ANALYSIS OF DEPENDENCY

José Antônio Cordeiro

RELATÓRIO TÉCNICO Nº 48/87

ABSTRACT. In this paper we use the Hellinger distance to define the dependency of a column (row) in a contigency table, as well as the codependency between two columns (rows). We can explore the association structure in the data through a principal components analysis carried out on the columns (and rows) dependency and codependency matrix, and plotting the projections of the column (row) profiles on subspaces of interest. Some calculations were done on two well known data sets to show applications of the methodology.

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
IMECC – UNICAMP
Caixa Postal 6065
13.081 - Campinas, SP
BRASIL

Novembro – 1987

Analysis of Dependency

José Antônio Cordeiro
Departamento de Estatística
Universidade Estadual de Campinas
CP 6065, 13081 Campinas, SP - Brazil

ABSTRACT

In this paper we use the Hellinger distance to define the dependency of a column (row) in a contigency table, as well as the codependency between two columns (rows). We can explore the association structure in the data through a principal components analysis carried out on the columns (and rows) dependency and codependency matrix, and plotting the projections of the column (row) profiles on subspaces of interest. Some calculations were done on two well known data sets to show applications of the methodology.

Keywords: ASSOCIATION, CODEPENDENCY, CORRESPONDENCE ANALYSIS, DEPENDENCY, DEPENDENCY ANALYSIS, HELLINGER'S DISTANCE.

1. INTRODUCTION

The Hellinger's distance (Hellinger, 1909; Hellinger & Toeplitz, 1910) has been proved to be useful as a σ-finite measure of nearness between probability distributions (see, for example, Beran, 1977). Bhattacharyya (1946) used it to define the divergence between two multinomial distributions without any references to Mr Hellinger's paper. Freeman & Tukey (1950) used the square root transformations of functions of counts in binomial trials in an attempt to stabilize variances, and Matusita (1951) gave the same Hellinger's metric definition of distance between probability distributions in a general probability space. It has, therefore, been called Freeman-Tukey goodness-of-fit statistic (Bishop-Feinberg-Holland, 1975, pg. 513), Matusita's distance (Goldstein-Wolf-Dillon, 1976; Khan & Ali, 1973), and Hellinger's distance (Beran, 1977). Because of

This research was sponsored by FAPESP-Fundação de Amparo à Pesquisa do Estado de São Paulo, São Paulo, Brazil, grant no. 85/0957-1, and by the Department of Statistics, University of Washington, USA, through the usage of computational and structural facilities, during a sabbatical year provided by my university.

precedence, we are going to call it, as we already did, as Hellinger's metric or distance.

Let's then consider an $I \times J$ contingency table $T = (n_{ij})$, where n_{ij} is the count of individuals with the caracteristics defined by the cell (i, j) out of a random sample of size n, $f_{ij} = n_{ij}/n$ and p_{ij} being, respectively, the corresponding relative frequency and the true probability of cell (i, j). The lexicographically ordered vector $(n_{ij} ; 1 \le i \le I, 1 \le j \le J)$ has then a multinomial distribution $M(n, (p_{ij} ; 1 \le i \le I, 1 \le j \le J))$. Let $p_{ij} > 0$ for all pair (i, j).

Khan & Ali (1973) proposed

$$M(X,Y) = \{\tfrac{1}{2}\int[\sqrt{f(x,y)} - \sqrt{f_x(x) f_y(y)}]^2 m(dx, dy)\}^{1/2} \tag{1.1}$$

as a measure of association between two random variables X and Y both absolutely continuous with respect to a measure m on the product space $X \times Y$. Here $f(x,y)$, $f_x(x)$ and $f_y(y)$ are, respectively, the joint and marginals Radon-Nykodin derivatives of the joint distribution of (X,Y) with respect to m, and of X and Y with respect to the respective sections of m.

They also proposed in the same paper

$$M = \{\tfrac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}[\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}}]^2\}^{1/2} \tag{1.2}$$

as a measure of association between rows and columns in a $I \times J$ contingency table, in which $p_{i.}$ is the marginal probability of the $i-th$ row and $p_{.j}$ of the $j-th$ column.

Goldstein et al. (1976) generalized this association measure to higher dimensional tables and did some Monte Carlo comparisons between this, the chi-squared and the likelihood ratio test statistic for independence. Their study showed that these statistics behave quite similar in spike alternatives for some particular tables.

In this paper we are going to deal with measure of association between rows and between columns of a rectangular table, and show that if we add up the column (or row) dependencies, we get the measure of association given in (1.2), and apply the technique on two examples, without do any comparisons with othes analysis of the same data.

2. DEPENDENCY AND CODEPENCENCY

The definitions and results given below refer to columns of a contingency table, but they are extendible to the rows in the same fashion.

definition 1. Let $p = (p_{ij})$ a probability distribution on a $I \times J$ table as defined above, and let $p_{+.}^t = (p_{1.}, p_{2.}, ..., p_{I.})$ and $p_{.+}^t = (p_{.1}, p_{.2}, ..., p_{.J})$ be, respectively, the column and row true marginal distributions. Then we call

$$d^c_{jj^*} = \frac{1}{2}\sum_{i=1}^{I}(\sqrt{p_{ij}} - \sqrt{p_{i.}p_{.j}})(\sqrt{p_{ij^*}} - \sqrt{p_{i.}p_{.j^*}}) \tag{2.1}$$

the *codependency* between columns j and j^* $(j \neq j^*)$ for $1 \leq j \leq J$,

$$d^c_{jj} = \frac{1}{2}\sum_{i=1}^{I}[\sqrt{p_{ij}} - \sqrt{p_{i.}p_{.j}}]^2 \tag{2.2}$$

the *dependency* of column j, and

$$\partial^c_j = \sqrt{d^c_{jj}} \tag{2.3}$$

the *standard dependency* of column j, for $1 \leq j \leq J$,

$$\delta^c_{jj^*} = \frac{d^c_{jj^*}}{\partial^c_j \partial^c_{j^*}} \tag{2.4}$$

the *codependency coefficient* of columns j, j^*, and

$$d^2 = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\sqrt{p_{ij}} - \sqrt{p_{i.}p_{.j}})^2 = \sum_{j=1}^{J} d^c_{jj} \tag{2.5}$$

the total dependency of the table, which equals the sum of the columns (rows) dependencies and the square of the association measure defined in (1.2).

All these definitions, but the last one, are given along the rows (columns) and have their own interpretations. So as the covariance between two random variables, the codependency of columns j and j^* can be viewed as a weighted mean product deviation of the square roots of the expected j and j^* coordinates of the row profiles in the general case, from the square roots of their expected coordinates in the case of independence, with the weights given by the row marginal probabilities as in

$$d^c_{jj^*} = \frac{1}{2}\sum_{i=1}^{I}p_{i.}(\sqrt{p_{ij}/p_{i.}} - \sqrt{p_{.j}})(\sqrt{p_{ij^*}/p_{i.}} - \sqrt{p_{.j^*}}), \tag{2.6}$$

or as a more geometrical one, the inner product of the deviations of the square roots of the column profiles j and j^*, both from the square root of the columns marginal, weighted by the product of the square roots of the respective row marginal coordinates

$$d^c_{jj^*} = \frac{1}{2}\sqrt{p_{.j}p_{.j^*}}\sum_{i=1}^{I}(\sqrt{p_{ij}/p_{.j}} - \sqrt{p_{i.}})(\sqrt{p_{ij^*}/p_{.j^*}} - \sqrt{p_{i.}}) \tag{2.7}$$

The same interpretations are given for the dependencies, and also that

$$d^c_{jj} = 2p_{.j}(\sin(\theta_j/2))^2 \tag{2.8}$$

where θ_j is the angle between the vectors square roots of the $j-th$ column profile, and of the column marginal p_{+}.

definition 2. We will say that the column j $(1 \le j \le J)$ is independent of the rows if

$$p_{ij} = p_{i.}p_{.j} \tag{2.9}$$

for all $i = 1, 2, ..., I$, and that the row i $(1 \le i \le I)$ is independent of the columns if (2.9) holds for all $j = 1, 2, ..., J$.

proposition 1. The codependency of two columns j, j^* (rows i, i^*) is zero, if one of them is independent of the rows (columns).
proof : trivial from the definitions (2.1) and (2.9).QED

proposition 2. The dependency of a column j (row i) is zero if and only if it is independent of the rows (columns).
proof : if the column j is independent of the rows, then by the definitions (2.2) and (2.9), $d_{jj}^c = 0$. Let us now suppose that $d_{jj}^c = 0$. Then each term of the sum (2.2) is zero, for it is a sum of nonnegative numbers, hence the result follows. QED

The total dependency can be written as

$$d^2 = 1 - \sum_{i=1}^{I}\sum_{j=1}^{J}\sqrt{p_{ij}p_{i.}p_{.j}}\,, \tag{2.10}$$

and it is allways less than 1 in a contingency table.(1) The double sum in (2.10) is the *affinity* (Matusita, 1951) between the distributions p in the general case, and in the case of independence.

proposition 3. The absolute value of the codependency coefficient $\delta_{jj^*}^c$, at most equals 1.
proof : By the Schwartz inequality we have that $|d_{jj^*}^c| \le \partial_j^c \partial_{j^*}^c$. Then the result follows. QED

This coefficient has an interpretation which parallels that of correlation coefficient between real valued variables, that is, if it is positive it means that when one column tends to depart from independence (the expected situation), along the rows on one side, the other column does the same, and if it is negative, the departures are side opposite.

(1)- in the definition (1.1) M(X,Y) assumes the value 1 if and only if the supports of X and Y are disjoints.

3. SAMPLE DEPENDENCY AND CODEPENDENCY

So far with the definitions of the dependency parameters, we have to deal now with their estimation, and with significance tests for them.

As pointed out in the introduction we will consider the multinomial model on the $I \times J$ table T. So, out of a random sample of size n, n_{ij} is the number of sample items which satisfy the properties defining the cell (i,j), and $f_{ij} = n_{ij}/n$ is the respective relative frequency. As allways, $\sum n_{ij} = n$ and $\sum f_{ij} = 1$.

Each parameter will be estimated by its maximum likelihood estimator just by substituting p_{ij} by f_{ij}, and the marginal probabilities by their respective marginal relative frequencies. Hence we have

$$d^c_{jj^*} = \tfrac{1}{2}\sum_{i=1}^{I}(\sqrt{f_{ij}} - \sqrt{f_{i.}f_{.j}})(\sqrt{f_{ij^*}} - \sqrt{f_{i.}f_{.j^*}}) \ , \quad (3.1)$$

$$d^c_{jj} = \tfrac{1}{2}\sum_{i=1}^{I}(\sqrt{f_{ij}} - \sqrt{f_{i.}f_{.j}})^2 \ , \quad (3.2)$$

and

$$\partial^c_j = \sqrt{d^c_{jj}} \ , \quad (3.3)$$

and

$$\delta^c_{jj^*} = \frac{d^c_{jj^*}}{\partial^c_j \partial^c_{j^*}} \quad (3.4)$$

$$d^2 = \sum_{j=1}^{J} d^c_{jj} \quad (3.5)$$

as estimators of the parameters defined in (2.1)-(2.5). The estimators for the row parameters are defined in the same fashion.[2] Under independence of rows and columns in the table, $8nd^2$ is asymptotically distributed chi-squared with $(I-1)(J-1)$ degrees of freedom (Bhattacharrya, 1946; Khan & Ali, 1973).

Let us apply these definitions on some examples before we deal with hypotheses testing.

example 1. The table bellow refer to a cross-classification of 7,477 women according to their right

(2)- for simplicity, we are using the same notation for the parameters and their estimators.

eye grade and left eye grade with respect to unaided distance vision. It was analysed earlier by Stuart (1953, 1955), Bishop et al. (1975, page 284), Tukey (1977, page 519), Gokhale and Kullback (1978, page 255), Kendall and Stuart (1979, page 618), Plackett (1981, page 25), Agresti (1984, page 215) and Goodman (1985):

Table 1

Cross-classification of 7,477 women according to their right eye grade and left eye grade with respect to unaided distance vision

right eye grade	left eye grade: Best	Second	Third	Worst	Total
Best	1520	266	124	66	1976
Second	234	1512	432	78	2256
Third	117	362	1772	205	2456
Worst	36	82	179	492	789
Total	1907	2222	2507	841	7477

The estimated matrices of column (left eye grade) and row (right eye grade) codependency coefficients are

Table 2

Left eye grade codependency coefficients

	Best	*Second*	*Third*	*Worst*
Best	1.000	-.174	-.666	-.302
Second	-.174	1.000	-.157	-.336
Third	-.666	-.157	1.000	.230
Worst	-.302	-.336	.230	1.000

Table 3

Right eye grade codependency coefficients

	Best	*Second*	*Third*	*Worst*
Best	1.000	-.223	-.679	-.320
Second	-.223	1.000	-.138	-.265
Third	-.679	-.138	1.000	.278
Worst	-.320	-.265	.278	1.000

We note in tables 2 and 3 the similar patterns of the codependencies for the left and right eye grades.

The dependencies for both left and right eye grades are shown in the

Table 4

Dependency of eye grade

eye	*grade*			
	Best	Second	Third	Worst
Left	.041	.022	.030	.018
Right	.040	.023	.030	.019

We note also that the dependencies of the grades for the left and right eyes have almost equal values.

example 2. This example was considered earlier by Srole et al. (1962, page 213), Haberman (1974; 1979, page 375), Gross (1981), Escoufier (1982), Gigula and Haberman (1984) and Goodman (1985).

Table 5

Cross-classification of 1,660 people according to their mental health and parents' socioeconomic status

mental health status	parents' socioeconomic status A	B	C	D	E	F	Total
Well	64	57	57	72	36	21	307
Mild S.F.	94	94	105	141	97	71	602
Moderate S.F.	58	54	65	77	54	54	362
Impaired	46	40	60	94	78	71	389
Total	262	245	287	384	265	217	1660

Note: S.F. for Symptom Formation

Tables 6 and 7 bellow give the sample codependency coefficients for columns and rows of table 5, respectively.

Table 6

Parents' socioeconomic status codependency coefficients

	A	*B*	*C*	*D*	*E*	*F*
A	1.000	.958	.934	-.284	-.989	-.931
B	.958	1.000	.963	-.313	-.940	-.914
C	.934	.963	1.000	-.551	-.951	-.802
D	-.284	-.313	-.551	1.000	.387	-.040
E	-.989	-.940	-.951	.387	1.000	.892
F	-.931	-.914	-.802	-.040	.892	1.000

Table 7

Mental health status codependency coefficients

	Symptom Formation			
	Well	*Mild*	*Moderate*	*Impaired*
Well	1.000	.710	-.281	-.956
Mild S.F.	.710	1.000	-.653	-.657
Moderate S.F.	-.281	-.653	1.000	.070
Impaired	-.956	-.657	.070	1.000

Note: S.F. for Symptom Formation

We note that the socioeconomic status *D*, in Table 6, and mental health statuses *symptom formation*, in Table 7, are less codependent with the other statuses than these ones are codependent amongst themselves. We shall test the significance of the codependencies and dependencies later in this paper.

Tables 8 and 9 show the dependencies for the categories in this example.

Table 8

Parents' socioeconomic status dependency ($\times 10^{-4}$)

	A	*B*	*C*	*D*	*E*	*F*
Dependency	6.612	6.967	.909	.584	5.942	15.562

Table 9

Mental health status dependency ($\times 10^{-3}$)

	Symptom Formation			
	Well	Mild	Moderate	Impaired
Dependency	1.776	.085	.137	1.660

Here we also note that the socioeconomic statuses C and D, and the mental health *symptom formation* statuses have smaller dependencies than the other levels of their respective categories.

The significance of the column and row dependencies can be tested, as well as the codependencies. It do not exist yet exact tests for doing this, but if the sample size is big enough, we can use weighted least square (see Grizzle et al., 1969) for testing those hypotheses.

The hypothesis $H_j^c : d_{jj}^c = 0$ is equivalent to

$$H_{oj}^c : p_{ij} = p_{i.}p_{.j} \ , \text{for } i = 1, 2, ..., I-1, \tag{3.6}$$

then if we use the model

$$F_j = Ap = I_{I-1}\beta = \beta \ , \tag{3.7}$$

where the vector $p = (p_{ij})$ is lexicographically ordered, I_a is the a-dimensional identity matrix, and

$$A = C_3(exp\,(C_2(logC_1))) \tag{3.8}$$

with

$$C_1 = \begin{bmatrix} I_{I-1} \otimes e_j^t & O_{I-1,J} \\ I_{I-1} \otimes 1_J^t & O_{I-1,J} \\ & 1_I^t \otimes e_j^t \end{bmatrix} : (2I-1)\times IJ \ , \tag{3.9}$$

where e_j is the $j-th$ vector of the ordered canonical base of R^J, 1_k a k-dimensional vector of ones, $O_{a,b}$ a $(a \times b)$-dimensional matrix of zeros, and $\otimes$ is the Kronecker product,

$$C_2 = \begin{bmatrix} I_{I-1} & O_{I-1,I-1} & O_{I-1,1} \\ O_{I-1,I-1} & I_{I-1} & 1_{I-1} \end{bmatrix} : 2(I-1)\times(2I-1) \ , \tag{3.10}$$

$$C_3 = \begin{bmatrix} 1 & -1 \end{bmatrix} \otimes I_{I-1} : (I-1)\times 2(I-1). \tag{3.11}$$

Then $F_j^t = (p_{ij} - p_{i.}p_{.j} \ , \ldots , p_{(I-1)j} - p_{(I-1)}p_{.j})$, and the hypothesis is $H_{oj}^c : C\beta = 0$, with $C = I_{I-1}$.

The weighted least square statistic for H_{oj}^c is

$$(Cf)^t V_f^{-1} Cf \ , \tag{3.12}$$

which is distributed χ_{I-1}^2, under H_{oj}^c, where $V_f = HV_pH^t$, with

$$V_p = (diag\,(f) - ff^t)/n \tag{3.13}$$

the maximum likelihood estimator of the covariance matrix of f, the vector of relative frequencies lexicographically ordered, and

$$H = C_3 D_{a2} C_2 D_{a1}^{-1} C_1 \tag{3.14}$$

with

$$D_{a1} = diag(C_1 f) \tag{3.15}$$

$$D_{a2} = diag(exp(C_2 log(C_1 f))) \tag{3.16}$$

being diagonal matrices.

Let us test $H_{jj*}^c : d_{jj*}^c = 0 \; (j \neq j^*)$. By (2.1) the hypothesis H_{jj*}^c is then equivalent to

$$\sum_{i=1}^{I} (\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}})(\sqrt{p_{ij*}} - \sqrt{p_{i.} p_{.j*}}) = 0 \tag{3.17}$$

which can be tested by fitting the model

$$F_{jj*} = A_{jj*} p = \beta_{jj*} \tag{3.18}$$

where

$$F_{jj*}^t = ((\sqrt{p_{1j}} - \sqrt{p_{1.} p_{.j}})(\sqrt{p_{1j*}} - \sqrt{p_{1.} p_{.j*}}), \ldots, (\sqrt{p_{Ij}} - \sqrt{p_{I.} p_{.j}})(\sqrt{p_{Ij*}} - \sqrt{p_{I.} p_{.j*}})) \tag{3.19}$$

is the vector of the terms of the sum (3.16), and

$$A = exp(C_4 log(C_3 sqrt(I_I exp(C_2 log(C_1^{jj*}))))) \tag{3.20}$$

with

$$C_1^{jj*} = (C_j^t \,|\, C_{j*}^t \,|\, C_o^t)^t \; : (3I+2) \times IJ \; , \tag{3.21}$$

$$C_j = \begin{bmatrix} I_I \otimes e_j^t \\ 1_I^t \otimes e_j^t \end{bmatrix} : (I+1) \times IJ \; , \tag{3.22}$$

$$C_o = I_I \otimes 1_J^t \; : I \times IJ \; , \tag{3.23}$$

$$C_2 = \begin{bmatrix} I_I & O_{I,1} & O_{I,J} & O_{I,1} & O_{I,J} \\ O_{I,J} & 1_I & O_{I,J} & O_{I,1} & I_I \\ O_{I,1} & O_{I,1} & I_I & O_{I,1} & O_{I,J} \\ O_{I,J} & O_{I,1} & O_{I,J} & 1_I & I_I \end{bmatrix} : 4I \times (3I+2) \; , \tag{3.24}$$

$$C_3 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \otimes I_I \; : 2I \times 4I \; , \tag{3.25}$$

and

$$C_4 = [1\ \ 1] \otimes I_I \ : I \times 2I \ , \tag{3.26}$$

The hypothesis is equivalent to

$$H^c_{ojj*} : C\beta = 0 \tag{3.27}$$

with $C = 1^t_{I-1}$.

The testing statistic for (3.27) is

$$Q = (Cf)^t (CV_f C^t)^{-1} Cf \tag{3.28}$$

which is asymptotically distributed χ^2_1 under H^c_{ojj*}, where

$$V_f = HV_p H^t \tag{3.29}$$

and

$$H = D_{a5} C_4 D^{-1}_{a4} C_3 D^{-1}_{a3} D_{a2} C_2 D^{-1}_{a1} C_1 \tag{3.30}$$

with

$$D_{a1} = diag\,(C_1 f) \tag{3.31}$$

$$D_{a2} = diag\,(\exp C_2 (logC_1 f)) \tag{3.32}$$

$$D_{a3} = 2diag\,(sqrt\,(I \exp C_2 (logC_1 f))) \tag{3.33}$$

$$D_{a4} = diag\,(C_3 sqrt\,(Iexp\,(C_2 (log\,(C_1 f))))) \tag{3.34}$$

and

$$D_{a5} = diag\,(F_{jj*}) \tag{3.35}$$

are diagonal matrices. For more details on the above, see for example, Grizzle et al. (1969), or Koch et al. (1984, page 2.46).

Let us now test the dependency significance in the examples we are dealing with.

example 1 - (continued) - all the dependencies and codependencies of the unaided eye grade data are significant with p-values less than 10^{-4}.

example 2 - (continued)

Table 10

significance test for dependency of Mental Health Status

	Symptom Formation			
	Well	*Mild*	*Moderate*	*Impaired*
p-value	.000	.880	.792	.000

Note: degrees of freedom = 5

Table 11

significance test for dependency of parents' socioeconomic status

	A	*B*	*C*	*D*	*E*	*F*
p-value	.002	.014	.750	.752	.019	.000

Note: degrees of freedom = 3

Tables 10 and 11 above show that there is no evidence that *symptom formation* statuses of mental health are dependent of parents' socioeconomic status, and on the other side, there is also no evidence that parents' socioeconomic statuses C and D are dependent of mental health status.

Table 12

p-value for WLS test for significance of codependency

parents' socioeconomic status (1)

	A	*B*	*C*	*D*	*E*
B	.000				
C	.025	.022			
D	.590	.552	.422		
E	.000	.000	.044	.443	
F	.000	.000	.067	.937	.000

(1)- WLS for weighted least squares

Note: degrees of freedom = 1

Table 13
p-value for WLS test for significance of codependency
mental health status (1)

	Well	*Mild S.F.* (2)	*Moderate S.F.*
Mild S.F.	.057		
Moderate S.F.	.412	.244	
Impaired	.000	.111	.834

(1) WLS for weighted least squares; (2) S.F. for Symptom Formation
Note: degrees of freedom =1

Table 12 shows that there is no evidence that the parents' socioeconomic status D is codependent to the other statuses. This is, in fact, a consequence of the no significance of its dependency to mental health status (cf. table 11), nonetheless we know that logic implications are in general not confirmed by statistical tests. The status C has its codependencies with the statuses A, B, E and F reasonably significant, despite not been significantly dependent of mental health status.

4. GRAPHICS DISPLAYS

The most important feature of correspondence analysis (see, for example, Greenacre, 1984; Lebart et al., 1984) is that we can display simultaneously projections of rows and columns profiles into bidimensional subspaces, and see association structures between them. This analysis uses the chi-squared metric.

With the Hellinger's metric we can do an analog to correspondence analyis.

Suppose $J \leq I$, and let u be a unit vector in R^J, and $p_i^r = (p_{i1}/p_{i.}, \ldots, p_{iJ}/p_{i.})^t$ for $i = 1, 2, ..., I$ the row profiles in R^J. Under the general model the inner product of u and $\sqrt{p_i^r}$ is

$$u^t\sqrt{p_i^r} = \sum_{j=1}^{J} u_j\sqrt{p_{ij}/p_{i.}} = 1, 2, ..., I \qquad (4.1)$$

and under independence in the table T

$$u^t\sqrt{p_i^r} = \sum_{j=1}^{J} u_j\sqrt{p_{.j}} = u^t\sqrt{p_{.+}}\,. \qquad (4.2)$$

definition 3. We will call dependency of the unit vector u in R^J the value

$$d_u = \sum_{i=1}^{I} p_{i.}[\sum_{j=1}^{J} u_j(\sqrt{p_{ij}/p_{i.}} - \sqrt{p_{.j}})]^2 \qquad (4.3)$$

To better "see" the cloud of square roots of row profiles on the unit sphere in R^J, we ask which vectors would give us an orthonormal basis which show them from the side of their biggest dependency, that is, which orthonormal vector set $\{ \gamma_1, \ldots, \gamma_J \}$ maximize (4.3). The answer to this question is given by

proposition 4. The vectors $\gamma_1, \ldots, \gamma_J$ which maximize (4.3), subject to $\|\gamma_j\| = 1$, $j = 1, 2, \ldots, J$, and $\gamma_j^t \gamma_{j^*} = 0$, $j \neq j^*$, are the eigenvectors of de column dependency-codependency matrix

$$\Delta^t \Delta = [d^c_{jj^*}] : J \times J \tag{4.4}$$

related to its ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_{J-1}, \lambda_J$, where

$$\sqrt{2}\Delta = [\sqrt{p_{ij}} - \sqrt{p_{i.}p_{.j}}] : I \times J \tag{4.5}$$

is the matrix of the deviations of the square roots of the cell probabilities in the general case from that in the case of independence.

proof : the dependency of the unit vector u in (4.3) can be written as the quadratic form

$$d_u = u^t \Delta^t \Delta u \ . \tag{4.6}$$

Hence the maximization problem subject to orthonormality is reduced to the eigen-decomposition of $\Delta^t \Delta$. Since this matrix is symmetric and positive semi-definite, we have $\lambda_1 \geq, \ldots, \geq \lambda_{J-1} \geq 0, \lambda_J \geq 0$, and the set $\{\gamma_1, \ldots, \gamma_J\}$ is orthonormal.QED

In the same fashion we deal with in R^I, but as $J \leq I$ it suffices to work in R^J, and solve the eigenproblem in R^I as shown by

proposition 5. If γ_j is eigenvector of $\Delta^t \Delta$ related to λ_j, then $\psi_j = \Delta \gamma_j$ is eigenvector of $\Delta \Delta^t$, related to the same eigenvalue λ_j.

proof :

$$\Delta^t \Delta \gamma_j = \lambda_j \gamma_j \ \Rightarrow \ \Delta \Delta^t (\Delta \gamma_j) = \lambda_j (\Delta \gamma_j). \ QED$$

Let us set $\psi_j = \Delta \gamma_j / \sqrt{\lambda_j}$. Then we have $\psi_j^t \psi_j = 1$.

The maximum likelihood estimator of matrix Δ in (4.5) is

$$\sqrt{2}D = [\sqrt{f_{ij}} - \sqrt{f_{i.}f_{.j}}] : I \times J \tag{4.7}$$

and then, $D^t D$ is the sample dependency-codependency matrix.

The sample total dependency (3.5) equals the sum of the eigenvalues of $D^t D$, for $d^2 = trace\,(D^t D) = \sum_{j=1}^{J} \lambda_j$, and $d_{\gamma_j} = \lambda_j$.

The eigenvalue λ_J refers to the size of the smallest sector of the unit sphere in R^J containing all the square roots of the row profiles. The more scattered are these points on the sphere, the bigger is λ_J. Its corresponding eigenvector has all the coordinates of same sign. In correspondence analysis (Lebart et al., 1984) we have at most $J-1$ factors (eigenvectors), and here also we have the same number of factors, but the $J-th$ eigenvector has its usefulness by showing *how much* the points are scattered on the sphere, meanwhile the others $J-1$ show *how* they are scattered on the sphere. The same interpretation is given when we are looking to the points on the sphere in R^I.

definition 2: Let $\lambda_1 \geq, \ldots, \geq \lambda_k$, $k \leq J-1$ be k the first eigenvalues of $D'D$ not related to the size of the minimal sphere sector containing the row profiles. The value

$$ed(k) = \sum_{j=1}^{k} \frac{\lambda_j}{d^2} \times 100\% \tag{4.8}$$

is called *percentage of the total dependency explained by the k first factors*.

The tables bellow show the eigenvectors of the examples we are examining, as well as their eigenvalues and explained percentages.

example 1 – (*continued*):

Table 14

eigen-decomposition of dependency-codependency matrices

left eye grade and right eye grade

				Left				*Right*			
j	λ_j	%	*ed(j)(%)*	Bst	2nd	3rd	Wst	Bst	2nd	3rd	Wst
1	.062	55.6	55.6	.771	-.006	-.599	-.214	.757	-.041	-.607	-.239
2	.028	25.4	81.0	.256	-.846	.185	.430	.259	-.852	.224	.396
3	.015	13.1	94.1	.239	-.186	.579	-.757	.142	-.204	.514	-.821
4	.007	5.9	100.0	.531	.500	.521	.444	-.583	-.481	-.563	-.334

Note: Bst = Best; Wst = Worst

The first factor, which explain 55.6% of the total dependency, opposes the *Best* to the *Third* and *Worst* grades on both eyes, and also on both eyes the second factor (25.4%) weights the *Second* grade much more than the others. If we go back to tables 2 and 3, we see that the *Second* is that one with the smaller dependency with the other grades, on both eyes, and that the *Best* and *Third* grades have the biggest dependencies. Factor 3 apposes the *Third* and *Worst* grades, which are the only grades positively codependent (cf. Tables 2 and 3). We could then say that the vision intensity is the most important factor of variation between the individuals of the population under analysis, from the point of view of both eyes. The second and third factors are also important, but of difficult interpretations.

The last factor (5.9%) is the one determined by the size of the sphere sector containing the square roots of the row profiles in R^J, and the square roots of the column profiles in R^I. Note that the weight given by this factor to the grade *Worst* is quite different from those given to the others, which have more or less the same size.

Table 15

eigen-decomposition of dependency-codependency matrices

mental health status

j	$\lambda_j(\times 10^{-4})$	%	*ed(j)(%)*	*Well*	*Mld.S.F.*(1)	*Mod.S.F.*	*Impaired*
1	34.068	93.1	93.1	-.715	-.111	.040	.689
2	2.000	5.5	98.6	-.297	-.327	.799	-.408
3	.446	1.2	99.8	-.658	.645	-.190	-.475
4	.001	.2	100.0	.278	.682	.569	.366

(1) Mld.S.F. = Mild Symptom Formation; Mod.S.F. = Moderate Symptom Formation

Table 16

eigen-decomposition of dependency-codependency matrices

parents' socioeconomic status

j	$\lambda_j(\times 10^{-4})$	%	*ed(j)(%)*	*A*	*B*	*C*	*D*	*E*	*F*
1	34.068	93.1	93.1	-.434	-.439	-.149	.024	.403	.658
2	2.000	5.5	98.6	.225	.258	.261	-.524	-.378	.630
3	.446	1.2	99.8	-.350	.769	.172	-.035	.505	.012
4	.001	.2	100.0	-.699	-.176	.165	-.498	-.299	-.340

In this example the first two factors explain 98.6% of the total dependency, and the first one alone explains 93.1%, and it opposes the mental health statuses *Well* against *Impaired* in terms of their association to parents' socioeconomic status. It also opposes the best parents' socioeconomic status against the worst'ones.

Given the eigenvectors $\gamma_1, \ldots, \gamma_J$ of $D^t D$ we have the matrix $\Gamma = (\gamma_1 | \gamma_2 | \ldots | \gamma_{J-1} | \gamma_J)$, and respectively $\Psi = (\psi_1 | \psi_2 | \cdots | \psi_{J-1} | \psi_J)$ as the matrix of the eigen-vectors of DD^t, and do the principal component transformation (PCT)

$$q^r = \Gamma^t (p^r - p_{+}), \qquad (q^c = \Psi^t (p^c - p_{+})), \tag{4.9}$$

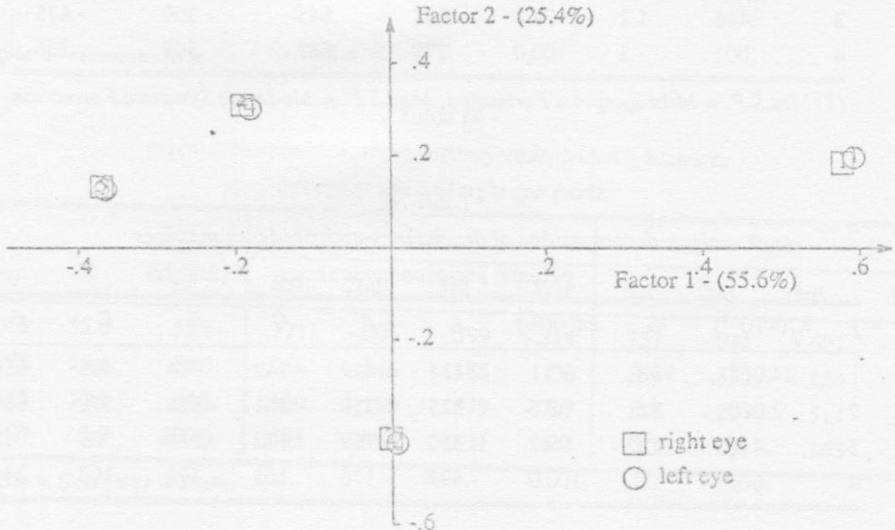
and plot the coordinates of the q^r and q^c projections simultaneously on the planes $F_1 \times F_2, F_1 \times F_3, \ldots, F_{k-1} \times F_k$ where $F_1, \ldots, F_k$ are factors representing the first k columns of Γ and of $D\Gamma$ which achieve certain percentage of explainded total dependency (ETD).

The figures bellow show $F_1 \times F_2$ projections of the examples we are dealing with.

example 1 – (continued)

Figure 1

Unaided Eye data - Simultaneous projections of Right and Left eye grade profiles on the two first factors subspace - percentage of ETD = 81.0%

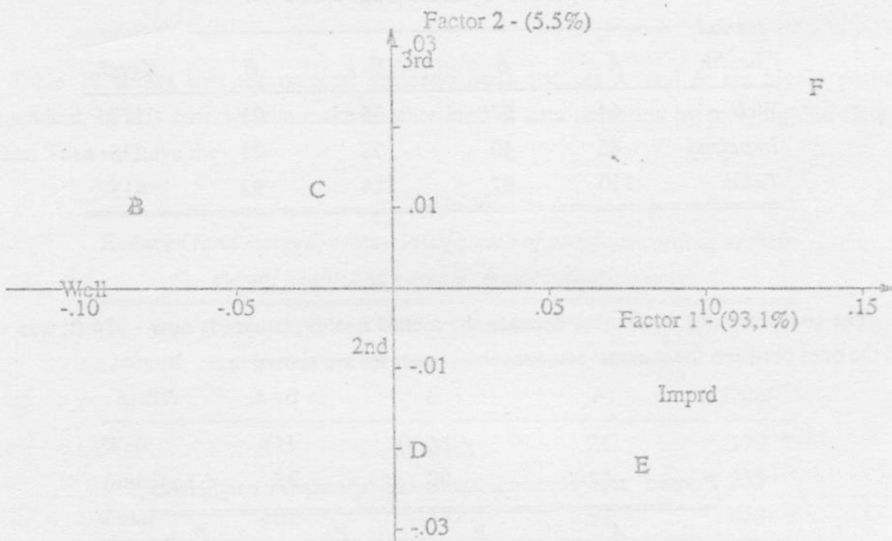


This figure shows that from the point of view of the two first factors, the *Left* and *Right* eye grades are almost symmetric in terms of profiles. The model of symmetry doesn't fit the data (see Bishop et al., 1975 page 284). This lack of symmetry is more visible when we look at the projections of the profiles on the fourth eigenvector, the factor of spreadness of the points on the unit sphere (not shown here).

example 2 – (continued)

Figure 2

Mental Health & Parents' Socioeconomic Status - Simultaneous projections of the mental health and parents' socioeconomic statuses on the two first factors space - percentage of ETD = 98.6%



Looking at the Figure 2 we see that parents' socioeconomic status decreases (from A to F) along the first factor, and so does mental healt status. This allows us to interprete this factor as the socioeconomic influence on mental health, when the classifications of both categories are done as they were in this example. And this factor is the far more important amongst the four available (93,1% of ETD). We also see that the nondependent classes of parents' socioeconomic status and mental health status (cf. Tables 10 and 11) have their projections on this factor much nearer zero than the others. The second factor explain much less of the total dependency (5,5%), being therefore more difficult to interprete.

Because we have nondependent classes in both categories in this example, we can merge them into the respective marginal distributions, i.e, we can reduce the data dimensionality by the *symptom formation* statuses of mental health and parents' socioeconomic statuses C and D, as in

Table 17

Reduced cross-classification of people according to their mental health and parents' socioeconomic status

mental health	*parents' socioeconomic status* A	B	E	F	Total
Well	64	57	36	21	178
Impaired	46	40	78	71	235
Total	110	97	114	92	413

The codependency coefficient between the mental health statuses is now -.984 (it was -.956), and the ones between the parents' socioeconomic statuses are shown in

Table 18

Parents' socioeconomic status - codependency coefficients

	A	B	E	F
A	1.000	1.000	-.991	-.983
B	1.000	1.000	-.990	-.982
E	-.991	-.990	1.000	.999
F	-.983	-.982	.999	1.000

All the codependencies are significant with p-values less than 10^{-4}.

We are not going to show the dependencies values, but instead the p-values for tests of their significance.

Table 19

Significance tests for dependency - mental health and parents' socioeconomic status --- REDUCED DATA

	mental health		parents' socioecnomic status			
	Well	Imp	$A\ (\times 10^{-4})$	$B\ (\times 10^{-4})$	$E\ (\times 10^{-3})$	$F \times 10^{-6})$
p-value	0	0	2.229	4.507	2.594	1.788

Note: Imp = Impaired

Table 18 shows that the parents' socioeconomic statuses A and B are highly positively codependent. In this case we can make another kind of data reduction by merging one class into another. Then we have the

Table 20

Reduced (and merged) cross-classification of people according to their mental health and parents' socioeconomic status

mental health	parents' socioeconomic status $A+B$	E	F	Total
Well	121	36	21	178
Impaired	86	78	71	235
Total	207	114	92	413

The codependency coefficients between the classes of both categories remain the same (the differences are less than 10^{-3}), and highly significant. The p-values for the tests of the dependencies are shown in the

Table 21

Significance tests for depedency - mental health and parents' socioeconomic status --- REDUCED AND MERGED DATA

	mental health		parents' socioeconomic status		
	Well	Imp	$A+B\ (\times 10^{-4})$	$E\ (\times 10^{-3})$	$F\ (\times 10^{-6})$
p-value	0	0	0	2.594	1.788

Note: Imp = Impaired

We see that by merging the classes A and B into another, the dependencies remain the same (not shown), and also the p-values for their tests of significance remains pratically unaffected, except for the merged classes.

The only factor of association between the two categories of this example hold 99.237% of ETD in both reductions, and also its interpretations are the same as the one given for the complete data set.

The decomposition $D'D = \Gamma\Lambda\Gamma'$ such that $\Gamma\Gamma' = \Gamma'\Gamma = I$, gives us the interpretation of the composition of column codependencies and dependendencies through the identities:

$$d^c_{jj^*} = \sum_{m=1}^{J} \lambda_m \gamma_{jm} \gamma_{j^*m} \quad \text{for } j \neq j^* \tag{4.10}$$

and

$$d^c_{jj} = \sum_{m=1}^{J} \lambda_m \gamma^2_{jm} \tag{4.11}$$

as being, respectively, a kind of contrast between the various factors with positive $\lambda's$, for $\sum_{m=1}^{J} \gamma_{jm}\gamma_{j^*m} = 0$, such that the weight of the $k-th$ factor is $\lambda_k \gamma_{jk} \gamma_{j^*k}$, which explain how the dependency between columns j and j^* is decomposable, and a proportional (or mean) factor decomposition of the dependency of column j, for $\sum_{m=1}^{J} \gamma^2_{jm} = 1$.

The same has to be said of the row codependencies and dependencies as functions of the entries of $\Psi = R\,\Gamma$.

We can then call

$$\alpha^c_{jj^*,m} = \frac{\lambda_m \gamma_{jm} \gamma_{j^*m}}{d^c_{jj^*}} \times 100\% \tag{4.12}$$

the percentual index of contribution of the $m-th$ factor to the column codependency $d^c_{jj^*}$, and

$$\alpha^c_{jj,m} = \frac{\lambda_m \gamma^2_{jm}}{d^c_{jj}} \times 100\% \tag{4.13}$$

the percentual contribution of the $m-th$ factor to the column dependency d^c_{jj}.

The dependency of factor m $(m=1, 2, ..., J)$ is

$$\lambda_m = \gamma'_m D' D \gamma_m = \psi'_m D D' \psi_m \tag{4.14}$$

or

$$= \sum_{k=1}^{J} \gamma^2_{km} d^c_{kk} + 2 \sum_{k<k'} \gamma_{km} \gamma_{k'm} d^c_{kk'} \tag{4.15}$$

or

$$= \sum_{k=1}^{J} \psi_{km}^2 d_{kk}^r + 2 \sum_{k<k'} \psi_{km} \psi_{k'm} d_{kk'}^r \ , \ m = 1, 2, \ldots, J. \qquad (4.16)$$

The equations (4.15) and (4.16) give us decompositions of λ_m in terms of linear combinations of columns and of rows dependencies and codependencies. We still do not have a interpretation for this decompositions, but they could give us out some light on the composition of the $m-th$ factor of association between the two categories which compose the table.

Acknowledgment – Many thanks are given to Dr. R. Douglas Martin, and Dr. Michael Perlman of the Department of Statistics of University of Washington, Seattle, Wa, for providing the necessary environment I needed for work out my ideas.

REFERENCES

Agresti, A.A. (1984) *Analysis of Ordinal Categorical Data*, Wiley, New York

Beran, R. (1977) Robust location estimates, Ann. of Statist., 5 No.3 431-444

Bhapkar, V.P. (1966) A note on the equivalence of two criteria for hypotheses in categorical data, J. Amer. Statist. Assoc., 61 228-235

Bhattacharryya, A. (1946) On a measure of divergence betwee two multinomial populations, Sankhya, 7 part IV 401-406

Bishop, Y.M.M., Feiberg, S.E. and Holland, P.W. (1975) *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts, USA

Bowker, A.H. (1948) A test for symmetry in contingency tables, J. Amer. Statist. Assoc., 43 572-574

Caussinus, H. (1966) Contribution à l'analyse statistique des tableaux de corrélations, Ann. Fac. Sci. Univ. Toulouse, France 29 77-182

Escoufier, Y. (1982) L'analyse des tableaux de contingence simples et multiples. (*Proc. Internat. Meeting on the Analysis of Multimensional Contingency Tables* (*Rome*, 1981),Metron, 40 53-77

Freeman, M.F. and Tukey, J.W. (1950) Transformations related to the angular and the square root, Ann. Math. Statist., 21 607-611

Gigula, Z. and Haberman, S. (1984) Canonical analysis of contingency tables by maximum likelihood, Manuscript

Gokhale, D.V. and Kullback, S. (1978) *The Information in Contingency Tables*, Dekker, New

York, NY, USA

Goldstein, M., Wolf, E. and Dillon, W. (1976) On a test of independence for contingency tables, Comm. Statist.- Theor. Meth., A5(2) 159-169

Goodman, L.A. (1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, Ann. Statist., 13 No.1 10-69

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

Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969) Analysis of categorical data by linear models, Biometrics, 25 489-504

Gross, S.T. (1981) On asymptotic power and efficiency of tests of independence in contingency tables with ordered classifications, J. Amer. Statist. Assoc., 76 935-941

Haberman, S.J. (1974) *The Analysis of Frequency Data*, University of Chicago Press, Chicago,Ill, USA

------------- (1979) *The Analysis of Qualitative Data, Vol. 2: New Developments*, Academic Press, New York, NY, USA

Hellinger, E. (1909) Quadratischen Formen von unendlichvielen Veränderlichen, J. reine u. ang. Math.,136 210-271

Hellinger, E. and Toeplitz, T.H. (1910) Grundlagen für eine Theorie der unendlichen Matrizen, Math. Annalen, 69 289-330

Ireland, C.T. and Kullback, S. (1969) Symmetry and marginal homogeneity of an $r \times r$ contingency table, J. Amer. Statist. Assoc., 64 1323-1341

Khan, A.H. and Ali, S.M. (1973) A new coefficient of association, Ann. Inst. Statist. Math., 25 41-50

Kendall, M.G. and Stuart, A. (1979) *The Advanced Theory of Statistics*, Vol. 2, Fourth Edition, Hafner, New York, NY, USA

Koch, G.G. and Reinfurt, D.W. (1971) The analysis of categorical data from mixed models, Biometrics, 27 157-173

Koch, G.G., Imrey, P.B., Singer, J.M., Atkinson, S.S. and Stokes, M.E. (1984) *Lecture Notes for Analysis of Categorical Data*, University of North Carolina, Chapel Hill, NC, USA

Lebart, L., Morineau, A. and Warwick, K.M. (1984) *Multivariate Descriptive Statistical Analysis, Correspondence Analysis and Related Techniques for Large Matrices*, Wiley, New York, NY, USA

Matusita, K. (1951) On the theory of statistical decision functions, Ann. Inst. Statist. Math., 3 17-35

Plackett, R.L. (1981) *The Analysis of Categorical Data*, 2nd Edition, Macmillan, New York, NY,USA

Srole, L., Langner, T.S., Michael, S.T., Opler, M.K. and Rennie, T.A.C. (1962) *Mental Health in the Metropolis: The Midtown Manhattan Study*, McGraw-Hill, New York, NY, USA

Stuart, A. (1953) The estimation and comparison of strenghts of association in contingency tables, Biometrika, 40 105-110

Stuart, A. (1955) A test for homogeneity of the marginal distributions in a two-way classification, Biometrika, 42 412-416

Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts, USA

NY, USA

Matusita, K. (1951) On the theory of statistical decision functions, Ann. Inst. Statist. Math., 3 17-35

Plackett, R.L. (1981) *The Analysis of Categorical Data*, 2nd Edition, Macmillan, New York, NY,USA

Srole, L., Langner, T.S., Michael, S.T., Opler, M.K. and Rennie, T.A.C. (1962) *Mental Health in the Metropolis: The Midtown Manhattan Study*, McGraw-Hill, New York, NY, USA

Stuart, A. (1953) The estimation and comparison of strenghts of association in contingency tables, Biometrika, 40 105-110

Stuart, A. (1955) A test for homogeneity of the marginal distributions în a two-way classification, Biometrika, 42 412-416

Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts, USA