The Distribution of the Ratios of Characteristic Roots (Condition Numbers) and Their Applications in Principal Component or Ridge Regression*

C. G. Troskie*
Department of Statistics
University of Cape Town
Rondebosch 7700, South Africa

and

W. J. Conradie
University of Stellenbosch
Stellenbosch 7600, South Africa

Submitted by 18 June 1985; revised 12 August 1985

ABSTRACT

In regression analysis large condition numbers indicate the presence of multicollinearity. Principal component regression and ridge regression are used to correct for the ill effects of such collinearities. In this paper some distributional properties of the condition number are considered.

1. INTRODUCTION

Consider a random $(p+1)$ vector $z' = (y, x_1, \dots, x_p)'$ having a multivariate normal distribution with mean vector

$$\mu' = (\mu_y, \mu'_x) = (\mu_y, \mu_1, \dots, \mu_p)$$

*Part of this research was done while the author was a Visiting Professor at the University of Stanford. The CSIR is thanked for financial assistance.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/86/$3.50

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}.$$

The regression function is the conditional expectation of y given fixed values of $(x_1, \ldots, x_p)$:

$$E(y|x) = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x) = \beta_0 + \beta'_{(2)}x, \tag{1.1}$$

where $\beta_{(2)} = \Sigma_{xx}^{-1}\Sigma_{xy}$. The conditional variance is

$$V(y|x) = \sigma^2 = \sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}. \tag{1.2}$$

For a random sample of size n,

$$\begin{pmatrix} Z'_1 \\ \vdots \\ Z'_n \end{pmatrix} = \begin{pmatrix} Y_1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & & \vdots \\ Y_{n1} & X_{n1} & \cdots & X_{np} \end{pmatrix} = (Y, X), \tag{1.3}$$

the maximum likelihood estimates of $\beta' = (\beta'_0 \;\; \beta'_{(2)})$ and σ^2 are given by

$$\hat{\beta} = \begin{pmatrix} \bar{Y} - \hat{\beta}'_{(2)}\bar{X} \\ A_{xx}^{-1}A_{xy} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_{(2)} \end{pmatrix}, \qquad \bar{Z} = \begin{pmatrix} \bar{Y} \\ \bar{X} \end{pmatrix} = \frac{1}{n}\sum Z_\alpha, \tag{1.4}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left(a_{yy} - A_{yx}A_{xx}^{-1}A_{xy}\right), \tag{1.5}$$

where

$$A = \sum (Z_\alpha - \bar{Z})(Z_\alpha - \bar{Z})' = \begin{pmatrix} a_{yy} & A_{yx} \\ A_{xy} & A_{xx} \end{pmatrix},$$

$$A_{xx} = \sum (X_\alpha - \bar{X})(X_\alpha - \bar{X})', \tag{1.6}$$

If the variables $(x_1, \ldots, x_p)$ are considered to be fixed, it is often convenient to write the above model as

$$Y = X\beta + e, \tag{1.7}$$

where $E(e) = 0$, $E(ee') = \sigma^2 I$, and e is distributed $N(0, \sigma^2 I)$. Here X is an $n \times (p+1)$ matrix with first column consisting of the unit vector $1' = (1, \ldots, 1)$.

The model (1.7) is the standard linear model when the variables $(x_1, \ldots, x_p)$ are either considered to be fixed or known independent variables. The estimates of β and σ^2 are usually given by

$$\hat{\beta} = (X'X)^{-1}X'Y, \qquad s^2 = \frac{1}{n-p-1}(Y - X\hat{\beta})'(Y - X\hat{\beta}). \tag{1.8}$$

In what follows a brief review of some of the distributional results will be given. The purpose is mainly to highlight the differences between the *fixed* model given by (1.7) and the *random* model given by (1.1), where $Z \sim N(\mu, \Sigma)$. For detailed discussions of the differences between these two models see [29] or [27].

For the fixed case,

$$\hat{\beta} \sim N\big(\beta, \sigma^2(X'X)^{-1}\big), \qquad \frac{(n-p-1)s^2}{\sigma^2} \sim \chi^2_{n-p-1} \tag{1.9}$$

are independently distributed.

For the random case, if we condition on the variables $(X_1, \ldots, X_p)$, then the distribution of $\hat{\beta}$ and $(n-p-1)s^2/\sigma^2$ will, of course, again be given by (1.9). Although the unconditional distribution of $(n-p-1)s^2/\sigma^2$ remains unchanged for the random case, the unconditional distribution of $\hat{\beta}$ as given by (1.9) or (1.4) is entirely different and very complicated. The density is given by Kabe [17] and is not repeated here. The marginal density of $\hat{\beta}^{(2)} = A_{xx}^{-1}A_{xy}$ is in fact a multivariate t-density.

For most practical purposes the model (1.7) is sufficient for both the fixed or conditional cases. The difference arises in the power functions. For example, suppose we test at level α the hypothesis $H_0 : \beta_{(2)} = 0$ against $H_1 : \beta_{(2)} \neq 0$. Then the F-statistic for both cases would be

$$F = \frac{\hat{\beta}'_{(2)}C_{22}^{-1}\hat{\beta}_{(2)}}{Y'Y - \hat{\beta}'X'Y} \cdot \frac{n-p-1}{p}, \tag{1.10}$$

which has an F-distribution under H with p and $n-p-1$ degrees of freedom.

For the fixed case under H_1, F has a noncentral F-distribution with noncentrality parameter

$$\lambda = \frac{1}{2\sigma^2}\beta'_{(2)}C_{22}^{-1}\beta_{(2)},$$

where

$$(X'X)^{-1} = C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

The density of

$$u = \frac{p}{n-p-1} F$$

is then given by

$$cu^{\frac{1}{2}p-1}(1+u)^{-\frac{1}{2}(n-p-1+p)} e^{-\lambda} {}_1F_1\left(\left(\tfrac{1}{2}n - p - 1 + p\right); \tfrac{1}{2}p\lambda \frac{u}{1+u}\right). \quad (1.11)$$

For the random case under H_1 the density of F given by (1.10) is no longer that of a noncentral F statistic. The density of $u = [p/(n-p-1)]F$ is now given by

$$cu^{\frac{1}{2}p-1}(1+u)^{-\frac{1}{2}(n-p-1+p)}(1-\rho^2)^{\frac{1}{2}n}$$
$$\times {}_2F_1\left(\tfrac{1}{2}(n-1), \tfrac{1}{2}(n-1); \tfrac{1}{2}p; \rho^2 \frac{u}{1+u}\right), \quad (1.12)$$

where $\rho^2 = \beta'_{(2)} \Sigma_{xx} \beta_{(2)} / \sigma_{yy}$ is the population multiple correlation coefficient between y and $(x_1, \ldots, x_p)$. For both cases

$$c = \frac{\Gamma\left(\frac{1}{2}(n-p-1+p)\right)}{\Gamma\left(\frac{1}{2}p\right)\Gamma\left(\frac{1}{2}(n-p-1)\right)}.$$

Whether the x-variables are fixed or random, it often happens that there are linear dependencies between the x-variables, causing multicollinearities to exist in the matrix X. In the following we assume that X is centered and standardized. If X is centered and random, then $X'X = A$ [compare with (1.3)] has a Wishart distribution. If $X'X$ is standardized, then $X'X = R$ is the correlation matrix between the x-variables, whereas $X'Y$ is the correlation matrix between y and $(x_1, \ldots, x_p)$, assuming that Y is also standardized.

The effects of multicollinearities have been discussed by several authors in the literature, notably Hoerl and Kennard [12], Marquardt [21], and Webster, Gunst, and Mason [32, 9].

The effects could best be seen by examining the latent roots and vectors of $X'X$.

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$ be the latent roots of $X'X$ (correlation matrix), and $V_1, \ldots, V_p$ the corresponding vectors. If there are near-multicollinearities, some of the roots will be small. For convenience assume only one near-multicollinearity, i.e., assume that λ_p is close to zero.

Several authors have suggested a correction to the least squares estimates if some or one of these roots are "too small." In the following we give a brief outline of some of the more important procedures.

The ridge procedure proposed by Hoerl and Kennard [12] is to add a small constant k to the correlation matrix $X'X$. Then

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y, \qquad 0 < k < 1.$$

The method of principal components proposed by Marquardt [21] computes a "generalized inverse" for $X'X$ by considering only the so-called large characteristic roots and associated characteristic vectors of $X'X$.

A method called latent root regression (LRRA) was also proposed by Webster, Gunst, and Mason [32] and independently by Hawkins [11]. They argue that the dependent variable y may be involved in the multicollinearity. For example, if constants $a_0, \ldots, a_p$ exist such that $a_0 y + a_1 x_1 + \cdots + a_p x_p \simeq 0$ and if $a_0 \neq 0$, then there is a perfect predictor for y. On the other hand, if $a_0 \simeq 0$, then the multicollinearity exists only among the x's and should be eliminated. Their procedure is first to calculate the characteristic roots $(\lambda_0, \lambda_1, \ldots, \lambda_p)$ and vectors $(V_0, V_1, \ldots, V_p)$ of the correlation matrix

$$\begin{pmatrix} Y'Y & Y'X \\ X'Y & X'X \end{pmatrix} = R_{yx}. \tag{1.13}$$

Then if λ_p is "too small" and $|V_{p0}|$ is "too small" (where $|V_{p0}|$ is the first element of the vector V_p), then the last root and vector are eliminated to get the LRRA estimate of β.

Several authors (e.g. Forsythe and Moler [7], Marshall and Olkin [22], Forsythe, Malcolm, and Moler [6], and Vinod [31] proposed the computation of the "condition number" to measure the instability of a matrix when solving for a system of linear equations. Since $X'X$ is symmetric, the condition number of $X'X$ is λ_1/λ_p, where $\lambda_1 > \lambda_2 \cdots > \lambda_p$ are the characteristic roots of $X'X$. The condition number is a better measure of the nearness to singularity than the determinant of a matrix A. For example, if A is a 100×100 matrix with 0.1 on the diagonal, then $|A| = 10^{-100}$, which is usually regarded as a small number. But the condition number of A is

$\lambda_1/\lambda_p = 0.1/0.1 = 1$. For systems of the type $Ax = b$ an A as above behaves more like the identity matrix than like a singular matrix.

Recently Belsley, Kuh, and Welsch [2] stated:

Most of the experimental evidence shows that weak dependencies begin to exhibit themselves with "condition indices" around 10. A number in the neighbourhood of 15–30 tends to result from an underlying near dependency with an associated correlation of 0.9. Condition indices of 100 or more appear to be large indeed causing substantial variance inflation and great potential harm to regression estimates.

(Note that the condition index as defined by Belsley, Kuh, and Welsch is $\sqrt{\lambda_1/\lambda_p}$.)

If the x-variables are random variables, then the roots $(\lambda_1, \ldots, \lambda_p)$ are also random variables, and then one will be interested in the distribution of the condition number λ_1/λ_p. Another measure which can be considered as a measure of the condition of a matrix is $\Sigma\lambda_i/\lambda_p = \text{tr}(X'X)/p$.

The purpose of this paper is to focus attention on the distributions of the condition numbers λ_1/λ_p and $\Sigma\lambda_i/p$. The distributional results will obviously depend on the underlying distributional assumptions on y and $(x_1, \ldots, x_p)$. This is again briefly discussed in the next section to highlight the effect these distributional assumptions will have on the distributions of the condition numbers. In Section 3 the exact distributions of the condition numbers are considered. The exact distribution of λ_1/λ_p is derived, but that of $\Sigma\lambda_i/\lambda_p$ appears to be intractable. These results are very complicated, and some asymptotic results are considered in Section 5. The exact distribution of λ_1/λ_p for a circular population covariance matrix is given in Section 6. Some practical examples of the condition numbers and the effects of large condition numbers on the estimates of β are given in Section 4.

2. THE EFFECT OF THE UNDERLYING DISTRIBUTIONS ON THE DISTRIBUTIONS OF THE CONDITION NUMBERS

The distributions of the condition numbers depend on whether we deal with the fixed or the random case.

First consider the fixed case, where $(x_1, \ldots, x_p)$ are known (fixed) variables. Thus $R = X'X$ is a known fixed correlation matrix, and the roots $(\lambda_1, \ldots, \lambda_p)$ are known fixed quantities and their distributional results are no longer of interest. For this case it is probably sufficient to determine a cutoff value for λ_1/λ_p and to correct the least squares estimators if λ_1/λ_p is larger than the cutoff value. A cutoff value of 25 to 100 was suggested by Belsely, Kuh, and Welsch.

With the LRRA estimates the roots $\lambda_0,\ldots,\lambda_p$ of the correlation matrix R_{yx} [see (1.13)] between y and $(x_1,\ldots,x_p)$ is of interest. Although the x's are fixed, y is random and R_{yx} is a random matrix, so that $\lambda_0,\ldots,\lambda_p$ are random variables. Thus we need the conditional distribution of the roots $(\lambda_0,\ldots,\lambda_p)$ given the x-variables. This appears to be a very difficult derivation.

For the random case when y and $(x_1,\ldots,x_p)$ are distributed $N(\mu,\Sigma)$, both R and R_{yx} are random correlation matrices with the same distributions except that R has dimension p and R_{yx} dimension $p+1$. The distributions of these matrices will depend on the assumptions made on Σ.

Now

$$\Sigma=\begin{pmatrix}\Sigma_{yy} & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_{xx}\end{pmatrix},$$

and if $\Sigma_{yx}=0$, then $\beta_{(2)}=\Sigma_{xx}^{-1}\Sigma_{xy}=0$, and no linear relationship exists among y and $(x_1,\ldots,x_p)$, and we shall not be interested in estimating $\beta_{(2)}$. Thus we need the distribution of $\lambda_0,\ldots,\lambda_p$ for general Σ.

The distribution of the roots $(\lambda_1,\ldots,\lambda_p)$ of R depends on Σ_{xx}. If $\Sigma_{xx}=\sigma^2 I_p$, then all the x's are orthogonal and no collinearities exist. This is not the case of interest in this paper, so we again need distributional results for general Σ_{xx}.

Thus for both R and R_{yx} we need distributional results for general Σ.

To avoid notational difficulties we assume that we have p variables $(x_1,\ldots,x_p)$, where y could be assumed to be one of the x-variables. Thus R is a correlation matrix of p dimensions, and we are interested in the distribution of λ_1/λ_p and $\Sigma\lambda_i/\lambda_p$ for general Σ.

The distribution of the correlation matrix for general Σ is not known. If $\Sigma=I$, i.e. all the variables are independent (and orthogonal), then the distribution of R is known but the joint distribution of the roots of R is not known. As mentioned before, this case is not of interest in this paper.

If $a_{ii}=\Sigma_{j=1}^n(x_{ij}-\bar{x}_i)^2$ and $D=\operatorname{diag}(a_{11},\ldots,a_{pp})$, then

$$A=D^{1/2}RD^{1/2}=[a_{ij}],\qquad i,j=1,\ldots,p,$$

is a Wishart matrix. If $(\lambda_1,\ldots,\lambda_p)$ are now the characteristic roots of A, then λ_1/λ_p and $\Sigma\lambda_i/\lambda_p$ will measure the condition of A. If some of the x-variables are collinear, then we expect large condition numbers, and corrections should then be made in the least squares estimates.

Thus we are interested in the condition numbers of A where A has a Wishart distribution. The distribution of λ_1/λ_p is derived in the next section. The distribution of $\Sigma\lambda_i/\lambda_p$ appears to be intractable.

3. THE DISTRIBUTION OF CONDITION NUMBERS

In this section we consider the distribution of the condition numbers. Many of the results will be very theoretical and extremely complicated. For the interested reader the applications of these results will be considered in a separate section.

Assume that A has a Wishart distribution, $W(\Sigma, n)$, and that the characteristic roots of A are $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$. The joint distribution of the roots $(\lambda_1, \ldots, \lambda_p)$ has been derived by James [14] and is given by

$$K(p,n)|\Sigma|^{-\frac{1}{2}n}|\Lambda|^{\frac{1}{2}(n-p-1)}\operatorname{etr}\left(-\tfrac{1}{2}\Lambda\right)\alpha_p(\Lambda)\,{}_0F_0\left(\tfrac{1}{2}\left(I_p-\Sigma^{-1}\right),\Lambda\right), \quad (3.1)$$

for $0 < \lambda_p \leqslant \cdots \leqslant \lambda_1 < \infty$, where

$$K(p,n) = \frac{\pi^{\frac{1}{2}p^2}}{2^{\frac{1}{2}pn}\Gamma_p\left(\frac{1}{2}n\right)\Gamma_p\left(\frac{1}{2}p\right)},$$
$$\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p), \qquad \alpha_p(\Lambda) = \prod_{i<j}(\lambda_i - \lambda_j), \quad (3.2)$$

and

$$\begin{aligned}&{}_pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; Q, T\right)\\ &= \sum_{k=0}^{\infty}\sum_{\kappa}\frac{\left[(a_1)_\kappa, \ldots, (a_p)_\kappa\right]}{\left[(b_1)_\kappa, \ldots, (b_q)_\kappa\right]}\cdot\frac{C_\kappa(Q)C_\kappa(T)}{C_\kappa(I)k!},\end{aligned} \quad (3.3)$$

where $a_1, \ldots, a_p, b_1, \ldots, b_q$ are real or complex constants and the multivariate coefficient $(a)_\kappa$ is given by $(a)_\kappa = \prod_{i=1}^{p}(a - \frac{1}{2}(i-1))_{k_i}$, where $(a)_k = a(a+1)\cdots(a+k-1)$. The partition of κ of k is such that

$$\kappa = (k_1, k_2, \ldots, k_p), \qquad k_1 \geqslant k_2 \geqslant \cdots \geqslant k_p \geqslant 0,$$

$k_1 + k_2 + \cdots + k_p = k$, and the zonal polynomials $C_\kappa(T)$ are expressible in terms of elementary symmetric functions of the latent roots of T [14].

Let

$$u_i = \frac{\lambda_i}{\sum_{j=1}^{p}\lambda_i}, \qquad i = 2, \ldots, p, \quad (3.4)$$

$$S = \sum_{i=1}^{p}\lambda_i. \quad (3.5)$$

The Jacobian of the transformation is easily seen to be S^{p-1}.

The joint density of $(u_2,\ldots,u_p)$ and S is seen to be

$$f(u_2,\ldots,u_p,S)=K(p,n)|\Sigma|^{-\frac{1}{2}n}|U|^{\frac{1}{2}(n-p-1)}\operatorname{etr}\left(-\tfrac{1}{2}S\right)$$
$$\times S^{\frac{1}{2}p(n-p-1)+\frac{1}{2}p(p-1)+p-1}\alpha_p(U)$$
$$\times\sum_{k=0}^{\infty}\sum_{\kappa}S^kC_k\left(\tfrac{1}{2}(I-\Sigma^{-1})\right)\frac{C_\kappa(U)}{C_\kappa(I)k!},\tag{3.6}$$

where

$$U=\operatorname{diag}\left(1-\sum_{i=2}^{p}u_i,u_2,\ldots,u_p\right).\tag{3.7}$$

Integration over S, $0<S<\infty$, yields the joint density of $(u_2,\ldots,u_p)$ as

$$\frac{\pi^{\frac{1}{2}p^2}}{\Gamma_p\left(\frac{1}{2}n\right)\Gamma_p\left(\frac{1}{2}p\right)}|\Sigma|^{-\frac{1}{2}n}|U|^{\frac{1}{2}(n-p-1)}\alpha_p(U)$$
$$\times\sum_{k=0}^{\infty}\sum_{\kappa}\frac{C_\kappa\left(I_p-\Sigma^{-1}\right)C_\kappa(U)\Gamma\left(\frac{1}{2}np+k\right)}{C_\kappa(I)k!}.\tag{3.8}$$

James [16] has considered the joint distribution of $(\lambda_i/\bar{\lambda})$, $i=2,\ldots,p$, where $\bar{\lambda}=(1/p)\Sigma\lambda_i$. Krishnaiah and Waikar [19] have derived the joint density of $(u_2,\ldots,u_p)$ in the above form.

Since $u_p=\lambda_p/\Sigma\lambda_i$, the marginal density of u_p can be found by integrating out $u_2,\ldots,u_{p-1}$ over the range $0\leqslant u_p\leqslant u_{p-1}\leqslant\cdots\leqslant u_2\leqslant u_1=1-\Sigma_{i=2}^{p}u_i$. To find an explicit expression for u_p does not seem feasible. The condition number $\Sigma\lambda_i/\lambda_p$ is of course given by

$$C_1=\frac{1}{u_p}=\sum\frac{\lambda_i}{\lambda_p}.\tag{3.9}$$

We now derive the density of the condition number $C_2=\lambda_1/\lambda_p$.

If the joint density of $(\lambda_1,\ldots,\lambda_p)$ is given by (3.1), make the transformation

$$l_i=\frac{\lambda_1-\lambda_i}{\lambda_1},\qquad i=2,\ldots,p,\tag{3.10}$$

so that

$$C_2 = \frac{\lambda_1}{\lambda_p} = \frac{1}{1 - l_p}. \tag{3.11}$$

The Jacobian is $J(\lambda_1, \lambda_2, \ldots, \lambda_p \to \lambda_1, l_2, \ldots, l_p) = \lambda_1^{p-1}$. If $\Lambda_l = \operatorname{diag}(l_2, \ldots, l_p)$ with $1 > l_p \geqslant \cdots \geqslant l_2 > 0$, then

$$\operatorname{etr}\left(-\frac{1}{2}\Lambda\right) = e^{-\frac{1}{2}\lambda_1 p}\operatorname{etr}\left(\tfrac{1}{2}\lambda_1\Lambda_l\right),$$

$$\prod \lambda_i^{\frac{1}{2}(n-p-1)} = \lambda_1^{\frac{1}{2}(n-p-1)p}|I - \Lambda_l|^{\frac{1}{2}(n-p-1)},$$

$$\alpha_p(\Lambda) = \prod_{i<j}(\lambda_i - \lambda_j) = \lambda_1^{\frac{1}{2}p(p-1)}|\Lambda_l| \prod_{1<i<j}^{p}(l_j - l_i),$$

and

$$C_\kappa(\Lambda) = C_\kappa(\lambda'_1(I - \Lambda_l)) = \lambda_1^k C_\kappa('(I - \Lambda_l)),$$

where

$$'(I - \Lambda_l) = \operatorname{diag}(1, 1 - l_2, \ldots, 1 - l_p).$$

We use the well-known expansion [3]

$$C_\kappa('(I - \Lambda_l)) = \sum_{t=0}^{k} \sum_{\tau} b_{\kappa,\tau} C_\tau(I - \Lambda_l),$$

where the $b_{\kappa,\tau}$'s are constants depending on κ and τ, and are tabulated in [18]. The joint density is then given by

$$f(\lambda_1, l_2, \ldots, l_p) = k|\Sigma|^{-\frac{1}{2}n}\lambda_1^{\frac{1}{2}np-1}e^{-\frac{1}{2}\lambda_1 p}\sum_k\sum_\kappa \frac{C_\kappa\left(\frac{1}{2}(I - \Sigma^{-1})\right)\lambda^k}{k!C_\kappa(I)}$$

$$\times \sum_{t=0}^{k}\sum_\tau b_{\kappa,\tau}C_\tau(I - A_l)\operatorname{etr}\left(\tfrac{1}{2}\lambda_1\Lambda_l\right)$$

$$\times |I - \Lambda_l|^{\frac{1}{2}(n-p-1)}|\Lambda_l| \prod_{1<i<j}(l_j - l_i). \tag{3.12}$$

But [15]

$$C_\tau(I-\Lambda_l)=\sum_{\tau'\preccurlyeq\tau}\alpha_{\tau'}C_{\tau'}(\Lambda_l),$$

where τ' is a partition of t into not more than p parts. The ordering $\preccurlyeq$ is lexicographic and is described in [15], and

$$\operatorname{etr}\left(\tfrac{1}{2}\lambda_1\Lambda_l\right)=\sum_m\sum_\mu\frac{\left(\frac{1}{2}\lambda_1\right)^m C_\mu(\Lambda_l)}{m!},$$

$$C_{\tau'}(\Lambda_l)C_\mu(\Lambda_l)=\sum_\delta g^{\delta}_{\tau',\mu}C_\delta(\Lambda_l) \qquad [4],$$

where τ', μ, and δ are partitions of t, m and $d=t+m$, respectively, into not more than p parts. The coefficients $g^{\delta}_{\tau',\mu}$ are tabulated in [18] for all partititions up to order 7. Also

$$|I-\Lambda_l|^{-\frac{1}{2}(p+1-n)}C_\delta(\Lambda_l)=\sum_s\sum_\sigma\frac{\left[\frac{1}{2}(p+1-n)\right]_\sigma C_\sigma(\Lambda_l)C_\delta(\Lambda_l)}{\frac{1}{2}s!}$$

$$=\sum_s\sum_\sigma\sum_\nu\frac{\left[\frac{1}{2}(p+1-n)\right]_\sigma g^{\nu}_{\sigma,\delta}C_\nu(\Lambda_l)}{s!}.$$

$$|I-\Lambda_l|^{-\frac{1}{2}(p+1-n)}\operatorname{etr}\left(\tfrac{1}{2}\lambda_1\Lambda_l\right)C_\tau(I-\Lambda_l)$$

$$=\sum_{\tau'\preccurlyeq\tau}\sum_m\sum_\mu\sum_\delta\sum_s\sum_\sigma\sum_\nu\left(\tfrac{1}{2}\lambda_1\right)^m\alpha_{\tau'}g^{\delta}_{\tau',\mu}g^{\nu}_{\sigma,\delta}\frac{\frac{1}{2}(p+1-n)_\sigma C_\nu(\Lambda_l)}{m!s!}$$

$$(3.13)$$

and

$$f(\lambda_1,l_2,\ldots,l_p)=K|\Sigma|^{-\frac{1}{2}n}\lambda_1^{\frac{1}{2}np-1}e^{-\frac{1}{2}\lambda_1 p}$$

$$\times\sum{}^{*}C_\kappa\left(\tfrac{1}{2}(I-\Sigma^{-1})\right)\lambda_1^{k+m}b_{\kappa,\tau}\alpha_{t'}\tfrac{1}{2}(p+1-n)_\sigma$$

$$\times g^{\delta}_{\tau',\mu}g^{\nu}_{\sigma,\delta}C_\nu(\Lambda_l)|\Lambda_l|\prod_{1<i<j}(l_j-l_i), \qquad (3.14)$$

where

$$\sum{}^{*}=\sum_{k}\sum_{\kappa}\sum_{t=0}^{k}\sum_{\tau}\sum_{\tau\preccurlyeq\tau'}\sum_{m}\sum_{\mu}\sum_{\delta}\sum_{s}\sum_{\sigma}\sum_{\nu}. \tag{3.15}$$

Now, let $r_i = l_i/l_p$, $i = 2,\ldots, p-1$, in (3.14) with Jacobian equal to l_p^{p-2}; then $|\Lambda_l| = l_p^{p-1}|\Lambda_r|$ and

$$C_\nu(\Lambda_l) = l_p^{s+t+m}C_\nu('\Lambda_r),$$

$$\prod_{1<i<j}(l_j - l_i) = l_p^{\frac{1}{2}(p-1)(p-2)}|I_{p-2} - \Lambda_r|\prod_{1<i<j}^{p-1}(r_j - r_i).$$

The joint density of $(\lambda_1, r_2,\ldots, r_{p-1}, l_p)$ is then

$$\begin{aligned}f(\lambda_1, r_2,\ldots, r_{p-1}, l_p) &= K|\Sigma|^{-\frac{1}{2}n}\lambda_1^{\frac{1}{2}np-1}e^{-\frac{1}{2}\lambda_1 p}\\ &\times\sum{}^{*}C_\kappa\left(\tfrac{1}{2}(I-\Sigma^{-1})\right)\lambda^{k+m}\left(\tfrac{1}{2}(p+1-n)\right)_\sigma\\ &\times b_{\kappa,\tau}\alpha_{t'}g^{\delta}_{\tau',\mu}g^{\nu}_{\sigma,\delta}l_p^{\frac{1}{2}(p-1)(p+2)+s+t+m-1}\\ &\times|\Lambda_r|C_\nu('\Lambda_r)|I_{p-2}-\Lambda_r|\prod_{1<i<j}^{p-1}(r_j-r_i).\end{aligned}$$

To find the marginal density of λ_1 and l_p we integrate over $(r_2,\ldots, r_{p-1})$, i.e. (see also [26])

$$\begin{aligned}&\int_{1>r_p>\cdots>r_2>0}|\Lambda_r|C_\nu('\Lambda_r)|I_{p-2}-\Lambda_r|\prod_{1<i<j}(r_j-r_i)\prod_{i=2}^{p-1}dr_i\\ &=\left[\tfrac{1}{2}(p-1)(p+2)+s+t+m\right]\\ &\times\Gamma_{p-1}\left(\tfrac{1}{2}(p-1)\right)\Gamma_{p-1}\left(\tfrac{1}{2}(p+2)\right)\Gamma_{p-1}\left(\tfrac{1}{2}p\right)C_\nu(I_{p-1})\\ &\times\frac{\left(\frac{1}{2}(p+2)\right)_\nu}{\pi^{\frac{1}{2}(p-1)^2}\Gamma_{p-1}(p+1)(p+1)_\nu}.\end{aligned} \tag{3.17}$$

The marginal density of (λ_1, l_p) is then

$$f(\lambda_1, l_p) = \pi^{p-\frac{1}{2}}\Gamma_{p-1}\left(\tfrac{1}{2}(p-1)\right)\Gamma_{p-1}\left(\tfrac{1}{2}(p+2)\right)$$

$$\times \frac{|\Sigma|^{-\frac{1}{2}n}\Gamma_{p-1}\left(\frac{1}{2}p\right)}{\Gamma_p\left(\frac{1}{2}p\right)\Gamma_p\left(\frac{1}{2}n\right)2^{\frac{1}{2}np}\Gamma_{p-1}(p+1)}\lambda_1^{\frac{1}{2}np-1}e^{-\frac{1}{2}\lambda_1 p}$$

$$\times \sum{}^* C_\kappa\left(\tfrac{1}{2}(I-\Sigma^{-1})\right)\lambda_1^{k+m}\lambda_p^{\frac{1}{2}(p-1)(p+2)+s+t+m-1}$$

$$\times\left(\tfrac{1}{2}(p+1-n)\right)_\sigma b_{\kappa,\tau}\alpha_{\tau'}g^{\delta}_{\tau'\mu}g^{\nu}_{\sigma,\delta}\left[\tfrac{1}{2}(p-1)(p+2)+s+t+m\right]$$

$$\times\frac{\left(\frac{1}{2}(p+1)\right)_\nu C_\nu(I_{p-1})}{k!s!m!2^m C_\kappa(I_p)(p+1)_\nu}. \tag{3.18}$$

Finally integrating over λ_1 yields the marginal density of l_p as

$$g(l_p) = \frac{\Pi^{p-\frac{1}{2}}\Gamma_{p-1}\left(\frac{1}{2}(p-1)\right)\Gamma_{p-1}\left(\frac{1}{2}(p+2)\right)\Gamma_{p-1}\left(\frac{1}{2}p\right)|\Sigma|^{-\frac{1}{2}n}}{\Gamma_p\left(\frac{1}{2}p\right)\Gamma_p\left(\frac{1}{2}n\right)2^{\frac{1}{2}np}\Gamma_{p-1}(p+1)}$$

$$\times\sum{}^* C_\kappa\left(\tfrac{1}{2}(I-\Sigma^{-1})\right)l_p^{\frac{1}{2}(p-1)(p+2)+s+t+m-1}\left(\tfrac{1}{2}(p+1-n)\right)_\sigma$$

$$\times b_{\kappa,\tau}\alpha_{\tau'}g^{\delta}_{\tau'\mu}g^{\nu}_{\sigma,\delta}\left(\tfrac{1}{2}(p+2)\right)_\sigma\left[\tfrac{1}{2}(p-1)(p+2)+s+t+m\right]$$

$$\times\frac{\Gamma\left(\frac{1}{2}np+k+m\right)C_\nu(I_{p-1})}{k!m!s!C_\kappa(I_p)(p+1)_\nu 2^m\left(\frac{1}{2}p\right)^{\frac{1}{2}np+k+m}} \tag{3.19}$$

and Σ^* is given by (3.15).

Krishnaiah and Waikar [19] also report that they have derived the density of λ_1/λ_p but consider the expression so complicated that it is of no practical value, and hence the density was not published.

It is quite clear from the above expressions that the densities of the condition numbers $C_1 = \Sigma\lambda_i/\lambda_p$ and $C_2 = \lambda_1/\lambda_p$ are extremely complicated for general Σ and would be of little practical value. By relaxing the assumptions on Σ or by using asymptotic results, some of the complications may disappear. These alternatives are considered in the next two sections.

4. THE DISTRIBUTIONS OF THE CONDITION NUMBERS WITH RESTRICTIONS ON THE COVARIANCE MATRIX Σ

As stated in Section 2, the distribution of the condition number is of interest for general Σ. If $\Sigma = \sigma^2 I$, it would imply that the variables $(x_1, \ldots, x_p)$ (assuming that y is one of the x-variables) are independent. If $(1/n)A$ and R are the maximum likelihood estimates of Σ and $P = I$ (the population correlation matrix), respectively, then it is unlikely that multicollinearities will be present in the matrix A or R. If such collinearities are however present, then the condition number $C_1 = \Sigma\lambda_i/\lambda_p$ and $C_2 = \lambda_1/\lambda_p$ will, of course, be much larger than they would have been for an orthogonal system. Thus rejection of hypothesis of the type $\Sigma = \sigma^2 I$ is an indication that collinearities are present; however, such collinearities may not be harmful.

The joint distribution of the roots $(\lambda_1, \ldots, \lambda_p)$ for the case $\Sigma = \sigma^2 I$ is now much easier to handle and percentage points of the condition number $C_1 = \Sigma\lambda_i/\lambda_p$, can be found from the tables by Schuurmann, Krishnaiah, and Chattopadhyay [25]. They tabulated the percentage points of $\lambda_p/\Sigma\lambda_i = 1/C_1$ for significance level $\alpha = (0.05; 0.01)$. More extensive tables are available in a technical report (ARL-730010) by the authors. The tables are given for $p = 3(1)5$ and $r = \frac{1}{2}(n - p - 1) = 0(1)25$ but restricted to $r = 0(1)16$ for $p = 6$.

Tables of the percentage points of the condition number $C_2 = \lambda_1/\lambda_p$ are given by Krishnaiah and Schuurmann [20].

We now consider two examples to illustrate some of the uses of these tables. The first example was reported by Troskie [30]. To illustrate the effect of multicollinearity, Troskie performed a regression on eight variables with $n = 104$ observations, one variable being the dependent variable. The characteristic roots of the sample covariance matrix

$$S = \frac{1}{n-1} A$$

of the seven independent variables was given by

$$\lambda_{1S} = 26585.002, \qquad \lambda_{2S} = 6107.498, \qquad \lambda_{3S} = 4340.910,$$

$$\lambda_{4S} = 864.108, \lambda_{5S} = 303.634, \qquad \lambda_{6S} = 63.763, \qquad \lambda_{7S} = 23.537.$$

The condition number $C_{2S} = \lambda_{1S}/\lambda_{7S} = 1129$, while $C_{1S} = \Sigma\lambda_i/\lambda_7 = 1626$.

For the correlation matrix R the characteristic roots are as follows:

$$\lambda_{1R} = 3.393, \qquad \lambda_{2R} = 1.224, \lambda_{3R} = 1.004, \lambda_{4R} = 0.645,$$

$$\lambda_{5R} = 0.363, \qquad \lambda_{6R} = 0.067, \lambda_{7R} = 0.004.$$

The condition numbers are $C_{2R} = \lambda_{1R}/\lambda_{7R} = 948$, $C_{1R} = \Sigma\lambda_i/\lambda_7 = 1750$.

Notice that by inspection one would not consider that the smallest root of S, i.e. $\lambda_{7S} = 23.537$, is small. On the other hand, the smallest root of R, i.e. $\lambda_{7R} = 0.004$, can certainly be considered to be too small. It is very interesting, however, to note that the condition numbers for the two matrices are hardly different in magnitude and, in fact, extremely large, reflecting the ill conditioning of these matrices. Under the assumption that $\Sigma = \sigma^2 I$ the approximate critical values for these condition numbers with $n = 104$ are (from the tables reported above) $C_{1S}(0.01) = 21$ and $C_{2S}(0.01) = 15$.

Now obviously the assumption that $\Sigma = \sigma^2 I$ is not feasible (and of course will be rejected if tested on the sample roots). Nevertheless the difference in magnitude of the observed condition numbers and the critical values, under the assumption $\Sigma = \sigma^2 I$, is so large that one would immediately expect that the least squares estimates will be seriously affected if the matrices S or R are used without adjustment.

The second example is taken from data supplied by Thompson [28]. The characteristic roots of the covariance matrix S and correlation matrix R are given by Table 1 ($p = 9$ independent variables). All the condition numbers are extremely large, indicating that the matrices S and R are probably ill conditioned. It is remarkable how much larger the condition numbers for the matrix S are compared to that of the matrix R. There is a definite indication that a strong multicollinearity exists among the independent variables and that a correction procedure is necessary for the least squares estimates. Because of the large number of variables ($p = 9$), critical values for the condition numbers are available in the cited technical report.

Table 2 illustrates the difference in magnitude of the least squares estimates and some correction procedures as suggested in Section 1.

For both examples the effect of only one small root on the principal component and LRRA estimates has been eliminated. Investigating the effect

TABLE 1

λ_{iS}	λ_{iR}
722.272	3.296
201.667	3.154
22.338	1.021
6.829	0.808
0.562	0.363
0.443	0.215
0.134	0.106
0.029	0.034
0.001	0.004
$C_{2S} = \lambda_1/\lambda_p = 722{,}272$	$C_{2R} = 849$
$C_{1S} = \Sigma\lambda_i/\lambda_p = 953{,}110$	$C_{1R} = 2250$

TABLE 2

Troskie

	O.L.S.	Ridge trace $k = 0.2$	Princ. comp. rank$(X'X) = 6$	LRRA
β_0	4625.7	3605.9	1382.1	892.46
β_1	6.9807	4.372	6.3295	6.2215
β_2	− 8.8436	− 3.3826	− 9.6453	− 9.7290
β_3	0.4058	− 1.3295	0.6796	0.7086
β_4	7.8023	7.1952	11.0451	11.5130
β_5	− 5.2807	− 0.2072	0.4712	1.3622
β_6	0.4217	0.8562	0.3622	0.3536
β_7	5.7400	− 0.5619	− 0.1760	− 1.0969

Thompson

	O.L.S.	Ridge trace $k = 0.1$	Princ. comp. rank$(X'X) = 8$	LRRA
β_0	2.1016	0.0835	6.1883	6.6231
β_1	0.7088	0.4017	0.9558	0.9729
β_2	− 1.9220	− 0.1071	− 4.6870	− 4.9338
β_3	1.1256	0.0944	0.0650	− 0.0774
β_4	0.1130	0.2485	0.0922	0.0931
β_5	0.1795	0.1369	0.2218	0.2258
β_6	− 0.0183	0.0158	− 0.0187	− 0.0183
β_7	− 0.0776	− 0.0610	− 0.0818	− 0.0878
β_8	− 0.0852	− 0.0059	− 0.0116	− 0.0015
β_9	− 0.3432	− 0.0906	− 0.2740	− 0.2617

of the second smallest root, one finds for the first example

$$\lambda_{1S} = 26585.002, \qquad \lambda_{1R} = 3.393,$$
$$\lambda_{6S} = 63.763, \qquad \lambda_{6R} = 0.067,$$
$$C_{2S} = \lambda_{1S}/\lambda_{6S} = 417, \qquad C_{2R} = \lambda_{1R}/\lambda_{6R} = 51.$$

Although C_{2S} is large, C_{2R} is not. Perhaps a modified fractional rank estimate (between 6 and 7) as suggested by Marquardt [21] would be better.

For the second example we have

$$\lambda_{1S} = 722.272, \qquad \lambda_{1R} = 3.396,$$
$$\lambda_{8S} = 0.029, \qquad \lambda_{8R} = 0.034,$$
$$C_{2S} = 24905, \qquad C_{2R} = 99.$$

Here again C_{2S} is very large, while C_{2R} is not that large.

For LRRA, Webster, Gunst, and Mason [32] suggested that vectors should be eliminated for which the roots $\lambda_i \leqslant 0.3$ and the weights of the characteristic vectors with respect to the dependent variable are too small—say $V_{yi} \leqslant 0.1$. In the first of the above two examples we have for the roots and vectors of the augmented matrix

$$\begin{pmatrix} Y'Y & Y'X \\ X'Y & X'X \end{pmatrix}$$

the values

$$\lambda_{1R} = 4.478, \qquad V_{y1} = 0.431,$$
$$\lambda_{7R} = 0.025, \qquad V_{y7} = -0.661,$$
$$\lambda_{8R} = 0.003, \qquad V_{y8} = -0.077;$$

in the second,

$$\lambda_{1R} = 3.893, \qquad V_{y1} = 0.445,$$
$$\lambda_{9R} = 0.024, \qquad V_{y9} = 0.360,$$
$$\lambda_{10R} = 0.004, \qquad V_{y10} = 0.083.$$

Thus, as suggested by Webster et al., only the last root and characteristic vector were eliminated for both examples.

The differences in magnitude between the OLS estimates and the other estimates are quite alarming. For both examples the principal component and LRRA estimates are very close to each other.

One important point which emerges from the example given by Thompson [28] is the very large condition numbers of the matrix S compared to that of the matrix R. The matrix S has very small roots and is possibly very unstable. It is clear that the first four roots of S virtually explain all the variation in the matrix S, with the result that the remaining roots are small.

A third example of the effect of multicollinearity and the application of ridge regression has been reported by Hadgu [10]. The model is the following:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e,$$

where

Y = reported annual rates of congenital syphilis in the United States,

X_1 = reported annual rates of primary and secondary syphilis in the United States,

X_2 = reported annual rates of early latent syphilis in the United States,

X_3 = reported annual rates of late and late latent syphilis in the United States

(all rates are calculated per 500,000).

The correlation matrix for the 13 observations reported by Hadgu (years 1957–1969) is

	Y	X_1	X_2	X_3
Y	1	− 0.629	0.867	0.990
X_1	− 0.629	1	− 0.259	− 0.625
X_2	0.867	− 0.259	1	0.895
X_3	0.990	− 0.625	0.895	1

The condition number for the Wishart matrix is 1184, while the condition number for the correlation matrix is 122. Both these numbers are very large, indicating severe multicollinearity.

The estimated regression coefficients are

	$k=0$	$k=0.1$	$k=0.2$	$k=0.3$
Constant	− 2.7753	− 2.4091	− 1.8204	− 1.2784
X_1	0.0334	− 0.0465	− 0.0516	− 0.0523
X_2	− 0.1710	0.1623	0.1808	0.1821
X_3	0.1047	0.0438	0.0374	0.0343

It is clear that even small ridging has considerable effect on the regression coefficients.

5. SOME ASYMPTOTIC RESULTS

Anderson (1965) gives the following expansion for the joint distribution of $(\lambda_1, \ldots, \lambda_p)$, the roots of $(1/n)A = S$, when the roots $\alpha_1, \ldots, \alpha_p$ of Σ, are assumed to be all distinct:

$$f(\lambda_1, \ldots, \lambda_p) = M(A) \prod_{i=1}^{p} \lambda_i^{n-p-1} e^{-n/2(\lambda_i/\alpha_i)} \prod_{i<j} (\lambda_i - \lambda_j)^{1/2} F, \quad (5.1)$$

where

$$M(A) = \frac{\prod_{i=1}^{n} (\alpha_i)^{-\frac{1}{2}n} \prod_{i<j} \left(\frac{1}{\alpha_j} - \frac{1}{\alpha_i} \right)^{-1/2}}{\left(\frac{n}{2} \right)^{-[np/2 - p(p-1)/4]} \prod_{i=1}^{p} \Gamma(\frac{1}{2}(n-i+1))} \quad (5.2)$$

and

$$F = 1 + \frac{1}{2n}\sum_{i<j}\frac{1}{C_{ij}} + \frac{9}{8n^2}\sum_{i<j}\frac{1}{C_{ij}^2} + \frac{1}{4n^3}\sum\frac{1}{C_{ij}C_{kl}} + \cdots \tag{5.3}$$

and

$$C_{ij} = \left(\frac{1}{\alpha_i} - \frac{1}{\alpha_j}\right)(\lambda_i - \lambda_j). \tag{5.4}$$

Assume n large enough so that $F = 1$.

Then by again making the transformation $u_i = \lambda_i / \Sigma\lambda_i$, $i = 2, \ldots, p$, and $l_i = \lambda_i/\lambda_p$, $i = 1, \ldots, p-1$, we get the marginal densities of $(u_2, \ldots, u_p)$ and $(l_1, \ldots, l_{p-1})$ as

$$f(u_2, \ldots, u_p) = M(A)|U|^{\frac{1}{2}(n-p-1)}\left(\frac{u_1}{\alpha_1} + \cdots + \frac{u_p}{\alpha_p}\right)^{-\frac{1}{4}(2n-p+1)}$$
$$\times \prod_{i<j}(u_i - u_j)^{1/2}\Gamma\left(\tfrac{1}{4}(2n-p+1)\right)\left(\frac{n}{2}\right)^{\frac{1}{4}(2n-p+1)}, \tag{5.5}$$

where $U = \operatorname{diag}(u_1, \ldots, u_p)$ and $u_1 = 1 - \Sigma_{i=2}^{p} u_i$, and

$$f(l_1, \ldots, l_{p-1}) = M(A)|L|^{\frac{1}{2}(n-p-1)}\left(\frac{l_1}{\alpha_1} + \cdots + \frac{l_{p-1}}{\alpha_{p-1}} + \frac{1}{\alpha_p}\right)^{-\frac{1}{4}(2n-p+1)}$$
$$\times |L - I| \prod_{i<j<p}(l_i - l_j)^{1/2}\Gamma\left(\tfrac{1}{4}(2n-p+1)\right), \tag{5.6}$$

where $L = \operatorname{diag}(l_1, \ldots, l_{p-1})$. Again it appears difficult to find the marginal densities of the condition numbers $C_1 = 1/u_p$ and $C_2 = l_1$.

Anderson [1] also showed that for large n,

$$\prod_{i<j}\left(\frac{\lambda_i - \lambda_j}{\alpha_j - \alpha_i}\right)^{1/2} \to 1$$

with probability one, and hence the joint density of $(\lambda_1, \ldots, \lambda_p)$ then becomes

$$C\prod \lambda_i^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}n\lambda_i/\alpha_i} \tag{5.7}$$

with

$$C = \frac{1}{\Gamma\left(\frac{1}{2}(n-p+1)\right)} \prod_{i=1}^{p} \left(\left(\frac{n}{2\alpha_i} \right)^{\frac{1}{2}(n-p+1)} \right). \tag{5.8}$$

Thus $n\lambda_i/\alpha_i$ are independently distributed as X^2-variates with $(n-p+1)/2$ degrees of freedom. Thus the density of $C_2 = l_1 = \lambda_1/\lambda_p$ is that of (α_1/α_p) $F(\frac{1}{2}(n-p+1), \frac{1}{2}(n-p+1))$, and therefore knowledge of the ratio α_1/α_p is needed. Under the assumption $\Sigma = \sigma^2 I$, critical values for λ_1/λ_p of the order of 20 appear to be large. Assuming therefore $\alpha_1/\alpha_p = 20$, a cutoff value of $20F^{\alpha}(\frac{1}{2}(n-p+1), \frac{1}{2}(n-p+1))$ seems reasonable. For the example in Section 4 with $n = 104$, we have $F^{0.01}_{(49,49)} \simeq 2$, so that the cutoff value is 40.

The joint density of $u_2, \ldots, u_{p-1}$ with the condition number $C_1 = 1/u_p$ is more complicated and is given by

$$f(u_2, \ldots, u_p) = C' \prod_{i=1}^{p} u_1^{\frac{1}{2}(n-p-1)} \left(\frac{u_1}{\alpha_1} + \cdots + \frac{u_p}{\alpha_p} \right)^{-\frac{1}{2}(n-p-1)p+p} \tag{5.9}$$

with

$$C' = C\left(\tfrac{1}{2}n\right)^{\frac{1}{2}(n-p-1)p+p} \Gamma\left(\tfrac{1}{2}(n-p-1)p+p\right)$$

and

$$u_1 = 1 - \sum_{i=2}^{p} u_i.$$

It is again difficult to find the marginal density of $u_p = 1/C_2$.

Girshick [8] has given the following normal approximation for the joint distribution of the roots $(\lambda_1, \ldots, \lambda_p)$ as n becomes large: $\sqrt{n-1}\,(\lambda_i - \alpha_i)$ is normally distributed with mean zero and variance $2\alpha_i^2$, and independent of

$$\lambda_i \sim N\left(\alpha_i, \frac{2\alpha_i^2}{n-1}\right),$$

so that

$$\frac{(n-1)\lambda_i^2}{2\alpha_i^2} \sim X_i^{2\prime}(\mu) \qquad \text{with} \quad \mu = \frac{\frac{1}{2}\alpha_i^2}{2\alpha_i^2/(n-1)} = \frac{n-1}{4}. \tag{5.10}$$

Thus

$$\frac{\lambda_1^2/\alpha_1^2}{\lambda_p^2/\alpha_p^2} \sim F_{1,1}''(\mu,\mu). \tag{5.11}$$

Now the doubly noncentral F can be approximated by a central F as follows:

$$F_{1,1}''(\mu,\mu) = F_{\nu,\nu'}, \tag{5.12}$$

where

$$\nu = \nu' = \left(1 + \frac{n-1}{4}\right)^2 \left(1 + \frac{2(n-1)}{4}\right)^{-1}.$$

For the example with $n = 104$ we obtain $\nu \simeq 14$ with $F_{(14,14)}^{0.01} \simeq 3.66$. Assuming $\alpha_1/\alpha_p = 20$, a cutoff value of $(20)^2 \times (3.66) = 1464$ for λ_1^2/λ_p^2 or 38 for λ_1/λ_p seems reasonable.

From the fact that $\lambda_i \sim N(\alpha_i, 2\alpha_i^2/(n-1))$, confidence limits can be found for the smallest root α_p. These are

$$a = \frac{\lambda_p}{1 + Z_{1/2}(\alpha)\sqrt{2/n-1}} \leqslant \alpha_p \leqslant \frac{\lambda_p}{1 - Z_{1/2}(\alpha)\sqrt{2/n-1}} = b. \tag{5.13}$$

where $Z_{1/2}(\alpha)$ is the $\frac{1}{2}\alpha$ 100% critical value of the standard normal distribution. Press [24] states that for large n it is approximately true that

$$\frac{a}{\Sigma\lambda_j} \leqslant \frac{\alpha_p}{\Sigma\alpha_i} \leqslant \frac{b}{\Sigma\lambda_j}. \tag{5.14}$$

Thus for large n an approximate $1-\alpha$ confidence interval for the population condition number $\Sigma\alpha_i/\alpha_p$ is given by

$$\frac{\Sigma\lambda_j}{b} \leqslant \frac{\Sigma\alpha_i}{\alpha_p} \leqslant \frac{\Sigma\lambda_j}{a}. \tag{5.15}$$

6. SOME EXACT RESULTS FOR A CIRCULAR COVARIANCE MATRIX

Let $(X_{1\alpha},\ldots,X_{p\alpha})$, $\alpha=1,2,\ldots,N$, be a sample from a p-variate $N(\mu,\Sigma)$ distribution. Σ is uniform and is given by

$$\Sigma=\sigma^2(1-\rho)I+\sigma^2\rho ee',$$

where $e'=(1,1,\ldots,1)$. Olkin and Pratt [23] have shown that the minimal sufficient statistic of (μ,τ_1,τ_2) is given by

$$\left(N^{1/2}\bar{x}\Gamma,v_1,v_2\right)=\left(N^{1/2}\bar{x}\Gamma,v_{11},\sum_{i=2}^{p}v_{ii}\right),$$

where

$$\tau_1=\sigma^2\left[1+(p-1)\rho\right],\qquad \tau_2=\sigma^2(1-\rho)$$

are the characteristic roots of Σ (the last $p-1$ roots being equal to τ_2),

$$\Gamma=\left[\gamma_{ij}\right]=\left[p^{-1/2}(\cos 2\pi p^{-1}(i-1)(j-1)+\sin 2\pi p^{-1}(i-1)(j-1)\right],$$

and

$$V=\Gamma'S\Gamma=\left[v_{ij}\right],\qquad i,j=1,\ldots,p.$$

Also $\bar{x}=(\bar{x}_1,\ldots,\bar{x}_p)'$ and S/n, $n=N-1$, is the sample covariance matrix, where $S=[s_{ij}]$, $i,j=1,\ldots,p$, and

$$s_{ij}=\sum_{\alpha=1}^{N}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j).$$

The distribution of ν_1/τ_1 is χ^2_n, and of ν_2/τ_2 is $\chi^2_{n(p-1)}$ and independent. Furthermore

$$\hat{\tau}_1=\lambda_1=\frac{\nu_1}{n}\quad\text{and}\quad \hat{\tau}_2=\lambda_2=\frac{\nu_2}{p(p-1)}.$$

Now τ_1/τ_2 is the population condition number and

$$\frac{\tau_1}{\tau_2} = \frac{1+(p-1)\rho}{1-\rho}$$

Thus $[\nu_1/\tau_1 n]/[\nu_2/\tau_2 n(p-1)]$ is distributed as $F_{n,n(p-1)}$, so that the sample condition number $C_2 = \lambda_1/\lambda_2$ is distributed as

$$\frac{1+(p-1)\rho}{1-\rho} F_{n,n(p-1)}.$$

The maximum likelihood estimate of ρ is given by

$$\hat{\rho} = \frac{p\left(\sum_{i=1}^{p} \sum_{\substack{j=1 \\ j\neq 1}}^{p} S_{ij}\right)}{(p-1)\sum_{i=1}^{p} S_{ii}}.$$

CONCLUSIONS

If $(\lambda_1, \ldots, \lambda_p)$ are the roots of either the Wishart matrix A or the correlation matrix R, then the condition numbers are defined to be $\Sigma\lambda_i/\lambda_p$ or λ_1/λ_p. It is demonstrated that large condition numbers have severe effects on the usual least squares estimates in regression. The distributions of these condition numbers for general Σ are extremely complicated and in some cases intractable.

If Σ is assumed to be of the form $\Sigma = \sigma^2 I$, then percentage points of the condition numbers are available. Rejection of the hypothesis $\Sigma = \sigma^2 I$ is an indication that collinearity is present, but the collinearity may not be severe. The condition numbers should be much larger than the critical points under the assumption $\Sigma = \sigma^2 I$. If $\alpha_1 > \alpha_2 > \cdots > \alpha_p$ are the roots of Σ and α_1/α_p the population condition number, then if α_1/α_p can be assumed to be some large number (say larger than 20), then the asymptotic results given in Section 5 can be used to compute a cutoff value for the condition number λ_1/λ_p. Some such correction as the ridge or principal component or LRRA should be made to the least squares estimates if large condition numbers are present.

REFERENCES

1 G. A. Anderson, An asymptotic expansion for the distribution of the latent roots of the estimated covariance matrix, *Ann. Math. Statist.* 36:1153–1173 (1965).

2 D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics*, Wiley, New York, 1977.

3 A. G. Constantine, Some non-central distribution problems in multivariate analysis, *Ann. Math. Statist.* 34:1270–1285 (1963).

4 A. G. Constantine, The distribution of Hotelling's generalized T_0^2, *Ann. Math. Statist.* 37:215–225 (1966).

5 R. A. Fisher, Simultaneous distribution of correlation coefficients, *Sankhyā* 24:1–8 (1962).

6 G. E., Forsythe, M. A. Malcolm and C. E. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, N. J., 1977.

7 G. E. Forsythe and C. B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs, N. J., 1967.

8 M. A. Girshick, Roots of determinantal equations, *Ann. Math. Statist.* 10:203–224 (1939).

9 R. F. Gunst, J. T. Webster, and R. L. Mason, A comparison of least squares and latent root regression estimators, *Technometrics* 18:75–83 (1976).

10 A. Hadgu, An application of ridge regression analysis in the study of syphilis data, *Statist. Med.* 3:293–299 (1984).

11 D. M. Hawkins, On the investigation of alternative regression by principal components analysis, *Appl. Statist.* 22:275–286 (1973).

12 A. E. Hoerl, and R. W. Kennard, Ridge regression: Application to nonorthogonal problems, *Technometrics* 12:69–82 (1970).

13 A. E. Hoerl, and R. W. Kennard, Ridge regression. Iterative estimation of biasing parameter, *Comm. Statist. A—Theory Methods* 5 (1):77–88 (1976).

14 A. T. James, The distribution of the latent roots of the covariance matrix, *Ann. Math. Statist.* 31:151–158 (1960).

15 A. T. James, Distributions of matrix variates and latent roots derived from normal samples, *Ann. Math. Statist.* 35:475–501 (1964).

16 A. T. James, Inference on latent roots by calculation of hyper-geometric functions of matrix argument, in *Multivariate Analysis* (P. K. Krishnaiah, Ed.), Academic, New York, 1966, pp. 209–235.

17 D. G. Kabe, On the distribution of the regression coefficient matrix of a normal distribution, *Austral. J. Statist.* 10:21–23 (1968).

18 C. G. Khatri, and K. C. S. Pillai, On the non-central distributions of two test criteria in multivariate analysis of variance, *Ann. Math. Statist.* 39:215–226 (1968).

19 P. R. Krishnaiah, and V. B. Waikar, Simultaneous tests for equality of latent roots against certain alternatives–II, *Ann. Inst. Statist. Math.* 24:81–85 (1972).

20 P. R. Krishnaiah, and F. J. Schuurmann, On the evaluation of some distributions that arise in simultaneous tests for the equality of the latent roots of the covariance matrix. *J. Multivariate Anal.* 4:245–282 (1974).
21 D. W. Marquardt, Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation, *Technometrics* 12:591–612 (1970).
22 A. W. Marshall, and I. Olkin, Norms and inequalities for condition numbers, III, Technical Report No. 53, Stanford Univ.
23 I. Olkin, and J. W. Pratt, Unbiased estimation of certain correlation coefficients, *Ann. Math. Statist.* 29:201–211 (1958).
24 S. J. Press, *Applied Multivariate Analysis*, Holt, Rhinehart and Winston, New York, 1972.
25 F. J. Schuurmann, P. R. Krishnaiah, and A. K. Chattopadhyay, On the distributions of the ratio of the extreme roots to the trace of the Wishart matrix, *J. Multivariate Anal.* 3:445–453 (1973).
26 T. Sugiyama, Distribution of the largest latent root and the smallest latent root of the generalized B statistic and F statistic in multivariate analysis, *Ann. Math. Statist.* 38:1152–1159 (1967).
27 A. L. Sampson, A tale of two regressions, *J. Amer. Statist. Assoc.* 69:682–689 (1974).
28 M. Thompson, Selection of variables in multiple regression. Part II. Chosen procedures, computation and examples, *Internat. Statist. Rev.* 46:129–146 (1978).
29 C. G. Troskie, Regression and correlation, Proceedings of the Third Symposium on Mathematical Statistics, NRIMS, *Wisk.* 89:21–50 (1971).
30 C. G. Troskie, Multicollinearity, ridge regression and principal components, Unpublished report, Univ. of Cape Town (1977).
31 H. D. Vinod, A survey of ridge regression and related techniques for improvement over ordinary least squares, *Rev. Econom.* and *Statist.* LX:121–131 (1978).
32 J. T. Webster, R. F. Gunst, and R. L. Mason, Latent root regression analysis, *Technometrics* 16:513–522 (1974).

Received 18 June 1985; revised 12 August 1985