Eigenvalues, Eigenvectors, Matrix Factoring, and Principal Components

The *eigenvalues* and *eigenvectors* of a square matrix play a key role in some important operations in statistics. In particular, they are intimately connected with the determination of the rank of a matrix, and the “factoring” of a matrix into a product of matrices.

Determinant of a Square Matrix

The determinant of a matrix **A**, denoted $|\mathbf{A}|$ is a scalar function that is zero if and only if a matrix is of deficient rank. This fact is sufficient information about the determinant to allow the reader to continue through much of the remainder of this book. As needed, the reader should consult the more extensive treatment of determinants in the class handout on matrix methods.

Eigenvalues

Definition (Eigenvalue and Eigenvector of a Square Matrix).

For a square matrix $\mathbf{A}$, a scalar c and a vector $\mathbf{v}_v$ are an *eigenvalue* and associated *eigenvector*, $\mathbf{v}$**,** respectively, if and only if they satisfy the equation,

$$\mathbf{A}\mathbf{v} = c\mathbf{v} \qquad (1)$$

Comment. Note that if $\mathbf{Av} = c\mathbf{v}$, then of course $\mathbf{A}(\mathbf{v}k) = c(\mathbf{v}k)$ for any scalar k, so eigenvectors are not uniquely defined. They are defined only up to their *shape*. To avoid a fundamental indeterminacy, we normally assume them to be *normalized*, that is satisfy the restriction that $\mathbf{v}'\mathbf{v} = 1$.

Comment. If $\mathbf{A}\mathbf{v} = c\mathbf{v}$, then $\mathbf{A}\mathbf{v} - c\mathbf{v} = \mathbf{0}$, and $(\mathbf{A} - c\mathbf{I})\mathbf{v} = \mathbf{0}$. Look at this last equation carefully. Note that $\mathbf{A} - c\mathbf{I}$ is a square matrix, and a linear combination of its columns is null, which means $\mathbf{A} - c\mathbf{I}$ is not of full rank. This implies that its determinant must be zero. So an eigenvalue c of a square matrix $\mathbf{A}$ must satisfy the equation

$$\left|\mathbf{A} - c\mathbf{I}\right| = 0 \qquad (2)$$

Key Properties of Eigenvalues and Eigenvectors

For $N \times N$ matrix with eigenvalues c_i and associated eigenvectors $\mathbf{v}_i$, the following key properties hold:

1. $$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{N} c_i \qquad (3)$$

and

2. $$|\mathbf{A}| = \prod_{i=1}^{N} c_i \qquad (4)$$

Key Properties of Eigenvalues and Eigenvectors

3. Eigenvalues of a symmetric matrix with real elements are all real.

4. Eigenvalues of a positive definite matrix are all positive.

Key Properties of Eigenvalues and Eigenvectors

5. If a $N \times N$ symmetric matrix **A** is positive semidefinite and of rank r, it has exactly r positive eigenvalues and $p-r$ zero eigenvalues.

6. The nonzero eigenvalues of the product **AB** are equal to the nonzero eigenvalues of **BA**. Hence the traces of **AB** and **BA** are equal.

Key Properties of Eigenvalues and Eigenvectors

7. The eigenvalues of a diagonal matrix are its diagonal elements.

8. The scalar multiple $b\mathbf{A}$ has eigenvalue bc_i with eigenvector $\mathbf{v}_i$. *Proof*: $\mathbf{A}\mathbf{v}_i = c_i\mathbf{v}_i$ implies immediately that $(b\mathbf{A})\mathbf{v}_i = (bc_i)\mathbf{v}_i$.

Key Properties of Eigenvalues and Eigenvectors

9. Adding a constant b to every diagonal element of $\mathbf{A}$ creates a matrix $\mathbf{A} + b\mathbf{I}$ with eigenvalues $c_i + b$ and associated eigenvectors $\mathbf{v}_i$.

Proof.

$$\left(\mathbf{A} + b\mathbf{I}\right)\mathbf{v}_i = \mathbf{A}\mathbf{v}_i + b\mathbf{v}_i = c_i\mathbf{v}_i + b\mathbf{v}_i = \left(c_i + b\right)\mathbf{v}_i$$

Key Properties of Eigenvalues and Eigenvectors

10. $\mathbf{A}^m$ has c_i^m as an eigenvalue, and $\mathbf{v}_i$ as its eigenvector.

Proof: Consider

$$\begin{aligned}\left(\mathbf{A}^2\right)\mathbf{v}_i &= \mathbf{A}\left(\mathbf{A}\mathbf{v}_i\right) = \mathbf{A}\left(c_i\mathbf{v}_i\right) = \mathbf{A}\mathbf{v}_i c_i = \\ &= c_i\mathbf{v}_i c_i = \left(c_i^2\right)\mathbf{v}_i\end{aligned} \tag{5}$$

The general case follows by induction.

Key Properties of Eigenvalues and Eigenvectors

11. $\mathbf{A}^{-1}$, if it exists, has $1/c_i$ as an eigenvalue, and $\mathbf{v}_i$ as its eigenvector.

Proof:

$$\mathbf{A}\mathbf{v}_i = c_i\mathbf{v}_i = \mathbf{v}_i c_i$$
$$\mathbf{A}^{-1}\mathbf{A}\mathbf{v}_i = \mathbf{v}_i = \mathbf{A}^{-1}\mathbf{v}_i c_i$$

But the right side of the previous equation implies that

$$(1/c_i)\mathbf{v}_i = (1/c_i)\mathbf{A}^{-1}\mathbf{v}_i c_i = \mathbf{A}^{-1}\mathbf{v}_i, \text{ or}$$
$$\mathbf{A}^{-1}\mathbf{v}_i = \left(1/c_i\right)\mathbf{v}_i$$

Key Properties of Eigenvalues and Eigenvectors

12. For *symmetric* $\mathbf{A}$, for distinct eigenvalues c_i, c_j with associated eigenvectors $\mathbf{v}_i$, $\mathbf{v}_j$ we have $\mathbf{v}_i{}'\mathbf{v}_j$.

Proof:

$\mathbf{Av}_i = c_i\mathbf{v}_i$, and $\mathbf{Av}_j = c_j\mathbf{v}_j$. So $\mathbf{v}'_j\mathbf{Av}_i = c_i\mathbf{v}'_j\mathbf{v}_i$ and $\mathbf{v}'_i\mathbf{Av}_j = c_j\mathbf{v}'_i\mathbf{v}_j$. But, since a bilinear form $\mathbf{a}'\mathbf{Ab}$ is a scalar, it is equal to its transpose, and, remembering that $\mathbf{A} = \mathbf{A}'$, $\mathbf{v}'_i\mathbf{Av}_j = \mathbf{v}'_j\mathbf{A}'\mathbf{v}_i = \mathbf{v}'_j\mathbf{Av}_i$. So placing parentheses around $\mathbf{Av}$ expressions, we see that $c_i\mathbf{v}'_j\mathbf{v}_i = c_j\mathbf{v}'_i\mathbf{v}_j = c_j\mathbf{v}'_j\mathbf{v}_i$. If c_i and c_j are different, this implies $\mathbf{v}'_j\mathbf{v}_i = 0$.

Key Properties of Eigenvalues and Eigenvectors

13. For any real, symmetric $\mathbf{A}$, there exists a $\mathbf{V}$ such that $\mathbf{V}'\mathbf{A}\mathbf{V} = \mathbf{D}$, where $\mathbf{D}$ is diagonal. Moreover, any real, symmetric matrix $\mathbf{A}$ can be written as $\mathbf{V}\mathbf{D}\mathbf{V}'$, where contains the eigenvectors $\mathbf{v}_i$ of $\mathbf{A}$ in order in its columns, and $\mathbf{D}$ contains the eigenvalues c_i of $\mathbf{A}$ in the ith diagonal position.

Key Properties of Eigenvalues and Eigenvectors

14. Suppose that the eigenvectors and eigenvalues of $\mathbf{A}$ are ordered in the matrices $\mathbf{V}$ and $\mathbf{D}$ in descending order, so that the first element of $\mathbf{D}$ is the largest eigenvalue of $\mathbf{A}$, and the first column of $\mathbf{V}$ is its corresponding eigenvector. Define $\mathbf{V}^*$ as the first m columns of $\mathbf{V}$, and $\mathbf{D}^*$ as an $m \times m$ diagonal matrix with the corresponding m eigenvalues as diagonal entries. Then

$$\mathbf{V}^*\mathbf{D}^*\mathbf{V}^{*\prime} \tag{6}$$

is a matrix of rank m that is the best possible (in the least squares sense) rank m approximation of $\mathbf{A}$.

Key Properties of Eigenvalues and Eigenvectors

15. Consider all possible "normalized quadratic forms in **A**," i.e.,

$$q(\mathbf{x}_i) = \mathbf{x}_i'\mathbf{A}\mathbf{x}_i \qquad (7)$$

with $\mathbf{x}_i'\mathbf{x}_i = 1$. The maximum of all quadratic forms is achieved with $\mathbf{x}_i = \mathbf{v}_1$, where $\mathbf{v}_1$ is the eigenvector corresponding to the largest eigenvalue of **A**. The minimum is achieved with $\mathbf{x}_i = \mathbf{v}_m$, the eigenvector corresponding to the smallest eigenvalue of **A**.

Applications of Eigenvalues and Eigenvectors

1. Principal Components

From property 15 in the preceding section, it follows directly that the maximum variance linear composite of a set of variables is computed with linear weights equal to the first eigenvector of $\mathbf{\Sigma}_{yy}$, since the variance of this linear combination is a quadratic form in $\mathbf{\Sigma}_{yy}$.

2. Matrix Factorization

Diagonal matrices act much more like scalars than most matrices do. For example, we can define fractional powers of diagonal matrices, as well as positive powers. Specifically, if diagonal matrix $\mathbf{D}$ has diagonal elements d_i, the matrix $\mathbf{D}^x$ has elements d_i^x . If x is negative, it is assumed $\mathbf{D}^x$ is positive definite. With this definition, the powers of $\mathbf{D}$ behave essentially like scalars. For example, $\mathbf{D}^{1/2}\mathbf{D}^{1/2} = \mathbf{D}$.

Example.

Suppose we have

$$\mathbf{D} = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$$

Then

$$\mathbf{D}^{1/2} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Example.

Suppose you have a variance-covariance matrix $\mathbf{\Sigma}$ for some statistical population. Assuming $\mathbf{\Sigma}$ is positive semidefinite, then (from Property 13 on page 14 it can be written in the form $\mathbf{\Sigma} = \mathbf{VDV}' = \mathbf{FF}'$, where $\mathbf{F} = \mathbf{VD}^{1/2}$ is called a "Gram-factor of $\mathbf{F}$."

Comment. Gram-factors are not, in general, uniquely defined.

Example.

Suppose $\boldsymbol{\Sigma} = \mathbf{FF}'$. Then consider any orthogonal matrix $\mathbf{T}$, conformable with $\mathbf{F}$, such that $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$. There are infinitely many orthogonal matrices of order 2×2 and higher. Then for any such matrix $\mathbf{T}$, we have

$$\boldsymbol{\Sigma} = \mathbf{FTT}'\mathbf{F} = \mathbf{F}^{*}\mathbf{F}^{*\prime} \tag{8}$$

where $\mathbf{F}^{*} = \mathbf{FT}$.

Applications of Gram-Factors

Gram-factors have some significant applications. For example, in the field of random number generation, it is relatively easy to generate pseudo-random numbers that mimic p variables that are independent with zero mean and unit variance. But suppose we wish to mimic p variables that are not independent, but have variance-covariance matrix $\mathbf{\Sigma}$? The following result describes one method for doing this.

Result.

Given $p \times 1$ random vector $\mathbf{x}$ having variance-covariance matrix $\mathbf{I}$. Let $\mathbf{F}$ be a Gram-factor of $\mathbf{\Sigma} = \mathbf{FF}'$. Then $\mathbf{y} = \mathbf{Fx}$ will have variance-covariance matrix $\mathbf{\Sigma}$.

So if we want to create random numbers with a specific covariance matrix, we take a vector of independent random numbers, and premultiply it by $\mathbf{F}$.

Symmetric Powers of a Symmetric Matrix

In certain intermediate and advanced derivations in matrix algebra, reference is made to "symmetric powers" of a symmetric matrix $\mathbf{\Sigma}$, in particular the "symmetric square root" $\mathbf{\Sigma}^{1/2}$ of $\mathbf{\Sigma}$, a symmetric matrix which, when multiplied by itself, yields $\mathbf{\Sigma}$. Recall that $\mathbf{\Sigma} = \mathbf{VDV}' = \mathbf{VD}^{1/2}\mathbf{D}^{1/2}\mathbf{V}'$. Note that $\mathbf{VD}^{1/2}\mathbf{V}'$ is a symmetric square root of $\mathbf{\Sigma}$, i.e.,

$$\mathbf{VD}^{1/2}\mathbf{V}'\mathbf{VD}^{1/2}\mathbf{V}' = \mathbf{VDV}'$$

Orthogonalizing a Set of Variables

Consider a random vector $\mathbf{x}$ with $\mathrm{Var}(\mathbf{x}) = \mathbf{\Sigma} \neq \mathbf{I}$. What is $\mathrm{Var}\left(\mathbf{\Sigma}^{-1/2}\mathbf{x}\right)$? How might you compute $\mathbf{\Sigma}^{-1/2}$?

Suppose a set of variables $\mathbf{x}$ have a covariance matrix $\mathbf{A}$, and you want to linearly transform them so that they have a covariance matrix $\mathbf{B}$. How could you do that if you had a computer program that easily gives you the eigenvectors and eigenvalues of $\mathbf{A}$ and $\mathbf{B}$? (Hint: First orthogonalize them. Then transform the orthogonalized variables to a covariance matrix you want.)