

## Relative Eigenvalues and Eigenvectors

Relative eigenvalues and eigenvectors generalize ordinary eigenvalues and eigenvectors. They are quite important in multivariate analysis.

The terminology “relative eigenvalues” and “relative eigenvector” is not common. Sometimes they are called *generalized* eigenvalues and eigenvectors.

**Definition of ordinary eigenvectors and eigenvalues**

Suppose  $\mathbf{A} = \mathbf{A}'$  is a  $p$  by  $p$  symmetric matrix. If  $\mathbf{u} \neq \mathbf{0}$  is a  $p$  by 1 vector and  $\lambda$  a scalar such that

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u},$$

then  $\mathbf{u}$  is an (ordinary) *eigenvector* of  $\mathbf{A}$  with *eigenvalue*  $\lambda$ .

In MacAnova, you compute the eigenvalues and eigenvectors of a *symmetric* matrix  $\mathbf{a}$  by `eigen(a)`. This returns a structure with two components, `values`, a vector of eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , and `vectors`, a matrix whose columns are the eigenvectors  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ . `eigenvals(a)` computes just the eigenvalues.

**Caution** You *cannot* use `eigen(a)` or `eigenvals(a)` when  $\mathbf{a}$  is *not* symmetric.

A  $p$  by  $p$  symmetric matrix  $\mathbf{B}$  is *positive definite* if and only if  $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$ , for all  $p$  by 1 vectors  $\mathbf{x} \neq \mathbf{0}$ . A positive definite matrix  $\mathbf{B}$  is always invertible, that is  $\mathbf{B}^{-1}$  exists.  $\mathbf{B}$  is positive definite if and only if all of its eigenvalues are positive, that is  $\lambda_p = \lambda_{\min} > 0$ .

**Definition of relative eigenvectors and eigenvalues**

Now suppose  $\mathbf{A} = \mathbf{A}'$  and  $\mathbf{B} = \mathbf{B}'$  are *two*  $p$  by  $p$  symmetric matrices with  $\mathbf{B}$  positive definite and hence non-singular (invertible). It's perfectly OK if  $\mathbf{A}$  is *not* invertible.

Then if  $\mathbf{u}$  is a  $p$  by 1 vector and  $\lambda$  a scalar such that

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{B}\mathbf{u}, \tag{1}$$

then  $\mathbf{u}$  is an *eigenvector of A relative to B* (a *relative eigenvector*) with *relative eigenvalue*  $\lambda$ .

Because  $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_p$ , when you multiply both sides of (1) on the left by  $\mathbf{B}^{-1}$ , you get

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{u} = \lambda\mathbf{I}_p\mathbf{u} = \lambda\mathbf{u}. \tag{2}$$

This shows that  $\mathbf{u}$  is an *ordinary* eigenvector of the *non-symmetric* matrix  $\mathbf{B}^{-1}\mathbf{A}$  with *ordinary* eigenvalue  $\lambda$ .

Let  $\mathbf{v} \equiv \mathbf{B}\mathbf{u}$ , so that  $\mathbf{u} = \mathbf{B}^{-1}\mathbf{v}$ . Then you can rewrite eq. (1) as  $\mathbf{A}\mathbf{B}^{-1}\mathbf{v} = \lambda\mathbf{B}\mathbf{B}^{-1}\mathbf{v}$ . But  $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}_p$ , and hence  $\mathbf{A}\mathbf{B}^{-1}\mathbf{v} = \lambda\mathbf{v}$ , so  $\lambda$  is also an eigenvalue of the *non-symmetric* matrix  $\mathbf{A}\mathbf{B}^{-1}$  with eigenvector  $\mathbf{v} = \mathbf{B}\mathbf{u} \neq \mathbf{u}$ .

## Relative Eigenvectors and Eigenvalues

Although  $\mathbf{u}$  and  $\lambda$  are (ordinary) eigenvector and eigenvalue of the non-symmetric matrix  $\mathbf{B}^{-1}\mathbf{A}$ , you *cannot* compute them in MacAnova by `eigen(solve(b) %*% a)`, because `eigen()` requires that its argument is *symmetric*. The special algorithm provided by MacAnova function `releigenvals()` and `releigen()` is essential

Like `eigen(a)`, `releigen(a,b)` returns a structure with components `values`, a vector of relative eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  in decreasing order, and `vectors`, a matrix whose columns are the relative eigenvectors  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ . `releigen(a,b)` returns the vector of relative eigenvalues in decreasing order, that is, it's the same as `releigen(a,b)$values`.

When  $\mathbf{B} = \mathbf{I}_p$  is the identity matrix, the eigenvectors and eigenvalues of  $\mathbf{A}$  relative to  $\mathbf{B}$  are ordinary eigenvectors and eigenvalues of  $\mathbf{A}$ . Thus most of the properties of the ordinary eigenvectors and eigenvalues of a symmetric matrix are special cases of properties of relative eigenvalues and eigenvectors. You can obtain the corresponding properties by substituting  $\mathbf{I}_p$  for  $\mathbf{B}$  in most of the equations.

### Properties

- All eigenvalues  $\lambda_i$  of  $\mathbf{A}$  relative to  $\mathbf{B}$  are real (as opposed to imaginary or complex) numbers and the eigenvectors  $\mathbf{u}_i$  have real components.
- $\mathbf{A}$  is positive definite if and only if all  $\lambda_i > 0$ , that is  $\lambda_p = \lambda_{\min} > 0$ .
- There are always  $p$  linearly independent eigenvectors  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$  of  $\mathbf{A}$  relative to  $\mathbf{B}$  with relative eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  (the ordering we always assume). When the  $\lambda_i$ 's are all different ( $\lambda_i \neq \lambda_j$ , all  $i \neq j$ ), the  $\mathbf{u}_i$ 's are unique up to multiplication by scalars. In addition, when  $\lambda_i \neq \lambda_j$ ,

$$\mathbf{u}_i' \mathbf{B} \mathbf{u}_j = \mathbf{u}_j' \mathbf{B} \mathbf{u}_i = 0, \quad (3)$$

that is  $\mathbf{u}_i$  and  $\mathbf{u}_j$  are *orthogonal relative to  $\mathbf{B}$* .

- When  $\lambda_i = \lambda_j$  for some  $i \neq j$ , the eigenvectors are not unique but they may still always be computed so that (3) is true.
- By multiplying eigenvalues by a suitable scalar ( $1/\sqrt{\mathbf{u}_i' \mathbf{B} \mathbf{u}_i}$ ), you can always find  $\mathbf{u}_i$  such that

$$\mathbf{u}_i' \mathbf{B} \mathbf{u}_i = 1. \quad (4)$$

We will always assume (3) and (4) hold, even when there are equal  $\lambda$ 's.

When the  $\mathbf{u}$ 's satisfy (3) and (4), multiplying  $\mathbf{A} \mathbf{u}_j = \lambda_j \mathbf{B} \mathbf{u}_j$  on the left by  $\mathbf{u}_i'$ , you have, from (3) and (4),

$$\mathbf{u}_i' \mathbf{A} \mathbf{u}_j = \mathbf{u}_j' \mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_j' \mathbf{B} \mathbf{u}_i = \begin{cases} 0, & i \neq j \\ \lambda_i, & i = j \end{cases}, \quad (5)$$

You can summarize (1), (3), (4), and (5) in the matrix identities

$$\mathbf{A} \mathbf{U} = \mathbf{B} \mathbf{U} \mathbf{\Lambda}, \text{ with } \mathbf{U} \equiv [\mathbf{u}_1, \dots, \mathbf{u}_p] \text{ and } \mathbf{\Lambda} \equiv \text{diag}[\lambda_1, \dots, \lambda_p], \quad (6)$$

$$\mathbf{U}' \mathbf{B} \mathbf{U} = \mathbf{I}_p \quad (7)$$

and, from eq. (5)

$$\mathbf{U}'\mathbf{A}\mathbf{U} = \mathbf{\Lambda} \quad (8)$$

Eq. (6) and (7) imply that

$$\mathbf{U}^{-1} = \mathbf{U}'\mathbf{B} \quad (9)$$

and

$$\mathbf{A} = \mathbf{B}\mathbf{U}\mathbf{\Lambda}\mathbf{U}'\mathbf{B} = \sum_{1 \leq i \leq p} \lambda_i \mathbf{v}_i \mathbf{v}_i', \text{ where } \mathbf{v}_i \equiv \mathbf{B}\mathbf{u}_i. \quad (10)$$

$$\mathbf{B} = \mathbf{B}\mathbf{U}\mathbf{U}'\mathbf{B} = \sum_{1 \leq i \leq p} \mathbf{v}_i \mathbf{v}_i'$$

Eq. (10) expresses  $\mathbf{A}$  as a sum of *outer products*  $\mathbf{v}_i \mathbf{v}_i'$  weighted by  $\lambda_i$ . By the definition of the rank of a matrix, this implies that  $\text{rank}(\mathbf{A}) = s$  if and only if there are exactly  $s$  non-zero eigenvalues of  $\mathbf{A}$  relative to  $\mathbf{B}$ . In most statistical applications,  $\mathbf{A}$  is non-negative definite so that  $\lambda_i \geq 0$ ,  $i = 1, \dots, p$ . This means that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s \geq \lambda_{s+1} = \lambda_{s+2} = \dots = \lambda_p = 0$ , that is the last  $p - s$  relative eigenvalues are zero.

Eq. (11) expresses  $\mathbf{B}$  as the sum of the same outer products  $\mathbf{v}_i \mathbf{v}_i'$  equally weighted by 1 instead of by  $\lambda_i$ .

### Corresponding properties of ordinary eigenvalues and vectors

Ordinary eigenvalues and eigenvectors of a symmetric matrix  $\mathbf{A}$  satisfy  $\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ ,  $i = 1, \dots, p$  with  $\mathbf{u}_i' \mathbf{u}_i = \|\mathbf{u}_i\|^2 = 1$ ,  $\mathbf{u}_i' \mathbf{u}_j = 0$ ,  $i \neq j$ . It is helpful to compare (6) to (10) with the corresponding identities for ordinary eigenvalues and eigenvectors.

$$\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}, \quad \mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p], \quad \mathbf{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_p], \quad (6')$$

$$\mathbf{U}'\mathbf{U} = \mathbf{I}_p, \quad \text{that is } \mathbf{U} \text{ is an } \textit{orthogonal} \text{ matrix,} \quad (7')$$

$$\mathbf{U}'\mathbf{A}\mathbf{U} = \mathbf{\Lambda} \quad (8')$$

$$\mathbf{U}^{-1} = \mathbf{U}', \quad \text{another way of expressing (7'),} \quad (9')$$

and

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' = \sum_{1 \leq i \leq p} \lambda_i \mathbf{u}_i \mathbf{u}_i' \quad (10')$$

$$\mathbf{I}_p = \mathbf{U}\mathbf{U}' = \sum_{1 \leq i \leq p} \mathbf{u}_i \mathbf{u}_i' \quad (11')$$

### Other properties and identities

Because the eigenvalues of  $\mathbf{A}$  relative to  $\mathbf{B}$  are ordinary eigenvalues and eigenvectors of  $\mathbf{B}^{-1}\mathbf{A}$ , you also have (recall  $\text{tr}(\mathbf{C}) \equiv \text{trace}(\mathbf{C}) \equiv \sum_i c_{ii}$ )

$$\text{tr}\{\mathbf{B}^{-1}\mathbf{A}\} = \text{tr}\{\mathbf{A}\mathbf{B}^{-1}\} = \sum_{1 \leq i \leq p} \lambda_i = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (12)$$

and

$$\det\{\mathbf{B}^{-1}\mathbf{A}\} = \det\{\mathbf{A}\mathbf{B}^{-1}\} = \prod_{1 \leq i \leq p} \lambda_i \equiv \lambda_1 \lambda_2 \dots \lambda_p. \quad (13)$$

The following facts are sometimes useful.

Suppose  $\mathbf{u}$  is an eigenvector of  $\mathbf{A}$  relative to  $\mathbf{B}$  with relative eigenvalue  $\lambda$ . Then  $\lambda/(1+\lambda)$  is an eigenvalue of  $\mathbf{A}$  relative to  $\mathbf{A} + \mathbf{B}$  and  $1 + \lambda$  is an eigenvalue of  $\mathbf{A} + \mathbf{B}$  relative to  $\mathbf{B}$ , both with the same corresponding relative eigenvector  $\mathbf{u}$ . That is

$$\mathbf{A}\mathbf{u} = \{\lambda/(1+\lambda)\}(\mathbf{A} + \mathbf{B})\mathbf{u} \quad (14)$$

and

$$(\mathbf{A} + \mathbf{B})\mathbf{u} = (1 + \lambda)\mathbf{B}\mathbf{u} \quad (15)$$

However, if  $\mathbf{u}'\mathbf{B}\mathbf{u} = 1$ , then  $\mathbf{u}'(\mathbf{A} + \mathbf{B})\mathbf{u} = 1 + \lambda \neq 1$ , so that  $\mathbf{u}$  does not satisfy the usual normalizing condition (4) relative to  $\mathbf{A} + \mathbf{B}$ .

By applying eq. (12) to eq. (14) and eq. (13) to eq. (15), you obtain

$$\text{tr}\{(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}\} = \sum_{1 \leq i \leq p} \{\lambda_i/(1 + \lambda_i)\} \quad (16)$$

and

$$\det\{\mathbf{B}^{-1}(\mathbf{A} + \mathbf{B})\} = \prod_{1 \leq i \leq p} (1 + \lambda_i) = (1 + \lambda_1)(1 + \lambda_2)\dots(1 + \lambda_p). \quad (17)$$

$$\log(\det\{\mathbf{B}^{-1}(\mathbf{A} + \mathbf{B})\}) = \sum_{1 \leq i \leq p} \log(1 + \lambda_i) = \log(1 + \lambda_1) + \log(1 + \lambda_2) + \dots + \log(1 + \lambda_p) \quad (18)$$

### Extremal properties of relative eigenvalues and eigenvectors

The importance of relative eigenvalues and eigenvectors in multivariate analysis stems from the following facts:

$$\lambda_1 = \lambda_{\max} = \mathbf{u}_1'\mathbf{A}\mathbf{u}_1/\mathbf{u}_1'\mathbf{B}\mathbf{u}_1 = \max_{\mathbf{u}}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u}) \quad (19)$$

$$\lambda_2 = \mathbf{u}_2'\mathbf{A}\mathbf{u}_2/\mathbf{u}_2'\mathbf{B}\mathbf{u}_2 = \max_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_1=0}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u})$$

$$\lambda_3 = \mathbf{u}_3'\mathbf{A}\mathbf{u}_3/\mathbf{u}_3'\mathbf{B}\mathbf{u}_3 = \max_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_1=0, \mathbf{u}'\mathbf{B}\mathbf{u}_2=0}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u}), \text{ etc.}$$

$\max_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_1=0}(\dots)$  means the maximum over all choices of  $\mathbf{u}$  that satisfy  $\mathbf{u}'\mathbf{B}\mathbf{u}_1 = 0$ , that is all  $\mathbf{u}$  that are orthogonal to  $\mathbf{B}\mathbf{u}_1$  and  $\max_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_1=0, \mathbf{u}'\mathbf{B}\mathbf{u}_2=0}(\dots)$  means the maximum over all choices of  $\mathbf{u}$  that are orthogonal to both  $\mathbf{B}\mathbf{u}_1$  and  $\mathbf{B}\mathbf{u}_2$ .

Less important in statistics are minimization properties:

$$\lambda_p = \lambda_{\min} = \mathbf{u}_p'\mathbf{A}\mathbf{u}_p/\mathbf{u}_p'\mathbf{B}\mathbf{u}_p = \min_{\mathbf{u}}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u})$$

$$\lambda_{p-1} = \mathbf{u}_{p-1}'\mathbf{A}\mathbf{u}_{p-1}/\mathbf{u}_{p-1}'\mathbf{B}\mathbf{u}_{p-1} = \min_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_p=0}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u})$$

$$\lambda_{p-2} = \mathbf{u}_{p-2}'\mathbf{A}\mathbf{u}_{p-2}/\mathbf{u}_{p-2}'\mathbf{B}\mathbf{u}_{p-2} = \min_{\mathbf{u}, \mathbf{u}'\mathbf{B}\mathbf{u}_p=0, \mathbf{u}'\mathbf{B}\mathbf{u}_{p-1}=0}(\mathbf{u}'\mathbf{A}\mathbf{u}/\mathbf{u}'\mathbf{B}\mathbf{u}), \text{ etc.}$$

### Application to one-way MANOVA

The one-way multivariate analysis of variance (MANOVA) provides an important example of how relative eigenvalues and eigenvectors may be used.

Suppose you have  $g$  independent random samples of size  $n_1, \dots, n_g$  from  $g$  populations that are  $N_p(\mu_j, \Sigma)$ ,  $j = 1, \dots, g$  and are interested in testing the null hypothesis

$$H_0: \mu_1 = \mu_2 = \dots = \mu_g$$

One approach finds the linear combination of the response variables  $x_j$ ,  $j = 1, \dots, p$ , for which  $H_0$  appears to be most violated and then tests to see if an  $F$  statistic computed from the linear combination is significantly large.

For any particular length  $p$  vector  $\mathbf{u}$ ,  $y = y_{\mathbf{u}} \equiv \mathbf{u}'\mathbf{x} = \sum_j u_j x_j$  is a *univariate* random variable which is a linear combination of the  $x_j$ 's. Computing  $y_{\mathbf{u}}$  for every observation yields  $g$  independent *univariate* random samples from the  $g$  populations  $N(\mu_{1y}, \sigma_y^2)$ ,  $N(\mu_{2y}, \sigma_y^2), \dots$ , where  $\mu_{1y} = \mathbf{u}'\mu_1$ ,  $\mu_{2y} = \mathbf{u}'\mu_2$ , ..., and  $\sigma_y^2 = \mathbf{u}'\Sigma\mathbf{u}$ .

When  $H_0$  is true, the univariate null hypothesis

$$H_{0u}: \mu_{1y} = \mu_{2y} = \dots = \mu_{gy}$$

is certainly true also.

You can, in fact say more:  $H_0$  is true *when and only when*  $H_{0u}$  is true for *every*  $\mathbf{u}$ , that is, the expectations of any linear combination of the  $x$ 's is the same in each population.

For a fixed  $\mathbf{u}$ , a familiar test statistic for the univariate null hypothesis  $H_{0u}$  is the univariate  $F$  statistic

$$F_{\mathbf{u}} = \frac{SS_h / (g-1)}{SS_e / (N-g)} = \frac{\frac{1}{g-1} \sum_{j=1}^g n_j (\bar{y}_{.j} - \bar{y}_{..})^2}{\frac{1}{N-g} \sum_{j=1}^g \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2}, \text{ where } N = \sum_j n_j.$$

$SS_h$  and  $SS_e$  are often notated  $SS_B$  (B for "between") and  $SS_W$  (W for "within").

For any *fixed*  $\mathbf{u}$ , when  $H_{0u}$  is true  $F_{\mathbf{u}}$  is distributed as  $F_{f_h, f_e} = F_{g-1, N-g}$ ,  $f_h = g-1$ ,  $f_e = N-g$ .

Another expression for  $F_{\mathbf{u}}$  is

$$F_{\mathbf{u}} = \{\mathbf{u}'\mathbf{H}\mathbf{u} / (g-1)\} / \{\mathbf{u}'\mathbf{E}\mathbf{u} / (N-g)\} = \{f_e / f_h\} (\mathbf{u}'\mathbf{H}\mathbf{u} / \mathbf{u}'\mathbf{E}\mathbf{u}), \quad (20)$$

where the *hypothesis* and *error* matrices are

$$\mathbf{H} = \sum_{j=1}^g n_j (\bar{\mathbf{x}}_{.j} - \bar{\mathbf{x}}_{..})(\bar{\mathbf{x}}_{.j} - \bar{\mathbf{x}}_{..})' \text{ and } \mathbf{E} = \sum_{j=1}^g \sum_{i=1}^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})'.$$

$\mathbf{H}$  and  $\mathbf{E}$  are exact analogs of  $SS_h$ , the hypothesis sum of squares and  $SS_e$ , the error sum of squares in a univariate one-way ANOVA, except that terms of the form  $(\dots)^2$  become terms of the form  $(\dots)(\dots)'$ .

## Relative Eigenvectors and Eigenvalues

The diagonal elements of  $\mathbf{H}$  and  $\mathbf{E}$  are the hypothesis and error sums of squares in univariate one-way analyses of variance of  $x_1, x_2, \dots$ , and  $x_p$ .

Note that sometimes, as in Johnson and Wichern,  $\mathbf{H}$  is notated  $\mathbf{B}$ , since it is a matrix of *between* group sums of squares and products, and  $\mathbf{E}$  is notated  $\mathbf{W}$ , since it is a matrix of *within* group sums of squares and products. I will consistently use  $\mathbf{H}$  and  $\mathbf{E}$  because the notation is applicable to more situations.

In this one-way MANOVA situation,  $s = \text{rank}(\mathbf{H}) = \min(g-1, p)$ . In more general MANOVA situations, when  $\mathbf{H}$  has degrees of freedom  $f_h$ ,  $s = \text{rank}(\mathbf{H}) = \min(f_h, p)$ .

I use  $f_h$  as a standard notation for hypothesis degrees of freedom. Here  $f_h = g-1$ . There are never more than  $s = \min(f_h, p)$  non-zero relative eigenvalues of  $\mathbf{H}$  relative to  $\mathbf{E}$ .

### Multivariate test statistics based on relative eigenvalues

Because  $H_0$  is true if and only if the univariate null hypotheses  $H_{0u}$  are true for *all* vectors  $\mathbf{u}$ , a plausible test statistic for  $H_0$  is the maximum value of  $F_u$  over *all* choices for  $\mathbf{u}$ . Maximizing  $F_u = \{(N-g)/(g-1)\}(\mathbf{u}'\mathbf{H}\mathbf{u}/\mathbf{u}'\mathbf{E}\mathbf{u})$  (see eq. (20)) is equivalent to maximizing  $\mathbf{u}'\mathbf{H}\mathbf{u}/\mathbf{u}'\mathbf{E}\mathbf{u}$  and thus by eq. (19) the maximum value is

$$F_{\max} = \{(N-g)/(g-1)\} \frac{\hat{\mathbf{u}}_1' \mathbf{H} \hat{\mathbf{u}}_1}{\hat{\mathbf{u}}_1' \mathbf{E} \hat{\mathbf{u}}_1} = \{(N-g)/(g-1)\} \hat{\lambda}_1 = \{(N-g)/(g-1)\} \hat{\lambda}_{\max} \quad (21)$$

where  $\hat{\mathbf{u}}_j$  and  $\hat{\lambda}_j$ ,  $j = 1, \dots, p$  are the eigenvectors and eigenvalues of  $\mathbf{H}$  relative to  $\mathbf{E}$ . Thus Roy suggested  $\hat{\lambda}_{\max}$  as a test statistic for the multivariate null hypothesis  $H_0$ .

I must emphasize that  $\hat{\mathbf{u}}_1$  is *not* fixed, but is random because it is computed from the random matrices  $\mathbf{H}$  and  $\mathbf{E}$ . Moreover, it is specifically chosen so as to maximize  $F_u$ . This means that the null distribution of  $F_{\max}$  is *not*  $F_{g-1, N-g}$ . Use of ordinary F tables is *erroneous* with  $F_{\max}$ , although such use is a common mistake. Instead new tables or charts are required to get its critical values.

### Geometric interpretation

If  $s \equiv \text{rank}(\mathbf{H}) = \min(p, f_h) > 1$ , and the largest relative eigenvalue is dominant (that is,  $\hat{\lambda}_1 \gg \hat{\lambda}_2 \approx 0$ ), then all the sample mean vectors  $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_g$  lie close to a line (a 1 dimensional structure) in  $p$ -dimensional space. This suggests that  $\hat{\lambda}_{\max}$  might be a particularly good test statistic when the population means  $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g$  are different and near a line.

When the first two relative eigenvalues dominate ( $\hat{\lambda}_1 \geq \hat{\lambda}_2 \gg \hat{\lambda}_3 \approx 0$ ), then the sample mean vectors lie close to a two-dimensional plane in  $p$ -dimensional space.

## Relative Eigenvectors and Eigenvalues

All the  $\hat{\lambda}_j$ 's are indicators of the size of  $\mathbf{H}$  relative to  $\mathbf{E}$ . Since "large"  $\mathbf{H}$  is evidence against  $H_0$ , you can use the  $\hat{\lambda}_j$ 's to construct other test statistics.

- The **likelihood ratio test** of  $H_0$  against the alternative

$H_1$ : not all  $\mu_j$ 's the same, that is  $\mu_{j_1} \neq \mu_{j_2}$  for some  $j_1 \neq j_2$ ,

rejects  $H_0$  for large values of

$$1/\Lambda^* \equiv \det\{\mathbf{E}^{-1}(\mathbf{H} + \mathbf{E})\} = (1 + \hat{\lambda}_1)(1 + \hat{\lambda}_2) \dots (1 + \hat{\lambda}_p) \text{ (see eq. (15)).}$$

This is usually used in the form  $-2 m_1 \log \Lambda^* = 2m_1 \log(\det\{\mathbf{E}^{-1}(\mathbf{H} + \mathbf{E})\}) = 2m_1 \sum_i \log(1 + \hat{\lambda}_i)$  where  $m_1 = f_e - (p - f_h + 1)/2$  is a constant

- **Hotelling's generalized  $T_0^2$**  Hotelling suggested that  $H_0$  should be rejected for large values of

$$T_0^2 \equiv f_e \text{tr}(\mathbf{E}^{-1}\mathbf{H}) = f_e \sum_i \hat{\lambda}_i \text{ (see eq. (12))}$$

- **Pillai's trace statistic  $V$**  Pillai proposed rejecting  $H_0$  for large values of the statistic

$$V = m_3 \text{tr}\{(\mathbf{H} + \mathbf{E})^{-1}\mathbf{H}\} = m_3 \sum_i \frac{\hat{\lambda}_i}{1 + \hat{\lambda}_i}, m_3 = f_h + f_e = N-1 \text{ (see eq. (16))}$$

All these statistics, the maximum relative eigenvalue  $\hat{\lambda}_1$ ,  $1/\Lambda^*$ ,  $T_0^2$ , and  $V$ , can be considered as overall measures of the size of  $\mathbf{H}$  relative to  $\mathbf{E}$ . In general they are not equivalent and can lead to different conclusions when used in a hypothesis test. However, when  $p = 1$  (univariate case) or  $f_h = 1$ , they all depend only on  $\hat{\lambda}_1$  and are thus equivalent since  $\hat{\lambda}_1 = 1/\Lambda^* - 1 = T_0^2/f_e = V/(1 - V)$ . And in large samples, all but the test using only  $\hat{\lambda}_1$  are essentially identical.

### MANOVA canonical variables

The variable  $\hat{z}_i \equiv \hat{\mathbf{u}}_i' \mathbf{x}$  is often called the *first canonical variable* associated with the null-hypothesis. Since the F statistic in the univariate ANOVA based on  $\hat{z}_1$  is  $F_{\max} = F_{\hat{u}_1}$ ,  $\hat{z}_1 = \hat{\mathbf{u}}_1' \mathbf{x} = \sum_{i=1}^p \hat{u}_{i1} x_i$  is the linear combination of variables  $x_1, x_2, \dots, x_p$  for which the null hypothesis *appears* to be most violated. The elements of vector  $\hat{\mathbf{u}}_1$  are the coefficients defining the linear combination. Thus examination of the elements of  $\hat{\mathbf{u}}_1$  can sometimes clarify in what way a null hypothesis fails to be true.

The linear combination  $\hat{z}_2 = \hat{\mathbf{u}}_2' \mathbf{x} = \sum_{i=1}^p \hat{u}_{i2} x_i$  is the *second* MANOVA canonical variable. It is the linear combination of the original  $x$ 's that is *uncorrelated* with  $\hat{z}_1$  and for which the null hypothesis appears to be most violated. Similarly you can define MANOVA canonical variables  $\hat{z}_3, \hat{z}_4, \dots, \hat{z}_p$  by  $\hat{z}_j = \hat{\mathbf{u}}_j' \mathbf{x}$ . At most, only the first  $s = \min(f_h, p)$  canonical variables are of interest in the sense that they help describe the differences among the groups. This is because, for  $j > s$ , the F-statistic  $= (f_e/f_h) \hat{\lambda}_j$  associated with  $\hat{z}_j$  is exactly 0 and has no among-group information. When  $s = \min(p, f_h)$  is large, the hope is that only the first few relative eigenvalues, perhaps  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  or  $\hat{\lambda}_1, \hat{\lambda}_2$  and  $\hat{\lambda}_3$ , or even just  $\hat{\lambda}_1$ , will be dominant and then only those few corresponding canonical variables will have almost all the information about differences among the group.