribution of $\sum_{i=s}^{m} \psi_i$.

5.2 An Invariance Property

Consider the group of transformations $G = \{g_{a,P}: a > 0,$ $P(m\times m)$ is such that $PP' = aI\}$, where $g_{a,P}(b,H) = (ab, PHP')$. If $b \sim \sigma^2\chi^2_\beta$ and $H \sim W_m(\sigma^2 I+M,h,0)$, then $ab \sim a\sigma^2\chi^2_\beta$, $PHP' \sim W_m(a\sigma^2 I+PMP',h,0)$ and $\text{rank}(PMP') = \text{rank}(M)$. Hence, the problem of testing the hypothesis $H_0^{(s)}$: $\text{rank}(M) \leq s-1$ against $H_1^{(s)}$: $\text{rank}(M) = s$ is invariant under the group G.

Now consider the roots $\psi_1 > \psi_2 > \ldots > \psi_m$ of $|H-\psi bI| = 0$ and the roots $\theta_1 > \theta_2 > \ldots > \theta_m$ of $|PHP'-\theta abI| = 0$, where $a > 0$ and P is such that $PP' = aI$. Clearly,

$$|PHP' - \theta abI| = 0$$

implies

$$|PHP' - \theta bPP'| = 0,$$

so that

$$|H - \theta bI| = 0,$$

and thus, $\theta_i = \psi_i$: $i = 1,2,\ldots,m$. Suppose now that $\theta_i = \psi_i$: $i = 1,2,\ldots,m$ are the roots of $|H_1-\theta b_1 I| = 0$ and $|H_2-\psi b_2 I| = 0$, respectively, where $b_1 > 0$, $b_2 > 0$, and H_1 and H_2 are positive definite, symmetric matrices. There exist orthogonal matrices Q_1 and Q_2 such that

$$Q_1(H_1/b_1)Q_1' = \Psi,$$

$$Q_2(H_2/b_2)Q_2' = \Psi,$$

where $\Psi = \mathrm{diag}(\psi_1,\psi_2,\ldots,\psi_m)$. Take $a = b_2/b_1$ and $P = \sqrt{a}\ Q_2'Q_1$. It now follows that

$$\begin{aligned}
\text{(a)} \qquad g_{a,P}(b_1,H_1) &= (ab_1,PH_1P') \\
&= (b_2,aQ_2'Q_1H_1Q_1'Q_2) \\
&= (b_2,ab_1Q_2'Q_1(H_1/b_1)Q_1'Q_2) \\
&= (b_2,b_2Q_2'\Psi Q_2) \\
&= (b_2,b_2(H_2/b_2)) = (b_2,H_2),
\end{aligned}$$

$$\begin{aligned}
\text{(b)} \qquad PP' &= (\sqrt{a}Q_2'Q_1)(\sqrt{a}Q_2'Q_1)' \\
&= aQ_2'Q_1Q_1'Q_2 = aI,
\end{aligned}$$

and (c) $\qquad a > 0,$

so that $g_{a,P} \in G$. Therefore, by Definition 3.3.1 $\{\psi: |H-\psi bI| = 0\}$ is the maximal invariant with respect to G. The test statistic $\sum_{i=s}^{m} \psi_i$ is clearly a function of $(\psi_1,\psi_2,\ldots,\psi_m)$, and so, by Lemma 3.3.2, the test statistic $\sum_{i=s}^{m} \psi_i$ is an invariant test statistic for testing the hypothesis $H_0^{(s)}$ against the hypothesis $H_1^{(s)}$.

5.3 A Monotonicity Property of the Power Function

The test procedure which we have been investigating depends on the latent roots $\psi_1,\psi_2,\ldots,\psi_m$ of the random matrix $H(bI)^{-1} = (\sigma^{-2}H)((b/\sigma^2)I)^{-1}$. If we let $\theta_1>\theta_2>\ldots>\theta_m$ be the latent roots of $\sigma^{-2}H$, then $\psi_i = \sigma^2\theta_i/b$:

$i = 1,2,\ldots,m$. The distribution of the roots $\theta_1,\theta_2,\ldots,\theta_m$ (see, for example, James [1964]) depends upon the latent roots of the corresponding population matrix $I+\sigma^{-2}M$ as parameters. Let $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_m \geq 1$ be the latent roots of $I+\sigma^{-2}M$, and note that M has rank of at most s-1 if and only if $\delta_s = 1$. Thus, testing the hypothesis $H_0^{(s)}$: rank(M) $\leq$ s-1 against $H_1^{(s)}$: rank(M) = s is equivalent to testing the hypothesis $H_0^{(s)}$: $\delta_s = 1$ against $H_1^{(s)}$: $\delta_s > 1$. Since we are using $\sum_{i=s}^{m} \psi_i$ as a test statistic in testing the hypothesis $H_0^{(s)}$ against $H_1^{(s)}$, a desirable property would be that it stochastically increases in δ_s, and hence, the power function increases monotonically in δ_s. In this section we not only show that $\sum_{i=s}^{m} \psi_i$ stochastically increases in δ_s, but also that it stochastically increases in each δ_i: $i = 1,2,\ldots,m$. This more general result will be utilized in the following section.

We will need the following results from Anderson and Das Gupta [1964].

Lemma 5.3.1: Let $X(m\times h)$ $(h \geq m)$ be a random matrix having density

$$f(X;\Sigma,h) = (2\pi)^{-\frac{1}{2}hm}|\Sigma|^{-\frac{1}{2}h} \exp[-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}XX'],$$

where Σ is positive definite. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ be the latent roots of XX' and ω be a set in the space of $\lambda_1,\lambda_2,\ldots,\lambda_m$ such that when a point $(\lambda_1,\lambda_2,\ldots,\lambda_m)$ is in ω so is every point $(\lambda_1',\lambda_2',\ldots,\lambda_m')$ for which $\lambda_i' \leq \lambda_i$: $i = 1,2,\ldots,m$. Then the probability of the set ω depends on Σ only through the latent roots of Σ and is a monotonically decreasing function of each of the latent roots of Σ.

Lemma 5.3.2: Let A be a positive definite matrix of order m, and D and D_* be two diagonal matrices of order m such that $D_* - D$ is positive semidefinite, and D is positive definite. Then

$$ch_i(DAD) \le ch_i(D_*AD_*) \qquad \text{for } i = 1,2,\ldots,m.$$

Using these two results, we can now prove the main result of this section.

Lemma 5.3.3: Let $X(m\times h)$ be a random matrix having density

$$f(X;D,h) = (2\pi)^{-\frac{1}{2}hm}|D|^{-\frac{1}{2}h} \exp[-\tfrac{1}{2}\mathrm{tr}D^{-1}XX'],$$

where $D = \mathrm{diag}(d_1,d_2,\ldots,d_m)$. Let $V(m\times m)$ be a random, positive definite matrix independent of X. Let ω be a set in the space of the latent roots of $XX'V^{-1}$ satisfying the condition stated in Lemma 5.3.1. Then the probability of the set ω is a monotonically decreasing function of each of the elements of D.

Proof: Consider V as fixed, and let $V^{-1} = T'T$ where T is nonsingular. Then the density of $W = TX$ is $f(W;TDT',h)$, and

$$ch_i(XX'V^{-1}) = ch_i(TXX'T') = ch_i(WW')$$

for $i = 1,2,\ldots,m$. Thus, for any fixed V, we have

$$\int_{R(X)} f(X;D,h)\,dX = \int_{R(W)} f(W;TDT',h)\,dW \qquad (5.3.1)$$

where $R(X)$ denotes the region $\{X: (ch_1(XX'V^{-1}),\ldots,ch_m(XX'V^{-1})) \varepsilon\ \omega\}$, and $R(W)$ denotes the region $\{W: (ch_1(WW'),\ldots,ch_m(WW')) \varepsilon\ \omega\}$. Let D_* be a diagonal matrix for which D_*-D is positive semidefinite. It follows from Lemma 5.3.2 that

$$ch_i(TD_*T') = ch_i(D_*^{\frac{1}{2}}(T'T)D_*^{\frac{1}{2}}) \geq ch_i(D^{\frac{1}{2}}(T'T)D^{\frac{1}{2}}) = ch_i(TDT')$$

for $i = 1,2,\ldots,m$. Then from Lemma 5.3.1 and (5.3.1) we have

$$\int_{R(X)} f(X;D,h)dX \geq \int_{R(X)} f(X;D_*,h)dX$$

for any fixed V. Taking expectations with respect to V, we find that

$$P_D(\omega) \geq P_{D_*}(\omega).$$

Now recall that we are investigating the test statistic $\sum_{i=s}^{m} \psi_i$. Let P be the orthogonal matrix such that $P(I+\sigma^{-2}M)P' = \Delta = \text{diag}(\delta_1,\delta_2,\ldots,\delta_m)$. Then since $\sigma^{-2}H \sim W_m(I+\sigma^{-2}M,h,0)$, it follows that $P(\sigma^{-2}H)P' \sim W_m(\Delta,h,0)$, and we can write $P(\sigma^{-2}H)P' = XX'$, where $X(m\times h)$ has density $f(X;\Delta,h)$ given in Lemma 5.3.1. The latent roots of $\sigma^{-2}H((b/\sigma^2)I)^{-1}$ are the latent roots of $P(\sigma^{-2}H)P'((b/\sigma^2)I)^{-1}$, or equivalently, $XX'((b/\sigma^2)I)^{-1}$. Hence, with $V = (b/\sigma^2)I$, clearly V is independent of X, and $\psi_1,\psi_2,\ldots,\psi_m$ are the latent roots of $XX'V^{-1}$. In addition, if $\sum_{i=s}^{m} \psi_i \leq c$ and $\psi_i' \leq \psi_i$: $i = 1,2,\ldots,m$, then $\sum_{i=s}^{m} \psi_i' \leq c$, so that the set $\omega = \{(\psi_1,\psi_2,\ldots,\psi_m): \sum_{i=s}^{m} \psi_i \leq c\}$ satisfies the condition of Lemma 5.3.3. So it follows from Lemma 5.3.3 that the probability of the set ω is a monotonically decreasing function of each of the latent roots

$\delta_1, \delta_2, \ldots, \delta_m$ of $I+\sigma^{-2}M$; in other words, the power function of the test based on $\sum_{i=s}^{m} \psi_i$ is a monotonically increasing function of δ_i: $i = 1,2,\ldots,m$.

We now know that as $\delta_s \to \infty$ $P(\sum_{i=s}^{m} \psi_i > c)$ increases monotonically. We will show that, in fact, as $\delta_s \to \infty$ $P(\sum_{i=s}^{m} \psi_i > c) \to 1$, and thus, for sufficiently large values of δ_s the probability of rejecting $H_0^{(s)}$: $\delta_s = 1$ will be arbitrarily close to unity. Let $K_1(m\times m)$ be such that

$$K_1 = \text{diag}(\alpha k_1, \alpha k_2, \ldots, \alpha k_s, 1, \ldots, 1).$$

Note that $K_1 P(\sigma^{-2}H)P'K_1' \sim W_m(K_1 \Delta K_1', h, 0)$, and

$$K_1 \Delta K_1' = \text{diag}(\alpha^2 k_1^2 \delta_1, \alpha^2 k_2^2 \delta_2, \ldots, \alpha^2 k_s^2 \delta_s, 1, \ldots, 1),$$

so that as $\alpha \to \infty$, $ch_i(K_1 \Delta K_1') = \alpha^2 k_i^2 \delta_i \to \infty$ for $i = 1,2,\ldots,s$. Thus, we need to show that

$$P(\sum_{i=s}^{m} ch_i(K_1 P(\sigma^{-2}H)P'K_1'((b/\sigma^2)I)^{-1}) > c) \to 1$$

as $\alpha \to \infty$. However, clearly,

$$P(\sum_{i=s}^{m} ch_i(K_1 P(\sigma^{-2}H)P'K_1'((b/\sigma^2)I)^{-1}) > c)$$

$$\geq P(ch_s(K_1 P(\sigma^{-2}H)P'K_1'((b/\sigma^2)I)^{-1}) > c).$$

The result now follows from the following lemma.

<u>Lemma 5.3.4</u>: Let $V(m\times m)$ and $U(m\times m)$ be random matrices independently distributed such that both V and U are positive definite with probability one. Let

$$K_1(m\times m) = \text{diag}(\alpha k_1, \alpha k_2, \ldots, \alpha k_s, 1, \ldots, 1).$$

Then $P(ch_s(K_1 U K_1' V^{-1}) > c) \to 1$ as $\alpha \to \infty$.

<u>Proof</u>: The proof is identical to that of Lemma 3.5.6.

5.4 The Limiting Distribution of $\sum_{i=s}^{m} \psi_i$

If $\sum_{i=s}^{m} \psi_i$ is to be used as a test statistic in the test of the hypothesis $H_0^{(s)}$: rank(M) $\leq$ s-1 against $H_1^{(s)}$: rank(M) = s, it is necessary to compute the significance level, α, where

$$\alpha = \sup_{H_0^{(s)}} P\left(\sum_{i=s}^{m} \psi_i > c \middle| H_0^{(s)}\right).$$

Let $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_m$ be the latent roots of $I+\sigma^{-2}M$, and recall that the null hypothesis can be written $H_0^{(s)}$: $\delta_s = 1$, or more precisely, $H_0^{(s)}$: $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_{s-1} \geq 1, \delta_s = \delta_{s+1} = \ldots = \delta_m = 1$. We will write $\psi_{i:m}(\delta_1, \delta_2, \ldots, \delta_m)$ to indicate that ψ_i is the i^{th} largest root of m roots and depends on the population roots $\delta_1, \delta_2, \ldots, \delta_m$. Then we may write α, the significance level, as

$$\alpha = \sup_{\delta_1 \geq \delta_2 \geq \ldots \geq \delta_{s-1} \geq 1} P\left(\sum_{i=s}^{m} \psi_{i:m}(\delta_1, \delta_2, \ldots, \delta_{s-1}, 1, \ldots, 1) > c\right).$$

However, we have seen in the previous section that $\psi_{i:m}$ is stochastically increasing in each δ_j: j = 1,2,...,m. It then follows that

$$\alpha = P\left(\sum_{i=s}^{m} \psi_{i:m}(\infty, \infty, \ldots, \infty, 1, \ldots, 1) > c\right),$$

where $\psi_{i:m}(\infty, \infty, \ldots, \infty, 1, \ldots, 1)$ denotes the random variable which has the limiting distribution of $\psi_{i:m}(\delta_1, \delta_2, \ldots, \delta_{s-1}, 1, \ldots, 1)$ as $\delta_j \to \infty$: j = 1,2,...,s-1. Hence, we need to determine the distribution of $\sum_{i=s}^{m} \psi_{i:m}(\infty, \infty, \ldots, \infty, 1, \ldots, 1)$.

Recall that $b \sim \sigma^2\chi^2_\beta$ and $H \sim W_m(\sigma^2 I+M,h,0)$, and there exists an orthogonal matrix P such that $P(I+\sigma^{-2}M)P'$ $= \Delta = \text{diag}(\delta_1,\delta_2,\ldots,\delta_m)$. If we define $\tilde{B}$ and $\tilde{H}$ as

$$\tilde{B} = (b/\sigma^2)\Delta^{-1},$$

$$\tilde{H} = \Delta^{-\frac{1}{2}}P(\sigma^{-2}H)P'\Delta^{-\frac{1}{2}}$$

where $\Delta^{-\frac{1}{2}} = \text{diag}(\delta_1^{-\frac{1}{2}},\delta_2^{-\frac{1}{2}},\ldots,\delta_m^{-\frac{1}{2}})$, then, clearly, $\tilde{H} \sim W_m(I,h,0)$ and $\psi_{i:m}(\delta_1,\delta_2,\ldots,\delta_m) = ch_i((\sigma^{-2}H)((b/\sigma^2)I)^{-1})$ $= ch_i(\tilde{H}\tilde{B}^{-1})$. Then if we let $\tilde{B}_n = (b/\sigma^2)\Delta_n^{-1}$, where $\Delta_n = \text{diag}(n\delta_1,n\delta_2,\ldots,n\delta_{s-1},1,\ldots,1)$, we need to find the limiting distribution of $\sum_{i=s}^{m} ch_i(\tilde{H}\tilde{B}_n^{-1})$ as $n \to \infty$.

We will need the following result.

Lemma 5.4.1: Suppose $v \sim \chi^2_\alpha$. Then

$$\underline{x}_n(m\times 1) = \begin{bmatrix} c_1 v/n \\ c_2 v/n \\ \vdots \\ c_{s-1}v/n \\ v \\ \vdots \\ v \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \underline{x}_1 \\ v \\ \vdots \\ v \end{bmatrix} = \underline{x},$$

where $c_1,c_2,\ldots,c_{s-1}$ are constants and $\underline{x}_1((s-1)\times 1)$ is a degenerate random vector with all of its probability at $\underline{0}$.

Proof: Clearly $\underline{x}_1$ and v are independent, so the characteristic function of $\underline{x}$ is

$$E[\exp(i\ \underline{x}'\underline{t})] = E[\exp(i\ \underline{x}_1'\underline{t}_1)\exp(i\ v\sum_{j=s}^{m} t_j)]$$

$$= E[\exp(i\ \underline{x}_1'\underline{t}_1)]\quad E[\exp(i\ v\sum_{j=s}^{m} t_j)]$$

$$= E[\exp(i\ v\sum_{j=s}^{m} t_j)]$$

$$= (1 - i\ 2\sum_{j=s}^{m} t_j)^{-\frac{1}{2}\alpha}.$$

Now the characteristic function of $\underline{x}_n$ is

$$E[\exp(i\ \underline{x}_n'\underline{t})] = E[\exp(i(\sum_{j=1}^{s-1} c_j t_j/n + \sum_{j=s}^{m} t_j)v]$$

$$= (1 - i\ 2(\sum_{j=1}^{s-1} c_j t_j/n + \sum_{j=s}^{m} t_j))^{-\frac{1}{2}\alpha},$$

so $\lim_{n\to\infty} E[\exp(i\ \underline{x}_n'\underline{t})] = (1 - i\ 2\sum_{j=s}^{m} t_j)^{-\frac{1}{2}\alpha} = E[\exp(i\ \underline{x}'\underline{t})]$.

The result now follows from the continuity theorem (Lemma 3.6.2).

From Lemma 5.4.1 we observe that $\tilde{B}_n \xrightarrow{d} \tilde{B}$ with

$$\tilde{B} = \begin{pmatrix} \tilde{B}_1 & (0) \\ (0) & \tilde{B}_2 \end{pmatrix},$$

where $\tilde{B}_1((s-1)\times(s-1)) = (0)$ with probability one, $\tilde{B}_2((m-s+1) \times (m-s+1)) = \tilde{b}I$, and $\tilde{b} \sim \chi^2_\beta$. We now need to show that $\sum_{i=s}^{m} ch_i(\tilde{H}\tilde{B}^{-1})$ is continuous with probability one under the distribution of $(\tilde{H},\tilde{B})$. Put

$$\tilde{H} = \begin{pmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{pmatrix},$$

where $\tilde{H}_{11}$ is $(s-1)\times(s-1)$, $\tilde{H}_{12}$ is $(s-1)\times(m-s+1)$, $\tilde{H}_{21}$ is $(m-s+1)\times(s-1)$, and $\tilde{H}_{22}$ is $(m-s+1)\times(m-s+1)$. Then the roots of interest are the solutions to

$$\left| \begin{pmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{pmatrix} - \psi \begin{pmatrix} (0) & (0) \\ (0) & \tilde{B}_2 \end{pmatrix} \right| = 0. \qquad (5.4.1)$$

Since $\tilde{H}$ is nonsingular with probability one, we may put

$$\tilde{H}^{-1} = G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

so that (5.4.1) can be written

$$\left| I_m - \psi \begin{pmatrix} (0) & G_{12}\tilde{B}_2 \\ (0) & G_{22}\tilde{B}_2 \end{pmatrix} \right| = 0,$$

or

$$\begin{vmatrix} I_{s-1} & -\psi G_{12}\tilde{B}_2 \\ (0) & I_{m-s+1} - \psi G_{22}\tilde{B}_2 \end{vmatrix} = 0.$$

Hence, it must be true that

$$|I_{m-s+1} - \psi G_{22}\tilde{B}_2| = 0,$$

or

$$|G_{22}^{-1} - \psi\tilde{B}_2| = 0. \qquad (5.4.2)$$

Thus, with probability one $ch_1(\tilde{H}\tilde{B}^{-1}), ch_2(\tilde{H}\tilde{B}^{-1}), \ldots, ch_{s-1}(\tilde{H}\tilde{B}^{-1})$ are undefined and $ch_s(\tilde{H}\tilde{B}^{-1}), ch_{s+1}(\tilde{H}\tilde{B}^{-1}), \ldots, ch_m(\tilde{H}\tilde{B}^{-1})$ are the solutions to (5.4.2); that is, since $\tilde{B}$ is of rank m-s+1 with probability one, there are only m-s+1 solutions to $|\tilde{H}-\psi\tilde{B}| = 0$. Now since $\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12}$ is nonsingular with probability one, it follows that $G_{22} = (\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12})^{-1}$, and so (5.4.2) can be rewritten

$$|\tilde{H}_{22} - \tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12} - \psi\tilde{B}_2| = 0.$$

Clearly $\tilde{B}_2$ is also nonsingular with probability one, and thus, by Lemma 3.6.6, $ch_i(\tilde{H}\tilde{B}^{-1})$ is continuous with probability one under the distribution of $(\tilde{H},\tilde{B})$ for $i = s,s+1,\ldots,m$. This implies that $\sum_{i=s}^{m} ch_i(\tilde{H}\tilde{B}^{-1})$ is also continuous with probability one under the distribution of $(\tilde{H},\tilde{B})$. Note that the set of discontinuity points, R, is closed, since $R = \{(\tilde{H},\tilde{B}): |\tilde{B}_2| = 0\}$, and also recall that $\tilde{H}_{22}-\tilde{H}_{21}\tilde{H}_{11}^{-1}\tilde{H}_{12} \sim W_{m-s+1}(I,h-s+1,0)$. Therefore, from Lemma 3.6.7, since $(\tilde{H},\tilde{B}_n) \xrightarrow{d} (\tilde{H},\tilde{B})$, it follows that for $i = s,s+1,\ldots,m$,

$$\psi_{i:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) \sim \psi_{i-s+1:m-s+1}(1,1,\ldots,1),$$

where $\psi_{i-s+1:m-s+1}(1,1,\ldots,1)$ denotes the distribution of the $i-s+1^{th}$ largest root of $|W-\psi vI| = 0$, with $W \sim W_{m-s+1}(I,h-s+1,0)$ and $v \sim \chi^2_\beta$, independently. Now if we let $\theta_1>\theta_2>\ldots>\theta_{m-s+1}$ be the solutions to $|W-\theta I| = 0$, then we can put $\psi_{i:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) = \theta_{i-s+1}/v$, so that

$$\sum_{i=s}^{m} \psi_{i:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) = \sum_{j=1}^{m-s+1} \theta_j/v = (\mathrm{tr}\ W)/v.$$

But $\mathrm{tr}\ W \sim \chi^2_\nu$, where $\nu = (m-s+1)(h-s+1)$, so

$$\sum_{i=s}^{m} \psi_{i:m}(\infty,\infty,\ldots,\infty,1,\ldots,1) \sim \frac{\nu}{\beta} F^{\nu}_{\beta}.$$

Hence, in testing $H_0^{(s)}$: rank(M) $\leq$ s-1 against $H_1^{(s)}$: rank(M) = s, we choose $\frac{\nu}{\beta} F(\nu,\beta,\alpha)$ as our critical value, where $F(\nu,\beta,\alpha)$ is the constant for which $P(F(\nu,\beta) > F(\nu,\beta,\alpha)) = \alpha$ when $F(\nu,\beta) \sim F_{\beta}^{\nu}$. By so doing we will guarantee

$$\sup_{H_0^{(s)}} P(\sum_{i=s}^{m} \psi_{i:m}(\delta_1,\delta_2,\ldots,\delta_m) > \frac{\nu}{\beta} F(\nu,\beta,\alpha) \mid H_0^{(s)}) = \alpha.$$

In order to determine the rank of M, we will again use a sequential procedure. To illustrate this procedure, we will return to the example presented in Section 4.2. Recall that D = diag(94.1065, 34.8845, 1.01721, .618312), h = 20, and β = 420, so that since $\psi_i = hd_i/\beta$: i = 1,2,3,4, ψ_1 = 4.4813, ψ_2 = 1.6612, ψ_3 = .04844, and ψ_4 = .029443. We will first consider testing the hypothesis $H_0^{(4)}$: rank(M) $\leq$ 3 against $H_1^{(4)}$: rank(M) = 4. We reject the null hypothesis, $H_0^{(4)}$, if $\psi_4 >$ 17 F(17, 420, .05)/420. Now 17 F(17, 420, .05)/420 is approximately equal to .066 and ψ_4 = .029443 <.066, so that we do not reject $H_0^{(4)}$ and, instead, consider testing the hypothesis $H_0^{(3)}$: rank(M) $\leq$ 2 against $H_1^{(3)}$: rank(M) = 3. The quantity 36 F(36, 420, .05)/420 is approximately equal to .122, and clearly $\psi_3 + \psi_4$ = .07788 < .122, so that the null hypothesis, $H_0^{(3)}$, is not rejected. Since $H_0^{(3)}$ is not rejected, we next consider testing the hypothesis $H_0^{(2)}$: rank(M) $\leq$ 1 against $H_1^{(2)}$: rank(M) = 2. We find that 57 F(57, 420, .05)/420 is approximately equal to .181, and therefore, since

$\psi_2 + \psi_3 + \psi_4 = 1.7391 > .181$, we reject $H_0^{(2)}$ and conclude that the rank of M could very reasonably be taken as being two.

Note that this sequential procedure is open to the same objections, regarding the use of the significance level, α, at each step, mentioned earlier in Section 3.6. Again, however, it seems unlikely to cause serious error in practice. If the true rank of M is p, then there is a small probability, usually less than α, that the rank, s, determined by the sequential procedure will be greater than p. Also, if δ_p is sufficiently large, then the probability of s being less than p is also small.

BIBLIOGRAPHY

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proceedings of the American Mathematical Society 6, 170-176.

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York.

Anderson, T. W., and Das Gupta, S. (1964). A monotonicity property of the power functions of some tests of the equality of two covariance matrices. Annals of Mathematical Statistics 35, 1059-1063.

Barnard, G. A. (1963). Some logical aspects of the fiducial argument. Journal of the Royal Statistical Society B 25, 111-114.

Bellman, R. (1970). Introduction to Matrix Analysis. McGraw-Hill Book Company, Inc., New York.

Breiman, L. (1968). Probability. Addison-Wesley Publishing Company, Reading, Massachusetts.

Das Gupta, S., Anderson, T. W., and Mudholkar, G. S. (1964). Monotonicity of the power functions of some tests of the multivariate linear hypothesis. Annals of Mathematical Statistics 35, 200-205.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Company, Inc., Belmont, California.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. Annals of Mathematical Statistics 35, 475-501.

Kendall, M. G., and Stuart, A. (1963). The Advanced Theory of Statistics Vol. 1. Hafner Publishing Company, New York.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley & Sons, Inc., New York.

Mann, H. B.,and Wald, A. (1943). On stochastic limit and order relationships. Annals of Mathematical Statistics 14, 217-226.

Marshall, A. W., and Olkin I. (1974). Majorization in multivariate distributions. Annals of Statistics 2, 1189-1200.

Morrison, D. F. (1976). Multivariate Statistical Methods. McGraw-Hill Book Company, Inc., New York.

Ostrowski, A. M. (1973). Solution of Equations in Euclidean and Banach Spaces. Academic Press, Inc., New York.

Pillai, K. C. S. (1965). On the distribution of the largest characteristic root of a matrix in multivariate analysis. Biometrika 52, 405-414.

Pillai, K. C. S. (1967). Upper percentage points of the largest root of a matrix in multivariate analysis. Biometrika 54, 189-194.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Annals of Mathematical Statistics 24, 220-238.

BIOGRAPHICAL SKETCH

James Robert Schott was born on January 9, 1955, in Cincinnati, Ohio, where he spent the first twenty-two years of his life. Upon graduating from La Salle High School in June, 1973, he attended Xavier University, which is located in Cincinnati, and received the degree of Bachelor of Science with a major in mathematics in June, 1977.

In September, 1977, Jim enrolled in the graduate school at the University of Florida and was awarded the degree of Master of Statistics in March, 1979. Since that time he has been working toward the degree of Doctor of Philosophy. While at the University of Florida, Jim has been a recipient of a graduate fellowship and, in addition, he has been employed by the Department of Statistics as a graduate assistant for both teaching and consulting duties.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John G. Saw, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alan G. Agresti
Associate Professor of
Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ralph R. Grams
Professor of Pathology

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1981

Dean for Graduate Studies
and Research