NORTH-HOLLAND

The Derivative of an Orthogonal Matrix of Eigenvectors of a Symmetric Matrix

Tõnu Kollo
University of Tartu
Tartu, Estonia

and

Heinz Neudecker
Instituut voor Actuariaat en Econometrie
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands

Submitted by Richard William Farebrother

ABSTRACT

The authors supply the derivative of an orthogonal matrix of eigenvectors of a real symmetric matrix. To illustrate the applicability of their result they consider a real symmetric random matrix for which a more or less standard convergence in distribution is assumed to hold. The well-known delta method is then used to get the asymptotic distribution of the orthogonal eigenmatrix of the random matrix. © 1997 Elsevier Science Inc.

INTRODUCTION

Consider a real symmetric $p \times p$ matrix M, with distinct eigenvalues λ_i and associated normalized eigenvectors w_i $(i = 1, \dots, p)$. It is well known that then real-valued functions ψ_i and vector functions f_i are defined for all matrices Z in some neighborhood $\mathscr{N}(M) \subset \mathbb{R}^{p \times p}$ of M, such that

$$\psi_i(M) = \lambda_i, \qquad f_i(M) = w_i$$

655 Avenue of the Americas, New York, NY 10010
0024-3795/97/$17.00
PII S0024-3795(97)00016-5

and

$$Zf_i = \psi_i f_i, \qquad f_i' f_j = \delta_{ij} \qquad [Z \in \mathcal{N}(M)],$$

δ_{ij} being the Kronecker delta.

Kollo and Neudecker (1993) gave the derivatives of f_i for the case of symmetric perturbations dZ defined at M, viz. $\partial f_i(z)/\partial z'|_{z=m}$, where

$$z = D^+ \text{ vec } Z \quad \text{and} \quad m = D^+ \text{ vec } M,$$

D being the $p^2 \times \frac{1}{2}p(p+1)$ duplication matrix and D^+ its Moore-Penrose inverse. See Magnus and Neudecker (1995) for details on derivatives and the duplication matrix.

In the present paper we shall give a compact expression for the full derivative $\partial \text{ vec } F(z)/\partial z'|_{z=m}$, where $F = (f_1 \ldots f_p)$. It will be established by treating the full system of eigenequations.

DEVELOPMENT

Let

$$ZF = F\Psi, \qquad F'F = I \tag{1}$$

link the eigenvector and eigenvalue functions $f_i(z)$ and $\psi_i(z)$, where

$$F = F(z) = [f_1(z) \quad \cdots \quad f_p(z)], \qquad \Psi = \Psi(z) = \Sigma_i \psi_i(z) E_{ii},$$

$$E_{ii} = e_i e_i',$$

and e_i is the ith unit vector of dimension p ($i = 1, \ldots, p$). We then find the following result:

THEOREM.

$$\left.\frac{\partial \text{ vec } F(z)}{\partial z'}\right|_{z=m} = (I \otimes W)(\Lambda \otimes I - I \otimes \Lambda)^+ (W' \otimes W')D, \tag{2}$$

where $W = (w_1 \cdots w_p)$, $W'W = I$, *and* $\Lambda = \Sigma_i \lambda_i E_{ii}$.

Clearly w_i and λ_i ($i = 1, \ldots, p$) are orthogonal eigenvectors and eigenvalues of M.

Proof. Taking differentials of $ZF = Z\Psi$ at the point $z = m$ yields

$$(dZ)W + M\,dF = (dF)\Lambda + W\,d\Psi, \tag{3}$$

from which follows

$$\begin{aligned} d\Psi &= W'(dZ)W + W'M\,dF - W'(dF)\Lambda \\ &= W'(dZ)W + \Lambda W'\,dF - W'(dF)\Lambda = [W'(dZ)W]_d, \end{aligned} \tag{4}$$

as

$$(\Lambda W'\,dF)_d = \Lambda(W'\,dF)_d = (W'\,dF)_d\Lambda = [W'(dF)\Lambda]_d,$$

A_d generically being the diagonal matrix obtained from the (square) matrix A. Premultiplication of (3) by W' and vectorization give by (4) the equation

$$(\Lambda \otimes I - I \otimes \Lambda)(I \otimes W')d \operatorname{vec} F = (I - K_d)(W' \otimes W')\, d \operatorname{vec} Z. \tag{5}$$

We have used the property $K_d \operatorname{vec} A = \operatorname{vec} A_d$ for square A.

Define then $B := \Lambda \otimes I - I \otimes \Lambda$ and $C := I - K_d$. Clearly $C = B^+B$. The solution for (5) is now

$$d \operatorname{vec} F = (I \otimes W)B^+(W' \otimes W')\, d \operatorname{vec} Z + (I \otimes W)K_d \operatorname{vec} Q, \tag{6}$$

Q being an arbitrary matrix. Differentiation of $F'F = I$ at the point $z = m$ yields

$$(I \otimes W')\, d \operatorname{vec} F = -K(I \otimes W')\, d \operatorname{vec} F. \tag{7}$$

Clearly

$$KB^+ = -B^+K, \tag{8}$$

because $KB = -BK$ and $(KBK)^+ = KB^+K$. Using (6), (7) and (8) produces

$$(I \otimes W)K_d \operatorname{vec} Q = 0.$$

The solution (6) can then be written as

$$d \operatorname{vec} F = (I \otimes W)(\Lambda \otimes I - I \otimes \Lambda)^{+}(W' \otimes W')D\,dZ, \qquad (9)$$

which yields

$$\left.\frac{\partial \operatorname{vec} F(z)}{\partial z'}\right|_{z=m} = (I \otimes W)(\Lambda \otimes I - I \otimes \Lambda)^{+}(W' \otimes W')D. \quad \blacksquare$$

AN EXAMPLE

Assume the convergence

$$\sqrt{n}\, v\left(\hat{M}_{(n)} - M\right) \xrightarrow{D} \mathscr{N}(0, D^{+}GD^{+\prime})$$

holds for an estimator $\hat{M}_{(n)}$ of M from a sample size n, where $\hat{M}_{(n)}$ is a real symmetric $p \times p$ random matrix. Let $\hat{W}_{(n)}$ be orthogonal eigenmatrices of $\hat{M}_{(n)}$. We shall use the well-known convergence result due to Anderson (1958): Let $\{(y_{(n)})\}$ be a sequence of random vectors $y_{(n)}$, and b be a compatible fixed vector. Assume that

$$\sqrt{n}\,(y_{(n)} - b) \xrightarrow{D} \mathscr{N}(0, T) \qquad \text{(convergence in distribution)},$$

or equivalently that $\sqrt{n}\, y_{(n)}$ is asymptotically normally distributed with mean $\sqrt{n}\, b$ and variance T. Let $f(z)$ be a vector function of a vector variable z with first and second derivatives existing in a neighborhood of $z = b$. Then

$$\sqrt{n}\left[f(y_{(n)}) - f(b)\right] \xrightarrow{D} \mathscr{N}(0, \Delta T \Delta'),$$

where

$$\Delta = \left.\frac{\partial f(z)}{\partial z'}\right|_{z=b}$$

is a matrix derivative. This yields, in conjunction with the Theorem, for $n \to \infty$,

$$\sqrt{n}\, \operatorname{vec}\left(\hat{W}_{(n)} - W\right) \xrightarrow{D} \mathscr{N}(0, T),$$

where

$$T = (I \otimes W)(\Lambda \otimes I - I \otimes \Lambda)^{+} (W' \otimes W')G$$
$$\times (W \otimes W)(\Lambda \otimes I - I \otimes \Lambda)^{+} (I \otimes W').$$

Clearly T contains the asymptotic variances and covariances of the eigenvectors $\hat{w}_{(n)i}$ and $\hat{w}_{(n)j}$. The article by Kollo and Neudecker (1993) only gave the variances. See their Theorem 4.1. We find as asymptotic covariance between $\hat{w}_{(n)i}$ and $\hat{w}_{n(j)}$ the matrix

$$T_{ij} = \left[w_i' \otimes W(\lambda_i I - \Lambda)^{+} W'\right]G\left[w_j \otimes W(\lambda_j I - \Lambda)^{+} W'\right],$$

where $w_i = W_{\cdot i}$ and $\hat{w}_{(n)i} = (\hat{W}_n)_{\cdot i}$ $(i = 1, \ldots, p)$

REFERENCES

Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Kollo, T. and Neudecker, H. 1993. Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices, *J. Multivariate Anal.* 47:283–300. Corrigendum, *J. Multivariate Anal.* 51:210.

Magnus, J. R. and Neudecker, H. 1995. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.

Received 11 November 1995; final manuscript accepted 19 December 1996