Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
May 2024 5(5), 1482–1488

QUANTIFYING MONOTONE CORRELATION IN VISUAL PATTERN ANALYSIS WITH MATHEMATICAL STATISTICS

Gajraj Singh [1]

[1] Discipline of Statistics, School of Sciences, Indira Gandhi National Open University, Delhi, India

Corresponding Author

Gajraj Singh, gajrajsingh@ignou.ac.in

DOI
10.29121/shodhkosh.v5.i5.2024.6598

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

A two-way frequency table that categorizes all corporate bonds rated by Standard & Poor's and Moody's, or the well-known father-son social mobility data matrix that is frequently cited as an example of a Markov chain in human resources, are two examples to consider. Given that bond ratings and social class have a similar natural ordering, both are instances of contingency tables with rows and columns that represent ordinal categorical variables. The degree of association between such row and column variables is measured and quantified in this study. When it comes to cardinal variables, correlation provides a clear indicator of linkage that doesn't require category scaling. A measure of connection could be calculated using sup correlation, a scheme similar to eigen analysis, if the order relationships involved do not need to be respected. Methods for ordinal categorical data, like Spearman's or Kendall's tau, assign numerical values to the categories. We introduce a few novel methods for assessing the correlation between ordinal variables. The idea of monotone correlation is used to develop four new statistical measures of monotone relationships. The particular circumstance that produced the data determines whether each of these metrics is appropriate. An iterative process is the only way to obtain these metrics of association. These metrics are evaluated and the corresponding monotone scalings are obtained using a nonlinear optimization approach. Keywords: - Monotone correlation, Monotone scale invariance

Keywords: Monotone Correlation, Monotone Scale Invariance

1. INTRODUCTION

The intelligent use of statistics and a deeper understanding of the underlying bivariate probabilistic structures depend on the measurement and comprehension of the basis for the relationship between two random variables, X and Y. This work focuses on the relationship between ordinal random variables, or random variables whose observed values have a natural ordering but do not necessarily have numerical values that are naturally assigned. The values might be derived, for instance, from agree-based questionnaire responses.

It makes sense to require that the resulting numerical measure of association depend only on the orderings and not on the actual numerical values when assessing the association between two ordinal variables using a five-point scale: strongly disagree, disagree, no opinion, agree, and strongly measure. This characteristic is known as monotone scale invariance. The scale is invariant when the values 1... N are assigned to the scale levels in order to apply Pearson's correlation coefficient. A monotone invariant measure of association would not be obtained by computing the Pearson correlation for the five-point example, where 1 would represent strongly disagree and 5 would represent strongly agree. Gebelein [1941] introduced the sup correlation ρ', which was later developed by Sarmanov [1958a and 1958b], Rényi [1959], and Lancaster [1969]. It is defined as follows: $\rho'(X, Y) = \sup \rho(f(X), g(Y))$, where the supremum is taken over all Borel-measurable functions, f and g, such that $0 < \text{var } f(X) < \infty$ and 0. An important dependence concept between two random variables is that of complete dependence, introduced by Lancaster [1963]. A random variable Y is said to be completely dependent on a random variable X if there exists a function g such that

$$\text{prob}[Y = g(X)] = 1 \tag{1}$$

If Y is completely dependent on X and vice versa, then X and Y are said to be mutually completely dependent, in this case X and Y are perfectly predictable from each other. Observe that if X and Y are mutually completely dependent, then ρ (X, Y) = 1.

Kimeldorf and Sampson [1978] provided an example of random variables X and Y which were mutually completely dependent and yet were "almost" stochastically independent. To circumvent this difficulty, Kimeldorf and Sampson defined Y to be monotone increasing (decreasing) dependent on X if (1) holds for a monotone increasing (decreasing) function g. Furthermore, motivated by trying to measure the degree of monotone dependence, they defined the monotone correlation between random variables X and Y by

$$\rho^* (X, Y) = \sup \rho\, (\mathrm{f}(X), \mathrm{g}(Y)), \tag{2}$$

where the supremum is taken over all monotone functions f and g for which 0 < var f (X) < ∞ and 0<var g(Y)< ∞. The monotone correlation is a monotone scale-invariant measure of association and the maximizing functions (assuming they exist) for (2) are the best monotone scalings for cross linear predictability of X and Y. (Monotone scalings are order-preserving assignments of numerical values to ordinal data.) Kimeldorf and Sampson evaluated the monotone correlation in only two special situations: (i) X and Y bivariate normal, in which case $\rho^* = \rho$; and (ii) X and Y independent, in which case ρ^*=0.

In this study of the paper there are two-folds. One is to derive new measures associated with the monotone correlation and to study their applicability. A second is to provide a computational procedure and computer program to evaluate the monotone correlation and these derived measures for the case when X and Y assume a finite number of values. The approach is to find an equivalent nonlinear program and then employ a modification of the optimization algorithm of J.H.May [1979] to compute the maximizing values and the points at which they occur. In section 2 we introduce the concepts of concordancy, discordancy, and isoscaling for measuring the monotone association. The equivalent nonlinear programs are given in section 3. The specific algorithm and the computer program, which we call MONCOR, are described in section 4. A number of interesting applications and examples are considered in section 5 and discussion is given in section 6.

2. CONCORDANCY, DISCORDANCY, AND ISOSCALING

The concept of a monotone correlation can be refined by measuring separately the strength of the relationship between *X* and *Y* in a positive direction and the strength of the relationship in a negative direction, i.e. to measure separately the extent of concordancy and of discordancy between *X* and *Y*. These concepts are related to so-called measures of disagreement and measures of dissociation. If in (2) *f* and *g* are both restricted to be (nonstrictly) increasing (or equivalently both decreasing (nonstrictly)), the resulting measure is called the concordant monotone correlation (CMC). When *f* is restricted to be increasing and *g* decreasing (or equivalently *f* decreasing and *g* increasing), we find it convenient to examine -supρ (f(X), g(Y)), which in turn can be expressed as -sup ρ ($f(X)$, -$g(Y)$), where both *f* and *g* are increasing. This leads naturally to defining the discordant monotone correlation (DMC) by inf ρ ($f(X)$. $g(Y)$), where *f* and *g* are both restricted to be increasing.

The *DMC* and *CMC* have natural interpretations as measures of negative and positive association, respectively, for ordinal random variables. They also can be interpreted as providing bounds for the correlation between any arbitrary monotone scalings; specifically, for arbitrary increasing *f* and *g*.

$$DMC \le \rho\big(f(X).g(Y)\big) \le CMC \tag{3}$$

Suppose it is desired to impose numeric monotone scalings for a pair of new tests; if the CMC and DMC are close, then by (3) it makes little difference which monotone scales are used. Also, if CMC = DMC = 0, then X and Y are independent random variables; however, it is possible for DMC <CMC=0 and X and Y not to be independent. Furthermore, note that if X and Y are increasing monotone dependent then CMC = 1; and if X and Y are decreasing monotone dependent, then DMC=-1.

Sometimes the situation occurs when X and Y should have the same scaling. For example, suppose that X is a psychological test score pre-treatment and Y is the score post-treatment on the same test. This leads to another extension of the monotone correlation, which we refer to as isoscaling. If in (2) we restrict f=g, the resulting measure is called the isoconcordant monotone correlation (ICMC). Analogous to the DMC definition, the isodiscordant monotone correlation (IDMC) is given by inf (f(X), g(Y)), where f=g. Obviously, isoscaling is not appropriate in practice when X and Y have essentially different ranges of values. If X and Y are exchangeable ordinal random variables it might be conjectured, due to all the symmetries involved, that ICMC=CMC (and IDMC= DMC). However, as is shown in section 5, surprisingly this is not the case. The actual functions that maximize the correlations (assuming they exist) are generically called monotone variables; their specific interpretation depends upon which monotone correlation measure is used in their derivation. When measuring the monotone association using one of our monotone measures, we strongly advocate simultaneously examining the values of the corresponding monotone variables. Otherwise, there can be potential misinterpretations. For instance, Hall [1969, example 7] presents an example where the support of X and Y is three monotone disjunct pieces and the CMC=1. In this example, corresponding monotone variables are and , where is the indicator function of the set A. Also, as we note in section 6 below, the monotone variables themselves may be quite useful in constructing scales for ordinal data.

3. FORMULATION

The preceding extensions of the monotone correlation are applicable to all suitable pairs of random variables, continuous or discrete. We now focus on the case where X and Y jointly take on a finite number of values (a_i, b_j), $i=1,2,\ldots,I$, $j=1,2,\ldots,J$ and prob ($X = a_i, Y= b_j$)$=p_{ij}$, Then:

$$CMC = \max\left\{\left[\sum_{i=1}^{I}\sum_{j=1}^{J} f(a_i)\,p_{ij}\,g(b_j) - \left(\sum_{i=1}^{I} f(a_i)\,p_i\right)\left(\sum_{j=1}^{J} g(b_j)\,p_j\right)\right]\right.$$
$$\times\left[\left(\sum_{i=1}^{I} f^2(a_i)\,p_i - \left(\sum_{i=1}^{I} f(a_i)\,p_i\right)^2\right)^{1/2}\right. \tag{4}$$
$$\left.\left.\times\left(\sum_{j=1}^{J} g^2(b_i)\,p_j - \left(\sum_{j=1}^{j} g(b_i)\,p_j\right)^2\right)^{1/2}\right]^{-1}\right\}$$

Subject to f and g being increasing functions for which denominator in (4) is non-zero and where $p_{i\,.} = \sum_{j=1}^{J} p_{ij}$ and $p_{.\,j} = \sum_{i=1}^{I} p_{ij}$ denote the values $f(a_i)$ by xi, i=1,2,.....,I, and by yj ,j=1,2,.....,J, so that (4) can be formulated as:

$$CMC = \max \frac{x' P y - (x' P e)(y' P' e)}{\left(\sum x_i^2 p_i - (x' P e)^2\right)^{1/2}\left(\sum y_j^2 p_j - (y' P' e)^2\right)^{1/2}} \tag{5}$$

subject to

$x_1 \le \ldots\ldots \le x_I$ and $y_1 \le \ldots\ldots \le y_J$; $x \ne c_1 e$ and $y \ne c_2 e$, where $x = (x_1 \ldots, x_I)'$, $y = (y_1, \ldots, y_J)'$, $P=(p_{ij})$ and $e=(1,\ldots\ldots,1)$. Thus, to compute the *CMC* all that is required is the matrix *P* of probabilities. For instance, the values $a_1 \ldots\ldots a_5$ could be the five-point scale strongly disagree,...., strongly agree. The resultant monotone variable x would then provide a numerical scale assigning x_1 to strongly disagree, ……, x_5 to strongly agree. Analogous formulations of (5) can be given for *ICMC*, *DMC*, and *IDMC*. Again, the *ICMC* and *IDMC* are not defined when I≠J. When reporting the monotone variables, we standardize them without loss of generality so that in (5), for example, $x_1 = y_1 = 0$ and $x_I = y_J = 1$. Until this point, the CMC, etc. have been defined as population quantities. For data from finite discrete distributions, the joint probabilities can be estimated from the data viewed in ordinal contingency table form. Then the CMC can be evaluated based upon the estimated probabilities. In this situation the CMC can either be viewed as an estimate of the "true" CMC or be viewed as a measure of monotone association for the ordinal contingency table.

4. OPTIMIZATION APPROACH AND MONCOR DESCRIPTION

The nonlinear programming problem (5) involves the optimization of a nonlinear fractional form subject to linear constraints. Note that if it were not for the monotone constraints, (5) would be an eigenvalue problem. The objective function in (5) is not pseudoconcave. To see this, consider the simple case of evaluating the

$$ICMC = \max\left(x^{'}P\,x-\left(x^{'}P\,e\right)^{2}\right)\Big/\left(\sum x_i^2\,p_i-\left(x^{'}P\,e\right)^{2}\right)$$

for a symmetric probability matrix P. While both numerator and denominator are continuously differentiable on the feasible region, and $\left(\sum x_i^2 p_i-\left(x^{'}P\,e\right)^{2}\right)$ is a positive convex function of x, (x'Px - x'Pe)2 would have to be non-negative and concave for pseudoconcavity (see Avriel [1968]). This latter condition does not hold in general for symmetric P. Hence, in general, the CMC, and ICMC, DMC and IDMC will involve the optimization of a function with local optima. Although much work is presently being done in the area of global optimization (see, for example, Dixon and Szego [1975, 1978]), we follow the standard procedure of using various starting points, computing the optima, and then choosing the best result based upon the different starting points.

Note that since correlation is unique in x and y only up to location and scale change, we could express (5) as

maximize x'Py

Subject to:

$$\sum_i x_i p_i = 0; \qquad \sum_j y_i p_{.j} = 0; \qquad \sum_i x_i^2 p_i = 1; \qquad \sum_i y_j^2 p_{.j} = 1,$$

$$x_1 \le x_2 \le \ldots\ldots.. \le x_I \tag{6}$$

$$\text{and }\; y_1 \le y_2 \le \ldots\ldots.. \le y_J$$

The formulation of (6), because of its nonlinear constraints, is not a desirable formulation since complexity in the objective function is much easier to deal with than complexity in the constraints. The constraints $x \neq c_1 e$ and $y \neq c_2 e$ in (5) are not computationally implementable in continuous variables. However, without loss of generality, we eliminate those constraints by fixing x_1 and y_1 at zero and x_I and y_J at one.

Specifically, the computation of the *CMC* (*DMC*) involves optimizing a nonconcave (nonconvex) function in I+J-4 independent variables subject to monotonicity constraints. (The *ICMC* and *IDMC* involve I-2 independent variables.) Since P is envisioned to be not much larger than 10x10, a modified Newton method was considered desirable because it should converge in a small number of iterations. *QRMNEW* (see [1979]), an optimization method not requiring analytical derivatives, was employed because of its ease of adaptation and computational use. *QRMNEW* is a hybrid local variations-modified Newton method, using orthogonal (*QR*) matrix factorization to derive a representation for the locally feasible region. It has been shown by May that starting from any initial point, *QRMNEW* converges to a point satisfying both first and second-order necessary optimality conditions, so that any solution generated is at least a local optimum. Superlinear and order 2 convergence rates can be established under somewhat stronger conditions. Denote by $\{(x, y)^k\}$ the iterative sequence of points generated by the algorithm. In general, because of the lack of pseudo-concavity (pseudoconvexity) for the *CMC* and *ICMC* (*DMC* and *IDMC*), an iterate $(x, y)^k$ will usually be in a region not locally concave (convex). The algorithm does have a sophisticated method for dealing with the indefinite projected matrix of second derivatives implied by the lack of local concavity (convexity). *MONCOR* is an interactive package designed to analyze probability matrices, *P*, of dimension less than or equal to 20×20. The user may input a single starting point for an optimization run, or allow the program to generate its own multiple starting points. In both cases the constraint set corresponding to the correlation measure requested is generated internally, and *QRMNEW* is used to compute an optimum. Additionally, two different strategies are employed in seeking an optimal solution. Numerical experience indicates that optimum values sometimes lie at monotone extreme points, i.e. points where all the x and y entries are either zero or one. This appears to be especially the case when computing the *DMC* or *IDMC* for a matrix with highly positive *CMC*. and vice versa. In fact, for certain cases the optima for all four monotone correlation measures might be achieved only at such points. Additionally. because nonoptimal monotone extreme points can be local optima (satisfying the Karush-Kuhn-Tucker second-order necessary optimality conditions (see Fiacco and McCormick [1968]). QRMNEW starting from a random point might well be trapped by these local optima. Note that for an I×J matrix, there are only (I-1) (J-1) monotone extreme points to consider for the *CMC* and *DMC* (n - 1) for *ICMC* and *IDMC*, assuming (I = J = n) Hence, in order to avoid at a local optimum when the global optimum is a monotone extreme point, *MONCOR* evaluates the correlation of all monotone extreme points. Moreover, *MONCOR* generates ten random monotone points, with coordinates selected on (0, 1), using the DEC random number generator (see [1958]), and calls *QRMNEW* to compute an optimum starting from each of them. The user may select to see only the final output, or an iteration-by-iteration output of the monotone correlations and monotone variables.

5. APPLICATION

By means of the algorithm and the *MONCOR* program, we now compute the *CMC*, etc. for several insightful examples. Let (*X, Y*) be a discrete bivariate random vector taking values in a 6×6 lattice: $\{a_1,\dots\dots a_6\} \times \{b_1\dots\dots\dots b_6\}$ Furthermore, suppose prob($X = a_i$) =1/6, for all i, and prob ($Y= b_j$)=1/6, for all j; i.e. *X* and *Y* have uniform marginals. If *X* and *Y* are monotone increasing dependent then *P* = (1/6) I, where *P*=prob (X = i, Y = j) , and I is the 6×6 identity matrix are monotone dependent, then P= (1/6), where I* ={ δ (i + j - 7)} and $\delta(x)$ is 1 if *x*= 0 and is 0, otherwise. Now consider a one-parameter family of distributions indexed by θ i.e. for a given θ, prob (*X= i,Y=j*) is the (i, j)th element of P_θ, where

$$P_\theta = \left(\frac{1+\theta}{2}\right)\left(\frac{1}{6}\right)I + \left(\frac{1-\theta}{2}\right)\left(\frac{1}{6}\right)I^* \tag{7}$$

where - 1 < θ < 1. Note that X and Y still have uniform marginal distributions for all θ =1(-1), P_θ corresponds to the most monotone increasing (decreasing) dependent case; and intermediate values of θ describe varying degrees of mixtures of the two dependent extremes.

Now consider (*X, Y*) defined on a 3×3 lattice with,

$$P = \begin{pmatrix} 0 & 1/4 & 0 \\ 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 \end{pmatrix} \tag{8}$$

so that, for example, prob ($X = a_1$ $Y = b_2$) =1/4. Note that P is a symmetric probability matrix, so that and Y are exchangeable random variables. It follows in this case, by direct computation or by use of MONCOR, that the *ICMC* is 0 and the monotone variables for *X* and *Y* are (0, 0.5, 1). However, the *CMC* is 1/3 and the monotone variables for X and Y, respectively, are either (0, 1, 1)' and (0, 0, 1)' or (0, 0, 1)' and (0, 1, 1)'. Thus, (8) provides an example of exchangeable random variables where *ICMC≠ CMC*.

We now consider applying these monotone measures to an actual data example, taken from Bishop, Fienberg and Holland [1975, p. 100], which in turn was adapted from Glass and Hall [1969. p. 183]. These data are given in table 1.

Table 1 British Mobility Data (3500 Father-Son Data Values)

Sr. No.	Father's occupational status	Father's occupational status				
		S1	S2	S3	S4	S5
1	S1	50	45	8	18	8
2	S2	28	174	84	154	55
3	S3	11	78	110	223	96
4	S4	14	150	185	714	447
5	S5	03	42	72	320	411

Note: Status S1 is professional, and high administrative; status S2 is managerial, executive and higher grade supervisory; status S3 a lower grade supervisory, status S4 is skilled manual, and status S5 is semi-skilled and unskilled manual

Table 2 ICMC, IDMC and Monotone Variables for British Mobility Data

Measure	Value of measure	Monotone variable values				
ICMC	0.496	0	0.627	0.842	0.923	1.0
IDMC	0.242	0	0	0	0	1.0

Because the same categories are used to measure father's and son's occupational status, it is appropriate to use isoscaling. The ICMC, IDMC and the associated monotone variables were computed by the MONCOR program based on the empirical probability matrix specified by table 1. The values of the ICMC and IDMC as well as the monotone variables are presented in table 2.

The analogous version of (3) for isoscaling, namely *IDMC* ≤ *ρ* [*f(X), f(Y)*] ≤*ICMC*, shows that regardless of the assignment of numerical values to the five ordinal categories, the resultant correlation is between 0.242 and 0.496.

6. CONCLUSION

One important use of monotone variable theory is the ability to develop meaningful scales for ordinal variables. For example, suppose the five-point scale response to some question is elicited pre- and post- some experimental intervention. Through the use of the *ICMC*, we can provide a numerical scale for this five-point response, this numerical scale has the property that among all possible such ordinal scalings, the post-response for this scaling is most linearly predictable from the pre-response. In table 2 the row corresponding to *ICMC* provides this scaling for the occupational status variable based on the British mobility data. Specifically, the numerical values for S1, S2, S3, S4, and S5 are 0, 0.627, 0.842, 0.923, and 1.0, respectively.

Often, the number of distinct values for the numerically scaled variables is substantially less than the number of values for the original ordinal variables This reduction occurs when the optimizing *f, g* in (2) are not one-to-one functions. To illustrate this phenomenon, we consider the following example a 10x10 matrix is generated where each entry is a randomly generated number on (0, 1), each generated independently of the other entries. In order to generate a "slightly" positive dependent distribution, the constant 2 was added to each diagonal term and the entire matrix scaled so as to add to one. The resultant matrix is given in table 3.

Table 3

Sr. No.	X	Y									
		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
1	a_1	0.0331	0.0111	0.0092	0.0049	0.0016	0.0028	0.0009	0.0108	0.0096	0.0007
2	a_2	0.0101	0.0361	0.0057	0.0081	0.0133	0.0062	0.0121	0.0066	0.0003	0.0022
3	a_3	0.0102	0.0059	0.0347	0.0027	0.0055	0.0020	0.0124	0.0046	0.0069	0.0056
4	a_4	0.0144	0.0018	0.0065	0.0342	0.0006	0.0071	0.0055	0.0066	0.0084	0.0113
5	a_5	0.0006	0.0016	0.0087	0.0132	0.0435	0.0061	0.0100	0.0046	0.0044	0.0053
6	a_6	0.0022	0.0035	0.0151	0.0015	0.0056	0.0427	0.0062	0.0035	0.0089	0.0125
7	a_7	0.0002	0.0084	0.0026	0.0020	0.0005	0.0086	0.0387	0.0007	0.0034	0.0111
8	a_8	0.0084	0.0100	0.0079	0.0036	0.0100	0.0128	0.0044	0.0303	0.0121	0.0063
9	a_9	0.0028	0.0079	0.0141	0.0008	0.0133	0.0077	0.0064	0.0139	0.0402	0.0068
10	a_{10}	0.0009	0.0149	0.0042	0.0108	0.0022	0.0144	0.0130	0.0151	0.0146	0.0438

The CMC for the matrix in table 3 is 0.443, and the monotone variables for a1……,a10 and b1,…..,b10 are, respectively (0.000, 0.461, 0.461, 0.461, 0.872, 0.872, 0.872, 0.872, 0.873, 1.000)' and (0.000, 0.537, 0.541, 0.541, 0.842, 0.842. 0.842, 0.842, 0.842, 1.000). Note that while the original variables each had ten separate values, there are only five distinct monotonely scaled values for X and five for Y. While this scale reduction phenomenon is based upon empirical observation, it is clear that it has great potential value in deriving simplified scales for large data sets.

A description of the MONCOR program, and examples of its input and output procedures, is given in Kimeldorf, May and Sampson [10]. The FORTRAN program itself, and a user's manual, are available for distribution. For specific details contact Professor Jerrold May, Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

A. Fiacco and G. MeCormick, Nonlinear Programming Sequential Unconstrained Minimiza tion Techniques (Wiley, New York, 1968).

A. Renyi, "Measures of Dependence". Acta Math Acad. Sci. Hungar. 10 (1959) 441-451

D.V. Glass and 1.R. Hall, "Social Mobility in Britain: A Study of Intergeneration Changes in Status, in:D.V. Glass, ed., Social Mobility in Britain (Routledge and Kegan Paul Lid. London, 1954)

Fujita A, Takahashi DY, Balardin JB, Sato JR. Correlation between graphs with an application to brain

networks analysis. 2015. arXiv:1512.06830 [q-bio, stat]. Accessed 12 Jan 2020.
G. Kimeldorf and A.R. Sampson, "Monotone Dependence", Annals of Statistics 6 (1978) 895-903.
G. Kimeldorf, J.H. May and A.R. Sampson, "MONCOR A Program to Compute Concordant and Other Monotone Correlations, in: W.F. Eddy, ed. Proceedings of Computer Science and Statistics: 13th Symposium on the Interface (Springer-Verlag, New York, 1981)
H. Gebelein, "Das Statistiche Problem der Korrelation als Variations and Eigenwert problem und sein Zusammenhang mit der Ausgleichungsrechnung", Z. Angew Math Mech 21 (1941) 364-379
H.O. Lancaster, "Correlation and Complete Dependence of Random Variables". Ann. Mask Statist. 34 (1963) 1315-1321.
H.O. Lancaster, The Chi-Squared Distribution (Wiley, New York, 1969)
Iwasaki Y, Kusne AG, Takeuchi I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. NPJ Comput Mater. 2017; 3:4.
Jay JJ, Eblen JD, Zhang Y, Benson M, Perkins AD, Saxton AM, et al. A systematic comparison of genome-scale clustering algorithms. BMC Bioinform. 2012;13(Suppl 10): S7.
J.H. May, "Solving Nonlinear Programs Without Using Analytic Derivatives, Operations Research 27 (1979) 457-484.
L.C.W. Dixon and G.P. Szego, Towards Global Optimisation (North-Holland Publishing Company, New York, 1975)
LC.W. Dixon and GP. Szego, Towards Global Optimization, 2 (North-Holland Publishing Company, New York, 1978)
Lin W-T, Wu Y-C, Cheng A, Chao S-J, Hsu H-M. Engineering properties and correlation analysis of fiber cementitious materials. Materials. 2014;7: 7423–35.
M. Avriel, Nonlinear Programming: Analysis and Methods (Prentice-Hall. Inc., Englewood Cliffs, NJ, 1976).
Neto AM, Victorino AC, Fantoni I, Zampieri DE, Ferreira JV, Lima DA. Image processing using Pearson's correlation coefficient: applications on autonomous robotics. In: 2013 13th international conference on autonomous robot systems. Lisbon, Portugal: IEEE; 2013. p. 1–6. https://doi.org/10.1109/Robotica.2013.6623521.
OV Sarmanov, "The Maximal Correlation Coefficient (Symmetric Case)", Doki Akal Nana SSSR 120 (1958) 715-718 (English translation in Sel. Transl Math Statist. Probability 4271-2751
OV. Sarmanov, "The Maximal Correlation Coefficient (Non-Symmetric Case), Doki Akal Nak SSSR 121 (1955) 52-55 (English translation in Sel Transi Mach Statist Probability 4207-2101
Preacher KJ, Zhang Z, Zyphur MJ. Multilevel structural equation models for assessing moderation within and across levels of analysis. Psychol Methods. 2016;21: 189–205.
WJ. Hall, "On Characterizing Dependence in Joint Distributions", in: Bose et al, eds., Ernyt in Probability and Statistics (University of North Carolina Press, Chapel Hill, 1969)
WH. Payne J.R. Rabung and T.P. Bogyo, "Coding the Lehmer Pseudo-Random Number Generator, Communications of the ACM 12 (1969) 85-86.
Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, Discrete Multivariate Analysis: Theory and Practice (MIT Press, Cambridge, 1975).