The Annals of Statistics
1986, Vol. 14, No. 2, 418–430

ASYMPTOTIC THEORY FOR COMMON PRINCIPAL COMPONENT ANALYSIS[1]

By Bernard N. Flury

University of Berne

Under the common principal component model k covariance matrices $\mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_k$ are simultaneously diagonalizable, i.e., there exists an orthogonal matrix $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\mathbf{\Sigma}_i\boldsymbol{\beta} = \mathbf{\Lambda}_i$ is diagonal for $i = 1, \dots, k$. In this article we give the asymptotic distribution of the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\mathbf{\Lambda}_i$. Using these results, we derive tests for (a) equality of eigenvectors with a given set of orthonormal vectors, and (b) redundancy of $p - q$ (out of p) principal components. The likelihood-ratio test for simultaneous sphericity of $p - q$ principal components in k populations is derived, and some of the results are illustrated by a biometrical example.

1. Introduction. Common principal component analysis (CPCA) is a generalization of principal component analysis (PCA) to k groups (Flury, 1984). The key assumption is that the $p \times p$ covariance matrices $\mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_k$ of k populations can be diagonalized by the same orthogonal transformation, i.e., there exists an orthogonal matrix $\boldsymbol{\beta}$ such that

$$H_C\colon \boldsymbol{\beta}'\mathbf{\Sigma}_i\boldsymbol{\beta} = \mathbf{\Lambda}_i \quad \text{(diagonal)} \qquad (i = 1, \dots, k) \tag{1.1}$$

holds. H_C is called the hypothesis of common principal components (CPC's). Flury (1984) derives the normal theory maximum likelihood estimates of $\boldsymbol{\beta}$ and $\mathbf{\Lambda}_i$ and gives numerical examples.

In the one sample case $k = 1$, CPC's reduce to ordinary principal components (PC's). In this case the ML estimates of $\boldsymbol{\beta}$ and $\mathbf{\Lambda} = \mathbf{\Lambda}_1$ are the eigenvectors and eigenvalues of a Wishart matrix $\mathbf{S}_1$. The asymptotic distribution theory for this situation has been developed by Girshick (1939), Lawley (1953, 1956) and Anderson (1963). The present paper gives essentially generalizations of results obtained by Anderson.

In one-group PCA, the eigenvectors $\boldsymbol{\beta}_i$ forming the orthogonal matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$ are usually ordered according to the associated eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. In CPCA no obvious fixed order of the columns of $\boldsymbol{\beta}$ need be given, since the rank order of the diagonal elements of the $\mathbf{\Lambda}_i$ is not necessarily the same for all $\mathbf{\Lambda}_i$. However, we can use some convention, e.g., that the columns of $\boldsymbol{\beta}$ be arranged according to the first group, i.e., such that $\boldsymbol{\beta}_1'\mathbf{\Sigma}_1\boldsymbol{\beta}_1 > \boldsymbol{\beta}_2'\mathbf{\Sigma}_1\boldsymbol{\beta}_2 >$

Received September 1984; revised July 1985.

[1] This work was done under contract 82.008.0.82 of the Swiss National Science Foundation at Stanford University. Partial support was also provided by NSF grant DMS 84-11411.

AMS 1980 *subject classifications.* Primary 62H25; secondary 62H15, 62E20.

Key words and phrases. Maximum likelihood, covariance matrices, eigenvectors, eigenvalues.

$\cdots > \beta_p' \Sigma_1 \beta_p$ (assuming that the p characteristic roots of Σ_1 are all distinct). This will enable us to speak about first, second, or last principal components also in the k-group case.

Tests for various hypotheses about β and Λ in PCA have been proposed by Anderson (1963). We are going to construct analogous tests for CPCA. More specifically, we will treat the following problems:

1. Is the jth eigenvector β_j identical with a given (normalized) vector β_j^0? More generally, for q different eigenvectors $\beta_j, \beta_{j+1}, \ldots, \beta_{j+q-1}$, are they identical with q given (orthonormal) eigenvectors $\beta_j^0, \ldots, \beta_{j+q-1}^0$? This problem will be treated in Section 3.
2. As an associate editor handling the previous paper (Flury, 1984) has pointed out, the most useful applications of CPCA would probably be those in which some relatively small number q of rotated axes are sufficient to recover most of the variability in each of the k groups. It is therefore useful to have a criterion for neglecting CPC's with small contributions. A solution to this problem is given in Section 4.1.
3. When PC's are interpreted, it is important to make sure that the roots λ_j and λ_k (say) are not identical, because otherwise the associated eigenvectors β_j and β_k are not uniquely defined. Similarly in CPCA two eigenvectors β_j and β_k are uniquely defined if *in at least one population* the two associated eigenvalues are not identical. A likelihood ratio test dealing with this problem is given in Section 4.2.

We will from now on always assume that the matrices $\Sigma_1, \ldots, \Sigma_k$ are positive definite symmetric (p.d.s.). The diagonal elements of Λ_i will be denoted by λ_{ij}, i.e., $\Lambda_i = \operatorname{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$ $(i = 1, \ldots, k)$. All results will be based on k independent sample covariance matrices $\mathbf{S}_i$ with n_i degrees of freedom, respectively, such that $n_i \mathbf{S}_i$ has the Wishart distribution $W_p(n_i, \Sigma_i)$. The ML estimates of $\beta = (\beta_1, \ldots, \beta_p)$ and Λ_i are denoted by $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ and $\hat{\Lambda}_i = \operatorname{diag}(\hat{\lambda}_{i1}, \ldots, \hat{\lambda}_{ip})$.

2. Asymptotic distribution of the maximum likelihood estimates. In this section we are using general properties of ML-estimates under regularity conditions; see, e.g., Silvey (1975, Chapters 4 and 7) and Wilks (1944, Chapter 6). In particular we will use the fact that the joint asymptotic distribution of the parameter estimates is multivariate normal, the covariance matrix being given by the inverse of the Fisher information matrix. The log-likelihood function of the k samples, up to an additive constant, is given by

$$(2.1)\quad g(\Lambda_1, \ldots, \Lambda_k, \beta | \mathbf{S}_1, \ldots, \mathbf{S}_k) = -\tfrac{1}{2} \sum_{i=1}^{k} n_i \left[\sum_{j=1}^{p} \left(\log \lambda_{ij} + \beta_j' \mathbf{S}_i \beta_j / \lambda_{ij} \right) \right]$$

(Flury 1984, formula 2.5). Assume that the β_j are well defined, i.e., for each pair (j, l) there is at least one $i \in \{1, \ldots, k\}$ such that $\lambda_{ij} \neq \lambda_{il}$. Let $\boldsymbol{\lambda}_{(i)}' = (\lambda_{i1}, \ldots, \lambda_{ip})$, $s = p(p-1)/2$, and denote by β^* a vector composed of s functionally independent elements of β. Put $n = n_1 + \cdots + n_k$ and $r_i = n_i/n$ $(i =$

$1, \dots, k$). Then the information matrix is

$$
(2.2)\qquad
\begin{array}{c|cccc|c}
 & \boldsymbol{\lambda}'_{(1)} & \boldsymbol{\lambda}'_{(2)} & \cdots & \boldsymbol{\lambda}'_{(k)} & \boldsymbol{\beta}^{*\prime} \\
\hline
\boldsymbol{\lambda}_{(1)} & \frac{1}{2} n r_1 \boldsymbol{\Lambda}_1^{-2} & \mathbf{0} & \cdots & \mathbf{0} & \\
\boldsymbol{\lambda}_{(2)} & \mathbf{0} & \frac{1}{2} n r_2 \boldsymbol{\Lambda}_2^{-2} & \cdots & \mathbf{0} & \\
\vdots & \vdots & \vdots & & \vdots & \mathbf{G}' \\
\boldsymbol{\lambda}_{(k)} & \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{2} n r_k \boldsymbol{\Lambda}_k^{-2} & \\
\hline
\boldsymbol{\beta}^{*} & & & \mathbf{G} & & \mathbf{A}
\end{array},
$$

where $\mathbf{A}$ and $\mathbf{G}$ are not yet determined.

Since $\hat{\lambda}_{ij} = \hat{\boldsymbol{\beta}}_j' \mathbf{S}_k \hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\beta}}_j$ is a consistent estimate of $\boldsymbol{\beta}_j$, we can use the asymptotic ($n_i \to \infty$) normality of $n_i \mathbf{S}_i$ (Muirhead, 1982, page 19) to get the asymptotic univariate distribution of $\hat{\lambda}_{ij}$ as

$$
(2.3)\qquad \sqrt{n_i}\left(\hat{\lambda}_{ij} - \lambda_{ij}\right) \sim N\left(0, 2\lambda_{ij}^2\right).
$$

From (2.2), the joint asymptotic distribution of $(\hat{\boldsymbol{\lambda}}'_{(1)}, \dots, \hat{\boldsymbol{\lambda}}'_{(k)})'$ has covariance matrix

$$
(2.4)\qquad \frac{1}{n}\mathbf{V}_\lambda = \left[\begin{pmatrix} \frac{1}{2} n r_1 \boldsymbol{\Lambda}_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{2} n r_k \boldsymbol{\Lambda}_k^{-2} \end{pmatrix} - \mathbf{G}'\mathbf{A}^{-1}\mathbf{G} \right]^{-1}.
$$

Since, by (2.3), the diagonal elements of $\mathbf{V}_\lambda$ are $(2/r_1\lambda_{11}^2, 2/r_1\lambda_{12}^2, \dots, 2/r_k\lambda_{kp}^2)$, and $\mathbf{A}$ is p.d.s., it follows that $\mathbf{G} = \mathbf{0}$. Thus we get:

THEOREM 1. *The statistics* $\sqrt{n_i}(\hat{\lambda}_{ij} - \lambda_{ij})$ *are asymptotically* ($\min_{1 \le i \le k} n_i \to \infty$) *distributed as* $N(0, 2\lambda_{ij}^2)$, *independent of each other and independent of the* $\hat{\boldsymbol{\beta}}_j$.

The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ requires more work. First, from the log-likelihood function (2.1) it is clear that the matrix $\mathbf{A}$ can be written as the sum of k matrices $\mathbf{A}_1, \dots, \mathbf{A}_k$, where $\mathbf{A}_i$ is associated with the ith sample. Let $\mathbf{V}_i$ denote the asymptotic covariance matrix of $\sqrt{n_i}\operatorname{vec}\hat{\boldsymbol{\beta}} = \sqrt{n_i}(\hat{\boldsymbol{\beta}}'_1, \dots, \hat{\boldsymbol{\beta}}'_k)'$ as obtained from the ith sample alone. Following Anderson (1963, page 130), and writing

$$
(2.5)\qquad g_{jh}^{(i)} = \frac{1}{r_i} \frac{\lambda_{ij}\lambda_{ih}}{\left(\lambda_{ij} - \lambda_{ih}\right)^2} \qquad (h \neq j),
$$

we get

$$
(2.6)\qquad
\begin{array}{c|cccc}
 & \hat{\boldsymbol\beta}_1' & \hat{\boldsymbol\beta}_2' & & \hat{\boldsymbol\beta}_p' \\
\hline
\hat{\boldsymbol\beta}_1 & \sum_{\substack{h=1\\h\neq 1}}^{p} g_{1h}^{(i)}\boldsymbol\beta_h\boldsymbol\beta_h' & -g_{12}^{(i)}\boldsymbol\beta_2\boldsymbol\beta_1' & \cdots & -g_{1p}^{(i)}\boldsymbol\beta_p\boldsymbol\beta_1' \\
\hat{\boldsymbol\beta}_2 & -g_{21}^{(i)}\boldsymbol\beta_1\boldsymbol\beta_2' & \sum_{\substack{h=1\\h\neq 2}}^{p} g_{2h}^{(i)}\boldsymbol\beta_h\boldsymbol\beta_h' & \cdots & -g_{2p}^{(i)}\boldsymbol\beta_p\boldsymbol\beta_2' \\
\vdots & \vdots & \vdots & & \vdots \\
\hat{\boldsymbol\beta}_p & -g_{p1}^{(i)}\boldsymbol\beta_1\boldsymbol\beta_p' & -g_{p2}^{(i)}\boldsymbol\beta_2\boldsymbol\beta_p' & \cdots & \sum_{\substack{h=1\\h\neq p}}^{p} g_{ph}^{(i)}\boldsymbol\beta_h\boldsymbol\beta_h'
\end{array}
= \mathbf{V}_i.
$$

LEMMA 1. *The matrices* $\mathbf{V}_i$ $(i = 1,\dots,k)$ *as given by* (2.6) *are simultaneously diagonalizable.*

This lemma is easily proved by showing that $\mathbf{V}_i\mathbf{V}_h = \mathbf{V}_h\mathbf{V}_i$ for all pairs (i, h), using the equivalence of simultaneous diagonalizability and commutativity under multiplication.

Thus there exists an orthogonal $p^2 \times p^2$-matrix $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$, where $\mathbf{H}_1 = (\mathbf{h}_1,\dots,\mathbf{h}_s)$, $s = p(p-1)/2$, such that

$$(2.7)\qquad \mathbf{H}'\mathbf{V}_i\mathbf{H} = \begin{pmatrix} \mathbf{E}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \qquad (i = 1,\dots,k),$$

where $\mathbf{E}_i = \operatorname{diag}(e_{i1},\dots,e_{is})$, $e_{ij} > 0$. ($\mathbf{E}_i$ has rank s because there are s functionally independent elements in $\boldsymbol\beta$.) For the transformed variables $\mathbf{u} = \mathbf{H}_1' \operatorname{vec}\hat{\boldsymbol\beta}$ the information from the ith sample is therefore

$$(2.8)\qquad \mathbf{A}_i^* = n\operatorname{diag}\left(e_{i1}^{-1}, e_{i2}^{-1},\dots,e_{is}^{-1}\right)$$

with

$$e_{ij} = \mathbf{h}_j'\mathbf{V}_i\mathbf{h}_j.$$

The sum of these k information matrices is

$$(2.9)\qquad \begin{aligned} \mathbf{A}^* = \sum_{i=1}^{k}\mathbf{A}_i^* &= n\operatorname{diag}\left(\sum_{i=1}^{k} e_{i1}^{-1},\dots,\sum_{i=1}^{k} e_{is}^{-1}\right) \\ &= n\operatorname{diag}\left(e_1^{-1},\dots,e_s^{-1}\right), \end{aligned}$$

where e_j is the harmonic mean of $e_{1j},\dots,e_{kj}$. The asymptotic covariance matrix of $\mathbf{u}$ is therefore $\operatorname{diag}(e_1,\dots,e_s)/n$. Transforming back to $\hat{\boldsymbol\beta}$, we get the asymptotic covariance matrix of vec $\hat{\boldsymbol\beta}$ as

$$\frac{1}{n}\mathbf{V} = \frac{1}{n}\sum_{j=1}^{s} \mathbf{e}_j \mathbf{h}_j \mathbf{h}'_j. \tag{2.10}$$

To establish the final result, we need now the explicit form of $\mathbf{H}_1$. The $\mathbf{h}_j$ are vectors of dimension p^2. If $\boldsymbol\eta = (\boldsymbol\eta'_1, \dots, \boldsymbol\eta'_p)'$ is a p^2 vector partitioned into p vectors of dimension p, we will, for simplicity, refer to $\boldsymbol\eta_j$ as the jth position of $\boldsymbol\eta$ (which corresponds to the scalar positions $(j-1)p+1$ through jp).

LEMMA 2. *The $s = \binom{p}{2}$ normalized characteristic vectors of $\mathbf{V}_i$ associated with positive roots are as follows: For each pair of indices (j, l), with $1 \le j < l \le p$, there exists a characteristic vector having $\boldsymbol\beta_l/\sqrt{2}$ in position j and $-\boldsymbol\beta_j/\sqrt{2}$ in position l. All other positions are zero, and the associated eigenvalues are $2g_{jl}^{(i)}$.*

The proof of Lemma 2 is straightforward and need not be given. The eigenvectors defined in the lemma form the matrix $\mathbf{H}_1$. Assuming that all k matrices $\boldsymbol\Sigma_i$ have p distinct eigenvalues and that $r_i > 0$, the $g_{jl}^{(i)}$ are all positive. From (2.8) it is now seen that $e_{im} = g_{jl}^{(i)}$ for some pair (j, l). Thus, putting

$$g_{jl} = \left(\sum_{i=1}^{k} g_{jl}^{(i)-1} \right)^{-1} \tag{2.11}$$

and writing $\mathbf{h}_{jl}$ for the eigenvector associated with the roots $2g_{jl}^{(i)}$, we get the asymptotic covariance matrix of $\sqrt{n}$ vec $\hat{\boldsymbol\beta}$ as

$$\mathbf{V} = 2\sum_{j<l}^{p} g_{jl}\mathbf{h}_{jl}\mathbf{h}'_{jl}. \tag{2.12}$$

Writing this in terms of the $\boldsymbol\beta$-vectors, using Lemma 2, we get therefore:

THEOREM 2. *The asymptotic distribution of $\sqrt{n}\,\mathrm{vec}(\hat{\boldsymbol\beta} - \boldsymbol\beta)$ is normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{V}$ given by*

$$\begin{array}{c|cccc} & \boldsymbol\beta'_1 & \boldsymbol\beta'_2 & \cdots & \boldsymbol\beta'_p \\ \hline \hat{\boldsymbol\beta}_1 & \sum_{h=1, h\neq 1}^{p} g_{1h}\boldsymbol\beta_h\boldsymbol\beta'_h & -g_{12}\boldsymbol\beta_2\boldsymbol\beta'_1 & \cdots & -g_{1p}\boldsymbol\beta_p\boldsymbol\beta'_1 \\ \hat{\boldsymbol\beta}_2 & -g_{21}\boldsymbol\beta_1\boldsymbol\beta'_2 & \sum_{h=1, h=2}^{p} g_{2h}\boldsymbol\beta_h\boldsymbol\beta'_h & \cdots & -g_{2p}\boldsymbol\beta_p\boldsymbol\beta'_2 \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\boldsymbol\beta}_p & -g_{p1}\boldsymbol\beta_1\boldsymbol\beta'_p & -g_{p2}\boldsymbol\beta_2\boldsymbol\beta'_p & \cdots & \sum_{h=1, h\neq p}^{p} g_{ph}\boldsymbol\beta_h\boldsymbol\beta'_h \end{array} = \mathbf{V}, \tag{2.13}$$

where the g_{jl} are defined in (2.5), *and the* $\boldsymbol{\beta}_j$ *are the* (*common*) *eigenvectors of the k matrices* $\boldsymbol{\Sigma}_i$.

The foregoing proof of Theorem 2 is based on the assumption that all k matrices $\boldsymbol{\Sigma}_i$ have p distinct eigenvalues. However, since $g_{jl}^{(i)-1} = r_i(\lambda_{ij} - \lambda_{il})^2/\lambda_{ij}\lambda_{il}$, we can take $g_{jl}^{(i)-1} = 0$ if $\lambda_{ij} = \lambda_{il}$. In order for g_{jl} to be defined, it suffices to have *at least one* $\boldsymbol{\Sigma}_i$ with $\lambda_{ij} \neq \lambda_{il}$. It is therefore assumed that Theorem 2 holds whenever CPC's are well defined.

3. An asymptotic test for q hypothetical eigenvectors. Using the asymptotic distribution theory, Anderson (1963, Appendix B) constructs a test for the hypothesis that the jth eigenvector of $\boldsymbol{\Sigma}$ is identical with a specified vector $\boldsymbol{\beta}_j^0$ ($\boldsymbol{\beta}_j^{0\prime}\boldsymbol{\beta}_j^0 = 1$), under the assumption that this eigenvector corresponds to a root of multiplicity 1. In this section we are going to generalize Anderson's result in two ways by deriving an analogous test for q specified vectors ($1 \leq q \leq p$) and k groups. Without loss of generality we can order the CPCs such that the q vectors to be tested are labeled 1 thru q. The null hypothesis is

$$H_q\colon (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q) = (\boldsymbol{\beta}_1^0, \dots, \boldsymbol{\beta}_q^0), \tag{3.1}$$

where the $\boldsymbol{\beta}_j^0$ are specified, mutually orthogonal and have unit length.

The test of H_q will be based on the asymptotic covariance matrix of $\operatorname{vec}(\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_q)$, that is, the upper left $pq \times pq$ portion of $\mathbf{V}$. Call this submatrix $\mathbf{V}(q)$. The eigenstructure of $\mathbf{V}(q)$ is given by the following theorem, which is actually a generalization of Lemma 2. We are again using the convention that the p scalar elements in positions $(j-1)p + 1$ through jp of a vector are referred to as the jth position.

THEOREM 3. *The upper left* $pq \times pq$ *submatrix of* $\mathbf{V}$ *has the following eigenvectors and eigenvalues*:

1. $\binom{q}{2}$ *eigenvectors* (*one for each pair* j, l *with* $1 \leq j < l \leq q$) *have* $\boldsymbol{\beta}_l/\sqrt{2}$ *in position* j *and* $-\boldsymbol{\beta}_j/\sqrt{2}$ *in position* l. *All other positions are zero*, *and the associated roots are* $2g_{jl}$.
2. $(p-q)q$ *eigenvectors* (*one for each combination of indices* j, l *such that* $1 \leq j \leq q < l \leq p$) *have* $\boldsymbol{\beta}_l$ *in position* j *and* $\mathbf{0}$ *in all other positions*; *the associated roots are* g_{jl}.
3. $\binom{q}{2}$ *eigenvectors* (*one for each pair of indices* j, l *such that* $1 \leq j \leq l \leq q$) *can be chosen to have* $\boldsymbol{\beta}_l/\sqrt{2}$ *in position* j, $\boldsymbol{\beta}_j/\sqrt{2}$ *in position* l, *and zeros in all other positions. The associated roots are zero.*
4. q *eigenvectors* (*one for each* j *with* $1 \leq j \leq q$) *can be chosen to have* $\boldsymbol{\beta}_j$ *in position* j *and zeros elsewhere. The associated roots are zero.*

The proof of Theorem 3 is easy and is therefore omitted. We see that if $g_{jl} > 0$ for $1 \leq j \leq q$, $1 \leq l \leq p$, then $\mathbf{V}(q)$ has rank $t = \binom{q}{2} + q(p-q)$.

Let now $\mathbf{\Phi}$ denote a $t \times t$ diagonal matrix with diagonal elements equal to the nonzero roots of $\mathbf{V}(q)$, i.e., $\boldsymbol{\phi} = \operatorname{diag}(2g_{12}, \dots, 2g_{q-1,q}, g_{1,q+1}, \dots, g_{qp})$, and let the columns of the $pq \times t$ matrix $\mathbf{\Gamma}$ be given by the characteristic vectors associated with the nonzero roots. Putting $\mathbf{b}_q = \operatorname{vec}(\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_q) - \operatorname{vec}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q)$, the random vector $z = \sqrt{n}\,\mathbf{\Phi}^{-1/2}\mathbf{\Gamma}'\mathbf{b}_q$ has a limiting normal distribution with mean zero and covariance matrix $\mathbf{I}_t$. Thus $\mathbf{z}'\mathbf{z} = n\mathbf{b}_q'\mathbf{\Gamma}\mathbf{\Phi}^{-1}\mathbf{\Gamma}'\mathbf{b}_q$ has a limiting chi square distribution with t degrees of freedom. Using Theorem 3, this expression can be written as

$$\mathbf{z}'\mathbf{z} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q \end{pmatrix}' \mathbf{\Gamma}\mathbf{\Phi}^{-1}\mathbf{\Gamma}' \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q \end{pmatrix}$$

$$(3.2) \qquad = \tfrac{1}{4}\sum_{j=1}^{q-1}\sum_{l=j+1}^{q} g_{jl}^{-1}\left(\boldsymbol{\beta}_l'\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j'\hat{\boldsymbol{\beta}}_l\right)^2 + \sum_{j=1}^{q}\sum_{l=q+1}^{p} g_{jl}^{-1}\left(\boldsymbol{\beta}_l'\hat{\boldsymbol{\beta}}_j\right)^2.$$

In practical applications the asymptotic covariance matrix $\mathbf{V}(q)$ is not known, but can be consistently estimated by substituting the $\hat{\lambda}_{ij}$ and $\hat{\beta}_j$ for the respective parameters. This does not affect the validity of the asymptotic chi square approximation. Using $\hat{\boldsymbol{\beta}}_j$ instead of $\boldsymbol{\beta}_j$ and $\hat{g}_{il}^{-1} = \sum_{i=1}^{k} r_i(\hat{\lambda}_{ij} - \hat{\lambda}_{il})/\hat{\lambda}_{ij}\hat{\lambda}_{il}$ in the expression $\mathbf{\Gamma}\mathbf{\Phi}^{-1}\mathbf{\Gamma}'$ of (3.2) we get therefore:

THEOREM 4. *Under H_q as defined in* (3.1), *the statistic*

$$(3.3) \qquad X^2(H_q) = n\left[\tfrac{1}{4}\sum_{j=1}^{q-1}\sum_{l=j+1}^{q} \hat{g}_{jl}^{-1}\left(\hat{\boldsymbol{\beta}}_l'\boldsymbol{\beta}_j^0 - \hat{\boldsymbol{\beta}}_j'\boldsymbol{\beta}_l^0\right)^2 + \sum_{j=1}^{q}\sum_{l=q+1}^{p} \hat{g}_{jl}^{-1}\left(\hat{\boldsymbol{\beta}}_l'\boldsymbol{\beta}_j^0\right)^2\right]$$

is asymptotically distributed as chi square with $q(p - (q+1)/2)$ *degrees of freedom.*

It is worth noting that (3.3) has a geometrical interpretation. The first sum ranges over all pairs of eigenvectors fixed under H_q. If H_q holds, we would expect $\hat{\boldsymbol{\beta}}_l$ and $\boldsymbol{\beta}_j^0$ to be nearly orthogonal for $l \neq j$, and the cosines $\hat{\boldsymbol{\beta}}_l'\boldsymbol{\beta}_j^0$ would be expected close to zero. Similarly, the second sum extends over the squared cosines between the hypothetical vectors $\boldsymbol{\beta}_j^0$ and the observed vectors $\hat{\boldsymbol{\beta}}_l$ for the $p - q$ eigenvectors not considered under H_q. The $\hat{g}_{jl}^{-1}$ serve as weights—which seems intuitively reasonable, regarding the fact that $\hat{g}_{jl}^{-1}$ is large if (at least in one group) $\hat{\lambda}_{ij}$ and $\hat{\lambda}_{il}$ are far apart. It may also be noted that Theorem 4 does not require that the eigenvectors $\boldsymbol\beta_{q+1},\dots,\boldsymbol\beta_p$ are well defined.

Two special cases deserve more attention.

CASE $q = 1$. If only one hypothetical eigenvector $\boldsymbol\beta_1^0$ is specified, the first sum in (3.3) is empty, and the test statistic becomes

$$
\begin{aligned}
X^2(H_1) &= n\sum_{l=2}^{p}\hat g_{1l}^{-1}\left(\hat{\boldsymbol\beta}_l'\boldsymbol\beta_1^0\right)^2\\
&= n\boldsymbol\beta_1^{0\prime}\left[\sum_{i=1}^{k} r_i\sum_{l=2}^{p}\left(\frac{\hat\lambda_{i1}}{\hat\lambda_{il}}+\frac{\hat\lambda_{il}}{\hat\lambda_{i1}}-2\right)\hat{\boldsymbol\beta}_l\hat{\boldsymbol\beta}_l'\right]\boldsymbol\beta_1^0\\
&= n\boldsymbol\beta_1^{0\prime}\left[\sum_{j=1}^{k} r_i\left(\hat\lambda_{i1}\left(\hat{\boldsymbol\Sigma}_i^{-1}-\hat\lambda_{i1}^{-1}\hat{\boldsymbol\beta}_1\hat{\boldsymbol\beta}_1'\right)\right.\right.\\
&\qquad\left.\left.+\hat\lambda_{i1}^{-1}\left(\hat{\boldsymbol\Sigma}_i-\hat\lambda_{i1}\hat{\boldsymbol\beta}_1\hat{\boldsymbol\beta}_1'\right)-2\left(\mathbf I_p-\hat{\boldsymbol\beta}_1\hat{\boldsymbol\beta}_1'\right)\right)\right]\boldsymbol\beta_1^0\\
&= n\left[\sum_{i=1}^{k} r_i\left(\hat\lambda_{i1}\boldsymbol\beta_1^{0\prime}\hat{\boldsymbol\Sigma}_i^{-1}\boldsymbol\beta_1^0+\hat\lambda_{i1}^{-1}\boldsymbol\beta_1^{0\prime}\hat{\boldsymbol\Sigma}_i\boldsymbol\beta_1^0-2\right)\right],
\end{aligned}
\tag{3.4}
$$

where $\hat{\boldsymbol\Sigma}_i=\sum_{j=1}^{p}\hat\lambda_{ij}\hat{\boldsymbol\beta}_j\hat{\boldsymbol\beta}_j'$ is the ML estimate of $\boldsymbol\Sigma_i$. The number of degrees of freedom associated with (3.4) is $p-1$. If $k=1$, then (3.4) reduces to the well-known result given by Anderson (1963, page 145). If we replace $\hat{\boldsymbol\Sigma}_i$ by $\mathbf S_i$ and $\hat\lambda_{i1}$ by l_{i1} (the first eigenvalue of $\mathbf S_i$), we get a test for the hypothesis H_1^* that $\boldsymbol\beta_1^0$ is the first principal component of $\boldsymbol\Sigma_1,\dots,\boldsymbol\Sigma_k$ without specifying the CPC model. The test statistic

$$
X^2(H_1^*)=\sum_{i=1}^{k} n_i\left(l_{i1}\boldsymbol\beta_1^{0\prime}\mathbf S_i^{-1}\boldsymbol\beta_1^0+l_{i1}^{-1}\boldsymbol\beta_1^{0\prime}\mathbf S_i\boldsymbol\beta_1^0-2\right)
\tag{3.5}
$$

is merely the sum of k independent statistics of the form given by Anderson, and its asymptotic null distribution is chi square with $k(p-1)$ degrees of freedom.

CASE $q = p$. If all common eigenvectors $\boldsymbol\beta_1^0,\dots,\boldsymbol\beta_1^p$ of the $\boldsymbol\Sigma_i$ are specified, the second sum in (3.3) is empty. Since $\boldsymbol\beta_p^0$ is completely determined by $\boldsymbol\beta_1^0,\dots,\boldsymbol\beta_{p-1}^0$, the hypotheses H_{p-1} and H_p are equivalent. The two associated statistics $X^2(H_{p-1})$ and $X^2(H_p)$ are in general not identical (unless $p=2$), but the degrees of freedom are $p(p-1)/2$ for both statistics.

4. Asymptotic inference for eigenvalues.

4.1. *A criterion for neglecting common principal components with relatively small variances.* Anderson (1963, page 133ff.) has shown how asymptotic confidence intervals for individual roots or sums of roots in one-sample PCA can be constructed. Since the maximum likelihood estimates $\hat\lambda_{ij}$ in CPCA are asymptotically independent (cf. Theorem 1), the generalization of Anderson's results to the k-sample case is straightforward and need not be given here.

If the main purpose of CPCA is data reduction, it is useful to have some criterion for discarding CPC's with relatively small variances. Let

$$c_i = \sum_{j=1}^{q} \lambda_{ij}, \qquad \hat{c}_i = \sum_{j=1}^{q} \hat{\lambda}_{ij}, \tag{4.1}$$

and put $d_i = \operatorname{tr} \mathbf{\Sigma}_i - c_i$, $\hat{d}_i = \operatorname{tr} \hat{\mathbf{\Sigma}}_i - \hat{c}_i$. Suppose that we wish to discard the last $p - q$ CPC's in population i if their relative contribution to the trace of $\mathbf{\Sigma}_i$ is not larger than a given fraction f_0 $(0 < f_0 < 1)$. Putting

$$f_i = d_i / \operatorname{tr} \mathbf{\Sigma}_i, \tag{4.2}$$

the asymptotic distribution of $\sqrt{n_i}[(1 - f_i)\hat{d}_i - f_i\hat{c}_i]$ is normal with mean zero and variance $2[f_i^2\Sigma_{j=1}^q\lambda_{ij}^2 + (1 - f_i)^2\Sigma_{j=q+1}^p\lambda_{ij}^2]$. (The use of this criterion has been proposed by Anderson (1963, page 135).) Estimating this variance consistently by putting in the corresponding ML estimates $\hat{\lambda}_{ij}$ yields

$$z_i = \frac{\sqrt{n_i}\,[(1 - f_0)\hat{d}_i - f_0\hat{c}_i]}{\left(2\left[f_0^2\Sigma_{j=1}^q\hat{\lambda}_{ij}^2 + (1 - f_0)^2\Sigma_{j=q+1}^p\hat{\lambda}_{ij}^2\right]\right)^{1/2}} \sim N(0,1) \tag{4.3}$$

approximately for large n_i and under the hypothesis $f_i = f_0$. For testing the hypothesis that *all* f_i $(i = 1, \dots, k)$ are less than or equal to f_0, a possible procedure is to reject the hypothesis if

$$\max_{1 \le i \le k} z_i > z_\beta \quad \text{with } \beta = 1 - (1 - \alpha)^{1/k}, \tag{4.4}$$

where z_β is the upper β quantile of the standard normal distribution. This test has asymptotic level α if all f_i equal f_0.

4.2. A likelihood ratio test for sphericity of $p - q$ common principal components. In PCA, the main motivation for testing for equality of $p - q$ (out of p) characteristic roots stems from the model $\mathbf{\Sigma} = \psi + \sigma^2\mathbf{I}_p$, where ψ is positive semidefinite of rank q. In this model the last $p - q$ characteristic roots are all σ^2. In CPCA, we can study the model $\mathbf{\Sigma}_i = \psi_i + \sigma_i^2\mathbf{I}_p$ $(i = 1, \dots, k)$, where the ψ_i are simultaneously diagonalizable and of rank q. Then the $\mathbf{\Sigma}_i$ satisfy H_c, and the last $p - q$ CPC's are spherical, i.e.,

$$H_S\colon \lambda_{i,q+1} = \cdots = \lambda_{ip} \qquad (i = 1, \dots, k). \tag{4.5}$$

We will refer to H_S as "hypothesis of partial sphericity."

It may be noted that the following derivation of the likelihood ratio test holds as well for any subset of CPC's, but for notational simplicity it is given in terms of the hypothesis H_S as defined in (4.5).

Putting $\lambda_{i,q+1} = \cdots = \lambda_{ip} = \lambda_i^*$ ($i = 1, \dots, k$), we get from (2.1)

$$\begin{aligned} &-2g(\Lambda_1, \dots, \Lambda_k, \beta | \mathbf{S}_1, \dots, \mathbf{S}_k) \\ (4.6)\qquad &= \sum_{i=1}^{k} n_i \left[\sum_{j=1}^{q} \left(\log \lambda_{ij} + \beta_j' \mathbf{S}_i \beta_j / \lambda_{ij} \right) \right. \\ &\qquad \left. + (p-q) \log \lambda_i^* + \left(\sum_{j=q+1}^{p} \beta_j' \mathbf{S}_i \beta_j \right) \Big/ \lambda_i^* \right]. \end{aligned}$$

Using the same technique as Flury (1984) the likelihood equations are obtained as

$$\begin{aligned} &\beta_l' \left(\sum_{i=1}^{k} n_i \frac{\lambda_{il} - \lambda_{ij}}{\lambda_{il}\lambda_{ij}} \mathbf{S}_i \right) \beta_j = 0 \qquad (1 \le l < j \le q), \\ &\beta_l' \left(\sum_{i=1}^{k} n_i \frac{\lambda_{il} - \lambda_i^*}{\lambda_{il}\lambda_i^*} \mathbf{S}_i \right) \beta_j = 0 \qquad (1 \le l \le q < j \le p), \\ (4.7)\qquad &\lambda_{ij} = \beta_j' \mathbf{S}_i \beta_j \qquad (i = 1, \dots, k;\ j = 1, \dots, q), \\ &\lambda_i^* = \left(\sum_{j=q+1}^{p} \beta_j' \mathbf{S}_i \beta_j \right) \Big/ (p-q) \qquad (i = 1, \dots, k), \end{aligned}$$

with the orthogonality restrictions $\beta_l' \beta_j = 0$ ($l \neq j$), $\beta_j' \beta_j = 1$. The equation system (4.7) can be solved using an appropriate modification of the FG algorithm (Flury and Gautschi, 1986).

In contrast to the unrestricted CPC model the vectors $\beta_{q+1}, \dots, \beta_p$ are not uniquely determined by the likelihood equations. In fact, only the subspace spanned by $\beta_{q+1}, \dots, \beta_p$ is determined. Let us denote by $\tilde{\beta}_1, \dots, \tilde{\beta}_q, \tilde{\beta}_{q+1}, \dots, \tilde{\beta}_p$ a set of orthonormal vectors solving (4.7); then the same maximum of the likelihood is obtained if we replace $(\tilde{\beta}_{q+1}, \dots, \tilde{\beta}_p)$ by $(\tilde{\beta}_{q+1}, \dots, \tilde{\beta}_p)\mathbf{H}$, where $\mathbf{H}$ is an arbitrary orthogonal matrix of dimension $(p-q) \times (p-q)$. With $\tilde{\lambda}_{ij}$ and $\tilde{\lambda}_i^*$ denoting the ML estimates of the eigenvalues, the log-likelihood ratio statistic for H_S can be written as

$$(4.8)\qquad X_S^2 = \sum_{i=1}^{k} n_i \log \frac{(\tilde{\lambda}_i^*)^{p-q} \prod_{j=1}^{q} \tilde{\lambda}_{ij}}{\prod_{j=1}^{p} \hat{\lambda}_{ij}},$$

where the $\hat{\lambda}_{ij}$ are the ML estimates for the ordinary (unrestricted) CPC model. Under H_S, the number of parameters determining $\Sigma_1, \dots, \Sigma_k$ is $q(2p - q - 1)/2 + k(q+1)$, compared with $p(p-1)/2 + kp$ parameters for the ordinary CPC model (see, e.g., Mardia, Kent, and Bibby (1979, page 235ff.) for a discussion of this problem in the one-group situation). Thus the null distribution of (4.8) is asymptotically chi square with $(p - q - 1)(p - q + 2k)/2$ degrees of freedom.

It may be noted that, unless $k = 1$, the ML estimates $\tilde{\beta}_j$ and $\tilde{\lambda}_{ij}$ for $j \le q$ are not identical with $\hat{\beta}_j$ and $\hat{\lambda}_{ij}$. However, we can approximate X_S^2 without computing the restricted solution by replacing $\tilde{\lambda}_{ij}$ by $\hat{\lambda}_{ij}$ ($j = 1, \dots, q$) and $\tilde{\lambda}_i^*$

by $\hat{\lambda}_i^* = (\hat{\lambda}_{i,q+1} + \cdots + \hat{\lambda}_{ip})/(p-q)$. This yields

$$X_S^2(\text{approx}) = \sum_{i=1}^{k} n_i \log \frac{(\hat{\lambda}_i^*)^{p-q}}{\prod_{j=q+1}^{p} \hat{\lambda}_{ij}}. \tag{4.9}$$

Since, under H_S, the likelihood is maximized for the $\tilde{\lambda}$'s, we have always $X_S^2(\text{approx}) \geq X_S^2$. Thus the approximate statistic can be used to accept H_S, but not necessarily to reject it.

It should be noted that H_S does not necessarily imply a model of the form $\mathbf{\Sigma}_i = \psi_i + \sigma_i^2 \mathbf{I}_p$, and partial sphericity may also occur for those CPC's associated with large roots. In practical applications of CPCA it is important to make sure that those components that are to be interpreted are not spherical, since a coefficient should be interpreted only if it is well-defined.

5. Applications. In this section some of the preceding theory is illustrated by a numerical example. The data used have been published by Jolicoeur and Mosimann (1960) and have served as an example of PCA in various textbooks (e.g., Morrison, 1976; Mardia, Kent, and Bibby, 1979). The main appeal of this example is its simplicity—the data are only three-dimensional, yet illustrate the purpose of CPCA clearly, which outweighs the disadvantage of rather small sample sizes.

Table 1(a) displays covariance matrices $\mathbf{S}_i$ of samples of $n_1 + 1 = 24$ male and $n_2 + 1 = 24$ female individuals of the species *Chrysemys picta marginata* (painted turtle). The variables are (1) log(carapace length); (2) log(carapace width); (3) log(carapace height). The logarithms are used instead of the measured variables because of their relationship to allometry; see Morrison (1976, page 295). Table 1(b) shows the eigenvalues of the $\mathbf{S}_i$ and the ML estimates $\hat{\lambda}_{ij}$. The value of the chi square statistic for H_C (Flury, 1984) is $X^2 = 7.93$ with three degrees of freedom, which is close to the 95% quantile of the asymptotic null distribution of the criterion. Regarding the relatively small sample sizes it may be reasonable to assume that H_C holds.

Table 1(c) shows the estimated eigenvectors $\hat{\boldsymbol{\beta}}_j$ and the estimated asymptotic standard errors of their coefficients. The standard errors were obtained from the main diagonal of the sample analog of (2.13). It is obvious that $\hat{\boldsymbol{\beta}}_1$ has stable coefficients, while $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\beta}}_3$ seem rather poorly defined.

The hypothesis of allometric growth of an organism (Jolicoeur, 1963) implies that the first principal component of the covariance matrix of the logarithms of the measured dimensions is $\boldsymbol{\beta}_1' = (1, \ldots, 1)/\sqrt{p}$. Let us therefore test the hypothesis H_1 (3.1) for $\boldsymbol{\beta}_1^0 = (1,1,1)'/\sqrt{3}$. The statistic $X^2(H_1)$ is obtained from (3.3) or (3.4) as $X^2(H_1) = 46.17$ with two degrees of freedom. At any reasonable level α we would therefore conclude that the allometric model does not hold in this case.

Finally, let us see whether the second and third CPC's are well defined, i.e., let us test the hypothesis of simultaneous sphericity of the second and third CPC's. The null hypothesis is H_S: $\lambda_{i2} = \lambda_{i3}$ $(i = 1, 2)$. Without computing the ML estimates under H_S, we can easily calculate the approximation (4.9) from the

TABLE 1
Common principal component analysis of turtle carapace dimensions, transformed logarithmically.

(a) Sample covariance matrices[a]

males ($n_1 = 23$)

$$\mathbf{S}_1 = \begin{pmatrix} 1.1072 & 0.8019 & 0.8160 \\ 0.8019 & 0.6417 & 0.6005 \\ 0.8160 & 0.6005 & 0.6773 \end{pmatrix}$$

females ($n_2 = 23$)

$$\mathbf{S}_2 = \begin{pmatrix} 2.6391 & 2.0124 & 2.5443 \\ 2.0124 & 1.6190 & 1.9782 \\ 2.5443 & 1.9782 & 2.5899 \end{pmatrix}$$

(b) Variances of CPC's and eigenvalues of $\mathbf{S}_i$

males	$\hat{\lambda}_{1j}$	2.3148	0.0729	0.0385
	eigenvalues	2.3303	0.0599	0.0360
females	$\hat{\lambda}_{2j}$	6.7135	0.0807	0.0538
	eigenvalues	6.7200	0.0751	0.0530

(c) Coefficients of CPC's[b]

$$\hat{\boldsymbol{\beta}}_1 = \begin{pmatrix} 0.6406 \\ 0.4905 \\ 0.5907 \end{pmatrix} \begin{matrix} (0.013) \\ (0.015) \\ (0.016) \end{matrix} \qquad \hat{\boldsymbol{\beta}}_2 = \begin{pmatrix} -0.3839 \\ -0.4617 \\ -0.7997 \end{pmatrix} \begin{matrix} (0.182) \\ (0.201) \\ (0.032) \end{matrix} \qquad \hat{\boldsymbol{\beta}}_3 = \begin{pmatrix} -0.6650 \\ 0.7391 \\ 0.1075 \end{pmatrix} \begin{matrix} (0.105) \\ (0.126) \\ (0.218) \end{matrix}$$

[a] Multiplied by 10^2.
[b] Standard errors, given in parentheses, are based on large sample theory.

values displayed in Table 1(b). The resulting statistic is $X_S^2(\text{approx}) = 3.24$. Since this is far below the 95% quantile of the chi square distribution with three degrees of freedom, we conclude that H_S is reasonable. Taking into consideration the relative smallness of the second and third roots in both groups, we can thus think of the three shell dimensions as distributed about a single principal axis ("size") and two minor axes containing merely measurement errors, the main axis having the same orientation in space for both male and female turtles.

6. Conclusions. In this paper we have shown how asymptotic theory can be used for inference on CPC models. The methods given in Sections 3 and 4 merely reflect the author's opinion about which hypotheses might be important in practice. Other hypotheses and restrictions of the model can easily be formulated; we might for instance be interested in a model where some of the eigenvalues of two matrices $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_h$ are identical. Tests for such hypotheses could be constructed either by the likelihood ratio method or using the asymptotic results of Section 2.

One open problem deserves to be investigated: Suppose that we are interested only in the first q (out of p) CPC's and wish to neglect the last $p - q$ components. Then we would actually not care whether the $\boldsymbol{\Sigma}_i$ have *all* eigenvectors in common—it would be sufficient to know that $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$ are common to $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$. This could be called a *partial* CPC model.

Obviously a partial CPC model may hold even when the ordinary CPC model is wrong, and the test for H_C may in some situations reject the hypothesis although the first q eigenvectors are common to all matrices $\Sigma_1, \ldots, \Sigma_k$. This problem is currently under investigation.

Acknowledgments. I wish to thank Professors Ingram Olkin and T. W. Anderson for some helpful discussions. The comments of an associate editor and two referees helped considerably to improve the presentation of this article.

REFERENCES

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148.

FLURY, B. N. (1984). Common principal components in k groups. *J. Amer. Statist. Assoc.* **79** 892–898.

FLURY, B. N. and GAUTSCHI, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite matrices to nearly diagonal form. *SIAM J. Sci. Statist. Comput.* **6**.

GIRSHICK, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Statist.* **10** 203–224.

JOLICOEUR, P. (1963). The multivariate generalization of the allometry equation. *Biometrics* **19** 497–499.

JOLICOEUR, P. and MOSIMANN, J. E. (1960). Size and shape variation in the painted turtle: A principal component analysis. *Growth* **24** 339–354.

LAWLEY, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. *Uppsala Symposium on Psychological Factor Analysis*, 35–42. Almqvist and Wicksell, Uppsala.

LAWLEY, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43** 128–136.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic, New York.

MORRISON, D. F. (1976). *Multivariate Statistical Methods*. McGraw-Hill, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

SILVEY, S. D. (1975). *Statistical Inference*. Chapman and Hall, London.

WILKS, S. S. (1944). *Mathematical Statistics*. Princeton Univ. Press.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BERNE
SIDLERSTRASSE 5
CH 3012 BERNE
SWITZERLAND