Abstracta Botanica 15:43–46 (1991)

On the standardization of Procrustes statistics for the comparison of ordinations

J. Podani

Department of Plant Ecology and Taxonomy, L. Eötvös University, Ludovika tér 2, H-1083 Budapest, Hungary

Keywords: distributions, one-sided tests, randomization, simulation

Abstract: Procrustes analyses of ordinations yield a commensurable coefficient through standardization by the expected values for random configurations. Simulation of the distribution of these coefficients enables the user to test the significance of ordination dissimilarity.

Introduction

Ordination methods belong to the most widely used tools for analyzing multivariate data structures in biology. Comparison of ordination results is an important task in many cases when, for example, several equally interesting alternatives exist for the same set of data points (see e.g., Gower 1971, Digby & Kempton 1987, Podani 1989, for examples). Objective assessment is achieved by a family of methods known as *Procrustes Analysis* (PA), developed partly independently by various authors (Green 1952, Gower 1971, Schönemann & Carroll 1970). The name recalls *Procrustes*, the innkeeper in Greek mythology, who stretched his guests or cut their limbs to make them fit to the bed, an obvious reference to the transformation of ordinations before any meaningful comparisons can be made.

PA can be used for a multitude of purposes, such as the generation of average (consensus) configurations or the calculation of resemblance matrices through multiple PA to express dissimilarity relationships among many alternative ordinations. These strategies work without any standardization of Procrustes statistics because commensurability is not important. However, if we wish to evaluate individual values, standardization becomes crucial. This paper shows that although the standardizing manipulations suggested by Gower (1971) and Sibson (1978) do introduce a common scale for Procrustes statistics for any configurations, they are still inapplicable to compare situations when the number of points is different. Incorporation of a randomization procedure and standardization by random expectation provide a more meaningful measure of ordination dissimilarity and a test of its significance as well.

Procrustes statistics

A configuration of m points in an n-dimensional space is represented by an $n \times m$ matrix. Two such matrices, $\mathbf{X}$ and $\mathbf{Y}$ can be compared by the sum of squared positional differences given by

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - y_{ij})^2 = tr(\mathbf{X} - \mathbf{Y})'(\mathbf{X} - \mathbf{Y}).$$

Such a comparison is meaningless because usually there is no preferred position and orientation of the configurations; ordination scores are often arbitrary. We can move the points in $\mathbf{Y}$ relative to $\mathbf{X}$ through *rotation, reflection* and *translation* to optimize the goodness of fit of $\mathbf{Y}$ to $\mathbf{X}$. The first step is the centring of $\mathbf{X}$ and $\mathbf{Y}$ so that both sets of points have their centroid at the origin (for convenience, $\mathbf{X}$ and $\mathbf{Y}$ will denote *centered* matrices in the sequel). The required rotation of $\mathbf{Y}$ is obtained by its multiplication with the $n \times n$ orthogonal transformation matrix

$$\mathbf{H}=\mathbf{VU}',$$

where $\mathbf{U}$ and $\mathbf{V}$ result from the singular value decomposition of

$$\mathbf{X}'\mathbf{Y}=\mathbf{USV}'.$$

After finding these matrices, the goodness of fit statistic will be computed as

$$\mathbf{R}_E^2 = tr(\mathbf{X} - \mathbf{YH})'(\mathbf{X} - \mathbf{YH}) =$$

$$= tr(\mathbf{XX}') + tr(\mathbf{YY}') - 2tr(\mathbf{YX}'\mathbf{XY}')^{\frac{1}{2}},$$

which is a symmetric measure. $\mathbf{R}^2$ depends on the actual scales of $\mathbf{X}$ and $\mathbf{Y}$, therefore this statistic is not appropriate for arbitrarily scaled ordinations. Inclusion of a scaling parameter c in the transformation of $\mathbf{Y}$ to $c\mathbf{YH}$ resolves this problem

$$c = tr(\mathbf{YHX}')/tr(\mathbf{YY}'),$$

leading to

$$\mathbf{R}_S^2 = tr(\mathbf{XX}') - 2(tr(\mathbf{YX}'\mathbf{XY}')^{\frac{1}{2}})^2/tr(\mathbf{YY}').$$

This coefficient is unsymmetric, therefore Gower (1971) suggested to scale each configuration after centering so that

$$tr(\mathbf{XX}') = tr(\mathbf{YY}') = 1,$$

to yield a coefficient denoted by d^2. As Sibson (1978) pointed out, $\mathbf{R}_S^2$ may be standardized directly

$$\gamma_S = \mathbf{R}_S^2/tr(\mathbf{XX}'),$$

resulting a range between 0 and 1, and there is a simple relationship between this and Gower's measure

$$d^2 = 2\{1 - \sqrt{1 - \gamma_S}\}.$$

Then it becomes clear that d^2 ranges from zero to 2, but this simple fact is not mentioned in most texts. In the sequel I will use only d^2. without loss of generality of the discussion.

The problem

In a comprehensive bio-climatic survey (Sun & Feoli 1991; Feoli, Podani & Sun, in preparation). we analyzed the relationship between PCA ordinations of climatic stations in China and the actual geographic positions of the stations through Procrustes methods. Comparisons were made for various phytogeographic regions of the country separately. One problem was to identify the region whose bioclimatic analysis best reflects actual geographic positions. The number of points was quite different for the regions (ranging from 30 to 175), and we suspected that the d^2 values are not directly comparable even though they were measured on the same scale. Is it right, for example, that a value of $d^2 = 1.191$ for 30 points indicates closer relationships and better predictability of ordinations than $d^2 = 1.343$ for 95 points?

Distributional properties of d^2

A solution of the above problem is that the actual d^2 scores are related to a standard reference basis, i.e., a value that would be expected for two ordinations if they were constructed completely at random. If $\mathrm{E}(d^2)_m$ denotes this expectation for m points, the resulting coefficient will be

$$\delta^2 = d^2/\mathrm{E}(d^2)_m,$$

with a range of 0 to $2/\mathrm{E}(d^2)_m$. As long as exact analytical solutions are not available to find $\mathrm{E}(d^2)_m$, I suggest to use computer simulated pairs of random ordinations in a large number of steps to obtain an estimate of this expectation.

Program ORDTEST has been written in FORTRAN to perform this task on Macintosh computers. In each simulation step, the program generates two configurations of m points in the specified number of dimensions using the random number generator routine described by Park & Miller (1988) and computes d^2 in turn. After, say, k comparisons the expectation of d^2 ($= d^2/k$), the maximum, minimum and a frequency histogram are printed. The output list also enables us to find probability points in the distribution for testing the significance of d^2.

The program was run for $m = 3, 10, 50$ and 100 for two dimensions, using 1000 iterations for each case (Fig. 1). The resulting distributions are asymmetric, for small m they are skewed to the left and with increases of m this skewness is shifted to the right. Simulations for other values of m provide answer to the question asked at the end of the previous section. For 30 points the expectation is $\mathrm{E}(d^2) = 1.569$, yielding $\delta^2 = .76$. The probability that this value results if the two ordinations are random is $p = .02$. How-

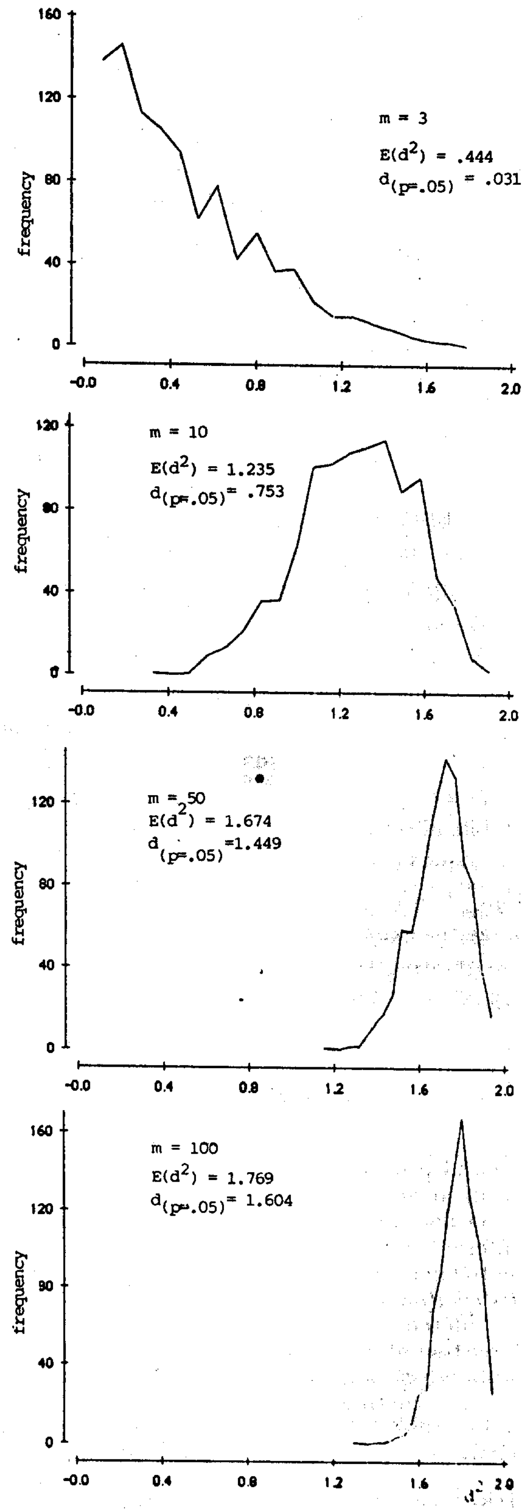

Fig. 1. Simulated frequency distributions for d^2. The range is divided into 20 intervals in each case. Points are connected for clarity only. Thresholds are presented for one-sided tests of similarity at $p = .05$.

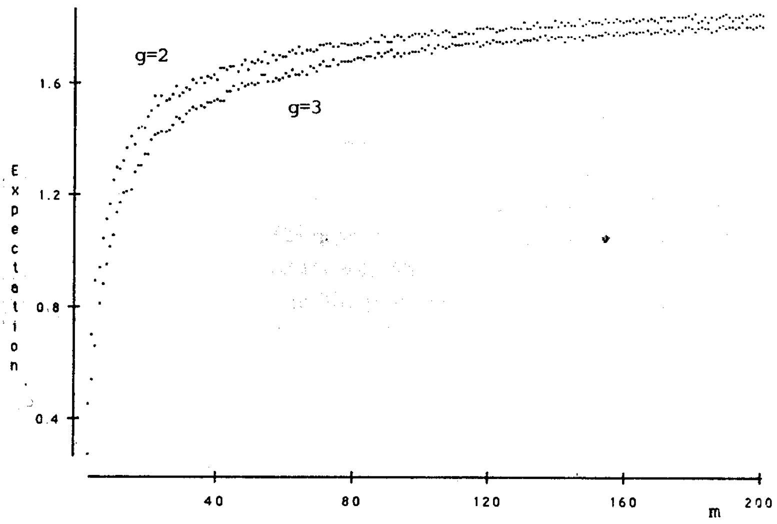

Fig. 2. The relationship between the number of points and E(d^2) for 2 and 3 dimensions. Each point is based on 100 pairs of randomly simulated ordinations.

ever, E(d^2) = 1.762 for m = 95 with δ^2 = .76 and p < .001. Contrary to what was suggested by comparing the original d^2 values, the standardized coefficients become similar. It is more likely that the actual coefficients relate to randomly generated ordinations for 30 points than for 95 points, and we conclude that the ordinations have higher mutual predictability in the 95-point case.

The effect of m and the dimensionality g on the expectation is reflected for m = 3 to 200 in Fig. 2, which shows that after a rapid initial stage the expectation increases very slowly, approximating the absolute maximum of 2. The difference between the expectations for g = 2 and g = 3 tends to disappear as m increases.

Concluding remarks

This paper attempted to draw attention to a general problem with interpreting scale, standardization and statistical significance of the same dissimilarity coefficient when applied to different data sets. The simple examples presented here sufficiently illustrate that *simulated distributions* and *random expectation* must be considered in analogous situations. The literature abounds in examples, suffice to mention the probabilistic approach taken by Goodall (1964) for constructing a general similarity measure, or by Hubert & Arabie (1985) for comparing partitions.

Acknowledgements. This research was partially completed during my stay at I.C.E.M., Trieste, Italy, in 1990. I am grateful to this institution for increasing the accessibility of a Macintosh computer to me.

References

DIGBY, P. G. N. & R. A. KEMPTON. 1987. Multivariate analysis of ecological data. Chapman & Hall, London.

FEOLI, E., J. PODANI & SUN CHENG-YONG. 1991. The relationship between geographic data and phytoclimatic ordinations: a Procrustes analysis approach (in preparation).

GOODALL, D. W. 1964. A probabilistic similarity index. Nature 203:1098.

GOWER, J. C. 1971. Statistical methods of comparing different multivariate analyses of the same data. In: Hodson, F. R., Kendall, D. G. & Tautu, P. (eds.), Mathematics in the archeological and historical sciences. Edinburgh Univ. Press, Edinburgh. pp. 138-149.

GREEN, B. F. 1952. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika 17:429-440.

HUBERT, L. & P. ARABIE. 1985. Comparing partitions. J. Classification 2:193-218.

PARK, S. K. & K. W. MILLER. 1988. Random number generators: good ones are hard to find. Communications of the ACM 31. No. 10.

PODANI, J. 1989. Comparison of ordinations and classifications of vegetation data. Vegetatio 83: 111-128.

SCHÖNEMANN, P. H. & R. M. CARROLL. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. Psychometrika 35:245-256.

SIBSON, R. 1978. Studies in the robustness of multidimensional scaling: Procrustes statistics. J. R. Statist. Soc. B. 40:234-238.

SUN CHENG-YONG & E. FEOLI. 1991. Numerical phytoclimatic classification of China. Int. J. Biometeorology (in press).

Received May 10, 1991