# MULTIPLE CORRESPONDENCE ANALYSIS OF COUPLED LANCASTER SEQUENCES

JAN DE LEEUW

ABSTRACT. Meet the abstract. This is the abstract.

## 1. A SEQUENCE OF STANDARD NORMALS

Suppose $\underline{x}$ and $\underline{y}$ are two standard normal random variables [1], with correlation $\rho$.

The bivariate normal density is $\phi_\rho(\bullet,\bullet)$ and the univariate normal density is $\phi(\bullet)$. The function ${}_s(\bullet)$ is the standardized Hermite-Chebyshev polynomial of order $s$, with $s = 0, 1, \cdots$.

Mehler's formula tells us that for all $s$

$$\int_{-\infty}^{+\infty} \phi_\rho(x,y)h_s(y)dy = \rho^s\phi(x)\mathscr{H}_s(x), \tag{1a}$$

$$\int_{-\infty}^{+\infty} \phi_\rho(y,x)h_s(x)dy = \rho^s\phi(y)h_s(y), \tag{1b}$$

or

$$\mathbf{E}(h_s(\underline{y}) \mid \underline{x}) = \rho^s h_s(\underline{x}), \tag{2a}$$

$$\mathbf{E}(h_s(\underline{x}) \mid \underline{y}) = \rho^s h_s(\underline{y}). \tag{2b}$$

Equation (1) says that the $h_s(\bullet)$ are the left and right singular functions of the kernel $\phi_\rho(\bullet,\bullet)$, normalized by the marginals, which are both equal to $\phi(\bullet)$. The successive powers $\rho^s$ are the singular values.

*Date*: Monday 15th March, 2010 — 12h 1min — Typeset in UTOPIA.

*Key words and phrases.* Correspondence Analysis, Principal Component Analysis, Nonlinear Multivariate Analysis.

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

Equation (2) says that for all $s$ the random variables $h_s(\underline{x})$ and $h_s(\underline{y})$ have linear regressions, with the $\rho^s$ as the regression coefficients.

Now suppose $\underline{x}_1, \underline{x}_2, \cdots$ is a sequence of standard normal random variables, with $\mathbf{E}(\underline{x}_j \underline{x}_\ell) = \rho_{j\ell}$. Suppose $\alpha$ is an eigenvector of the matrix $R^{(s)} = \{\rho^s_{j\ell}\}$, with corresponding eigenvalue $\lambda$. Thus

$$\sum_{\ell=1}^{\infty} \rho^s_{j\ell}\alpha_\ell = \lambda\alpha_j.$$

Now define $y_{s\ell}(\bullet) = \alpha_\ell h_s(\bullet)$. Then

$$\sum_{\ell=1}^{\infty} \mathbf{E}(y_{s\ell}(\underline{x_\ell}) \mid \underline{x}_j) = \left\{\sum_{\ell=1}^{\infty} \rho^s_{j\ell}\alpha^{(s)}_\ell\right\} h_s(\underline{x}_j) = \lambda y_{sj}(\underline{x}_j)$$

## 2. The Burt Table

Suppose $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_T$ is a sequence of $T$ discrete random variables which only take a finite number $\ell_1, \ell_2, \cdots, \ell_T$ of values.

Suppose $\Pi_{st}$ are the $\ell_s \times \ell_t$ bivariate marginals, and $\Pi_t$ are the univariate marginals, in diagonal matrices of order $\ell_t$.

Multiple Correspondence Analysis (MCA) is defined as solving the eigenvalue problem

$$\sum_{s=1}^{T} \Pi_{ts} y_s = \lambda \Pi_t y_t. \tag{3a}$$

The $T \times T$ block-matrix with elements $\Pi_{st}$ is known in the MCA literature as the *Burt Table*.

Using the conditional bivariate distributions $\Pi_{t|s} = \Pi_t^{-1}\Pi_{ts}$ equation (3a) can also be written as

$$\sum_{s=1}^{T} \Pi_{t|s} y_s = \lambda y_t, \tag{3b}$$

and if we think of the $y_t$ as functions of $\underline{x}_t$ also as

$$\sum_{s=1}^{T} \mathbf{E}[f_s(\underline{x}_s)|\underline{x}_t] = \lambda f_t(\underline{x}_t). \tag{3c}$$

In this paper we extend the definition of MCA to an infinite sequence of random variables, which possibly take an infinite number of values.

## 3. GAUSSIAN

Suppose $\underline{x}_1, \underline{x}_2, \cdots$, is a sequence of normally distributed random variables with zero means and unit variances. Define $\rho_{st} \stackrel{\Delta}{=} \mathbf{E}(\underline{x}_s \underline{x}_t)$.

Suppose $h_k$ is the standardized Hermite-Chebyshev polynomial or order $k \geq 0$. Then

$$\mathbf{E}[h_k(\underline{x}_s) \mid \underline{x}_t] = \rho_{st}^k h_k(\underline{x}_t).$$

We also have $\mathbf{E}[h_k^2(\underline{x}_s] = 1$ for all $s$ and $k$, and $\mathbf{E}[h_k(\underline{x}_s)] = 0$ for all $s$ and $k > 0$.

Now suppose $\alpha$ is an eigenvector of $\Gamma^{(k)} = \{\rho_{st}^k\}$, i.e.

$$\sum_{s=1}^{\infty} \rho_{ts}^k \alpha_s = \lambda \alpha_t, \tag{4}$$

with $\sum_{s=1}^{\infty} \alpha_s^2 = 1$. Then

$$\sum_{s=1}^{\infty} \mathbf{E}[\alpha_s h_k(\underline{x}_s) \mid \underline{x}_t] = \lambda \alpha_t h_k(\underline{x}_t),$$

in other words $\{\alpha_1 h_k, \alpha_2 h_k, \cdots\}$ is an eigenvector of the linear operator

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \end{bmatrix} \Rightarrow \begin{bmatrix} \sum_{s=1}^{\infty} \mathbf{E}[f_s(\underline{x}_s) \mid \underline{x}_1] \\ \sum_{s=1}^{\infty} \mathbf{E}[f_s(\underline{x}_s) \mid \underline{x}_2] \\ \vdots \end{bmatrix}$$

Note that $\sum_{s=1}^{\infty} \mathbf{E}[\alpha_s^2 h_k^2(\underline{x}_s)] = 1$ for all $s$ and $k$.

## 4. MULTIPLE CORRESPONDENCE ANALYSIS

Suppose $\underline{x}_1, \underline{x}_2, \cdots$, is a sequence of real valued random variables with finite variances. Also $\mathscr{L}_1, \mathscr{L}_2, \ldots$ are closed subspaces of real-valued functions with finite variance.

**Definition 4.1.** In *Multiple Correspondence Analysis (MCA)* we look for functions $f_t \in \mathscr{L}_t$ such that

$$\sum_{t=1}^{T}\omega_t\mathbf{E}(f_t(\underline{x}_t)\mid\underline{x}_s)=\lambda f_s(\underline{x}_s). \tag{5}$$

In words: for all $s$ the average regression of the $f_t(\underline{x}_t)$ on $\underline{x}_s$ is linear.

There are many ways in which data analysts, starting with Guttman [1941], have arrived at the MCA equations, but in this paper we are merely interested in characterizing and describing the solutions. For reviews of the data analysis background we refer to Gifi [1990] or Greenacre and Blasius [2006].

4.1. **Correlations and eigenvalues.** From the definition of conditional expectation we have

$$\mathbf{E}(f_t(\underline{x}_t)-\mathbf{E}(f_t(\underline{x}_t)\mid\underline{x}_s))h(\underline{x}_s)=0 \tag{6}$$

for any function $h$ [Whittle, 2000, Section 5.3]. By summing (6) over $t$, using $h=f_s$, and using (5), we find that at a solution

$$\sum_{t=1}^{T}\mathbf{E}(f_s(\underline{x}_s)f_t(\underline{x}_t))=T\lambda\mathbf{E}(f_s^2(\underline{x}_s)). \tag{7}$$

Suppose $\psi_0^t,\psi_1^t,\cdots\psi_{\ell_t}^t$ is an orthonormal basis for $\mathscr{L}_t$. Thus we can write $f_t=\sum_{u=0}^{\ell_t}\xi_u^t\psi_u^t$. In this new basis (7) becomes

$$\sum_{u=0}^{\ell_s}\xi_u^s\left\{\sum_{t=1}^{T}\sum_{v=0}^{\ell_t}c_{uv}^{st}\xi_v^t-T\lambda\xi_u^s\right\}=0,$$

where $c_{uv}^{st}\overset{\Delta}{=}\mathbf{E}(\psi_u^s(\underline{x}_s)\psi_v^t(\underline{x}_t))$. This is the eigenvalue problem

$$\sum_{t=1}^{T}\sum_{v=0}^{\ell_t}c_{uv}^{st}\xi_v^t=T\lambda\xi_u^s, \tag{8a}$$

or, equivalently, the eigenvalue problem

$$\sum_{t\neq s}^{T}\sum_{v=0}^{\ell_t}c_{uv}^{st}\xi_v^t=(T\lambda-1)\xi_u^s. \tag{8b}$$

4.2. **On Infinity.**

4.3. **Induced Correlation Matrices.** Normalize the $f_t$ by writing them as $\alpha_t \tilde{f}_t$, where $\mathbf{E}(\tilde{f}_t^2) = 1$. Then (7) becomes

$$\sum_{t=1}^{T} \mathbf{E}(\tilde{f}_s(\underline{x}_s)\tilde{f}_t(\underline{x}_t))\alpha_t = T\lambda\alpha_s \tag{9}$$

Thus each normalized solution of (5) defines an *induced correlation matrix* $\Gamma(\tilde{f}_1, \cdots, \tilde{f}_T)$, and a vector $\alpha$, which is an eigenvector of the correlation matrix corresponding with the eigenvalue $T\lambda$. This implies $\lambda \leq 1$, with equality if and only if all $T$ random variables $\tilde{f}_t(\underline{x}_t)$ are perfectly correlated and the correlation matrix has rank one. One such solution, and in most cases the only one possible, is $f_t \equiv 1$ for all $t$. Other solutions with $\lambda = 1$ are only possible if the bivariate distributions are degenerate.

4.4. **Additional solutions.** If $g_1, \cdots, g_T$ is a second solution to (5), with eigenvalue $\omega \neq \lambda$, then from (6)

$$\sum_{s=1}^{T}\sum_{t=1}^{T} \mathbf{E}(f_s(\underline{x}_s)g_t(\underline{x}_t)) = T\lambda \sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)g_t(\underline{x}_t)),$$
$$\sum_{s=1}^{T}\sum_{t=1}^{T} \mathbf{E}(g_s(\underline{x}_s)f_t(\underline{x}_t)) = T\omega \sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)g_t(\underline{x}_t)),$$

and thus $\sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)g_t(\underline{x}_t)) = 0$.

In particular, if $\lambda < 1$ we have $\sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)) = 0$. This result can be strengthened somewhat. By definition $\mathbf{E}(f_t(\underline{x}_t)) = \mathbf{E}(\mathbf{E}(f_t(\underline{x}_t) \mid \underline{x}_s))$. If we sum this over $t$ at a solution with $\lambda < 1$ we find

$$\sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)) = T\lambda\mathbf{E}(f_s(\underline{x}_s)),$$

and thus $\mathbf{E}(f_t(\underline{x}_t)) = 0$ for all $i$.

Suppose $\tilde{f}_t$ is a normalized solution, with corresponding induced correlation matrix $\Gamma$, and $\alpha$ is the corresponding eigenvector of (9). Suppose $\beta$ is another eigenvector of the same $\Gamma$. It would be tempting to conclude that $g_t = \beta_t\tilde{f}_t$ is another solution to (5), especially because

$$\sum_{t=1}^{T} \mathbf{E}(f_t(\underline{x}_t)g_t(\underline{x}_t)) = \sum_{t=1}^{T} \alpha_t\beta_t = 0.$$

In general, however, these $g_t$ will *not* satisfy the MCA equations (5). In the remainder of the paper we will study conditions that guarantee $g_t$ constructed in this way do satisfy (5).

4.5. **All solutions combined.** If the random variables are discrete, or if they are continuous and have densities, then we can write the MCA equations as the eigenvalue problem

$$\sum_{t=1}^{T}\int \omega_{st}(x,y)\overline{f}_t(y)dy = T\lambda\overline{f}_s(x). \tag{10}$$

where we use the univariate and bivariate densities $\pi_s$ and $\pi_{st}$ to define

$$\omega_{st}(x,y)\overset{\Delta}{=}\frac{\pi_{st}(x,y)}{\sqrt{\pi_s(x)\pi_t(y)}}.$$

Also $\overline{f}_t\overset{\Delta}{=}\sqrt{\pi_t}f_t$. Note that if $s=t$ then $\omega_{st}(x,y)=1$ if $x=y$ and $\omega_{st}(x,y)=0$ otherwise.

Thus the eigenvalues are the eigenvalues of a block-operator, which has the $T\times T$ (possibly infinite) blocks $\Omega_{st}$ as its elements. We call this the *Burt operator*, giving credit to Burt [1950].

Because of the nature of the diagonal blocks, we can replace them by zeroes, and instead study the eigenvalue problem

$$\sum_{t\neq s}^{T}\int \omega_{st}(x,y)\overline{f}_t(y)dy = (T\lambda-1)\overline{f}_s(x).$$

The vector with blocks $\sqrt{\pi_t}$ is an eigenvector with $\lambda=1$. We can deflate by substracting the effect of this solution, which will leave the deflated block-operator with blocks

$$\tilde{\omega}_{st}(x,y)\overset{\Delta}{=}\frac{\pi_{st}(x,y)-\pi_s(x)\pi_t(y)}{\sqrt{\pi_s(x)\pi_t(y)}}.$$

For $s\neq t$ define the Pearson Mean Square Contingency, the population version of the Chi-Square for Independence, as

$$\kappa^2_{st}=\int\int\tilde{\omega}^2_{st}(x,y)dxdy$$

Under regularity conditions on the operator, which we shall not discuss here, the sum of squares of the eigenvalues is the sum of squares of the blocks, and thus

$$\sum_{\nu=1}^{\infty}(T\lambda_\nu-1)^2=\sum_{1\le s\neq t\le T}\sum\kappa_{st}^2. \tag{11}$$

Formula (11) was first given by De Leeuw [1973, section 3.7]. Regularity conditions needed for these (and even more general) results are discussed in detail in Dauxois and Pousse [1976].

## 5. Strained Multinormals

Suppose $\underline{x}_1,\cdots,\underline{x}_T$ are jointly normal with mean zero and variances one. The correlation between $\underline{x}_s$ and $\underline{x}_t$ is $\rho_{st}$. For convenience we assume the correlations are not equal to $\pm1$.

Also suppose $\eta_1,\cdots,\eta_T$ are real-valued differentiable strictly monotone functions, with differentiable inverses, and define $\underline{y}_t\overset{\Delta}{=}\eta_t(\underline{x}_t)$. If the $\eta_t$ are non-linear the $\underline{y}_t$ are no longer normal, but they are *strained normal* in the sense of Yule [1912, page 612-613]. The major result in this section is that multiple correspondence analysis unstrains the normals, i.e. finds the inverses of the strains and transforms back to multivariate normality. We can also describe the MCA solutions in complete detail.

Use $\phi$ for the density of the standard normal and $\phi_{st}$ for the bivariate standard normal with correlation $\rho_{st}$. The marginal density of $\underline{y}_t$ is

$$\pi_t(y)=\phi(\eta_t^{-1}(y))\mathscr{D}\eta_t^{-1}(y),$$

and the joint density of $\underline{y}_s$ and $\underline{y}_t$ is

$$\pi_{st}(y_s,y_t)=\phi_{st}(\eta_s^{-1}(y_s),\eta_t^{-1}(y_t))\cdot\mathscr{D}\eta_s^{-1}(y_s)\cdot\mathscr{D}\eta_t^{-1}(y_t).$$

Thus

$$\mathbf{E}(f_t(\underline{y}_t)\mid\underline{y}_s)=$$
$$=\int_{\eta(-\infty)}^{\eta(+\infty)}\phi_{st}(\eta_t^{-1}(y_y)\mid\eta_s^{-1}(y_s))\cdot\mathscr{D}\eta_t^{-1}(y)f_t(y_t)dy_t.$$

Now make the change of variables $x_t = \eta_t^{-1}(y_t)$. Then

$$\mathbf{E}(f_t(\underline{y}_t) \mid \underline{y}_s) = \int_{-\infty}^{+\infty} \phi_{st}(x_t \mid x_s) f_t(\eta_t(x_t)) dx_t.$$

As a consequence the MCA problem for strained normals can be solved by computing MCA solution for a sequence of standard normals. If the solution for the standard normal sequence is $f_t$, then the corresponding solution for the strained problem is $f_t \circ \eta_t^{-1}$.

5.1. **Sequences of standard normals.** The solution for a sequence of standard normals is well known. See, for example, De Leeuw [1973, Section 3.8] or Gifi [1990, Section 11.3].

A complete set of solutions is of the form $g_{tv}^{(p)}(x) = \alpha_{tv}^{(s)} h^{(p)}(x)$, where $h^{(p)}(x)$ is the standardized Hermite polynomial of order $p$, and where $\alpha$ is a normalized eigenvector of the $T \times T$ matrix $\Gamma^{(p)}$ with elements

$$\gamma_{s,t}^{(p)} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi_{st}(x, y) h^{(p)}(x) h^{(p)}(y) dx dy = \rho_{j\ell}^{p}.$$

Indices $s$ and $t$ are used for variables, index $v$ for eigenvectors, and index $p$ for polynomial powers. Note that the $\Gamma^{(s)}$ are the Hadamard powers of the correlation matrix $\Gamma$.

Thus each polynomial order $p$ corresponds with $T$ solutions. For each of these solutions the $g_{tv}^{(p)}$ are proportional to the standardized Hermite polynomials or order $p$, with proportionality factors given by the eigenvectors of $\Gamma^{(p)}$. The values of the average regression coeficients $\lambda$ are the corresponding eigenvalues, which we can index as $\lambda_v^{(p)}$.

Note that straining the multivariate normal does not change the eigenvalues, only the functions. We could strain, for example, by $x^3$, or $(1 + \exp(-x))^{-1}$, of by the cumulative normal $\Phi$. Straining all variable by $\Phi$, for example, would give a multivariate distribution with marginals that are rectangular on the unit interval. The solutions for these three strains would be proportional to $h^{(s)}(\sqrt[3]{x})$ or $h^{(s)}(\log \frac{x}{1-x})$ or $h^{(s)}(\Phi^{-1}(x))$. Of course different strains can be mixed.

5.2. **Ordering the eigenvalues.** We can (partially) order the eigenvalues by $\lambda_1^{(p)} \geq \cdots \geq \lambda_T^{(p)}$ for all $p$. But this clearly does not produce a complete order. It is a simple consequence of Schur's bounds for the eigenvalues of Hadamard product that if $p < q$ then for all $v$

$$\lambda_T^{(p)} \leq \lambda_v^{(q)} \leq \lambda_1^{(p)}. \tag{12}$$

See, for example, Styan [1973, Corollary 3.1]. Thus the largest eigenvalues $\lambda_1^{(p)}$ form a non-increasing sequence, and the smallest eigenvalues $\lambda_T^{(p)}$ form a non-decreasing sequence. Both sequences converge to the eigenvalues of the identity matrix, i.e. to one.

Unfortunately the ordinal information given by (12) about the $\lambda_v^{(p)}$ is still rather incomplete. We know that $\lambda_1^{(1)}$ is the largest eigenvalue, but the second largest one can be either $\lambda_2^{(1)}$ or $\lambda_1^{(2)}$. If $\lambda_1^{(1)} > \lambda_2^{(1)} > \lambda_1^{(2)}$ then the two dominant solutions for $f$ are both linear in $x$. If $\lambda_1^{(1)} > \lambda_1^{(2)} > \lambda_2^{(1)}$ then the dominant solution is linear and the second solution is quadratic. If we plot the two dominant solutions against each other in the plane, the second case will produce the "horseshoe" familiar from many MCA solutions [Hill, 1974; Schriever, 1985; Van Rijckevorsel, 1987].

5.3. **An Example.** There is a nice example in [Gifi, 1990, page 382-383]. Consider the matrix

$$\Gamma = \begin{bmatrix} 1 & A & B & C \\ A & 1 & C & B \\ B & C & 1 & A \\ C & B & A & 1 \end{bmatrix}.$$

The eigenvectors are the columns of

$$\begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

and the eigenvalues are

$$\begin{aligned}\lambda_1&=1+A+B+C,\\\lambda_2&=1+A-B-C,\\\lambda_3&=1-A+B-C,\\\lambda_4&=1-A-B+C.\end{aligned}$$

Of course $\Gamma$ is a correlation matrix if and only if all four eigenvalues are non-negative. This means that $(A,B,C)$ must be a convex combination of the four vectors

$$\begin{bmatrix}+1\\+1\\+1\end{bmatrix}\begin{bmatrix}+1\\-1\\-1\end{bmatrix}\begin{bmatrix}-1\\+1\\-1\end{bmatrix}\begin{bmatrix}-1\\-1\\+1\end{bmatrix}.$$

At these four corners the correlation matrices are of rank one, which means that one eigenvalue is four and the others are zero. The region can be partitioned further in 24 convex regions with different orderings of the four eigenvalues.

The example is especially nice, because all Hadamard powers of $\Gamma$ have the same form, and thus the same eigenvectors. Just replace $A,B$, and $C$ by $A^p,B^p$ and $C^p$ to find the eigenvalues.

Gifi solves the problem of maximizing $\lambda_2^{(1)}-\lambda_1^{(2)}$. The maximum is attained at $A=\frac{1}{2}$ and $B=C=0$, for which the two largest eigenvalues of $\Gamma^{(p)}$ are $1+(\frac{1}{2})^p$ and the two smallest ones are $1-(\frac{1}{2})^p$. Thus $\lambda_2^{(1)}-\lambda_1^{(2)}=1.5-1.25=0.25$. The multinormal or strained multinormal analysis in this case will not show a horseshoe.

5.4. **All solutions combined.** Note that $\sum_{\nu=1}^T\lambda_\mu^{(p)}=1$ for all p. It follows that

$$\sum_{1\leq s\neq t\leq T}\sum\left(\rho_{st}^p\right)^2=\sum_{s=1}^T\sum_{t=1}^T\left(\rho_{st}^p\right)^2-T=\sum_{\nu=1}^T\left(T\lambda_\nu^{(p)}-1\right)^2.$$

Summing both sides over $p$ gives the multivariate normal version of (11), first given by De Leeuw [1973, Section 3.8],

$$\sum_{p=1}^{\infty}\sum_{\nu=1}^{T}\left(T\lambda_\nu^{(p)}-1\right)^2=\sum_{1\le s\neq t\le n}\sum\sum\sum_{p=1}^{\infty}\left(\rho_{st}^p\right)^2=$$

$$=\sum_{1\le s\neq t\le n}\sum\sum\frac{\rho_{st}^2}{1-\rho_{st}^2}=\sum_{1\le s\neq t\le n}\sum\sum\kappa_{st}^2.$$

## 6. Bivariate Distributions with Polynomial Expansions

The (strained) multinormal case has three characteristics which distinguish it from the general MCA case. The first is that the eigenvector are build up from a system of orthogonal polynomials. The second that there is a great deal of symmetry, in the sense that for all variables the orthogonal polynomials are the same. And third, the information in the multivariate distribution is concentrated in the bivariate marginals. These three characteristics can be generalized far beyond the multivariate normal.

### 6.1. **Markovian Processes.**

For a Markovian process we have

$$\mathbf{E}(f(\underline{x}_t)|\underline{x}_1)=\mathbf{E}(\mathbf{E}(f(\underline{x}_t)|\underline{x}_{t-1})|\underline{x}_1). \tag{13}$$

Equation (13) can be applied recursively. If there exists a function $f$ such that

$$\mathbf{E}(f(\underline{x}_t)|\underline{x}_{t-1})=\rho_{t-1}f(\underline{x}_{t-1})$$

for all $t$, we find

$$\mathbf{E}(f(\underline{x}_t)|\underline{x}_1)=\rho_{t-1}\mathbf{E}(f(\underline{x}_{t-1})|\underline{x}_1)=\left(\prod_{s=1}^{t-1}\rho_s\right)\cdot f(\underline{x}_1),$$

and thus

$$\sum_{t=1}^{T}\mathbf{E}(f(\underline{x}_t)|\underline{x}_1)=\left(\sum_{t=1}^{T}\prod_{s=1}^{t-1}\rho_s\right)\cdot f(\underline{x}_1).$$

In order to satisfy the MCA equation (5) we have to assume, in addition, that the chain is reversible.

6.2. **Sarmanov-Bratoeva Class.** Suppose we have $K$ standard multivariate normals with correlation matrices $\Gamma_1,\cdots,\Gamma_K$. Mix them with probabilities $\pi_k$. Then the Hermite polynomials satisfy

$$\mathbf{E}(h^{(p)}(\underline{x}_t)|\underline{x}_s)=\left(\sum_{k=1}^K\pi_k\rho_{stk}^p\right)\cdot h^{(p)}(\underline{x}_s),$$

and consequently they satisfy the MCA equations (5).

If we have two correlation matrices, for example, with $\rho_{st1}=-\rho_{st2}$ for all $s\neq t$ [2] and we choose $\pi_1=\pi_2=\frac{1}{2}$ then

$$\mathbf{E}(h^{(p)}(\underline{x}_t)|\underline{x}_s)=\begin{cases}\rho_{stk}^p h^{(p)}(\underline{x}_s) & \text{for } p \text{ is even,}\\ 0 & \text{for } p \text{ is odd.}\end{cases}$$

Thus all Hermite polynomials of odd order, including the ones of order one, disappear from the analysis (or, more precisely, get relegated to the zero eigenvalues). This approach [Sarmanov and Bratoeva, 1967] can be extended to continuous mixtures.

## 7. Coupled Lancaster Sequences

Suppose $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_T$ is a vector of real valued random variables with finite variances, defined on the same probability space.

**Definition 7.1.** The sequence $\{\underline{x}_i\}$ is a *Coupled Lancaster Sequence* if for each $i$ there is a (possible infinite) sequence of orthonormal functions $f_{i0},f_{i1},\cdots,f_{iR_i}$ such that

$$\mathbf{E}(f_{ir}(\underline{x}_i)\mid\underline{x}_j)=\lambda_{ijr}f_{jr}(\underline{x}_j). \tag{14}$$

In words: for each $r$ all bivariate regressions of the $T$ random variables $f_{1r}(\underline{x_1}),f_{2r}(\underline{x_2}),\cdots,f_{Tr}(\underline{x}_T)$ are linear.

Equation (14) implies, by the definition of conditional expectation,

$$\mathbf{E}(f_{ir}(\underline{x_i})f_{jr}(\underline{x}_j))=\lambda_{ijr}. \tag{15a}$$

[2] Thus $\Gamma_2=2I-\Gamma_1$, which means all eigenvalues of $\Gamma_1$ must be between 0 and 2.

See, for example, Whittle [2000, Section 5.3]. It follows that $\lambda_{ijr} = \lambda_{jir}$ and $\lambda_{iir} = 1$. Thus if $\Lambda^{(r)}$ is the (possibly infinite) matrix with the $\lambda_{ijr}$, then $\Lambda^{(r)}$ is a correlation matrix, i.e. it is symmetric, positive semi-definite, and it has diagonal elements equal to one.

If $r \neq s$ we conclude from (14) that

$$\mathbf{E}(f_{ir}(\underline{x_i})f_{js}(\underline{x}_j)) = 0. \tag{15b}$$

Choosing all $f_{i0} \equiv 1$ satisfies (14) with $\lambda_{ij0} = 1$. Thus for $r = 1,2,\cdots$ we have $\mathbf{E}(f_{ir}(\underline{x_i})) = 0$.

Suppose $\alpha_v^{(r)}$ is the eigenvector of $\Lambda^{(r)}$ corresponding with the $v^{th}$ largest eigenvalue $\gamma_v^{(r)}$. Define $g_{irv} = \alpha_{iv}^{(r)} f_{ir}$. Then

$$\sum_i \mathbf{E}(g_{irv}(\underline{x}_i) \mid \underline{x}_j) = \sum_i \alpha_{iv}^{(r)} \mathbf{E}(f_{ir}(\underline{x}_i) \mid \underline{x}_j) =$$
$$= \sum_i \lambda_{ijr} \alpha_{iv}^{(r)} f_{jr}(\underline{x}_j) = \gamma_v^{(r)} \alpha_{fv}^{(r)} f_{jr}(\underline{x}_j) = \gamma_v^{(r)} g_{jrv}(\underline{x}_j).$$

## 8. Discretization

The Burt Table.

## 9. Application to time series

Stationarity. 1 Trend.

## 10. Nonlinear Singular Spectrum Analysis

Hankelizing.

Combine the linear-quadratic-... solutions with the oscillatory eigenvectors of Toeplitz matrices.

## References

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

J. Dauxois and A. Pousse. *Les Analyses Factorielles en Calcul des Probabilités et en Statistique: Essai d'Étude Synthetique*. PhD thesis, Université Paul-Sabatier, Toulouse, France, 1976.

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.

A. Gifi. *Nonlinear Multivariate Analysis*. Wiley, Chichester, England, 1990.

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

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.

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.

O.V. Sarmanov and Z.N. Bratoeva. Probabilistic Propertgies of Bilinear Expansion of Hermite Polynomials. *Theory of Probability and Its Applications*, 12:470–481, 1967.

B.F. Schriever. *Order Dependence*. PhD thesis, University of Amsterdam, The Netherlands, 1985. Also published in 1985 by CWI, Amsterdam, The Netherlands.

G.P. Styan. Hadamard Products and Multivariate Statistical Analysis. *Linear Algebra and its Applications*, 6:217–240, 1973.

J.L.A. Van Rijckevorsel. *The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis*. PhD thesis, University of Leiden, The Netherlands, 1987. Also published in 1987 by DSWO-Press, Leiden, The Netherlands.

P. Whittle. *Probability via Expectation*. Springer Verlag, New York, Fourth edition, 2000.

G. U. Yule. On the Methods of Measuring Association Between Two Attributes. *Journal of the Royal Statistical Society*, 75:107–170, 1912.

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