BU-1400-M

USING PRINCIPAL COMPONENTS ANALYSIS AND CORRESPONDENCE ANALYSIS FOR ESTIMATION IN LATENT VARIABLE ECOLOGICAL MODELS

Henry S. Lynn

New England Research Institutes, Inc.

9 Galen Street, Watertown

MA 02172, U.S.A.

and

Charles E. McCulloch

Biometrics Unit, Department of Statistical Science

434 Warren Hall

Cornell University, Ithaca, NY 14853, U.S.A.

This work was partially supported by NSF DMS-9625476

Abstract

Correspondence analysis and principal components analysis are often used as descriptive tools in analyzing multivariate data. In certain applications, there have been attempts to justify the usefulness of these techniques through latent variable models. We assess the theoretical basis for such inference, examining the consistency of estimators of the modes of unimodal response curves defined in generalized linear models. In general, both principal components analysis and correspondence analysis lead to inconsistent estimators. However, with Gaussian responses and appropriate moment conditions on the latent variable, the principal components analysis estimator can be consistent up to a location and scale adjustment when response curves have constant widths. Simulations reveal that for finite sample sizes correspondence analysis can provide reasonable approximations when the modes of the response curves span a large range of the latent variable. But when dealing with incidence data where response curves are more clustered together, principal components analysis may approximate better.

Keywords: correspondence analysis; principal components analysis; ordination; incidental parameters; consistency.

1. Introduction

Principal component analysis (PCA) and correspondence analysis (CA) have a long history as techniques that provide low-dimension summaries of multivariate data. The latter is also mathematically equivalent to canonical correlation analysis, dual-scaling, and reciprocal averaging, with each method originating from a different context (Greenacre 1984). In many applications, the purpose for using PCA or CA is exploratory. In ecology, these methods are used for *ordination* to order sites and species in a manner that describes their ecological relationship, and the sites and species scores generated from these methods are typically plotted graphically. PCA and CA have also been compared to least squares and maximum likelihood estimation in models where the abundances (e.g. biomass, percentage, frequency, or incidence) of different species of flora/fauna are assumed to be specific functions of some latent environmental gradient (Gauch *et al.* 1974, 1977, Goodall and Johnson 1982, Ihm and van Groenewood 1984, Ter Braak 1985). However, most of these comparisons did not examine the properties of the estimators. We address the theoretical basis for statistical inference using PCA and CA. The models and discussion will be motivated within an ecological context since much of the research in modeling originated in this area, but as Ihm and van Groenewood (1984) noted, ordination is applicable to many other disciplines. For example, an economist may model different types of consumption for different socio-geographic groups as a function of a latent living standard index, or a medical researcher may arrange patients with different diagnoses (e.g. depression, chronic anxiety, stomach ulcer) along a hypothetical personality profile; and in archaeology ordination is known as seriation, a process by which artifacts excavated at different sites are typed and chronologically ordered. In section 2, we present a class of models where the expected response is a quadratic function of the latent variable, and also clarify identifiability issues associated with the parameters. Sections 3 and 4 introduce PCA and CA as eigenvector solutions obtained from singular value decompositions. The asymptotic property of the estimators are inspected in terms of consistency in section 5, while their finite sample properties are studied using simulations in section 6. Finally, section 7 suggests a possible alternative approach to the ordination problem.

2. Model

The typical ecological model assumes different species occupy different niches in their habitat and lets species abundances be unimodal functions of the environmental gradient. We model the species abundances, y_{ij}, as Gaussian, Poisson and Bernoulli random variables under the class of generalized linear models, with

$$\eta_{ij} = link(\mu_{ij}) = a_j - \frac{1}{2}\frac{(x_i - u_j)^2}{t_j^2}, \; (i=1,\dots,n \text{ sites}; \; j=1,\dots,m \text{ species}). \qquad (1)$$

Canonical links are assumed for μ_{ij}=E(y_{ij}); i.e. the linear, log, and logit links for the Gaussian, Poisson and Bernoulli cases respectively. a_j is the maximum of the expected response curve, t_j is the tolerance which measures the curve width, u_j is the optimum or mode of the response curve, and x_i can be some hypothetical gradient or it can be some environmental variable which is too expensive to measure or can no longer be measured as in paleontological studies. The Poisson model is often used when dealing with species counts, and the Bernoulli model applies to the case where only the presence/absence of the species is recorded (e.g. observations based on the distinct callings of songbirds which are difficult to locate visually). These two models were used by Ter Braak (1985), while the Gaussian case is included as a pedagogical model to illustrate the derivation of the asymptotic results.

The key feature of (1) is that the x_i's are treated as latent fixed effects to be estimated along with the other parameters, implying in essence a functional measurement error model (Fuller 1987). In ordination, the focus is on u_j's, the species optima, and x_i's, the site parameters. However, these parameters are intrinsically aliased (McCullagh and Nelder 1989) and identifiability constraints must be imposed to obtain unique estimates.

2.1 Identifiability

To illustrate what constraints are required, (1) will be examined as a series of simpler models. Consider first,

$$\eta_{ij} = (x_i - u_j)^2 , \tag{2}$$

which is similar to the ANOVA 2-way classification model ($\mu+\alpha_i+\beta_j$) but without the intercept parameter μ. In the ANOVA model, $n\mu$ is confounded with $\Sigma\alpha_i$ and $\Sigma\beta_j$ since the sum of the indicators vectors for α and β add up to a constant unit vector. Constraints like $\Sigma\hat{\alpha}_i=0$ and $\Sigma\hat{\beta}_j=0$ are usually applied to obtain unique estimates. For (2), only one of the constraints is required (e.g. $\Sigma\hat{x}_i=0$ *or* $\Sigma\hat{u}_j=0$) since there is no intercept parameter, and with the constraint $\hat{x}_i$ and $\hat{u}_j$ are determined up to a sign-change since $(x_i - u_j)$ is squared in (2). Now suppose

$$\eta_{ij} = a_j - (x_i - u_j)^2 . \tag{3}$$

Observe that a_j and u_j are aliased, which implies that for each j either $\hat{a}_j$ or $\hat{u}_j$ must be specified to obtain an unique estimate of the other parameter. Finally, with the additional tolerance parameters in (1), t_j and u_j are also aliased and thus for each j either $\hat{t}_j$ or $\hat{u}_j$ must be specified. (Note also that if all the $\hat{u}_j$'s were specified, then a constraint on the scale of either the $\hat{x}_i$'s or the $\hat{t}_j$'s is also required.) Consequently, a host of constraints are required to obtain unique estimates for (1). For example, by setting $\Sigma \hat{x}_i$=0 and specifying all the $\hat{a}_j$'s and $\hat{t}_j$'s, the $\hat{x}_i$'s and $\hat{u}_j$'s can be estimated up to a sign-change. It should be pointed out that these constraints are merely convenient rules for identifying the estimates, and are not part of the model. Although it does imply that two researchers may arrive at different estimates since there are an infinite number of possible constraints.

With PCA and CA, only the x_i's and u_j's are estimated, while the maxima and tolerances are ignored. This in effect assumes the a_j's and t_j's to be known, implying that the identifiability constraints required on $\hat{x}_i$ and $\hat{u}_j$ are the same as those for (2). However, since the solutions from PCA and CA are themselves only unique up to a scale change, $\hat{x}_i$ and $\hat{u}_j$ will only be identified up to a scale change when using PCA and CA.

3. Principal Component Analysis

PCA provides an orthogonal least squares approximation to the data via a singular value decomposition (Greenacre 1984). The popular method is species-centered PCA (Orlóci 1966), where the mean of each species is subtracted from the columns of the $n \times m$ site-by-species data matrix $\mathbf{Y}=\{y_{ij}\}$. Formally, if $\mathbf{W} \equiv \{y_{ij} - \bar{y}_{.j}\}$ has rank R, the singular value decomposition of $\mathbf{W}$ gives

$$\mathbf{W} = \sum_{r=1}^{R} \lambda_r \mathbf{p}_r \mathbf{q}'_r \text{ , such that} \tag{4}$$

$$\mathbf{WW'p}_r = \lambda_r^2 \mathbf{p}_r \text{ ,} \tag{5}$$

$$\mathbf{W'Wq}_r = \lambda_r^2 \mathbf{q}_r \text{ ,} \tag{6}$$

$$\mathbf{p}'_r \mathbf{p}_l = \mathbf{q}'_r \mathbf{q}_l = \delta_{rl} \text{ , where } \delta_{rl} \text{ is Kronecker's delta.}$$

Typically, the first two terms in (4) are selected and the coordinates $(\lambda_1 p_{1i}, \lambda_2 p_{2i})$ (i=1,...,n) and (q_{1j}, q_{2j}) (j=1,...,m) or their rescaled versions are treated as site and species scores respectively. These scores are then plotted together in a biplot (Gabriel 1971) for a visual appraisal of the relationship between species and sites. Equation (6) is also the usual definition of PCA, which can be derived by finding orthonormal vectors $\mathbf{q}_r$ that maximize the norm of $\mathbf{Wq}_r$.

4. Correspondence Analysis

CA is a multi-faceted technique (Nishisato 1980, Goodman 1986, Van der Heijden *et al.* 1989) dating back to the 1930s. In ecology, it is referred to as "reciprocal averaging" (Hill 1973) and was proposed as an extension to Whittaker's (1967) weighted averaging. Like PCA, CA applies a singular value decomposition to the data, except the data are weighted inversely by the square root of the row and column sums.

Let $\mathbf{C}$=diag($y_{\cdot j}$) be a $m\times m$ diagonal matrix with $y_{\cdot j}=\sum_i y_{ij}>0$, and $\mathbf{R}$=diag($y_{i\cdot}$) a $n\times n$ diagonal matrix with $y_{i\cdot}=\sum_j y_{ij}>0$. CA can be defined by

$$\mathbf{R}^{-1/2}\mathbf{Y}\mathbf{C}^{-1/2} = \sum_{r=1}^{R}\lambda_r\,\mathbf{p}_r\mathbf{q}_r'\text{, such that} \tag{7}$$

$$\mathbf{R}^{-1}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Y}'\mathbf{z}_r = \lambda_r^2\mathbf{z}_r\text{, and} \tag{8}$$

$$\mathbf{C}^{-1}\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y}\mathbf{v}_r = \lambda_r^2\mathbf{v}_r\text{,} \tag{9}$$

where $\mathbf{z}_r=\lambda_r\mathbf{R}^{-1/2}\mathbf{p}_r$, $\mathbf{v}_r=\mathbf{C}^{-1/2}\mathbf{q}_r$ are the site and species scores respectively. Using (7), the scores can be combined to give

$$\mathbf{z}_r = \mathbf{R}^{-1}\mathbf{Y}\mathbf{v}_r \text{ and } \lambda_r^2\mathbf{v}_r = \mathbf{C}^{-1}\mathbf{Y}'\mathbf{z}_r\text{, or}$$

$$z_{ri} = \frac{\sum_j y_{ij}v_{rj}}{\sum_j y_{ij}} \text{ and } \lambda_r^2 v_{rj} = \frac{\sum_i y_{ij}z_{ri}}{\sum_i y_{ij}}\,. \tag{10}$$

According to (10), site scores are weighted averages of the species scores and vice versa, hence the term "reciprocal averaging". It is easily verified that a solution to (10) is λ_1=1 with the corresponding site scores and species scores all equal to 1. In fact, Hill (1974) showed that λ_1 is the largest singular value. This trivial or zero order solution is generally ignored, and site and species scores are obtained from the higher order solutions.

4.1 Approximations

Ecologists have found CA to provide reasonable descriptions of the data when responses are unimodal, and this may be attributed to several explanations. Lancaster (1957) showed that when two variables X_1 and X_2 have a bivariate normal density, the CA scores correspond to Tchebycheff-Hermite polynomials of X_1 and X_2: the first order CA solutions are first order Hermite polynomials (which are x_i and u_j), and the second order CA solutions are second order Hermite polynomials (which are proportional to x_i^2 and u_j^2). In our model, x_i and u_j are not random variables but, for Poisson counts, μ_{ij} has a similar form as the bivariate normal density; i.e. $\mu_{ij} \propto \exp(-(x_i - u_j)^2)$. Therefore we may expect the first order CA scores to approximate x_i and u_j for the Poisson case. This also explains why a plot of the first and second order solutions of CA can exhibit an arch shape, sometimes referred to as the horseshoe effect.

There are also similarities between CA and ML estimation in our generalized linear models. The likelihood equations for x_i and u_j in model (1) can be rearranged as:

$$x_i = \sum_j \frac{y_{ij} u_j}{t_j^2} \Big/ \sum_j \frac{y_{ij}}{t_j^2} + \left[\sum_j \frac{(x_i - u_j)\mu_{ij}}{t_j^2} \Big/ \sum_j \frac{y_{ij}}{t_j^2} \right] \text{ and}$$

$$u_j = \sum_i \frac{y_{ij} x_i}{y_{i\cdot}} - \left[\sum_i \frac{(x_i - u_j)\mu_{ij}}{y_{i\cdot}} \right].$$

Ter Braak (1985) indicated that these reduce to the reciprocal averaging equations in (10) when t_j is constant across species and the terms in the square brackets are negligible. The latter condition holds under equal maxima, and uniformly distributed species optima (site scores) over an interval which is large compared to the range of site scores (species optima). Although Ter Braak (1987) further noted that the u_j's and x_i's cannot simultaneously be uniformly distributed about intervals beyond each others' range.

More importantly, it may not be desirable to approximate the maximum likelihood estimator since it can be inconsistent. In the scenario where the number of species m is fixed but the number of sites n increases with sample size, the number of site parameters increases with n, leading to the well known incidental parameters problem (Neyman and Scott 1948).

5. Consistency of Estimators

In the following sections, we examine the consistency of estimators of the species optima from PCA and CA assuming m fixed and $n\uparrow\infty$. This assumption seems appropriate in many ecological studies where there are a finite number of species but conceivably a much larger sampling frame. The x_i's are treated as nuisance parameters since, unlike the species parameters, they are not required to define the functional form of the predicted model. Ecological applications sometimes only require an accurate ordering of the species, therefore it suffices to consistently estimate $\mathbf{u}=(u_1,\ldots,u_m)'$ up to a location and scale change. In fact, we can at best only consistently estimate $\mathbf{u}$ up to a scale change when using PCA and CA. For both PCA and CA, consistency will first be examined theoretically via their exact or approximate eigenvector solutions, and the conclusions are then verified numerically.

5.1 PCA Estimator

The PCA estimator of $\mathbf{u}$ is taken to be the dominant eigenvector $\hat{\mathbf{w}}$ associated with the largest eigenvalue, $\lambda_{\max}$, of $\mathbf{W'W}$ defined in (6) with (j,k)th element equal to

$$\sum_i y_{ij}y_{ik} - \frac{1}{n}\sum_i y_{ij}\sum_i y_{ik}, \quad (j,k=1,\ldots,m). \tag{11}$$

To examine the asymptotic property of $\hat{\mathbf{w}}$, we first establish the conditions that $n^{-1}\mathbf{W'W}\to\Sigma$. If the y_{ij}'s are independent Gaussian random variables with mean μ_{ij} and variance σ_e^2, then

$$Var(\sum_i y_{ij}) = n\sigma_e^2 = o(n^2), \text{ and}$$

$$Var(\sum_i y_{ij}y_{ik}) = \sum_i \{\mu_{ij}^2\,Var(y_{ik}) + \mu_{ik}^2\,Var(y_{ij}) + Var(y_{ij})Var(y_{ik})\}$$

$$\le n\sigma_e^2(a_j^2 + a_k^2 + \sigma_e^2) = o(n^2);$$

if the y_{ij}'s are independent Poisson counts, then

$$Var(\sum_i y_{ij}) = \sum_i \mu_{ij} \le n\exp(a_j) = o(n^2), \text{ and}$$

$$Var(\sum_i y_{ij}y_{ik}) = \sum_i (\mu_{ij}^3 + \mu_{ik}^3 + \mu_{ij}\mu_{ik}) \le n(\exp(3a_j) + \exp(3a_k) + \exp(a_j+a_k)) = o(n^2);$$

and lastly if the y_{ij}'s are independent Bernoulli variables then y_{ij} and $y_{ij}y_{ik}$ are uniformly bounded. Therefore by the Weak Law of Large Numbers (WLLN)

$$\frac{1}{n}\sum_i y_{ij}y_{ik} \to \frac{1}{n}\sum_i E(y_{ij}y_{ik}) \text{ and } \frac{1}{n}\sum_i y_{ij} \to \frac{1}{n}\sum_i E(y_{ij}).$$

It follows according to (11) that $n^{-1}\mathbf{W'W}$ converges to Σ which has (j,k)th element equal to

$$\frac{1}{n}\sum_i (\,Var(y_{ij})\delta_{jk} + \mu_{ij}\mu_{ik}) - \frac{1}{n}\sum_i \mu_{ij}\frac{1}{n}\sum_i \mu_{ik}. \tag{12}$$

When $\lambda_{\max}$ of $\mathbf{W'W}$ is unique (i.e. has multiplicity one), $\hat{\mathbf{w}}$ is a continuous function of the elements of $\mathbf{W'W}$ (Ortega 1972 p.45). Let $\mathbf{w}$ be the dominant eigenvector of Σ satisfying the equation $\lambda_w\mathbf{w} = \Sigma\mathbf{w}$. If λ_w is also unique, then $\hat{\mathbf{w}}$ converges to $\mathbf{w}$ the dominant eigenvector of Σ. We declare $\hat{\mathbf{w}}$ to be a consistent (up to a scale and location change) estimator of $\mathbf{u}$ when $\mathbf{w}$ satisfies the equation $\lambda\mathbf{w} = k_1\mathbf{1} + k_2\mathbf{u}$ or equivalently

$$\lambda w_j = k_1 + k_2 u_j \quad (j=1,\dots,m), \tag{13}$$

where k_1, $k_2 \in \Re$, $k_2 \neq 0$, and $\mathbf{1}$ is a $m\times 1$ unit vector. This implies that asymptotically $\hat{\mathbf{w}}$ has perfect correlation with $\mathbf{u}$. Conversely, if $\lambda\mathbf{w} \neq c_1\mathbf{1}+c_2\mathbf{u}$, then $\hat{\mathbf{w}}$ cannot be a consistent estimator of $\mathbf{u}$.

Without loss of generality, consider the jth element of $\mathbf{w}$. Using (12), it can be shown after some algebraic simplifications that for the Gaussian case:

$$\begin{aligned}\lambda_w w_j &= \sum_k w_k\,\{\sigma_e^2\delta_{jk} + b_jb_k(\mathrm{v}_x - 2\mathrm{c}_x(u_j + u_k) + 4\sigma_x^2 u_ju_k)\}\\ &= \sigma_e^2 w_j + b_j(\mathrm{v}_x\sum_k b_k\, w_k - 2\mathrm{c}_x\sum_k b_k\, u_kw_k) + b_ju_j(4\sigma_x^2\sum_k b_k\, u_kw_k - 2\mathrm{c}_x\sum_k b_k\, w_k),\end{aligned} \tag{14}$$

where $\mathrm{v}_x = \dfrac{\Sigma x_i^4}{n} - \dfrac{(\Sigma x_i^2)^2}{n^2}$, $\mathrm{c}_x = \dfrac{\Sigma x_i^3}{n} - \dfrac{\Sigma x_i}{n}\dfrac{\Sigma x_i^2}{n}$, $\sigma_x^2 = \dfrac{\Sigma x_i^2}{n} - \dfrac{(\Sigma x_i)^2}{n^2}$, and $b_j = \dfrac{1}{2t_j^2}$.

For the Poisson case, we have

$$\begin{aligned}\lambda w_j = \frac{1}{n}\sum_i w_j \exp(a_j - b_j(x_i - u_j)^2) &+ \frac{n-1}{n^2}\sum_i \{\exp(a_j - b_j(x_i - u_j)^2)\sum_k w_k \exp(a_k - b_k(x_i - u_k)^2)\} - \\ &\frac{1}{n^2}\sum_{i\neq i'}\{\exp(a_j - b_j(x_i - u_j)^2)\sum_k w_k \exp(a_k - b_k(x_{i'} - u_k)^2)\}.\end{aligned} \tag{15}$$

While for the Bernoulli case,

$$\lambda w_j = \frac{1}{n}\sum_i w_j r_{ij}(1+r_{ij})^{-2} + \frac{n-1}{n^2}\sum_i (1+r_{ij})^{-1}\sum_k w_k (1+r_{ik})^{-1} -$$

$$\frac{1}{n^2}\sum_{i\neq i'}(1+r_{ij})^{-1}\sum_k w_k (1+r_{i'k})^{-1}, \text{ where } r_{ij} = \exp(-a_j + b_j(x_i - u_j)^2). \qquad (16)$$

Observe that for all three distributions, the asymptotic eigenvector equations (14), (15) and (16) do not have the simple linear form as in (13), suggesting that in general $\hat{\mathbf{w}}$ is inconsistent. However, for the Gaussian case, the eigenvector equation (14) can be linear in u_j if the tolerances are all equal. The result is stated in the following proposition:

Proposition. If $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|x_i^r| < \infty$ for $r=\{1,2,3,4\}$ and $\sigma_x^2\neq 0$, $b_j=b$ for all j, $\lambda_{\max}$ and λ_w (the largest eigenvalues of $\mathbf{W'W}$ and Σ respectively) are unique, and $\lambda_w > \sigma_e^2$, then

$$n^{-1}\hat{\mathbf{w}} \to c_1\mathbf{1} + c_2\mathbf{u},$$

where $c_1 = \mathrm{v}_x\mathbf{1}'\mathbf{w} - 2\mathrm{c}_x\mathbf{u}'\mathbf{w}$, and $c_2 = 2\sigma_x^2\mathbf{u}'\mathbf{w} - \mathrm{c}_x\mathbf{1}'\mathbf{w}$. Furthermore, if

$$2\sigma_x^2\mathbf{u}'\mathbf{w} \neq \mathrm{c}_x\mathbf{1}'\mathbf{w},$$

then $c_2\neq 0$ and $\hat{\mathbf{w}}$ the PCA estimator is consistent up to a location and scale change.

Proof. When all the tolerances t_j are equal, $b_j=b$ for all j and (14) implies that

$$\Sigma = \sigma_e^2\mathbf{I} + b^2\mathbf{1}(\mathrm{v}_x\mathbf{1}' - 2\mathrm{c}_x\mathbf{u}') + b^2\mathbf{u}(4\sigma_x^2\mathbf{u}' - 2\mathrm{c}_x\mathbf{1}'). \qquad (17)$$

Let $\Sigma_1=\Sigma-\sigma_e^2\mathbf{I}$. Observe that Σ_1 has rank 2, and thus it has 2 non-trivial eigenvectors and m–2 trivial eigenvectors with zero as their eigenvalue. Since

$$\Sigma_1\mathbf{w} = \Sigma\mathbf{w} - \sigma_e^2\mathbf{w} = (\lambda_w - \sigma_e^2)\mathbf{w},$$

the eigenvectors of Σ are equivalently those of Σ_1, and likewise Σ has m–2 trivial eigenvectors with the same eigenvalue σ_e^2. Therefore, if $\lambda_w > \sigma_e^2$, its associated eigenvector $\mathbf{w}$ is non-trivial. Moreover, $\mathbf{w}$ is proportional to

$$\Sigma_1\mathbf{w} = c_1\mathbf{1} + c_2\mathbf{u}, \text{ where } c_1 = \mathrm{v}_x\mathbf{1}'\mathbf{w} - 2\mathrm{c}_x\mathbf{u}'\mathbf{w},\ c_2 = 2\sigma_x^2\mathbf{u}'\mathbf{w} - \mathrm{c}_x\mathbf{1}'\mathbf{w}.$$

It follows that $n^{-1}\hat{\mathbf{w}} \to c_1\mathbf{1}+c_2\mathbf{u}$, which implies that $\hat{\mathbf{w}}$ is a location and scale consistent estimator of $\mathbf{u}$ when $c_2\neq 0$.

There are two interesting notes concerning this proposition. Firstly, the conditions

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|x_i^r| < \infty \text{ for } r=\{1,2,3,4\}$$

ensure that the terms v_x , c_x , and σ_x^2 are finite such that Σ is defined. These are similar to conditions used to establish consistency in functional measurement error models (Fuller 1987). Intuitively, this suggests that we regard x_i as coming from some population with finite moments, which is reminiscent of Keifer and Wolfowitz's (1956) approach. Secondly, the proof of the proposition actually showed that the two non-trivial eigenvectors of Σ are both linear in $\mathbf{u}$ when their associated eigenvalues are greater than σ_e^2. This implies that both the dominant (i.e. first) eigenvector and also the second eigenvector of $\mathbf{W'W}$ are consistent (up to a location and scale change) estimators of $\mathbf{u}$.

In a special case of the proposition, consistency up to a scale change can be established with the additional constraints that the x_i's be symmetric about zero and the u_j's sum to zero.

Corollary. If c_x=0, $\mathbf{u'1}$=0, and $mv_x < 4\sigma_x^2\mathbf{u'u}$, then $\hat{\mathbf{w}}$ is consistent up to a scale change.

Proof. When c_x=0, (17) implies that

$$\Sigma = \sigma_e^2\mathbf{I} + b^2\mathbf{1}(v_x\mathbf{1}') + b^2\mathbf{u}(4\sigma_x^2\mathbf{u}').$$

Thus with $\mathbf{u'1}$=0, the two non-trivial eigenvectors of Σ are easily verified to be $\mathbf{1}$ and $\mathbf{u}$ with eigenvalues $mb^2v_x + \sigma_e^2$ and $4b^2\sigma_x^2\mathbf{u'u} + \sigma_e^2$ respectively.

The above results imply that if one only requires consistency up to a location and scale change then both the first *and* second PCA eigenvectors are consistent under the conditions of the proposition. However, if consistency up to a scale change is required then either the first *or* the second PCA eigenvector is consistent under the conditions of the corollary. In particular, if $mv_x > 4\sigma_x^2\mathbf{u'u}$ then the second (rather than the first) PCA eigenvector should be used to estimate $\mathbf{u}$. Likewise, we will explore the performance of both the first and the second PCA eigenvectors in the simulations since in certain situations one may be preferred over the other.

We stated earlier that the eigenvector equations (14), (15) and (16) suggest that $\mathbf{w}$ cannot be linear in $\mathbf{u}$, implying that $\hat{\mathbf{w}}$ is inconsistent for model (1). This can also be verified numerically by calculating $\mathbf{w}$ for selected values of the parameters since $\mathbf{w}$ is simply the eigenvector of Σ given in (12), which is a function of μ_{ij}. However, the expression for Σ in (12) cannot be used directly since it depends on n, and we are interested in its numerical value as $n\uparrow\infty$. Therefore, in order to remove the dependency on n, we assume $x_i \sim$ iid $N(0,1)$ and essentially apply a conditioning argument. This assumption is also congruent with the "moment" conditions that the x_i's must satisfy to establish consistency. It can then be shown that the (j,k)th element Σ equals

$$\sigma_e^2\delta_{jk} + 2b_jb_k(1 + u_ju_k)$$

for the Gaussian case. For Poisson counts, the moment generating function for non-central chi-square distributions can be used to show that Σ has (j,k)th element

$$s_j\delta_{jk} + \exp(a_j - b_ju_j^2 + a_k - b_ku_k^2)\exp(d_{jk}/c_{jk} - d_{jk})(1 - c_{jk})^{-1/2} + s_js_k\,,$$

where $s_j = \exp(a_j)\exp(-b_ju_j^2(1 + 2b_j)^{-1})(1 - 2b_j)^{-1/2}$, $c_{jk} = 1 + 2(b_j + b_k)$, and

$$d_{jk} = (b_ju_j + b_ku_k)^2(-b_j - b_k)^{-1}.$$

In the Bernoulli case, there is no closed form solution but the components of Σ can be calculated using numerical integrations. For example,

$$\frac{1}{n}\sum_i E(y_{ij}) = \frac{1}{n}\sum_i E_X[E(y_{ij}\,|\,x_i)] = E_X[(1 + \exp(a_j - b_j(X - u_j)^2))^{-1}],$$

where $E_X(\cdot)$ is the expectation with respect to the standard Gaussian density. For the Poisson and Bernoulli cases, we set m=31, $\mathbf{a}=E(0.5,2)$, $\mathbf{b}=E(1,2.5)$, and $\mathbf{u}=E(-2,2)$, where $E(p,q)$ denotes a vector of equally spaced elements between (p,q). While for the Gaussian case, σ_e^2=1, $\mathbf{a}=E(30,32.25)$, $\mathbf{b}=E(1,4)$, and $\mathbf{u}=E(0,0.4)$. Gaussian and Uniform random variables were generated using built-in functions in GAUSS (Aptech Systems 1992) and numerical integration was performed using Simpson's method. The results are displayed in Figures 1a, 1b and 1c, where the residuals $w_j - u_j$ are plotted against the true values u_j. If $\mathbf{w}$ is linear in $\mathbf{u}$ then the plot would show a straight line. However, the PCA solutions produced curvilinear plots for all three distributions, confirming that $\hat{\mathbf{w}}$ is inconsistent.

5.2 CA Estimator

Before proceeding, we point out that $y_{i\cdot}$ and $y_{\cdot j}$ must be positive according to the definition of CA in (7), yet under model (1) there is a nonzero probability that $y_{i\cdot}$ and $y_{\cdot j}$ can be non-positive. Thus, in order to inspect the consistency of the CA estimator, we assume that y_{ij} has expectation and variance according to the specifications in model (1) but conditional on $y_{i\cdot}$ and $y_{\cdot j}$ being positive. Using these assumptions, we suggest by analogy that the asymptotic solution of the CA estimator is not linear in $\mathbf{u}$ by showing that its Taylor's series approximation cannot be written as $k_1\mathbf{1} + k_2\mathbf{u}$, $(k_1,k_2)\in\Re$.

Let $\hat{\mathbf{v}}$ be the CA estimator of $\mathbf{u}$ obtained from the first order solution in (9); i.e. the eigenvector associated with the second largest eigenvalue in (9). When the eigenvalue of $\hat{\mathbf{v}}$ is unique and $\mathbf{C}^{-1}\mathbf{Y'R}^{-1}\mathbf{Y}\to\Sigma$, $\hat{\mathbf{v}}$ converges to $\mathbf{v}$, the second eigenvector of Σ. First, we establish the conditions for $\mathbf{C}^{-1}\mathbf{Y'R}^{-1}\mathbf{Y}$ to converge to Σ. The (j,k)th element of $\mathbf{C}^{-1}\mathbf{Y'R}^{-1}\mathbf{Y}$ equals

$$\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}y_{\cdot j}} = \frac{n}{\sum_i y_{ij}}\frac{1}{n}\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}}.$$

Let $y_{i[m]}$ be the largest element in the series $y_{i\cdot}$. Since $y_{i\cdot}$ is by definition positive, $y_{i[m]}$ is also positive. This implies that for all three distributions

$$Var(\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}}) \le \sum_i E(\frac{y_{ij}^2y_{ik}^2}{y_{i\cdot}^2}) \le \sum_i E(\frac{y_{ij}^2y_{ik}^2}{y_{i[m]}^2}) \le \sum_i E(y_{ij}^2) \le o(n^2).$$

Therefore, by the WLLN and using a Taylor's series expansion,

$$\sum_i \frac{y_{ij}y_{ik}}{y_{i\cdot}y_{\cdot j}} \to \frac{1}{\sum_i E(y_{ij})}\sum_i E(\frac{y_{ij}y_{ik}}{y_{i\cdot}}) = \sum_i \frac{E(y_{ij}y_{ik})}{E(y_{\cdot j})E(y_{i\cdot})} + \sum_i R_{ijk}, \qquad (18)$$

where R_{ijk} is the remainder term. It follows that the jth element of $\mathbf{v}$ satisfies

$$\lambda v_j = \frac{1}{E(y_{\cdot j})}\sum_k v_k \sum_i \frac{Var(y_{ij})\delta_{jk} + \mu_{ij}\mu_{ik}}{E(y_{i\cdot})} + \sum_k v_k \sum_i R_{ijk}. \qquad (19)$$

The first term on the right hand side of (19) equals

$$\frac{1}{\mu_{\cdot j}}(v_j\sum_i \frac{\sigma_e^2}{\mu_{i\cdot}} + \sum_i \frac{1}{\mu_{i\cdot}}(a_j-b_j(x_i-u_j)^2)\sum_k v_k\,(a_k-b_k(x_i-u_k)^2))$$

for the Gaussian case,

$$\frac{1}{\mu_{\cdot j}}\left(\sum_i \frac{1}{\mu_{i\cdot}} \exp(a_j - b_j(x_i - u_j)^2)\right)(\mathrm{v}_j + \sum_k \mathrm{v}_k \exp(a_k - b_k(x_i - u_k)^2))$$

for the Poisson case, and

$$\frac{1}{\mu_{\cdot j}}(\mathrm{v}_j \sum_i \frac{1}{\mu_{i\cdot}} r_{ij}(1 + r_{ij})^{-2} + \sum_i \frac{1}{\mu_{i\cdot}}(1 + r_{ij})^{-1} \sum_k \mathrm{v}_k (1 + r_{ik})^{-1})$$

for the Bernoulli case, with $\mu_{\cdot j} = \sum_i \mu_{ij}$, $\mu_{i\cdot} = \sum_j \mu_{ij}$, and μ_{ij} defined according to the three distributions respectively. The above three expressions are all nonlinear functions of u_j, which suggest that, if the first order Taylor's series approximation were exact, v_j cannot be linear in u_j and likewise $\hat{\mathbf{v}}$ cannot be a consistent estimator of $\mathbf{u}$.

The hypothesis that $\hat{\mathbf{v}}$ is inconsistent is now verified by numerically calculating $\mathbf{v}$ under the assumption that $x_i \sim$ iid $N(0,1)$. According to (18), Σ has (j,k)th element

$$\frac{n}{\sum_i E(y_{ij})} \frac{1}{n} \sum_i E(\frac{y_{ij} y_{ik}}{y_{i\cdot}}).$$

The first expectation can be obtained using the same methods as described above for the numerical calculations of the PCA solutions, while the second expectation can be calculated using Monte Carlo integration. The same parameter inputs for calculating the PCA solutions are used here again, and Poisson and Bernoulli random variables were simulated using the inverse transform method. As shown in Figures 1a, 1b and 1c, the residuals $\mathrm{v}_j - u_j$ are nonlinear functions of v_j, indicating that the CA estimator $\hat{\mathbf{v}}$ is inconsistent for model (1).

6. Simulations

We compare the finite sample behavior of the CA and PCA estimates under different configurations of the parameter inputs. The focus is on the Poisson and Bernoulli cases since count data and incidence data are more common in ordination. We let m=31, n=41, $\mathbf{u}$ equal $E(0,0.4)$, $E(0,1.2)$, $E(-0.2,0.2)$, or $E(-0.6,0.6)$, $\mathbf{x}$ equal $E(-1,1)$ or $E(-1.5,1.5)$, $a_j \sim \mathrm{U}(0.5,1.25)$ or $\mathrm{U}(0.5,2.0)$, and $b_j \sim \mathrm{U}(1,4)$ or $\mathrm{U}(1,7)$, where $\mathrm{U}(p,q)$ is the discrete Uniform distribution with equally spaced elements between (p,q). The results are summarized in terms of the correlation between $\mathbf{u}$ and its estimate, which also indicates how well the ranks of the estimates approximate the ranks of the u_j's. The entries in tables 1 and 2 represent averages of 10 runs of 1000 replications each, with each run being one independent draw from the Uniform distributions. Overall, the largest effect on the correlations comes from changing the variance of the u_j's, while altering the variability of the other parameters has smaller impacts. The correlations are highest when the u_j's have large variance relative to the variance of the x_i's; e.g. $\mathbf{u}=E(-0.6,0.6)$ or $E(0,1.2)$. Plots of the response curves corresponding to these two cases are presented in Figures 2a and 2b respectively. In contrast, when $\mathbf{u}=E(-0.2,0.2)$ or $E(0,0.4)$ the u_j's have small variance relative to the x_i's and the response curves are clustered together (e.g. Figure 2c and 2d), resulting in weaker correlations. Changing the location of the u_j's seems to have negligible effect on the CA estimates. But for PCA, the behavior of its first and second eigenvectors switched when $\mathbf{u}=E(-0.6,0.6)$. In particular, the first PCA eigenvector has higher correlations when $\mathbf{x}=E(-1,1)$ but the second PCA eigenvector has higher correlations when $\mathbf{x}=E(-1.5,1.5)$. This may be compared to the Corollary in section 5.1, which showed that in certain circumstances the second eigenvector may be the preferred estimator.

In terms of the comparative performance between CA and PCA, the CA estimates have stronger correlations when the u_j's have large variance. However, interestingly, the PCA method can also produce reasonable estimates. The PCA second eigenvector did better in situations when the first eigenvector performed poorly. Their correlations are slightly weaker compared to those for the CA estimates, but can be stronger when the u_j's have small variance. The advantage of PCA becomes more pronounced in the Bernoulli model, where the second PCA eigenvector usually does better than the CA estimator when the u_j's have small variance, and is comparable to the CA estimator when the u_j's have large variance.

7. Discussion

Although the physical concept of a "latent" variable may be debatable, latent variable models are used extensively in statistics especially in applications of economics, sociology and psychology where factor analysis and structural equation models abound. In ecology, latent variable models have been introduced to better understand and compare with PCA and CA. For example, PCA is equivalent to ML estimation for a linear Gaussian latent variable model. Here, we focus on examining how PCA and CA perform when used for estimation in the quadratic latent variable model in (1). Two points are important in this progression from treating PCA and CA as descriptive tools to more inferential methods. Firstly, the problem of overparameterization or non-identifiability of the estimates needs to be addressed. As discussed earlier, unique estimates cannot be obtained for model (1) unless a host of constraints are imposed on the estimates. Secondly, under the scenario with m fixed and $n\uparrow\infty$, certain "moment" conditions must be imposed on the latent parameters before consistency can be ascertained for the estimators of $\mathbf{u}$. This incidental parameters problem is a consequence of treating the latent parameters as fixed effects. However, when sites are a random sample of ecological environments it may be reasonable to treat them as random effects, and apply methods used in analyzing generalized mixed effects models (McCulloch 1997). In addition, for some applications where m and n can conceptually both increase to infinity, consistency can also be attained without requiring any conditions on the x_i's (Haberman 1977, Portnoy 1988).

This research provides guidelines in deciding which estimator to consider when estimating $\mathbf{u}$, the modes or optima of the response curves. The PCA estimator is in general inconsistent, but can be a location and scale consistent for the Gaussian model when the b_j's are constant. Algebraically, this is similar to Kooijman's (1977) derivation of $\mathbf{u}$ as the PCA solution of the interaction matrix,

$$\mathbf{Z'Z}, \text{ where } \mathbf{Z}=\{z_{ij} - \bar{z}_{i\cdot} - \bar{z}_{\cdot j} + \bar{z}_{\cdot\cdot}\} \text{ and } z_{ij} = a_j - b(x_i - u_j)^2.$$

It is noteworthy and perhaps surprising that the PCA estimator can be consistent for a quadratic model since PCA implicitly assumes a linear model. On the other hand, although CA scores have been proposed as approximations to $\mathbf{x}$ and $\mathbf{u}$, the CA estimator is shown to be inconsistent. Nevertheless, consistency may be less important to ecologists, and for finite sample sizes the PCA and CA estimators can be satisfactory if one wishes only to obtain some approximate ordering of the u_j's. CA performs well when the optima of the different response curves are spread out over a large range of the x_i's, an assumption known as the "species packing model" in ecology (Gauch *et al.* 1974, Ter Braak 1985). However, when dealing with incidence data where response curves are clustered together, the PCA estimator (second eigenvector) may provide a better approximation. The advantage of PCA on incidence data has also been pointed out by Hill (1974), although he stated that "the reasons why the method is successful with such data have not been made clear". In situations where the response curves are very tightly clustered together, both CA and PCA provide poor approximations and are not recommended.

In section 5.2, we mentioned that model (1) is strictly speaking not suitable for CA. It would be more compatible to work with the conditional distributions $F(y_{ij}|y_{i\cdot}>0,\ y_{\cdot j}>0)$, although such a formulation introduces undue complications when examining the consistency of the estimators. Furthermore, it seems unproductive to apply this to PCA since it does not require $y_{i\cdot}$ or $y_{\cdot j}$ to be positive. Likewise, our approach is to use a previously adopted and relatively simple model (1) as a working model to examine the behavior of the estimators. Finally, the proposed ecological models assume species act independently of each other, when in reality correlation might be expected due to symbiosis or competition effects or it can be induced through site effects. Such non-independence increases the complexity of the problem, and further research is warranted since for categorical data there is less of a consensus as to how associations should be modeled and which method of estimation is most appropriate.

REFERENCES

Aptech Systems (1992), *Gauss Manual*, Version 3.0, Kent, WA: Author.

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley & Sons.

Gabriel, K. R. (1971), "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis," *Biometrika*, 58, 453-467.

Gauch, H. G., Chase, G. B. and Whittaker, R. H. (1974), "Ordinations of Vegetation Samples by Gaussian Species Distributions," *Ecology*, 55, 1382-1390.

Gauch, H. G., Whittaker, R. H., and Wentworth, T. R. (1977), "A Comparative Study of Reciprocal Averaging and Other Ordination Techniques," *Journal of Ecology*, 65, 157-174.

Goodall, D. W., and Johnson, R. W. (1982), "Non-linear Ordination in Several Dimensions. A Maximum Likelihood Approach," *Vegetatio*, 48, 197-208.

Goodman, L. A. (1986), "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-linear Models Approach in the Analysis of Contingency Tables," *International Statistics Review*, 54, 243-309.

Greenacre, M. J. (1984), *Theory and Application of Correspondence Analysis*, London: Academic Press.

Haberman, S. J. (1977), "Maximum Likelihood Estimation in Exponential Response Models," *The Annals of Statistics*, 5, 815-841.

Hill, M. O. (1973), "Reciprocal Averaging: an Eigenvector Method of Ordination," *Journal of Ecology*, 61, 237-249.

Hill, M. O. (1974), "Correspondence Analysis: a Neglected Multivariate Approach," *Applied Statistics*, 23, 340-354.

Ihm, P., and Van Groenewoud, H. (1984), "Correspondence Analysis and Gaussian Ordination," *COMPSTAT lectures*, 3, 5-60.

Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters," *Annals of Mathematical Statistics*, 27, 887-906.

Kooijman, S. A. L. M. (1977), Species Abundance with Optimum Relations to Environmental Factors," *Annals of System Research*, 6, 123-138.

Lancaster, H. O. (1957), "Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table," *Biometrika*, 44, 289-292.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models, 2nd Edition*, London: Chapman and Hall.

McCulloch, C. E. (1997), "Maximum Likelihood Algorithms for Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 92, 162-170.

Neyman, J., and Scott, E. L. (1948), "Consistent Estimators Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.

Nishisato, S. (1980), *Analysis of Categorical Data: Dual Scaling and Its Applications*, Toronto: University of Toronto Press.

Orlóci, L. (1966), "Geometric Models in Ecology. I. The Theory and Application of Some Ordination Methods," *Journal of Ecology*, 54, 193-215.

Ortega, J. M. (1972), *Numerical Analysis: A Second Course*, New York: Academic Press:

Portnoy, S. (1988), "Asymptotic Behavior of Likelihood Methods for Exponential Families When the Number of Parameters Tends to Infinity," *The Annals of Statistics*, 16, 356-366.

Ter Braak, C. J. F. (1985), "Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Model," *Biometrics*, 41, 859-873.

Ter Braak, C. J. F. (1987), "Ordination". In: *Data Analysis in Community and Landscape Ecology* (Ed. by R. H. G. Jongman, C. J. F. Ter Braak, and O. F. R. Van Tongeren), pp.91-173. Wageningen: Pudoc.

Van der Heijden, P. G. M., Falguerolles, A. D., and Leeuw, J. D. (1989), "A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Log-linear Analysis," *Applied Statistics*, 38, 249-292.

Whittaker, R. H. (1967), "Gradient Analysis of Vegetation," *Biological Reviews*, 49, 207-264.

Figures 1a, 1b. Asymptotic solutions of the CA (solid lines) and PCA (dotted lines) estimators for the Gaussian and Poisson models.

Figure 1a. Plot of PCA, CA Asymptotic Solutions for Gaussian Model

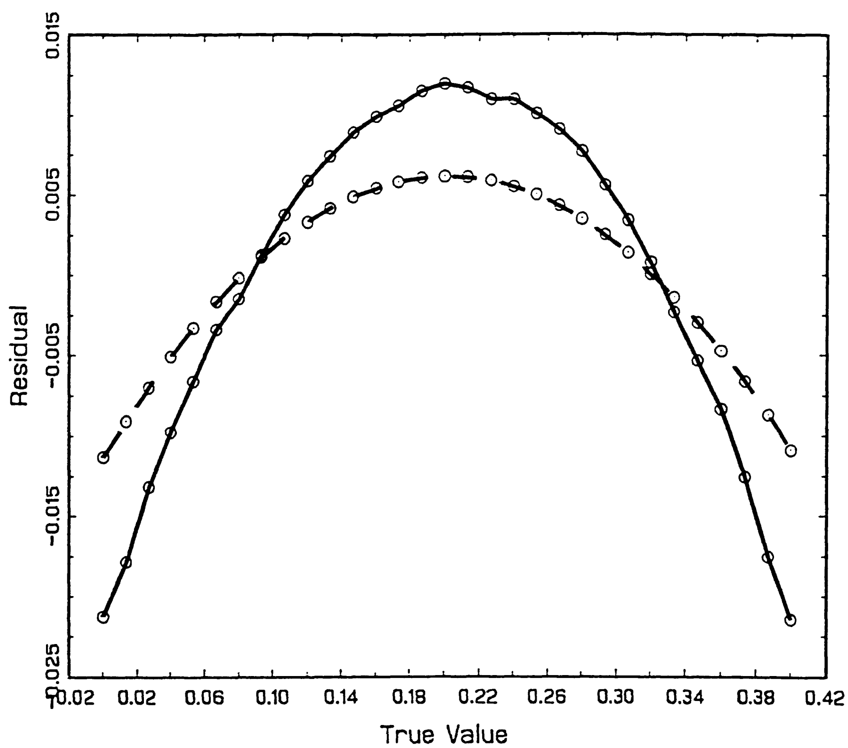


Figure 1b. Plot of PCA, CA Asymptotic Solutions for Poisson Model

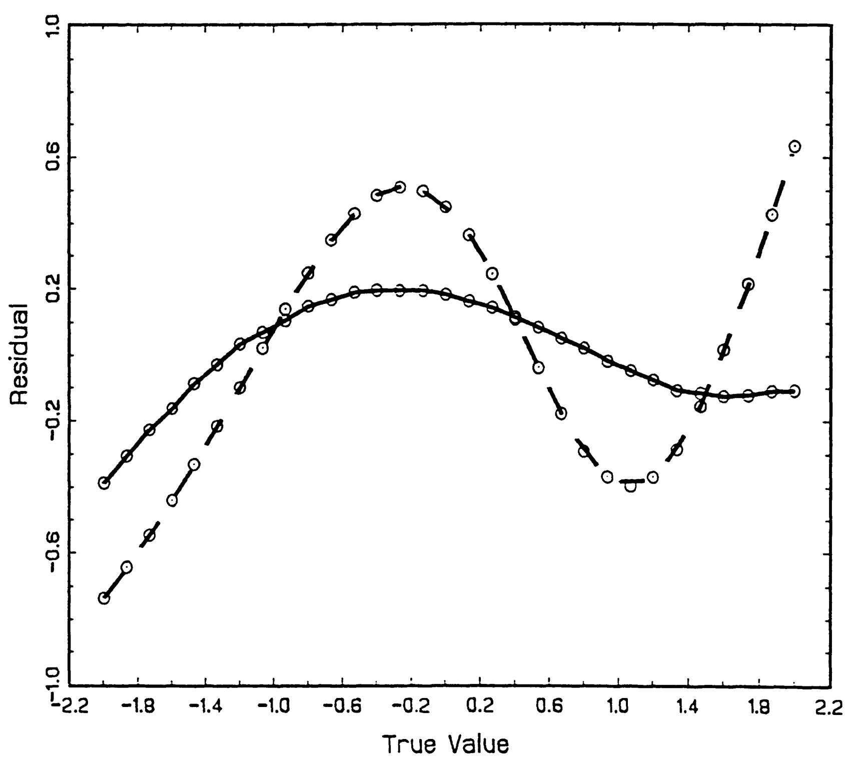


Figure 1c. Asymptotic solutions of the CA (solid lines) and PCA (dotted lines) estimators for the Bernoulli model.

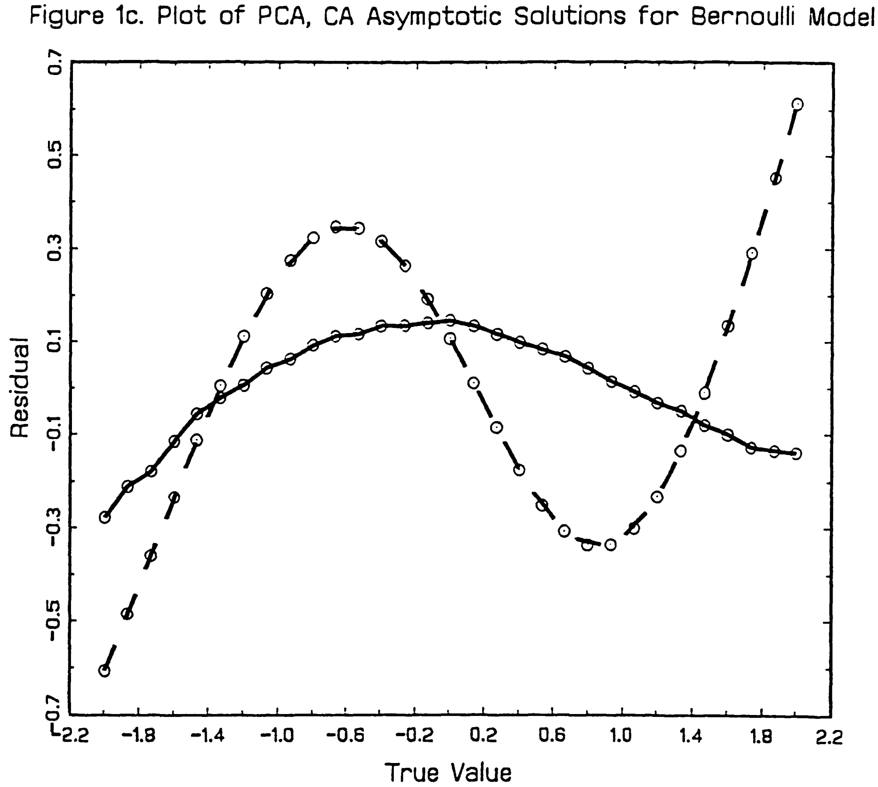


Table 1. Correlations ($\times 100$) between $\mathbf{u}$ and its CA and PCA estimates for different parameter inputs to the Poisson model. The entries are listed as r1/r2/r3, where r1, r2, r3 are the correlations of $\mathbf{u}$ with its CA, PCA (first eigenvector), and PCA (second eigenvector) estimates respectively. U(p,q) is the discrete Uniform distribution with equally spaced elements between (p,q), and $E(p,q)$ is a vector of equally spaced elements between (p,q).

Poisson Model	**u**	$E(0,0.4)$		$E(0,1.2)$	
	x	$E(-1,1)$	$E(-1.5,1.5)$	$E(-1,1)$	$E(-1.5,1.5)$
$b_j \sim$ U(1,4)	$a_j \sim$ U(0.5,1.25)	68/21/54	50/17/47	93/42/85	96/41/91
	$a_j \sim$ U(0.5,2.0)	76/19/66	55/15/61	94/40/82	97/38/86
$b_j \sim$ U(1,7)	$a_j \sim$ U(0.5,0.25)	60/21/64	38/17/55	94/42/85	96/44/90
	$a_j \sim$ U(0.5,2.0)	62/19/72	36/16/65	94/41/81	97/45/83

	u	$E(-0.2,0.2)$		$E(-0.6,0.6)$	
	x	$E(-1,1)$	$E(-1.5,1.5)$	$E(-1,1)$	$E(-1.5,1.5)$
$b_j \sim$ U(1,4)	$a_j \sim$ U(0.5,1.25)	70/19/56	49/17/47	97/91/31	97/43/90
	$a_j \sim$ U(0.5,2.0)	79/17/68	54/15/61	97/84/32	98/43/84
$b_j \sim$ U(1,7)	$a_j \sim$ U(0.5,0.25)	60/19/66	37/17/55	97/91/29	97/50/88
	$a_j \sim$ U(0.5,2.0)	58/18/74	35/16/65	97/83/34	98/49/80

Table 2. Correlations (×100) between **u** and its CA and PCA estimates for different parameter inputs to the Bernoulli model. The entries are listed as r1/r2/r3, where r1, r2, r3 are the correlations of **u** with its CA, PCA (first eigenvector), and PCA (second eigenvector) estimates respectively. U(p,q) is the discrete Uniform distribution with equally spaced elements between (p,q), and $E(p,q)$ is a vector of equally spaced elements between (p,q).

Bernoulli Model	**u**	$E(0,0.4)$		$E(0,1.2)$	
	x	$E(-1,1)$	$E(-1.5,1.5)$	$E(-1,1)$	$E(-1.5,1.5)$
b_j~U(1,4)	a_j~U(0.5,1.25)	35/29/29	26/17/31	85/54/66	88/46/86
	a_j~U(0.5,2.0)	37/33/30	25/17/35	85/60/67	88/53/87
b_j~U(1,7)	a_j~U(0.5,0.25)	35/23/40	23/17/35	87/45/79	91/41/89
	a_j~U(0.5,2.0)	34/24/45	19/16/40	86/48/80	91/44/89

	u	$E(-0.2,0.2)$		$E(-0.6,0.6)$	
	x	$E(-1,1)$	$E(-1.5,1.5)$	$E(-1,1)$	$E(-1.5,1.5)$
b_j~U(1,4)	a_j~U(0.5,1.25)	35/17/34	26/17/31	91/88/29	92/37/90
	a_j~U(0.5,2.0)	36/16/38	24/15/35	90/90/27	93/35/91
b_j~U(1,7)	a_j~U(0.5,0.25)	32/17/43	22/17/35	92/90/30	93/38/90
	a_j~U(0.5,2.0)	30/15/49	18/16/40	92/91/30	93/35/92

Figures 2a, 2b. Sample plots of μ_{ij}'s generated from the Bernoulli model of Table 2 using $\mathbf{u}=E(0,0.4)$ (Figure 2a) and $\mathbf{u}=E(0,1.2)$ (Figure 2b), when $\mathbf{x}=E(-1,1)$, $b_j \sim U(1,7)$, and $a_j \sim U(0.5,2.0)$.

Figure 2a. Expected Probabilities for Bernoulli Model

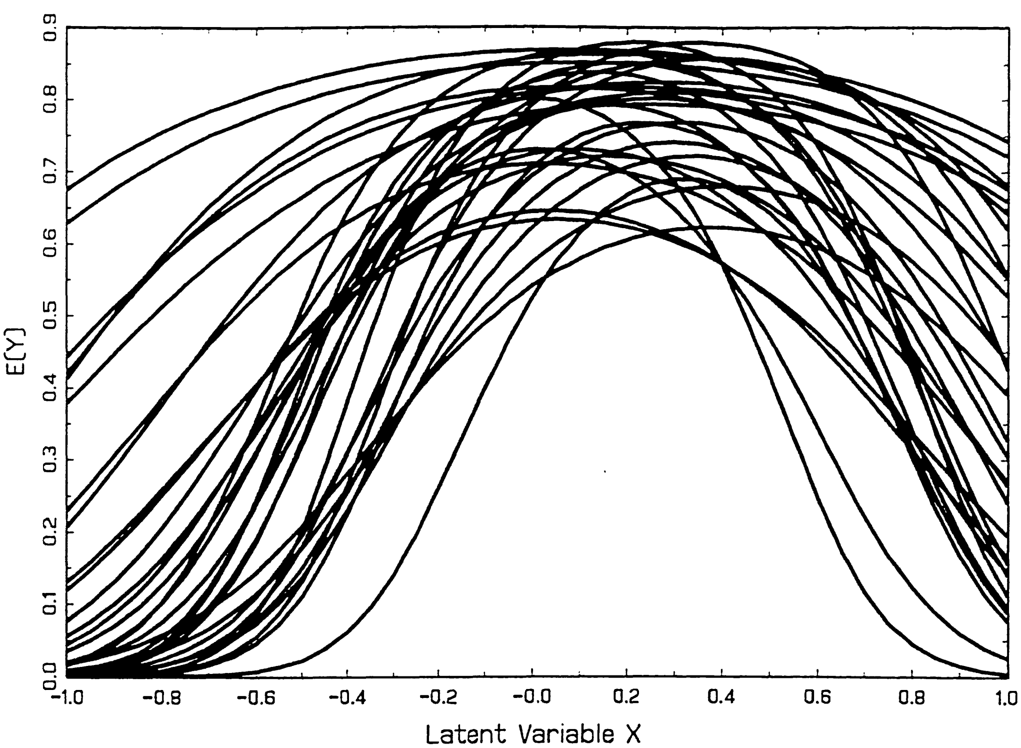
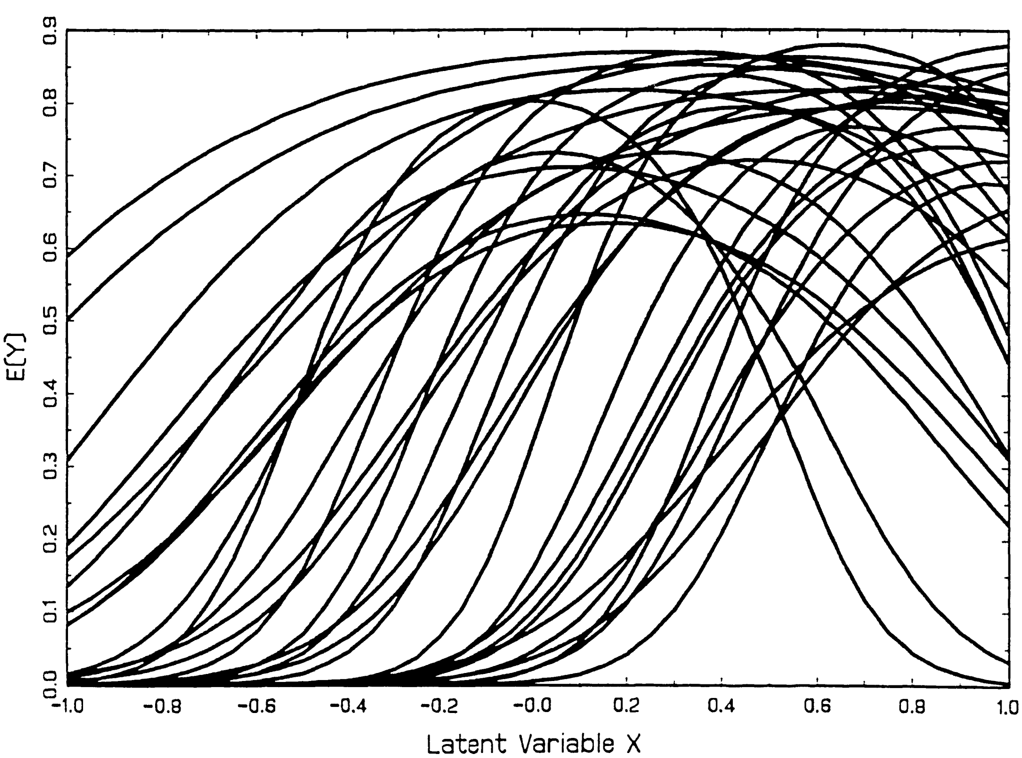


Figure 2b. Expected Probabilities for Bernoulli Model



Figures 2c, 2d. Sample plots of μ_{ij}'s generated from the Bernoulli model of Table 2 using $\mathbf{u}=E(-0.2,0.2)$ (Figure 2c) and $\mathbf{u}=E(-0.6,0.6)$ (Figure 2d), when $\mathbf{x}=E(-1,1)$, b_j~U(1,7), and a_j~U(0.5,2.0).

Figure 2c. Expected Probabilities for Bernoulli Model

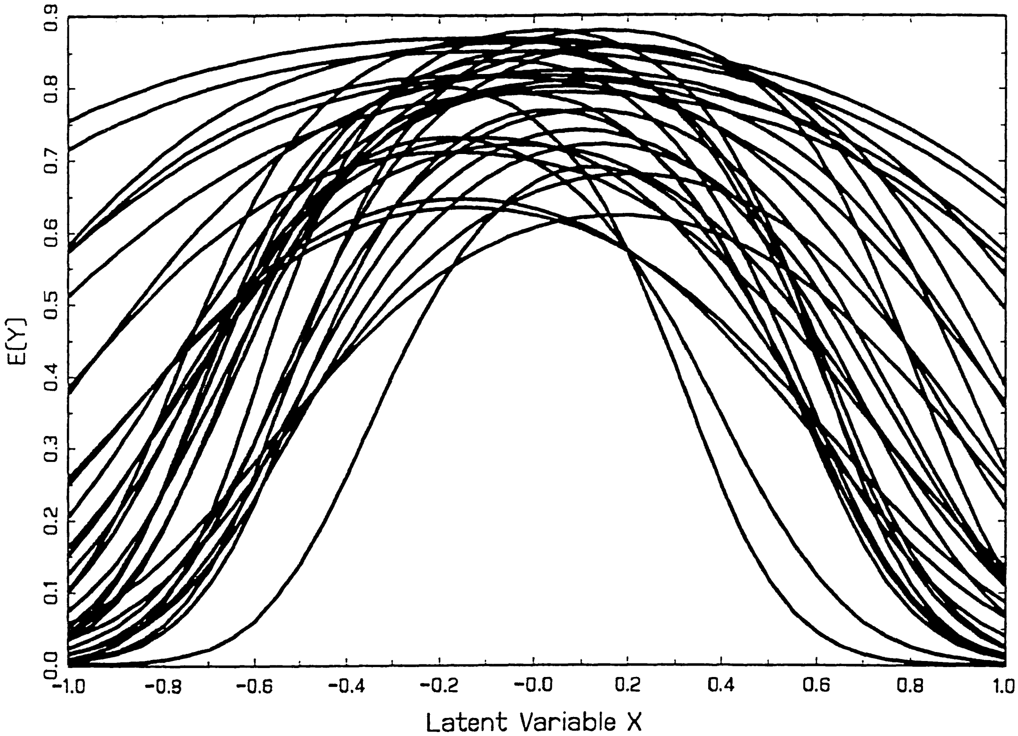


Figure 2d. Expected Probabilities for Bernoulli Model