BIAS AND VARIANCE OF MULTIPLE CORRESPONDENCE ANALYSIS

JAN DE LEEUW

ABSTRACT. Results on first and second derivatives of generalized eigenvalues and eigenvectors are used in Delta Method estimates for bias and variance of the statistics typically computed in Multiple and Ordinary Correspondence Analysis.

1. INTRODUCTION

Multiple Correspondence Analysis or *MCA* [Guttman, 1941, 1950; Burt, 1950; De Leeuw, 1973; Hill, 1974; Greenacre and Blasius, 2006] solves the generalized eigenvalue problem $Ax = \lambda Bx$, where A is the *Burt Matrix* of m categorical variables and B is m times the diagonal of the Burt Matrix. We normalize the solution by requiring $x'Bx = 1$.

The Burt matrix can be defined in terms of the profile vectors. If the m variables have $k_1, k_2, \cdots, k_m$ categories, then profile vectors are binary vectors z_ν of length $\sum k_j$, indicating the $K = \prod k_j$ possible patterns that can be observed. We have $A = \sum p_\nu z_\nu z_\nu'$, where summation is over all K possible profiles, and p_ν is the proportion of observed profile vectors in n trials equal to z_ν. Also $B = \sum p_\nu(mZ_\nu)$, where $Z_\nu = \mathbf{diag}(z_\nu z_\nu')$.

1.1. Summation over n.

If m is at all large, then $K = \prod k_j$ will tend to be very large, and many of the p_ν will be zero. In that case it makes more sense to translate the formulas from all K possible z_ν to only the n observed z_i. We

Date: Tuesday 12th May, 2009 — 10h 47min — Typeset in LUCIDA BRIGHT.

2000 *Mathematics Subject Classification.* 00A00.

Key words and phrases. Binomials, Normals, LaTeX.

1.2. **Ordinary Correspondence Analysis.** There is a simple relationship between MCA and ordinary *Correspondence Analysis (CA)* of a cross table [Benzécri, 1973; Greenacre, 1984]. CA is just the special case $m = 2$. The relationship between the eigenvalues of MCA and the canonical correlations ρ of CA is simply $\rho = 2\lambda - 1$. For CA we can write the stationary equations in partitioned form as

$$\begin{bmatrix} D & F \\ F' & E \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 2\lambda \begin{bmatrix} D & \varnothing \\ \varnothing & E \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

with $a'Da + b'Eb = 1$. This can also be written as

$$\begin{aligned} Fb &= \rho Da, \\ F'a &= \rho Db, \end{aligned}$$

which implies that $a'Da = b'Eb = \frac{1}{2}$.

2. Delta Method

Assume the profile proportions p_n are a realization of an asymptotically normal random vector[1] $\underline{p}_n$. More precisely, there is a vector π, with nonnegative elements π_v that add up to one, such that

$$n^{\frac{1}{2}}(\underline{p}_n - \pi) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, \Pi - \pi\pi'),$$

where Π is a diagonal matrix with π on the diagonal. By the *Delta Method* [Mann and Wald, 1943; Tiago De Olivera, 1982], if f is differentiable at π, and $G(\pi) = \mathcal{D}f(\pi)$, then we have the asymptotic distribution result

$$n^{\frac{1}{2}}(f(\underline{p}_n) - f(\pi)) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, G(\pi)(\Pi - \pi\pi')G(\pi)').$$

If f is real-valued, bounded, and two times continuously differentiable at π, with $H(\pi) = \mathcal{D}\mathcal{D}f(\pi)$, then we can approximate the bias by using

$$\lim_{n\to\infty} n\mathbf{E}(f(\underline{p}_n) - f(\pi)) = \frac{1}{2}\mathbf{tr}\, H(\pi)(\Pi - \pi\pi').$$

[1] Random variables are underlined [Hemelrijk, 1966].

3. MCA Results

We now combine the general results on the Delta Method from the previous section with the general results on differentiation of eigenvalues and eigenvectors from Appendix A. This extends results from O'Neill [1978a,b] and De Leeuw [1984]. Some of these results have been used recently in the `anacor` package on CRAN [De Leeuw and Mair, 2009].

First we determine the dispersion matrix of the asymptotic joint distribution of the vector of eigenvalues. Define $u_{s\nu}(\pi) = x_s(\pi)'(z_\nu z_\nu' - m\lambda_s(\pi)Z_\nu)x_s(\pi)$. Then

$$\lim_{n\to\infty} n\mathbf{E}\{(\lambda_s(\underline{p}_n) - \lambda_s(\pi))(\lambda_t(\underline{p}_n) - \lambda_t(\pi))\} = \sum_{\nu-1}^{K} \pi_\nu u_{s\nu} u_{t\nu}.$$

Our second result gives the dispersion matrix of the joint distribution of two eigenvectors. Define the vectors

$$v_{s\nu}(\pi) = -(A(\pi) - \lambda_s(\pi)B(\pi))^-(z_\nu z_\nu' - m\lambda_s(\pi)Z_\nu)x_s(\pi) - \frac{1}{2}m(x_s(\pi)'Z_\nu x_s(\pi))x_s(\pi).$$

Then

$$\lim_{n\to\infty} n\mathbf{E}\{(x_s(\underline{p}_n) - x_s(\pi))(x_t(\underline{p}_n) - x_t(\pi))'\} = \sum_{\nu=1}^{K} \pi_\nu v_{s\nu}(\pi)v_{t\nu}(\pi)' - \frac{1}{4}x_s(\pi)x_t(\pi)'.$$

And finally the expected value, and thus the bias correction, for an individual eigenvalue. Define the numbers

$$w_{s\nu}(\pi) = -u_{s\nu}(\pi)(mx_s(\pi)'Z_\nu x_s(\pi)) + 2x_s(\pi)'(z_\nu z_\nu' - m\lambda_s(\pi)Z_\nu)v_{s\nu}(\pi).$$

Then

$$\lim_{n\to\infty} n\mathbf{E}(\lambda_s(\underline{p}_n) - \lambda_s(\pi)) = \frac{1}{2}\sum_{\nu=1}^{K} \pi_\nu w_{s\nu}(\pi),$$

or, for CA,

$$\lim_{n\to\infty} n\mathbf{E}(\rho_s(\underline{p}_n) - \rho_s(\pi)) = \sum_{\nu=1}^{K} \pi_\nu w_{s\nu}(\pi).$$

3.1. **Summation over n.** If m is at all large, then $K = \prod k_j$ will tend to be very large, and many of the p_ν will be zero. In that case it makes more sense to translate the formulas from all K possible z_ν to only the n observed z_i. With obvious notation, we then have formulas of the form

$$\lim_{n\to\infty} n\mathbf{E}(\rho_s(\underline{p}_n) - \rho_s(\pi)) = \frac{1}{n}\sum_{i=1}^{n} w_{si}.$$

3.2. **Ordinary CA.** Note that if $m = 2$, we index the probabilities by the $I \times J$ cells of the bivariate table, and the two parts of the eigenvector are a and b, we have

$$u_{sij} = (1 - 2\lambda_s)(a_{is}^2 + b_{js}^2) + 2a_{is}b_{js} = 2a_{is}b_{js} - \rho_s(a_{is}^2 + b_{js}^2)$$

and

$$\lim_{n\to\infty} n\mathbf{E}\{\lambda_s(\underline{p}_n) - \lambda_s(\pi))(\lambda_t(\underline{p}_n) - \lambda_t(\pi))\} = \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_{ij}u_{sij}u_{tij}.$$

REFERENCES

J.P. Benzécri. *Analyse des Données: Correspondances*, volume 2. Dunod, Paris, 1973.

C. Burt. The Factorial Analysis of Qualitative Data. *British Journal of Statistical Psychology*, 3:166–185, 1950.

J. De Leeuw. *Canonical Analysis of Categorical Data*. PhD thesis, University of Leiden, The Netherlands, 1973. Republished in 1985 by DSWO-Press, Leiden, The Netherlands.

J. De Leeuw. Statistical Properties of Multiple Correspondence Analysis. Research Report RR-84-06, Department of Data Theory, University of Leiden, 1984. URL `http://www.datatheory.nl/pdfs/84/84_06.pdf`.

J. De Leeuw. Derivatives of Generalized Eigen Systems. Preprint Series 528, UCLA Department of Statistics, 2007. URL `http://preprints.stat.ucla.edu/528/derivatives.pdf`.

J. De Leeuw and P. Mair. Simple and Canonical Correspondence Analysis Using the R Package `anacor`. *Journal of Statistical Software*, (forthcoming), 2009.

M. Greenacre and J. Blasius, editors. *Multiple Correspondence Analysis and Related Methods*. Chapman and Hall, 2006.

M.J. Greenacre. *Theory and Applications of Correspondence Analysis*. Academic Press, New York, New York, 1984.

L. Guttman. The Quantification of a Class of Attributes: A Theory and Method of Scale Construction. In P. Horst, editor, *The Prediction of Personal Adjustment*, pages 321–348. Social Science Research Council, New York, 1941.

L. Guttman. The Principal Components of Scale Analysis. In S.A. Stouffer and Others, editors, *Measurement and Prediction*. Princeton University Press, Princeton, 1950.

J. Hemelrijk. Underlining random variables. *Statistica Neerlandica*, 20: 1–7, 1966.

M.O. Hill. Correspondence Analysis: a Neglected Multivariate Method. *Applied Statistics*, 23:340–354, 1974.

H.B. Mann and A. Wald. On Stochastic Limit and Order Relationships. *Annals of Mathematical Statistics*, 14:217–226, 1943.

M.E. O'Neill. Asymptotic Distribution of the Canonical Correlations from Contingency Tables. *Australian Journal of Statistics*, 20:75–82, 1978a.

M.E. O'Neill. Distributional Expansions for Canonical Correlations from Contingency Tables. *Journal of the Royal Statistical Society B*, 40:303–312, 1978b.

J. Tiago De Olivera. The Delta Method for Obtention of Asymptotic Distributions; Applications. *Publications de l'Institute de Statistique de l'Université de Paris*, 27:49–70, 1982.

Appendix A. Derivatives

A.1. **General Case.** Suppose A and B are positive semi-definite and depend on a vector of parameters θ. Suppose (x,λ) is a normalized eigenpair of (A,B), i.e. $Ax = \lambda Bx$ and $x'Bx = 1$. In a neighborhood where the eigenvalue is isolated we differentiate the eigen-equations and find

$$\mathcal{D}Ax + A\mathcal{D}x - \mathcal{D}\lambda Bx - \lambda\mathcal{D}Bx - \lambda B\mathcal{D}x = 0, \tag{1a}$$

$$x'\mathcal{D}Bx + 2x'B\mathcal{D}x = 0. \tag{1b}$$

Premultiplying both sides of (1a) by x' gives

$$\mathcal{D}\lambda = x'(\mathcal{D}A - \lambda\mathcal{D}B)x. \tag{2}$$

We can write (1b) as

$$x'B\mathcal{D}x = -\frac{1}{2}x'\mathcal{D}Bx, \tag{3}$$

and (1a) as

$$(A - \lambda B)\mathcal{D}x = -(\mathcal{D}A - \lambda\mathcal{D}B)x + \mathcal{D}\lambda Bx. \tag{4}$$

Suppose X is a non-singular matrix satisfying $X'BX = I$ and $X'AX = \Lambda$. Then x is one of the columns of X and λ is the corresponding diagonal element of Λ. Define, following De Leeuw [2007], the generalized inverse

$$(A - \lambda B)^- = X(\Lambda - \lambda I)^+X',$$

where $(\Lambda - \lambda I)^+$ is the Moore-Penrose inverse. Then $(A - \lambda B)^-Bx = 0$. Since x is the only vector in the null-space of $A - \lambda B$ we have from (4) that for some κ

$$\mathcal{D}x = -(A - \lambda B)^-(\mathcal{D}A - \lambda\mathcal{D}B)x + \kappa x.$$

Using (3) we find that

$$\mathcal{D}x = -(A - \lambda B)^-(\mathcal{D}A - \lambda\mathcal{D}B)x - \frac{1}{2}(x'\partial Bx)x. \tag{5}$$

A.2. **Linear Case.** Suppose $A = \sum \theta_\nu A_\nu$ and $B = \sum \theta_\nu B_\nu$. Switching to functional notation we find

$$\mathcal{D}_\nu \lambda = x'(A_\nu - \lambda B_\nu)x, \tag{6a}$$

as well as

$$\mathcal{D}_\nu x = -(A - \lambda B)^-(A_\nu - \lambda B_\nu)x - \frac{1}{2}(x'B_\nu x)x. \tag{6b}$$

If we differentiate (6a) again

$$\mathcal{D}_\xi \mathcal{D}_\nu \lambda = -(\mathcal{D}_\xi \lambda)x'B_\nu x + 2x'(A_\nu - \lambda B_\nu)(\mathcal{D}_\xi x). \tag{7}$$

Note that $\sum \theta_\nu(\mathcal{D}_\nu \lambda) = 0$ while $\sum \theta_\nu(\mathcal{D}_\nu x) = -\frac{1}{2}x$. Moreover, from (7), $\sum \theta_\nu \theta_\xi(\mathcal{D}_\xi \mathcal{D}_\nu \lambda) = 0$.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu