Ecology, 76(2), 1995, pp. 644–645

BOOTSTRAPPED PRINCIPAL COMPONENTS ANALYSIS—REPLY TO MEHLMAN ET AL.

Donald A. Jackson[1]

Mehlman et al. (1995) identify a condition that may arise in various multivariate procedures, i.e., the reflection or reversal of the axis direction. They suggest that this condition may have led to an underestimation of the correct multivariate dimensionality in Jackson's (1993*a*) study of principal components analysis (PCA). The reversal of a multivariate axis or component presents no problem in standard analyses and does not change the interpretation of the data summary. In Jackson (1993*a*) I was using bootstrapped PCA to evaluate the number of nontrivial or "significant" components from the frequency distribution and the confidence limits of bootstrapped eigenvalues and eigenvector coefficients. Mehlman et al. (1995) state that it is important to recognize that an axis can reverse in the bootstrapped PCAs. If left in this orientation, there would be a very real possibility of underestimating the number of components based on the criterion of whether the 95% confidence limits encompassed zero or not. It is possible that bimodal distributions could arise with one mode on either side of zero due to reversals, even with a strongly structured data set. One may falsely conclude the component to be uninformative unless some measure is taken to recognize and correct this condition.

During the course of the study by Jackson (1993*a*), axis reversals occurred frequently. This is a well-documented feature (e.g., Gauch et al. 1981, Knox 1989, Knox and Peet 1989, Jackson 1993*b*) of various multivariate methods, a point identified by Mehlman et al. (1995). Because of this documentation, I assumed it to be implicit in the approach and failed to state explicitly that one must examine for reversals in the coefficients. I am grateful to Mehlman et al. for making this clear to readers.

Mehlman et al. (1995) suggest an approach based on "fixing" the sign of the largest eigenvector coefficient on each axis when histograms suggest that reversals may have occurred. In their approach, a single variable having the largest eigenvector coefficient in one PCA is assumed to be representative of the results from each subsequent bootstrapped PCA. In some cases this may be true, but in others it may be incorrect. Depending on the relative magnitude of the larger coefficients, it is possible that reversals in the ordering of the coefficients may occur, i.e., the largest and subsequent coefficients may reverse their ordering. This reversal is analogous to the changes in the ordering of the components (Oksanen 1988, Jackson 1993*a, b*) that occur when eigenvalues are similar in magnitude. In such a case, the approach proposed by Mehlman et al. (1995) may fail. Their approach depends on the distributions of the eigenvector coefficients and the degree of structure inherent in the data set. Specifically, the more similar the magnitude of the coefficients and the less structured the data, the greater the possibility that their method may not work.

My approach (Jackson 1993*a*) was to correlate the eigenvector coefficients for each axis from the original PCA with those from each bootstrapped PCA. A negative correlation indicated a reversal or reflection of the axis orientation. Therefore, the coefficients were multiplied by −1. During the course of the study, each component from each data set was checked for such reversals and corrected, if necessary, prior to determining whether zero fell within the 95% confidence limits. In the course of my study, the bootstrapped eigenvector approach occasionally underestimated the number of dimensions simulated, i.e., underestimated the known dimensionality of the data and number of nontrivial components. Mehlman et al. (1995) suggest that failing to correct for the axis reversals might explain why the bootstrapped eigenvector approach underestimated the number of components. However, since this problem of axis reversal was recognized and corrected, clearly this is not the reason for any underestimation. However, regarding the underestimation of the nontrivial components, I originally stated (Jackson 1993*a*: 2212) "the combination of these two approaches, i.e., the bootstrapped eigenvalue and eigenvector coefficients, provides a better measure of the dimensionality than either approach alone." Undoubtedly there are conditions in certain analyses where this correlation-based approach will provide misleading results also. Although the method of determining axis reversals differs between Jackson (1993*a*) and Mehlman et al. (1995), the fundamental result of the correction process is identical, i.e., identify those axes having reversals and multiply them by −1.

An alternative to either of these approaches is that of Procrustean rotations (Gower 1971, Cliff 1987, Jackson and Somers 1989, Jackson 1993*b*). Knox (1989) choose this method in a study of the stability of detrended correspondence analysis in which he rotated several axes to maximize their concordance between solutions. Using this approach, one may find that results on a second or third component from a boot-

[1]Department of Zoology, University of Toronto, Toronto, Ontario Canada M5S 1A1.

strapped result may be rotated to fit with the first axis from the original data set. Clearly a Procrustean approach differs from the restrictions of both Jackson (1993*a*) and Mehlman et al. (1995), in which only a single axis from two PCA solutions are compared at one time. If the Procrustean rotation is restricted to rotating one component at a time, then the approach is equivalent to that of Jackson (1993*a*) because it is a least-squares fit of a bivariate relationship.

If the more generalized approach is used, with several components compared simultaneously, the Procrustean method may be compatible with the bootstrapped eigenvalue method (Jackson 1993*a*). With the eigenvalue approach and simulated data, I showed (in Jackson 1993*a*) that there could be considerable overlap among the first three PCs in several data sets. However, there was no overlap of these eigenvalues with the distributions of subsequent eigenvalues. Using the bootstrapped eigenvalues, one would conclude there were three nontrivial components. Taking these three nontrivial components, it would then be possible to use Procrustean rotations to look at the eigenvector coefficients. I am unaware of any formal development of methods using such a Procrustean approach as a means of defining the number of components to rotate, but it may provide a more accurate measure than that of either the correlational method I used (Jackson 1993*a*) or that proposed by Mehlman et al. (1995).

It should be apparent that there is no single approach available, nor has there been adequate work to determine which method of assessing and correcting reversals/reflections is best. However, given the findings of Jackson (1993*a*), it is likely that any of these approaches would be superior to the methods most commonly used, i.e., eigenvalues exceeding 1.0, scree plot, or total amount of variance.

Acknowledgments: I am grateful to Drs. D. H. Johnson, D. W. Mehlman, and M. W. Palmer for their comments on this study.

Literature Cited

Cliff, N. 1987. Analyzing multivariate data. Harcourt Brace Jovanovich, Fort Worth, Texas, USA.

Gauch, Jr., H. G., R. H. Whittaker, and S. B. Singer. 1981. A comparative study of nonmetric ordinations. Journal of Ecology **69**:135–152.

Gower, J. C. 1971. Statistical methods of comparing different multivariate analyses of the same data. Pages 138–149 *in* F. R. Hodson, D. G. Kendall, and P. Tautu, editors. Mathematics in the archaeological and historical sciences. Edinburgh University Press, Edinburgh, Scotland.

Jackson, D. A. 1993*a*. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology **74**:2204–2214.

———. 1993*b*. Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association, and ordination methods. Hydrobiologia **268**:9–26.

Jackson, D. A., and K. M. Somers. 1989. Putting things in order: the ups and downs of detrended correspondence analysis. American Naturalist **137**:704–712.

Knox, R. G. 1989. Effects of detrending and rescaling on correspondence analysis: solution stability and accuracy. Vegetatio **83**:129–136.

Knox R. G., and R. K. Peet. 1989. Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. Vegetatio **80**:153–165.

Mehlman, D. W., U. L. Shepherd, and D. A. Kelt. 1995. Bootstrapping principal components analysis—a comment. Ecology **76**:640–643.

Oksanen, J. 1988. A note on the occasional instability of detrending in correspondence analysis. Vegetatio **74**:29–32.

Manuscript received 22 June 1994;
revised 16 September 1994;
accepted 21 September 1994.