TESTING FOR EQUALITY OF MEANS, EQUALITY OF VARIANCES, AND EQUALITY OF COVARIANCES UNDER RESTRICTIONS UPON THE PARAMETER SPACE

LEON JAY GLESER* AND INGRAM OLKIN**

(Received July 23, 1968; revised Jan. 13, 1969)

1. Introduction

Suppose that the p-dimensional random row vector $x=(x_1, x_2, \cdots, x_p)$ has a multivariate normal distribution with mean vector $\mu=(\mu_1, \mu_2, \cdots, \mu_p)$ and covariance matrix $\Sigma=(\sigma_{ij})$. The test of the hypothesis that the distribution of x is symmetric with respect to the components $x_1, x_2, \cdots, x_p$ is equivalent to the hypothesis denoted by H_{mvc} that the components of the mean vector μ are equal ($\mu=(\eta, \cdots, \eta)$, η a constant), that the variances are equal ($\sigma_{ii}=\sigma^2$, $i=1, 2, \cdots, p$), and that the covariances are equal ($\sigma_{ij}=\sigma^2\rho$, $i\neq j$). The less restricted hypothesis H_{vc} that the variances are equal and that the covariances are equal (i.e., that Σ has the *intra-class correlational structure*) may also be of concern.

Likelihood ratio tests for testing H_{mvc} against the general alternatives for testing H_{vc} against general alternatives, and for testing H_{mvc} against H_{vc} have been derived by Wilks [7]. Wilks also considers methods of obtaining the null distributions of these test statistics.

Wilks' interest in the above statistical problem arose from a problem in psychological testing theory in which it is desired to test the hypothesis that p examinations are "parallel forms" of the same examination. To test this hypothesis, the p examinations are given to each of N subjects. If the hypothesis is true, and if test scores for each individual have a p-dimensional multivariate normal distribution, then the means, variances, and covariances of the test scores should obey H_{mvc}.

However, further consideration of the above psychological problem suggests that if the hypothesis of "parallel forms" holds, then a more restricted hypothesis concerning the means, variances, and covariances is appropriate. Since "parallel forms" of an examination presumably

* Supported in part by the Office of Naval Research, Contract No. 4010(09).
** Supported in part by the National Science Foundation, Grant No. 6681.

measure a common aptitude (or aptitudes) in each subject, one expects that the p examination scores for each subject are equally and *positively* correlated with one another. Thus, although under the hypothesis H_{mvc} (or the hypothesis H_{vc}) we only require that the common correlation ρ be restricted to the range from $-1/(p-1)$ to 1 (so as to guarantee that Σ is positive definite), we actually are interested in the restricted hypothesis H_{mvc}^r (or the restricted hypothesis H_{vc}^r) that requires ρ to fall in the range from ρ_0 to 1, where ρ_0 is some nonnegative number.

A hypothesis of the form H_{mvc}^r is also of concern when we are testing that the observations from a balanced one way analysis of variance design have a joint distribution obeying the assumptions of the Model II Analysis of Variance. If x is one complete replication of the design (one observation in each cell), then the Model II assumptions state that x can be written as

$$x=(v, v, \cdots, v)+\varepsilon=ve+\varepsilon ,$$

where $e=(1, 1, \cdots, 1)$, v is a scalar random variable, $v\sim N(\theta, \tau_0^2)$, $\varepsilon\sim N(0, \tau_1^2 I_p)$, and v and ε are independently distributed. Consequently, if the Model II assumptions hold, $x\sim N(\theta e, \tau_1^2[(1-\rho)I+\rho e'e])$, where $0\leqq\rho=\tau_0^2/(\tau_0^2+\tau_1^2)\leqq 1$. To test the fit of the Model II assumptions, we should test H_{mvc}^r (with $0=\rho_0\leqq\rho<1$) against general alternatives. Box [3] considers the problem of testing the Model II assumptions and obtains an approximate test by ignoring the restriction $\rho\geqq 0$. Herbach [4] has extensively studied certain tests related to the model H_{mvc}^r with $\rho_0=0$, and obtains some optimal properties of these tests.

In the present paper, we are interested in testing the hypotheses H_{mvc}^r and H_{vc}^r against general alternatives. Once it has been established that one of these hypotheses is an appropriate model for the data, we are also interested in the estimation of the parameters ρ and σ^2. (For the case $\rho_0=0$, Herbach [4] obtains the MLE and discusses completeness.)

In section 2, maximum likelihood estimators for ρ and σ^2 under H_{mvc}^r and H_{vc}^r are obtained. The exact and asymptotic distributions of these estimators are also discussed in section 2. Section 3 is concerned with deriving the likelihood ratio tests for H_{mvc}^r and H_{vc}^r against general alternatives, and with finding the asymptotic null distributions of these test statistics.

Before completing this introduction, it is useful to indicate the meanings of some of the notation used in the following sections. The symbol "$\sim$" means "is distributed as"; thus $x\sim N(\mu, \Sigma)$ means that x has the distribution of a multivariate normal random variable with mean vector μ and covariance matrix Σ. The notation $\mathcal{L}(x)=N(\mu, \Sigma)$ carries an equivalent meaning. We use both notations since the "$\mathcal{L}$"-notation is easier for use with convergences in law; thus, $\lim_{n\to\infty} \mathcal{L}(x_n)=x$ means that x_n converges in distribution to x. The symbol χ_d^2 denotes the law of a chi-squared variable having d degrees of freedom; the symbol $B(a, b)$ denotes either the law of a beta variable with a and b degrees of freedom, or the number $\Gamma(a)\Gamma(b)/\Gamma(a+b)$, depending on context. Finally, the symbol "$\propto$" means "is proportional to".

2. Maximum likelihood estimators for the parameters of the covariance matrix under H_{mvc}^r and H_{vc}^r

2.1) *Preliminaries.* Assume that we have N independent observations $x^{(1)}, x^{(2)}, \cdots, x^{(N)}$ upon the random p-dimensional row vector x. Recall that x has a multivariate normal distribution with mean vector μ and covariance matrix Σ. A sufficient statistic for (μ, Σ) is $(\bar{x}, S)$, where $\bar{x}=(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)=N^{-1}\sum_{j=1}^{N} x^{(j)}$ is the sample mean vector, and $S=(S_{ij})=\sum_{k=1}^{N}(x^{(k)}-\bar{x})'(x^{(k)}-\bar{x})$ is the sample cross-product matrix ($(N-1)^{-1}S$ is the sample covariance matrix). It is a well-known result that $\bar{x}$ and S are independent, that $\bar{x}\sim N(\mu, N^{-1}\Sigma)$, and that S has the Wishart distribution $W(\Sigma; p, n)$, $n\equiv N-1$. The joint density of $\bar{x}$ and S is thus:

$$(2.1)\qquad p(\bar{x}, S)=c(p, n)(2\pi)^{-p/2}N^{p/2}|\Sigma|^{-N/2}|S|^{(n-p-1)/2} \cdot \exp\left\{-\frac{1}{2}\{\mathrm{tr}\,\Sigma^{-1}S+N\,\mathrm{tr}\,\Sigma^{-1}(\bar{x}-\mu)'(\bar{x}-\mu)\}\right\},$$

where $c(p, n)\equiv\left[2^{np/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{n-i+1}{2}\right)\right]^{-1}$ and the range of definition for $p(\bar{x}, S)$ is S positive definite (i.e., $S>0$), $\bar{x}$ unrestricted.

We make use of the following lemma in section 2.2).

LEMMA 2.1. *For any matrix A having the form $A=aI+be'e$ where* $e=(1, 1, \cdots, 1)$:

(i) *A can be represented in the form $A=\Gamma'D\Gamma$, where Γ is any $p\times p$ orthogonal matrix having first row $p^{-1/2}e$, and $D=\mathrm{diag}\,((a+bp), a, a, \cdots, a)$,*

(ii) $|A-\lambda I|=(a-\lambda)^{p-1}[a+bp-\lambda]$.

Thus, under H_{vc} the covariance matrix Σ can be reduced to diagonal form by an orthogonal matrix whose elements are independent of the parameters. This fact allows us to considerably simplify our calculations.

Before deriving the maximum likelihood estimators of ρ and σ^2 under H_{mvc}^r and H_{vc}^r, we indicate the precise form of the parameter space for each of these two hypotheses. The parameter space Ω^r_{mvc} for H^r_{mvc} takes the form:

$$\text{(2.2)} \quad \Omega^r_{mvc}=\{(\mu, \Sigma):\ \mu=\eta e,\ \eta \text{ a constant},\ \Sigma=\sigma^2[(1-\rho)I+\rho e'e],\ -\infty<\eta<\infty,\ \sigma^2>0,\ \rho_0\leqq\rho<1\};$$

the parameter space Ω^r_{vc} for H^r_{vc} has the form:

$$\text{(2.3)} \quad \Omega^r_{vc}=\{(\mu, \Sigma):\ \mu \text{ unrestricted},\ \Sigma=\sigma^2[(1-\rho)I+\rho e'e],\ \sigma^2>0,\ \rho_0\leqq\rho<1\},$$

where ρ_0 is a given constant satisfying $-1/(p-1)<\rho_0<1$. Note that we have allowed ρ_0 to be negative. For applications to the Model II analysis of variance, ρ_0 is zero; for applications to the psychological testing problem mentioned by Wilks [7], ρ_0 is of moderate size and positive.

2.2) *Derivation of the maximum likelihood estimators of* ρ *and* Σ *under* Ω^r_{mvc} *and* Ω^r_{mc}. Since $(\bar{x}, S)$ is a sufficient statistic for (μ, Σ), we begin our derivation of the maximum likelihood estimators (MLE) of ρ and σ^2 by considering the density (2.1). Motivated by Lemma 2.1 (and the remark following that lemma), we let $y=\sqrt{N}\bar{x}\Gamma'$ and $V=\Gamma S\Gamma'$, where Γ is any $p\times p$ orthogonal matrix with first row $p^{-1/2}e$. Then, under Ω^r_{vc} (and under $\Omega^r_{mvc}\subset\Omega^r_{vc}$), we have that y and V are independently distributed, that $y\sim N(\xi, D)$, and that $V\sim W(D; p, n)$, where $\xi=\sqrt{N}\mu\Gamma'$, $D=\Gamma\Sigma\Gamma'=\text{diag}\,\{\sigma_1^2, \sigma_2^2, \sigma_2^2, \cdots, \sigma_2^2\}$, and

$$\text{(2.4)} \qquad \sigma_1^2=\sigma^2(1+(p-1)\rho)\,, \qquad \sigma_2^2=\sigma^2(1-p)\,.$$

Note that the joint distribution of y and V becomes

$$\text{(2.5)} \quad p(y, V)=c(p, n)(2\pi)^{-p/2}\sigma_1^{-N}\sigma_2^{-N(p-1)}\,|V|^{(n-p-1)/2} \cdot \exp\left\{-\frac{1}{2}\left[\sigma_1^{-2}(v_{11}+(y_1-\xi_1)^2)+\sigma_2^{-2}\left(\sum_{i=2}^p v_{ii}+\sum_{i=2}^p (y_i-\xi_i)^2\right)\right]\right\}.$$

From (2.5) it is easily verified that $\left(y, v_{11}, \sum_{i=2}^p v_{ii}\right)$ is a sufficient statistic for $(\xi, \sigma_1^2, \sigma_2^2)$. It may be directly verified that y, v_{11} and $\sum_{i=2}^p v_{ii}$ are independently distributed, that $\sigma_1^{-2}v_{11}\sim\chi^2_n$, and that $\sigma_2^{-2}\sum_{i=2}^p v_{ii}\sim\chi^2_{n(p-1)}$.

To find the MLE of ρ and σ^2, first maximize with respect to ξ in (2.5).

Case I. For $\xi=\sqrt{N}\mu\Gamma'$ subject to the restrictions of the parameter space Ω^r_{vc} (i.e., ξ unrestricted), the MLE of ξ is $\hat{\xi}=y$ so that

$$\text{(2.6a)} \quad \max_{\xi} p(y, V)\propto\sigma_1^{-N}\sigma_2^{-N(p-1)}\exp\left\{-\frac{1}{2}\left(\sigma_1^{-2}v_{11}+\sigma_2^{-2}\sum_{i=2}^p v_{ii}\right)\right\}.$$

Case II. For ξ subject to the restrictions of Ω^r_{mvc}, we have that $\xi=(\xi_1, 0, 0, \cdots, 0)$, $-\infty<\xi_1<\infty$. Thus $\hat{\xi}_1=y_1$, and

$$\max_{\substack{\xi=(\xi_1,0,0,\cdots,0)\\ -\infty<\xi_1<\infty}} p(y, V)\propto\sigma_1^{-N}\sigma_2^{-N(p-1)}\exp\left\{-\frac{1}{2}\left[\sigma_1^{-2}v_{11}+\sigma_2^{-2}\sum_{i=2}^{p}(v_{ii}+y_i^2)\right]\right\}. \tag{2.6b}$$

Case III. We can also consider the case Ω^r_{ovc} when μ is assumed to be zero. Then

$$\max_{\xi=0} p(y, V)\propto\sigma_1^{-N}\sigma_2^{-N(p-1)}\exp\left\{-\frac{1}{2}\left[\sigma_1^{-2}(v_{11}+y_1^2)+\sigma_2^{-2}\sum_{i=2}^{p}(v_{ii}+y_i^2)\right]\right\}. \tag{2.6c}$$

Expressions (2.6a), (2.6b) and (2.6c) have a common form, namely:

$$p(u, w)\equiv\max_{\xi} p(y, V)=c\sigma_1^{-N}\sigma_2^{-N(p-1)}\exp\left\{-\frac{1}{2}[\sigma_1^{-2}u+\sigma_2^{-2}w]\right\}, \tag{2.7}$$

where c is a normalizing constant and where u and w are given by

Case I. $u=v_{11}$, $\quad w=\sum_{i=2}^{p} v_{ii}$,

Case II. $u=v_{11}$, $\quad w=\sum_{i=2}^{p}(v_{ii}+y_i^2)$,

Case III. $u=v_{11}+y_1^2$, $\quad w=\sum_{i=2}^{p}(v_{ii}+y_i^2)$.

In all three cases, u and w are independently distributed, $u\sim\sigma_1^2\chi^2_{\nu_1}$, $w\sim\sigma_2^2\chi^2_{\nu_2}$ where in *Case* I (ν_1, ν_2) equals $(n, n(p-1))$, in *Case* II (ν_1, ν_2) equals $(n, N(p-1))$, and in *Case* III (ν_1, ν_2) equals $(N, N(p-1))$.

Let

$$\theta=\frac{1}{\sigma_1^2}, \qquad \delta=\frac{\sigma_1^2-\gamma_0\sigma_2^2}{\sigma_1^2\sigma_2^2}, \tag{2.8}$$

where

$$\gamma_0=\frac{1+(p-1)\rho_0}{1-\rho_0}.$$

Note that the transformation from (σ_1^2, σ_2^2) to (θ, δ) is one to one onto. In terms of the new parameters θ and δ, the condition $\rho\geqq\rho_0$ becomes $\delta\geqq0$, and

$$\sigma^2=\frac{\sigma_1^2+(p-1)\sigma_2^2}{p}=\frac{\delta+(\gamma_0+p-1)\theta}{p\theta(\delta+\gamma_0\theta)}, \tag{2.9}$$

$$\rho=\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2+(p-1)\sigma_2^2}=\frac{\delta+(\gamma_0-1)\theta}{\delta+(\gamma_0+p-1)\theta}. \tag{2.10}$$

Reparametrizing (2.7) in terms of θ and δ and taking logarithms, we obtain

$$(2.11)\quad 2\log p(u, w)=N\log\theta+N(p-1)\log(\delta+\theta\gamma_0)-\theta(u+w\gamma_0)-w\delta\,.$$

This function is strictly concave in θ and δ, so that its maximum is achieved either in the interior of the positive orthant $\theta>0$, $\delta>0$, or on the boundary $\theta>0$, $\delta=0$. Differentiating (2.11) with respect to θ and δ, we find that $p(u, v)$ is maximized for

$$\hat{\theta}=\frac{N}{u}\,,\qquad \hat{\delta}=\frac{N}{uw}[(p-1)u-\gamma_0 w]\,,$$

provided that $(p-1)u-\gamma_0 w>0$. Otherwise, the maximum is achieved when $\delta=0$, in which case

$$\hat{\theta}=\frac{pN}{u+w\gamma_0}\,,\qquad \hat{\delta}=0\,.$$

Consequently, we have

$$(2.12)\qquad \hat{\sigma}_1^2=\begin{cases}\dfrac{u}{N}\,, & \text{if } \dfrac{u}{w}>\dfrac{\gamma_0}{p-1}\,,\\[2ex] \dfrac{u+w\gamma_0}{pN}\,, & \text{if } \dfrac{u}{w}\leqq\dfrac{\gamma_0}{p-1}\,,\end{cases}$$

$$(2.13)\qquad \hat{\sigma}_2^2=\begin{cases}\dfrac{w}{(p-1)N}\,, & \text{if } \dfrac{u}{w}>\dfrac{\gamma_0}{p-1}\,,\\[2ex] \dfrac{u+w\gamma_0}{p\gamma_0 N}\,, & \text{if } \dfrac{u}{w}\leqq\dfrac{\gamma_0}{p-1}\,,\end{cases}$$

or in terms of $\hat{\sigma}^2$ and $\hat{\rho}$,

$$(2.14)\qquad \hat{\sigma}^2=\begin{cases}\dfrac{u+w}{pN}\,, & \text{if } \dfrac{u}{w}>\dfrac{\gamma_0}{p-1}\,,\\[2ex] \dfrac{(u+w\gamma_0)(\gamma_0+p-1)}{p^2\gamma_0 N}\,, & \text{if } \dfrac{u}{w}\leqq\dfrac{\gamma_0}{p-1}\,,\end{cases}$$

$$(2.15)\qquad \hat{\rho}=\begin{cases}\dfrac{(p-1)u-w}{(p-1)(u+w)}\,, & \text{if } \dfrac{u}{w}>\dfrac{\gamma_0}{p-1}\,,\\[2ex] \rho_0\,, & \text{if } \dfrac{u}{w}\leqq\dfrac{\gamma_0}{p-1}\,.\end{cases}$$

The expressions for $\hat{\sigma}^2$ and $\hat{\rho}$ in terms of $\bar{x}$ and S can now be worked out for each case in (2.7) by substituting the following values for u and w into (2.14) and (2.15):

Case I. $u=p^{-1}\sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}\,, \qquad w=\sum_{i=1}^{p} s_{ii}-u\,,$

Case II. $u=p^{-1}\sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}\,, \qquad w=\sum_{i=1}^{p}[s_{ii}+N(\bar{x}_i)^2]-Np^{-1}\left(\sum_{i=1}^{p}\bar{x}_i\right)^2-u\,,$

Case III. $u=p^{-1}\left[\sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}+N\left(\sum_{i=1}^{p}\bar{x}_i\right)^2\right],$

$$w=\sum_{i=1}^{p}[s_{ii}+N(\bar{x}_i)^2]-u\,,$$

where $\bar{x}=(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)$.

Remark. We note that

$$\hat{\rho}=\max(\tilde{\rho}, \rho_0)\,,$$

$$\hat{\sigma}^2=\begin{cases}\tilde{\sigma}^2\,, & \text{if } \tilde{\rho}>\rho_0\,,\\ \tilde{\sigma}^2(\rho_0)\,, & \text{if } \tilde{\rho}\leqq\rho_0\,,\end{cases}$$

where $\tilde{\rho}$ and $\tilde{\sigma}^2$ are the MLE of ρ and σ^2 when ρ is unrestricted, and $\tilde{\sigma}^2(\rho_0)$ is the MLE of σ^2 when it is known that $\rho=\rho_0$.

2.3) *A distributional representation for the MLE.* We now make use of the fact that if two random variables s_1 and s_2 are independently distributed, each having a chi-square distribution, then s_1+s_2 is independent of $s_1/(s_1+s_2)$. Thus if we let

$$s=\frac{u}{\sigma_1^2}+\frac{w}{\sigma_2^2}\,, \qquad r=\frac{w}{\sigma_2^2}\Big/\left(\frac{u}{\sigma_1^2}+\frac{w}{\sigma_2^2}\right), \tag{2.16}$$

it follows that r and s are independent with $s\sim\chi^2_{\nu_1+\nu_2}$ and $r\sim B\left(\frac{1}{2}\nu_2, \frac{1}{2}\nu_1\right)$, where (ν_1, ν_2) is equal to $(n, (p-1)n)$ in Case I, $(n, (p-1)N)$ in Case II and $(N, (p-1)N)$ in Case III.

The above transformation permits us to represent $\hat{\rho}$ as a function of r alone, and to represent $\hat{\sigma}^2$ as a function of r multiplied by a function of s, so that we can take advantage of independence. Thus, we have

$$\hat{\rho}=\max\left\{\frac{(p-1)\gamma(1-r)-r}{(p-1)[r+\gamma(1-r)]}, \rho_0\right\}, \tag{2.17}$$

where $\gamma=[1+(p-1)\rho]/(1-\rho)=\sigma_1^2/\sigma_2^2$. Also,

$$\hat{\sigma}^2=\sigma^2\frac{s}{Np}h(r)\,, \tag{2.18}$$

where

$$h(r)=\begin{cases}[1+(p-1)\rho](1-r)+(1-\rho)r\,, & 0\leqq r\leqq\lambda\,,\\ \dfrac{1+(p-1)\rho}{1+(p-1)\rho_0}(1-r)+\dfrac{1-\rho}{1-\rho_0}r\,, & \lambda<r\leqq 1\,,\end{cases}\tag{2.19}$$

and $\lambda=\gamma(p-1)/[\gamma(p-1)+\gamma_0]$.

2.4) *Remarks on expected values.* Using (2.18), we have

$$E\hat{\sigma}^2=\sigma^2E\left(\frac{s}{Np}\right)E[h(r)]=\sigma^2\frac{\nu_1+\nu_2}{Np}E[h(r)]\,,$$

$$E\hat{\sigma}^4=\sigma^4\frac{(\nu_1+\nu_2)(\nu_1+\nu_2+2)}{N^2p^2}E[h(r)]^2\,.$$

By a direct computation we have

$$\begin{aligned}E[h(r)]=&\frac{\nu_2}{\nu_1+\nu_2}\frac{1-\rho}{1-\rho_0}\left[1-\rho_0I_\lambda\left(\frac{\nu_2+2}{2},\frac{\nu_1}{2}\right)\right]\\&+\frac{1+(p-1)\rho}{1+(p-1)\rho_0}\left[\frac{\nu_1}{\nu_1+\nu_2}-(p-1)\rho_0I_\lambda\left(\frac{\nu_2}{2},\frac{\nu_1}{2}\right)\right.\\&\left.+\frac{(p-1)\nu_2\rho_0}{\nu_1+\nu_2}I_\lambda\left(\frac{\nu_2+2}{2},\frac{\nu_1}{2}\right)\right]\end{aligned}$$

for $I_\lambda(a,b)=[B(a,b)]^{-1}\int_0^\lambda y^{a-1}(1-y)^{b-1}dy$. The computation of $E[h(r)]^2$ is straightforward, though tedious, and results in a linear combination of $I_\lambda\left(\frac{\nu_2}{2},\frac{\nu_1}{2}\right)$, $I_\lambda\left(\frac{\nu_2+2}{2},\frac{\nu_1}{2}\right)$, and $I_\lambda\left(\frac{\nu_2+4}{2},\frac{\nu_1}{2}\right)$.

Since $\rho\geqq\rho_0$ and λ is strictly increasing in ρ, it follows that $\lambda\geqq(p-1)/p$ with equality if and only if $\rho=\rho_0$. Furthermore, when $\rho>\rho_0$, then $\lambda>(p-1)/p$ and

$$\lim_{N\to\infty}I_\lambda\left(\frac{\nu_2+d}{2},\frac{\nu_1}{2}\right)=1\,,$$

for every finite number $d\geqq 0$. A more refined result (see Appendix) can actually be obtained when $\rho>\rho_0$, namely

$$1-I_\lambda\left(\frac{\nu_2+d}{2},\frac{\nu_1}{2}\right)\leqq c_d\sqrt{N}e^{-\tau N}(1+O(N^{-1}))\,,\tag{2.20}$$

where, c_d is a constant varying with d, and $\tau>0$. It follows that

$$E[h(r)]=1+(p-1)\rho-p\rho\nu_2(\nu_1+\nu_2)^{-1}+O(\sqrt{N}e^{-\tau N})\,.\tag{2.21}$$

Hence

$$E\hat{\sigma}^2=\sigma^2\left(1+\frac{(Np-\nu_1-\nu_2)+\rho[(p-1)\nu_1-\nu_2]}{Np}\right)+O(Ne^{-\tau N}) . \tag{2.22}$$

Turning now to the expected value of $\hat{\rho}$, note that from (2.17), $\hat{\rho}\geqq\tilde{\rho}$ where

$$\tilde{\rho}=\frac{(p-1)\gamma(1-r)-r}{(p-1)[r+\gamma(1-r)]} \tag{2.23}$$

is the unrestricted MLE of ρ. Thus

$$E\tilde{\rho}\leqq E\hat{\rho}=E\tilde{\rho}+E(\hat{\rho}-\tilde{\rho})\leqq E\tilde{\rho}+\left(\rho_0+\frac{1}{p-1}\right)P\{\tilde{\rho}\leqq\rho_0\} , \tag{2.24}$$

the furthest right-hand inequality following since $\hat{\rho}\neq\tilde{\rho}$ only when $\tilde{\rho}<\rho_0$ $\left(\text{in which case } \hat{\rho}-\tilde{\rho}=\rho_0-\tilde{\rho}\leqq\rho_0+\frac{1}{p-1}\right)$. Now $\tilde{\rho}<\rho_0$ if and only if $\lambda<r\leqq1$. Consequently (2.20) and (2.24) imply that

$$E\tilde{\rho}\leqq E\hat{\rho}\leqq E\tilde{\rho}+\left(\rho_0+\frac{1}{p-1}\right)O(\sqrt{N}e^{-\tau N}) .$$

The result $E\tilde{\rho}=\rho+O(N^{-1})$ can be directly computed from the representation (2.23) by use of the delta method. Therefore,

$$E\hat{\rho}=\rho+O(N^{-1}) .$$

2.5) *Asymptotic distribution of the MLE.* For $\rho>\rho_0$, $\operatorname{plim}_{N\to\infty}(\hat{\rho}-\tilde{\rho})=0$. Indeed, the rate of convergence is very rapid as can be seen from (2.20). The limiting distribution of $\tilde{\rho}$ is known to be (c.f., Olkin and Pratt [5]*)

$$\lim_{N\to\infty}\mathfrak{L}(\sqrt{N}(\tilde{\rho}-\rho))=N(0, v_\infty) ,$$

where

$$v_\infty=\frac{2(1-\rho)^2(1+(p-1)\rho)^2}{p(p-1)} .$$

Hence, for $\rho>\rho_0$,

$$\lim_{N\to\infty}\mathfrak{L}(\sqrt{N}(\hat{\rho}-\rho))=N(0, v_\infty) . \tag{2.25}$$

For $\rho=\rho_0$, we still obtain asymptotic normality, but now the distribution is censored, i.e.,

* This paper contains a misprint in the omission of the factor 2 in v_∞ below.

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}(\hat{\rho}-\rho)) = N^{+}(0, v_\infty) , \tag{2.26}$$

where $N^{+}(0, v)$ represents the normal distribution censored at 0.

When $\rho>\rho_0$, $P\{\lambda\leqq r\leqq 1\}$ rapidly converges to 0 as $N\to\infty$. Thus

$$\lim_{N\to\infty} P\left\{\hat{\sigma}^2 \neq \frac{u+w}{Np}\right\} = \lim_{N\to\infty} P\{\lambda\leqq r\leqq 1\} = 0 .$$

It follows that

$$\begin{aligned}\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}(\hat{\sigma}^2-\sigma^2)) &= \lim_{N\to\infty} \mathfrak{L}\left(\sqrt{N}\left(\frac{u+w}{Np}-\sigma^2\right)\right) \\ &= N\left(0, \frac{2\sigma^4}{p}[1+(p-1)\rho^2]\right) .\end{aligned} \tag{2.27}$$

For $\rho=\rho_0$, the representation (2.18) of $\hat{\sigma}^2$ in terms of s^2 and r becomes

$$\hat{\sigma}^2 = \frac{\sigma^2 s^2}{Np} \max[1, 1+(p-1)\rho_0 - p\rho_0 r] \equiv R(s)g(r) .$$

Now

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}(R(s)-\sigma^2)) = N(0, 2p^{-1}\sigma^4) ,$$

while

$$\lim_{N\to\infty} \mathfrak{L}(\sqrt{N}(g(r)-1)) = \mathfrak{L}(\max(0, z)) ,$$

where $z\sim N(0, p^{-1}(p-1)\rho_0^2)$. Using the delta method, we conclude that the asymptotic distribution of $\sqrt{N}(\hat{\sigma}^2-\sigma^2)$ as $N\to\infty$ is the same as the distribution of $z^*+\sigma^2\max(0, z)$, where z^* and z are independently distributed, $z\sim N(0, p^{-1}(p-1)\rho_0^2)$, and $z^*\sim N(0, 2p^{-1}\sigma^4)$.

3. Tests of hypotheses for H^r_{mvc} and H^r_{vc}

3.1) *Derivation of the likelihood ratio tests.* We are interested in testing $H^r_{mvc}: (\mu, \Sigma)\in\Omega^r_{mvc}$ and $H^r_{vc}: (\mu, \Sigma)\in\Omega^r_{vc}$ against general alternatives $H: (\mu, \Sigma)\in\Omega$, where $\Omega=\{(\mu, \Sigma): \mu \text{ unrestricted}, \Sigma>0\}$. To obtain likelihood ratio tests for these hypotheses, we first obtain the supremum of the likelihood function (2.5) over each of the regions Ω^r_{mvc}, Ω^r_{vc} and Ω. From (2.7), (2.12) and (2.13), we find that

$$\sup_{\Omega^r_{mvc}} p(y, V) = \left[N^p e^{-p/2} T\left(v_{11}, \sum_{i=2}^{p}(v_{ii}+y_i^2)\right)\right]^{N/2} \tag{3.1}$$

$$\sup_{\Omega^r_{vc}} p(y, V) = \left[N^p e^{-p/2} T\left(v_{11}, \sum_{i=2}^{p} v_{ii}\right)\right]^{N/2} \tag{3.2}$$

where

$$T(a,b)=\begin{cases} a^{-1}\left(\dfrac{b}{p-1}\right)^{-(p-1)}, & \text{if } \dfrac{a}{b}>\dfrac{\gamma_0}{p-1}, \\ \left(\dfrac{a+b\gamma_0}{p}\right)^{-p}\gamma_0^{p-1}, & \text{if } \dfrac{a}{b}\leqq\dfrac{\gamma_0}{p-1}. \end{cases}$$

Finally, it is well known that

$$\sup_{\Omega} p(y, V)=\frac{N^{pN/2}\exp\left\{-\frac{1}{2}pN\right\}}{|V|^{N/2}}. \tag{3.3}$$

We find the likelihood ratio test statistic λ_1 for testing H_{mvc}^{γ} against general alternatives by forming the ratio of (3.1) with (3.3). The result can be expressed in the following form:

$$\lambda_1=\lambda_{mvc}\lambda_{0|mvc}, \tag{3.4}$$

where λ_{mvc} is Wilks' [7] likelihood ratio test statistic for testing H_{mvc} against general alternatives, that is,

$$\lambda_{mvc}^{2/N}=\frac{|V|}{v_{11}\left[\dfrac{\sum_{i=2}^{p}(v_{ii}+y_i^2)}{p-1}\right]^{p-1}}, \tag{3.5}$$

and where $\lambda_{0|mvc}$ is the likelihood ratio test of $\rho>\rho_0$ versus $\rho=\rho_0$ given that H_{mvc} is true, i.e.,

$$\lambda_{0|mvc}^{2/N}=\begin{cases} 1, & \text{if } \dfrac{v_{11}}{\sum_{i=2}^{p}(v_{ii}+y_i^2)}>\dfrac{\gamma_0}{p-1}, \\ \dfrac{v_{11}\left[\gamma_0\sum_{i=2}^{p}(v_{ii}+y_i^2)\right]^{p-1}p^p}{\left[v_{11}+\gamma_0\sum_{i=2}^{p}(v_{ii}+y_i^2)\right]^{p}(p-1)^{p-1}}, & \text{if } \dfrac{v_{11}}{\sum_{i=2}^{p}(v_{ii}+y_i^2)}\leqq\dfrac{\gamma_0}{p-1}. \end{cases} \tag{3.6}$$

Similarly, we can obtain the likelihood ratio test statistics λ_2 for testing H_{vc}^{γ} against general alternatives by dividing (3.2) by (3.3). We find that

$$\lambda_2=\lambda_{vc}\lambda_{0|vc}, \tag{3.7}$$

where λ_{vc} is Wilks' [7] likelihood ratio test statistic for testing H_{vc} against general alternatives, that is,

$$\lambda_{vc}^{2/N}=\frac{|V|}{v_{11}\left(\frac{\sum_{i=2}^{p} v_{ii}}{p-1}\right)^{p-1}}, \tag{3.8}$$

and where $\lambda_{0|vc}$ is the likelihood ratio test statistic for testing $\rho>\rho_0$ against $\rho=\rho_0$ given that H_{vc} is true, namely,

$$\lambda_{0|vc}^{2/N}=\begin{cases} 1, & \text{if } \frac{v_{11}}{\sum_{i=2}^{p} v_{ii}}>\frac{\gamma_0}{p-1}, \\ \frac{v_{11}\left[\gamma_0 \sum_{i=2}^{p} v_{ii}\right]^{p-1} p^p}{\left[v_{11}+\gamma_0 \sum_{i=2}^{p} v_{ii}\right]^{p}(p-1)^{p-1}}, & \text{if } \frac{v_{11}}{\sum_{i=2}^{p} v_{ii}}\leqq\frac{\gamma_0}{p-1}. \end{cases} \tag{3.9}$$

In terms of $\bar{x}$ and S, the likelihood ratio statistics can be computed by first substituting the equivalences

$$v_{11}=p^{-1}\sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}, \qquad \sum_{i=2}^{p} v_{ii}=\sum_{i=1}^{p} s_{ii}-v_{11},$$

$$|V|=|S|, \qquad \sum_{i=2}^{p} y_i^2=N\sum_{i=1}^{p}(\bar{x}_i)^2-Np^{-1}\left(\sum_{i=1}^{p}\bar{x}_i\right)^2,$$

into the formulas (3.5), (3.6), (3.8) and (3.9), and then using those results in formulas (3.4) and (3.7).

Remark. The likelihood ratio test statistic λ_3 of H_{mvc}^r versus H_{vc}^r may be of interest. This statistic can be obtained as the ratio of λ_1 to λ_2; that is, $\lambda_3=\lambda_1/\lambda_2$. From (3.4) and (3.7), we see that $\lambda_3=\lambda_m\lambda_{0|mvc}\lambda_{0|vc}^{-1}$, where λ_m is Wilks' [7] likelihood ratio test statistic for testing H_{mvc} when H_{vc} is known to be true. The ratio $\lambda_{0|mvc}\lambda_{0|vc}^{-1}$ is a complicated function of the observations, and the null distribution of λ_3 is quite involved even in the asymptotic case.

3.2) *Asymptotic null distributions of the likelihood ratio test statistics.* Since under H_{mvc} the triple $\left(y_1, v_{11}, \sum_{i=2}^{p}(v_{ii}+y_i^2)\right)$ constitutes a complete and sufficient statistic for the parameter $(\xi_1, \sigma_1^2, \sigma_2^2)$ (Herbach [4]), and since the distribution of λ_{mvc} is independent of $(\xi_1, \sigma_1^2, \sigma_2^2)$ while the distribution of $\lambda_{0|mvc}$ depends upon these parameters, the statistics λ_{mvc} and $\lambda_{0|mvc}$ are independent when H_{mvc} is true (Basu [1]). The distribution of λ_{mvc} under H_{mvc} is known (viz., Tukey and Wilks [6]) to be a product of independent Beta variables. Wilks [7] has shown that (under H_{mvc}),

$$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_{mvc})=\chi^2_{(p+3)(p-2)/2}. \tag{3.10}$$

The distribution of $\lambda_{0|mvc}$ is more complicated; $\lambda_{0|mvc}$ is equivalent to a one sided $F_{n,N(p-1)}$ test based on

$$\frac{v_{11}}{n}\Bigg/\frac{\sum_{i=2}^{p}(v_{ii}+y_i^2)}{N(p-1)}\,.$$

When $\rho>\rho_0$,

$$\lim_{N\to\infty} P\left\{\frac{v_{11}}{\sum_{i=2}^{p}(v_{ii}+y_i^2)}>\frac{\gamma_0}{p-1}\right\}=1\,,$$

so that $\mathfrak{L}(-2\log\lambda_{0|mvc})$ approaches to a distribution degenerate at 0. Thus, for $\rho>\rho_0$,

(3.11) $$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_1)=\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_{mvc})=\chi^2_{(p+3)(p-2)/2}\,.$$

When $\rho=\rho_0$,

$$\lim_{N\to\infty} P\left\{\frac{v_{11}}{\sum_{i=2}^{p}(v_{ii}+y_i^2)}>\frac{\gamma_0}{p-1}\right\}=\frac{1}{2}\,,$$

and it can be shown that

$$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_{0|mvc})\to\begin{cases}0\,, & \text{with probability } 1/2\,,\\ \chi_1^2\,, & \text{with probability } 1/2\,.\end{cases}$$

Therefore, when $\rho=\rho_0$,

(3.12) $$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_1)\to\begin{cases}\chi^2_{(p+3)(p-2)/2}\,, & \text{with probability } 1/2\,,\\ \chi^2_{(p+3)(p-2)/2+1}\,, & \text{with probability } 1/2\,.\end{cases}$$

Since the limiting law (3.12) is stochastically larger than the limiting law (3.11), it is recommended that the rejection region is chosen with reference to the limiting law (3.12). Indeed, under H_{mvc} the distribution of λ_1 has a monotone likelihood ratio in ρ, so that even in finite samples the rejection region should be selected with reference to the distribution of λ_1 when H_{mvc} holds and $\rho=\rho_0$.

When $\rho<\rho_0$, and H_{mvc} holds,

$$\lim_{N\to\infty} P\left\{\frac{v_{11}}{\sum_{i=2}^{p}(v_{ii}+y_i^2)}>\frac{\gamma_0}{p-1}\right\}=0\,,$$

so that

$$\operatorname*{plim}_{N\to\infty}(-2\log\lambda_{0|mvc})=\operatorname*{plim}_{N\to\infty}N\log\left\{\frac{F\left(\dfrac{\gamma_0}{p-1}\right)^{p-1}p^p}{(F+\gamma_0)^p}\right\},$$

where $F=v_{11}\Big/\sum_{i=2}^{p}(v_{ii}+y_i^2)$. But $\operatorname*{plim}_{N\to\infty}F=\gamma/p-1$, and consequently,

$$\operatorname*{plim}_{N\to\infty}(-2\log\lambda_{0|mvc})=\infty\ .$$

From this result and (3.10),

$$\operatorname*{plim}_{N\to\infty}(-2\log\lambda_1)=\infty\ .$$

Similar results hold for λ_2. Under H_{vc}, $-2\log\lambda_2$ is the convolution of the independent test statistics $-2\log\lambda_{vc}$ and $-2\log\lambda_{0|vc}$. The distribution of λ_{vc} under H_{vc} can be expressed as the product of independent Beta variables, while $\lambda_{0|vc}$ is equivalent to the one-sided $F_{n,n(p-1)}$-test based on $v_{11}\Big/\sum_{i=2}^{p}v_{ii}$ and has a monotone likelihood ratio in ρ. Consequently, the rejection region of λ_2 should be derived from the distribution for λ_2 when H_{vc} holds and $\rho=\rho_0$. The asymptotic distribution of λ_2 when H_{vc}^r holds and $\rho>\rho_0$ is

$$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_2)=\chi^2_{p(p-1)/2-2}\ ,$$

whereas the limiting distribution of λ_2 when H_{vc}^r holds and $\rho=\rho_0$ is

$$\lim_{N\to\infty}\mathfrak{L}(-2\log\lambda_2)=\begin{cases}\chi^2_{p(p-1)/2-2}\ , & \text{with probability } 1/2\ ,\\ \chi^2_{p(p-1)/2-1}\ , & \text{with probability } 1/2\ .\end{cases}$$

When H_{vc} holds and $\rho<\rho_0$,

$$\operatorname*{plim}_{N\to\infty}(-2\log\lambda_2)=\infty\ .$$

Appendix

4. An inequality for the incomplete Beta function

In this appendix we prove the following inequality.

THEOREM. *If* $\alpha_1=z$, $\alpha_2=az+b$, $a>0$, *and if* $\beta<1/(a+1)$, *then as* $z\to\infty$,

$$\text{(i)}\qquad I_\beta(\alpha_1,\alpha_2)=\frac{\int_0^\beta y^{\alpha_1-1}(1-y)^{\alpha_2-1}dy}{B(\alpha_1,\alpha_2)}\leq C\sqrt{z}e^{-\tau z}(1+0(z^{-1}))\ ,$$

for some $\tau > 0$.

PROOF. The maximum of $y^{\alpha_1-1}(1-y)^{a(\alpha_1-1)}$ is achieved for $y=1/(a+1)$. Further, this function is increasing for $y \leqq 1/(a+1)$, so that for $0 \leqq y \leqq \beta < 1/(a+1)$,

$$y^{\alpha_1-1}(1-y)^{a(\alpha_1-1)} \leqq \beta^{\alpha_1-1}(1-\beta)^{a(\alpha_1-1)} .$$

Consequently,

$$\text{(ii)} \qquad \int_0^\beta y^{\alpha_1-1}(1-y)^{\alpha_2-1}dy \leqq \left[\int_0^\beta (1-y)^{b+a-1}dy\right]\beta^{\alpha_1-1}(1-\beta)^{a(\alpha_1-1)} .$$

Furthermore, from the well-known Stirling Expansion of the Gamma function,

$$\text{(iii)} \qquad B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} = \sqrt{2\pi}\,\frac{e^{-(\alpha_1+\alpha_2)}\alpha_1^{\alpha_1-1/2}\alpha_2^{\alpha_2-1/2}}{e^{-(\alpha_1+\alpha_2)}(\alpha_1+\alpha_2)^{\alpha_1+\alpha_2-1/2}}[1+O(z^{-1})]$$

$$= \sqrt{2\pi}(\alpha_1+\alpha_2)^{-1/2}\left(\frac{\alpha_1}{\alpha_1+\alpha_2}\right)^{\alpha_1-1/2}\left(\frac{\alpha_2}{\alpha_1+\alpha_2}\right)^{\alpha_2-1/2}[1+O(z^{-1})]$$

$$= \sqrt{2\pi}\,e^{-b}\left(\frac{a}{a+1}\right)^{-1/2} z^{-1/2}\left(\frac{1}{a+1}\right)^{z}\left(\frac{a}{a+1}\right)^{az}[1+O(z^{-1})] .$$

Taking the ratio of (ii) to (iii), we obtain

$$I_\beta(\alpha_1, \alpha_2) \leqq \sqrt{2\pi}\,e^{-b}\left(\frac{a}{a+1}\right)^{-1/2}\left[\frac{1-(1-\beta)^{b+a}}{b+a}\right]\beta^{-1}(1-\beta)^{-a}z^{1/2}$$

$$\cdot\{[(a+1)\beta][a^{-1}(a+1)(1-\beta)]^a\}^z[1+O(z^{-1})] .$$

But since $\beta < 1/(a+1)$, the arithmetic-geometric mean inequality tells us that

$$[(a+1)\beta]^{1/(a+1)}[a^{-1}(a+1)(1-\beta)]^{a/(a+1)} < 1 .$$

Thus we can write

$$[(a+1)\beta][a^{-1}(a+1)(1-\beta)]^a = e^{-\tau} ,$$

for some $\tau > 0$. This completes the proof.

THE JOHNS HOPKINS UNIVERSITY
STANFORD UNIVERSITY

REFERENCES

[1] D. Basu, "On statistics independent of a complete sufficient statistic," *Sankhyā*, 18 (1955), 377-380.

[2] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, 36 (1949), 317-346.

[3] G. E. P. Box, "Problems in the analysis of growth and wear curves," *Biometrics*, 6 (1950), 362-389.

[4] L. H. Herbach, "Properties of model II-type analysis of variance tests, A: Optimum nature of the F-test for model II in the balanced case," *Ann. Math. Statist.*, 30 (1959), 939-959.

[5] I. Olkin and J. W. Pratt, "Unbiased estimation of certain correlation coefficients," *Ann. Math. Statist.*, 29 (1958), 201-211.

[6] J. W. Tukey and S. S. Wilks, "Approximation of the distribution of the product of beta variables by a single beta variable," *Ann. Math. Statist.*, 17 (1946), 318-324.

[7] S. S. Wilks, "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution," *Ann. Math. Statist.*, 17 (1946), 257-281.