Non-linear ordination in several dimensions

A maximum likelihood approach

D. W. Goodall\(^1\) & R. W. Johnson\(^2\)

\(^1\)CSIRO Division of Land Resources Management, Private Bag, Wembley, W.A. 6014, Australia
\(^2\)Department of Primary Industries, Queensland Herbarium, Meiers Road, Indooroopilly, Qld. 4068, Australia

Keywords: maximum likelihood approach, non-linear, ordination

Abstract

A method is described for fitting a model in which site characteristics are represented by a set of orthogonal axes, while the probability that a particular species will be present, and its quantity if it is present, are each related to the axes of the system by symmetrical hypersurfaces with an optimum, decreasing asymptotically to zero as conditions depart from the optimum. The method needs initial estimates of the positions of the sample sites within the axis system. Given these starting points, they can be progressively improved by an iterative procedure. Results are reported of an extensive series of tests using artificial data, and of an analysis of field data from brigalow woodland in Queensland.

Introduction

A description has been published (Johnson & Goodall, 1979) of a new procedure for ordinating data on the quantities of species in sample areas. The model underlying the procedure assumed that variation in characteristics of the sample areas could be expressed in terms of position along a single axis, while the probability of species presence, and the quantity if present, were related to this variable by a bell-shaped curve. The positions of the sample areas along the X-axis were estimated by a maximum-likelihood procedure, and successive iterations improved both the X estimates and the regressions. The underlying model was very similar to that assumed by Gauch et al. (1974), though these writers did not treat the probability of presence separately, and their fitting procedures were quite different.

It would clearly be appropriate and desirable to generalize this treatment to multiple dimensions. If in fact the variation in floristic composition between sites is dependent on environmental factors, there is no reason to suppose that these could be adequately represented, in the general case, by a single axis (although in particular sets of data a single factor or group of related factors may predominate). The same is true if the underlying model is not necessarily considered as representing environmental variation, but rather as a way of simplifying the relationships among the floristic components. In this case, too, a single dimension for the underlying model imposes an undesirable strait-jacket on the treatment.

Accordingly, the one-dimensional analysis already described had never been considered as anything more than a first step. In fact, where the method was first described (Johnson, 1973), it was envisaged that a single axis would first be fitted, and then residuals of the species quantities expected from this first fit would be used to fit a second axis. Prior to this, an attempt had been made to use the non-linear method of ordination introduced by McDonald (1962), involving polynominal regressions on principal components, but this had been abandoned in favour of direct maximum-likelihood estimation of the positions of samples in the postulated X-space, which seemed more directly

relevant to ecological data.

The successive fitting of residuals was soon abandoned because it became clear that the 'best' single axis, and the regressions on it, would not be related in any simple way to those in a multiple-axis system. Accordingly a more direct approach was adopted, in which the number of axes to be fitted was defined in advance, and the best fit for a model with this number of axes was sought. If one wished to, the result of fitting different numbers of axes could be compared, but each of the fits would be an independent operation.

It was explained in the previous paper that responses of organisms to an environmental factor can usually be represented by a curve which falls away from a maximum to values which become asymptotic to the axis at very high or low values; in other words, the curve is broadly bell-shaped, like a Gaussian curve, though not necessarily symmetrical. The same will also be true in multiple dimensions. There will be some vector of environmental factors permitting optimum development of the species; as any of the factors depart from this optimum, species performance will decline until, where the deviations are great, performance approaches zero.

Where a number of species are competing, the situation may become more complex. If two species have similar environmental optima, but one has a narrower range than the other, its competition in the vicinity of the optimum may lead to a depression in the response curve of the more eurytopic species, which thus may become bimodal. Some of the experimental curves for species mixtures reported by Austin & Austin (1980) suggest this, and some observations in the field are consistent with the idea that bimodal response curves may occur; so far, however, this has not been critically demonstrated. Multimodal curves could also conceivably occur, by an extension of the same argument.

As a first step towards biological reality from the linear models which have dominated ordination studies up to the present, we have selected a multidimensional bell-shaped model, unimodal, and symmetrical in all dimensions, as representing species response to the factors involved in the postulated X-space. Admittedly a model with possibilities of asymmetry and multimodality would be potentially a more adequate representation; but the numerous additional parameters involved in such a model would cause considerable practical difficulties, even though the present procedure could in principle be extended to cover these possibilities.

The multivariate bell-shaped model envisaged could be represented by a multivariate Gaussian curve; but this is only one choice among many (Austin, 1976). Though explanations are available for the appropriateness of the Gaussian curve to represent empirical statistical distributions, there are no physical or biological reasons for preferring it (or any other with similar descriptive properties) as a representation of the response of a species to environmental variables – particularly where these are ex hypothesi expressed as variables of which the scaling is entirely arbitrary. Accordingly we felt free to explore the use of other mathematical forms with the right descriptive properties. Our choice among possible curves was guided partly by computational convenience, but also by the fact that the dependent variables in which we were mainly interested – the probability of a species occurring in a sample, and the proportion of ground it covered if present – were constrained between the limits of 0 and 1, and attainment of neither of these limits was biologically realistic.

The relationship between each dependent variable \( y \) and the vector \( x \) representing the location of the site in the axis system was accordingly given by

\[
y' = a + \sum (b_i x_i + c_i x_i^2)
\]

where

\[
y' = \tan \pi (2y - 1)/2
\]

and where the constants \( c_i \) were negative. The hyperparaboloid fitted to the transformed variable \( y' \) thus could extend indefinitely far into negative values as \( y \) approached zero; and at the maximum, even with stochastic variation around the expected value of the transformed variable, the untransformed \( y \) could never do more than approach the theoretical limit of unity.

The treatment of presence/absence data needed reconsideration in generalizing from a one-dimensional to a multidimensional model. In the former, the sequence of sites along the X axis was divided into small groups of approximately equal size, and the proportion of sites within each group where the species was present was determined. A continuity correction was applied to avoid the frequency values of 0 or 100%, and a curve was then fitted to the frequencies and the mean \( x \) value for the group.

A similar procedure in the multidimensional case