IIMB-WP N0. 716/2025

WORKING PAPER NO: 716

A New Measure of Non-Parametric Correlation for Variables in the Likert Scale

Shubhabrata Das
Professor of Decision Sciences
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 560 076
shubho@iimb.ac.in

Year of Publication – March 2025

IIMB Working Paper 716, , 1–23
doi:

A New Measure of Non-Parametric Correlation for Variables in the Likert Scale

Shubhabrata Das

Decision Sciences, IIM Bangalore, Bannerghatta Road, Bangalore , 560076, Karnataka, India

Abstract

We propose a new measure of nonparametric correlation that is especially suited for measuring association between variables measured in the Likert scale where data is ordinal and tied observations are extremely common. The proposed general structure of the measure is based on graded level of concordance and discordance between the pairs of metrics. The general form of the measure has all the desirable properties except the measure is not necessarily zero for independent variables. This limitation is acceptable given only ordinal nature of the metrics. Three versions of the measure are studied. The first is based on simple equi-distant weights. In the other two variations, the measure attains the zero value under certain conditions of independence. In developing these two versions, linear and non-linear optimization techniques are adopted and their equivalence is demonstrated in finding the suitable weights.

Key words: Association, Constrained optimization, Fully-concordant, Fully-discordant, Greatly-concordant, Greatly-discordant, Highly-concordant, Highly-discordant, Independent, Mildly-concordant, Mildly-discordant, Nonparametric, Ties.

1. Introduction

1.1. Background

Correlation analysis is a fundamental statistical tool that is used to quantify the relationship between two variables. The most widely used measure, Pearson's [Pearson (1895)] correlation coefficient, has been a cornerstone in statistics since his seminal work in 1895. Pearson's correlation coefficient is a parametric measure that primarily assesses the strength of linear relationship between two continuous variables. However, it is particularly suited when variables are normally distributed. It is less applicable to ordinal or non-normally distributed data. When applied to Likert scale data, Pearson's correlation may yield misleading results due to its reliance on arithmetic means and variances that do not appropriately represent ordinal relationships.

1.2. Likert Scale

A Likert scale [Likert (1932)] is a widely used psychometric scale for measuring attitudes, opinions, or perceptions in survey research (e.g., measuring customer satisfaction, employee engagement, or public opinion), Psychology and social sciences (e.g. assessing personality traits, attitudes, or behavioural tendencies), in business and marketing (e.g. understanding consumer preferences or service feedback).

Likert-scale responses are statement-based; respondents evaluate a statement rather than answering a direct question. It is often based on statements where respondents indicate their level of agreement or disagreement using new graded response format. Likert scale commonly uses a 5-point or 7-point scale, though variations exist, as discussed later. In a 7-point agreement with intensity scale format, the respondent has to choose among the options: Strongly Disagree / Disagree / Slightly Disagree / Neutral / Slightly Agree / Agree /Strongly Agree, while in a 5-Point Scale, the options typically are Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree.

Likert scale can use symmetric or asymmetric scaling, with options ranging from one extreme attitude to another; often – but not always with a neutral midpoint. Indeed, this is an important variation having *odd vs. even* number of option points k. If k is odd, the scale includes a neutral option, while if k is even, it forces a choice from the respondent by removing neutrality. The other variations of Likert Scales are in terms of being *unipolar vs. bipolar.* Bipolar scale is the more common and standard one. In unipolar, intensity is measured mostly in one direction (e.g., Not at all $\rightarrow$ Extremely).

Beyond the traditional agreement-based Likert scale (Strongly Disagree $\rightarrow$ Strongly Agree), several other variations exist to measure different types of responses. Here are some commonly used alternatives:

- Frequency Scale (Measures how often something occurs) [options e.g./ Never/ Rarely / Sometimes / Often /Always]
- Importance Scale (Measures significance or priority) [options e.g./ Not Important at All / Slightly Important / Moderately Important / Very Important / Extremely Important]
- Satisfaction Scale (Measures contentment or approval) options e.g./ Very Dissatisfied / Dissatisfied / Neutral / Satisfied / Very Satisfied]
- Likelihood/Probability Scale (Measures the chance of something happening) [options e.g./ Highly Unlikely / Unlikely / Neutral / Likely / Highly Likely]
- Quality Scale (Measures the perceived quality of something) [options e.g./ Very Poor / Poor / Fair / Good / Excellent]
- Difficulty Scale (Measures the ease or difficulty of a task) [options e.g./ Very Easy / Easy / Neutral / Difficult / Very Difficult]
- Confidence Scale (Measures certainty or confidence in a statement) [options e.g./ Not at All Confident / Slightly Confident / Moderately Confident / Very Confident / Completely Confident]
- Effectiveness Scale (Measures how well something works) [options e.g./ Not Effective at All / Slightly Effective / Moderately Effective / Very Effective / Extremely Effective]

- Relevance Scale (Measures how applicable something is) [options e.g./ Not Relevant at All / Slightly Relevant / Moderately Relevant / Very Relevant / Extremely Relevant]

As is evident from the examples of different formats of the Likert scale, it provides *ordinal* measurement, based on a ranking order. Although often in the industry, analysts convert the responses into numbers 1,2,$\ldots, k$, this is highly problematic as the responses do not necessarily indicate equal intervals between consecutive different levels. This needs to be accounted for while measuring association in a suitable way.

1.3. Existing Nonparametric Correlations

Several nonparametric alternatives have been developed, starting from Spearman's rank correlation [Spearman (1904)] which measures the strength and direction of a monotonic relationship between two variables. It is computed based on rank transformations of the data, making it less sensitive to non-normality and outliers. However, it does not account for tied ranks very well, a common feature of Likert scale data. Kendall's Tau [Kendall (1938)] is another popular nonparametric measure of correlation that is based on the concordance and discordance between paired observations. It is robust for small sample sizes, but again is not particularly suited to deal with tied observations. Goodman and

Kruskal's gamma [Goodman and Kruskal (1954)] is specifically designed for ordinal data and is particularly useful in the presence of tied ranks. It evaluates the difference between concordant and discordant pairs relative to the total number of pairs. While this measure is more appropriate for Likert scale data, it still has limitations on that front. All the above non-parametric correlations suffer from a common limitation in that they do not have natural population counter-parts. Nešlehová [Nešlehová (2007)] made a significant contribution by extending nonparametric correlation measures to population-level interpretations. This work addressed the issue of defining nonparametric correlation measures in the context of broader statistical populations, bridging the gap between sample-based metrics and theoretical distributions. Chatterjee [Chatterjee (2021)] introduced an innovative correlation measure designed to capture nonlinear relationships. The population analogue of this measure, DSS correlation [Dette et al. (2013)], refines its applicability in theoretical contexts. However, Chatterjee's correlation, like Spearman's and Kendall's, does not handle tied observations well, making it suboptimal for Likert scale data, where ties are prevalent.

Given the limitations of existing correlation measures, there is a need for a new nonparametric measure that accounts for tied observations effectively, has a natural population analogue, and adheres to the value validity principle as proposed by Kvålseth [Kvålseth (2017)]. A correlation measure that satisfies these criteria would provide a more robust framework for analyzing Likert scale data, ensuring more reliable interpretations in social sciences, psychology, and other fields relying on ordinal data. While existing nonparametric correlation measures offer valuable tools for analyzing ordinal data, they each have limitations, particularly in handling tied observations.

1.4. Notation

Let X and Y be variables measured in k-point Likert scale. That is, X and Y have values in the range $\{1, 2, \ldots, k\}$. As mentioned earlier, typically, k would be an odd-integer: 3, or 5, or 7, may be at most 9.

Let $f_{i,j}$ or π_{ij} denote the relative frequency (in sample data context) or the probability (in population context) of $(X = i, Y = j)$.

$$f_{i\cdot} = \sum_j f_{i,j}, \qquad f_{\cdot j} = \sum_i f_{i,j}.$$

Thus, we use the following notations:

$$\pi_{ij} = P[(X = i) \cap (Y = j)], \quad i, j = 1, \ldots, k.$$

$$f_{ij} = \frac{\#[(X = i) \cap (Y = j)]}{n}, \quad i, j = 1, \ldots, k,$$

where $n = \sum_{i=1}^k \sum_{j=1}^k f_{ij}$.

Let $\pi_{i\cdot}$ and $\pi_{\cdot j}$ denote the marginals, i.e.

$$\pi_{i\cdot} = \sum_j \pi_{i,j}, \qquad \pi_{\cdot j} = \sum_i \pi_{i,j}.$$

2. The proposed measure of correlation

2.1. The general form of the measure and its properties

For a suitable weight matrix $W = ((w_{ij}))$, the general form of the proposed correlation between X and Y is given by:

$$\Psi = \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_{ij} w_{ij}. \tag{1}$$

The sample correlation is given by:

$$\hat{\Psi} = \sum_{i=1}^{k} \sum_{j=1}^{k} f_{ij} w_{ij}. \tag{2}$$

In (1) and (2), W is a symmetric matrix having all the diagonal elements as 1 and two non-diagonal extreme corner elements being -1; that is

$$w_{ii} = 1,\ \forall i = 1, \ldots, k; \quad w_{ij} = w_{ji}\ \forall i, j = 1, \ldots, k; \quad w_{1k} = w_{k1} = -1. \tag{3}$$

From the structure of (1) and (2), it is clear that the former is a population analogue of the latter. The consistency of f_{ij} to π_{ij} also establishes the consistency property of the latter. The other standard properties of a desirable estimator follows along the traditional path based on standard large sample inference, which we we do not divulge in this work. We focus our discussion on Ψ in this work, with analogous comments holding true for its estimate $\hat{\Psi}$.

Let us now discuss other properties of this proposed measure of correlation.

1. Symmetric measure of association, i.e. $\Psi(X, Y) = \Psi(Y, X)$.
2. It is bounded, scaled measure of association which captures the degree of association. In particular, $-1 \leq \Psi \leq 1$, with the sign of Ψ indicating direction of association/relation.
3. $\Psi = 1$ iff $\Pi_C = 1$ where $\Pi_C = \sum_{i=1}^{k} \pi_{ii}$ represents the probability of perfect concordance.
4. $\Psi = -1$ iff $\Pi_D = 1$ where $\Pi_D = \pi_{1k} + \pi_{k1}$ represents the probability of perfect discordance.
5. Intuitive connection and consistency between $\hat{\Psi}$ and Ψ
6. Adheres to the *Value-Validity principle*, as:

$$\Psi(\Pi^a_{2\times 2}) = 2 \times \frac{1+a}{4} - 2 \times \frac{1-a}{4} = a.$$

Depending on addition specification of the matrix W, we propose three versions of the measure of correlation, namely – Ψ_a, Ψ_b, Ψ_c. When, we wish to highlight the role of Likert scale k, we add the superfix and denote the measures by $\Psi_a^k, \Psi_b^k, \Psi_c^k$.

2.2. Ψ_a: Measure with the with first choice of weights

Under this version, we propose that weights reduce **linearly** from the diagonal entries in each row of W. That is:

$$w_{ij} = 1 - \frac{2|i-j|}{k-1}$$

Note that, in that case, for k odd (which is often/typical with 3 or 5 or 7 point Likert scale), the middle row (column) of W has two zero elements; the other rows (columns) has one zero element.

Thus, for $k = 3$, the other elements of W (non-specified by (3)) are zero's and hence

$$\Psi_a^3 = \pi_C - \pi_D \quad \text{where}$$

$$\pi_C = \pi_{11} + \pi_{22} + \pi_{33}; \qquad \pi_D = \pi_{13} + \pi_{31}.$$

For variables in 5-point Likert scale, we can express the measure as:

$$\Psi_a^5 = \Pi_{FC} + 0.5\Pi_{GC} - 0.5\Pi_{GD} - \Pi_{FD}$$

where

$$\Pi_{FC} = \sum_{i=1}^{5} \pi_{ii}$$

denotes the fully-concordant probability,

$$\Pi_{GC} = \Pi^5_{GC} = \sum_{\substack{i,j=1 \\ |i-j|=1}}^{5} \pi_{ij}$$

denotes the greatly-concordant probability,

$$\Pi_{GD} = \Pi^5_{GD} = \pi_{14} + \pi_{41} + \pi_{25} + \pi_{52}$$

denotes the greatly-discordant probability. The superscript in greatly-concordant and greatly-discordant probabilities Π^5_{GC} and Π^5_{GD} are omitted, when the scale is unambiguous from the context.

$$\Pi_{FD} = \pi_{15} + \pi_{51}$$

denotes the fully-discordant probability. Note that

The neutral probability

$$\Pi_N = 1 - \Pi_{FC} - \Pi_{GC} - \Pi_{GD} - \Pi_{FD}$$

does not contribute to the measure as it has associated weight equal. Similarly, the measure is defined for higher k.

2.3. Ψ_b: Measure with the second choice of weights

Since it is desirable that the measure of association is zero when the variables are independent, we explore enforcing this as a condition on the weights.

To begin with for $k = 3$, let us explore this with $\delta = W_{12}$. Let

$$\xi^+ = \pi_{1.}\pi_{.1} + \pi_{2.}\pi_{.2} + \pi_{3.}\pi_{.3}$$

$$\xi^- = \pi_{1.}\pi_{.3} + \pi_{3.}\pi_{.1}$$

$$\xi^0 = \pi_{1.}\pi_{.2} + \pi_{2.}\pi_{.1} + \pi_{2.}\pi_{.3} + \pi_{3.}\pi_{.2} = 1 - \xi^+ - \xi^-$$

If X and Y are independent, $\pi_C = \xi^+$ and $\pi_D = \xi^-$. Hence, $\Psi_b^3 = \pi_C - \pi_D + \delta(1 - \pi_C - \pi_D)$ is equal to zero under independence, provided:

$$\delta = \frac{\xi^- - \xi^+}{1 - \xi^- - \xi^+}. \tag{4}$$

We observe that (**??**) is meaningful only if $\xi^+ + \xi^- < 1$. Further, to ensure that $-1 < \delta < 1$, we must have:

$$\max(\xi^+, \xi^-) < 0.5, \tag{5}$$

which automatically satisfies $\xi^+ + \xi^- < 1$.

For higher values of k, conditions of feasibility is explored which shows that Ψ_b does not exist in full generality and even when it does, the weights depends on marginal probability distributions. Thus, we favour the version ψ_C as described in the next subsection.

2.4. Ψ_c: Measure with the third choice of weights

Under this approach, we choose weight matrix to ensure $\Psi = 0$ under uniformly distributed independent pairs. Thus, the weight matrix does not depend on specific distributions.

For $k = 3$, under this setup, note that $\xi^+ = \frac{3}{9}$, $\xi^- = \frac{2}{9}$; leading to $\delta = -\frac{1}{4}$. Thus, the weight Matrix W for $k = 3$

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	1	-0.25	-1
$X = 2$	-0.25	1	-0.25
$X = 3$	-1	-0.25	1

$$\Psi_c^3 = 1.25\Pi_C - 0.75\Pi_D - 0.25. \tag{6}$$

Ψ_c in 5-point Likert scale:

Let us now extend for larger k, starting with for $k = 5$. In this case, the weight matrix is of the structure:

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$
$X=1$	1	δ_1	δ_2	δ_3	-1
$X=2$	δ_1	1	δ_1	δ_2	δ_3
$X=3$	δ_2	δ_1	1	δ_1	δ_2
$X=4$	δ_3	δ_2	δ_1	1	δ_1
$X=5$	-1	δ_3	δ_2	δ_1	1

Thus,

$$\Psi_c^5 = \Pi_{FC} + \delta_1\Pi_{GC} + \delta_2\Pi_N + \delta_3\Pi_{GD} - \Pi_{FD},$$

The weights $\delta_1, \delta_2, \delta_3$ are to be decided such that

$$-1 < \delta_3 < \delta_2 < \delta_1 < 1;$$

$$\Psi_c = 0 \text{ when } \pi_{ij} = \frac{1}{25}\forall i, j$$

$$\Leftrightarrow \frac{5-2}{25} + \frac{8\delta_1}{25} + \frac{6\delta_2}{25} + \frac{4\delta_3}{25} = 0$$

We need to put additional conditions, to ensure (some of the) δ_i's being not too close to each other and -1 and 1. E.g. a solution with $\delta_1 = \delta_2 = \delta_3 = -\frac{1}{6}$ would not be acceptable.

Let us now discuss finding these weights for Ψ_c in 5-point Likert scale under two framework.

Approach 1: Non-linear Optimization We propose to maximize the segregation by considering this maxmin formation:

$$\text{Maximize } \min(1-\delta_1, \delta_1-\delta_2, \delta_2-\delta_3, \delta_3+1) \quad \text{subject to}$$

$$8\delta_1 + 6\delta_2 + 4\delta_3 = -3$$

$$-1 < \delta_3 < \delta_2 < \delta_1 < 1$$

Value of the objective function ≈ 0.375. Optimal weights are :

$$\delta_1 \approx 0.125, \quad \delta_2 \approx -0.25, \quad \delta_3 \approx -0.625.$$

Approach 2: Linear Optimization Note that if the δ_i's are spaced at equi-distant points (and from boundaries -1 and 1), the equal spacing would be equal to $\eta = \frac{2}{k-1}$. For example, in the 5-point Likert scale $\eta = 0.5$. We set the δ_i's for a certain fraction away from the nearest ones. Thus we consider additional constraints:

$$\delta_1 \leq 1-\theta\eta; \quad \delta_2 \leq \delta_1-\theta\eta; \quad \delta_3 \leq \delta_2-\theta\eta; \quad \delta_3 \geq -1+\theta\eta,$$

where θ may be maximized over the space (0,1).

For $k = 5$, $(\eta = 0.5)$, the solution of this LP (linear programming formulation:

$$Maximize \quad \theta$$

subject to:

$$8\delta_1 + 6\delta_2 + 4\delta_3 = -3,$$

$$\delta_1 + 0.5\theta \leq 1,$$

$$\delta_1 - \delta_2 - 0.5\theta \geq 0,$$

$$\delta_2 - \delta_3 - 0.5\theta \geq 0$$

$$\delta_3 - 0.5\theta \geq -1,$$

$$0 < \theta < 1$$

leads to an optimal value of $\theta = 0.75$ and values of

$$\delta_1 = 0.125, \quad \delta_2 = -0.25, \quad \delta_3 = -0.625.$$

Thus, either approach leads to the following measure when the variable are in 5-point Likert scale:

$$\Psi_c^5 = \Pi_{FC} + 0.125\Pi_{GC} - 0.25\Pi_N - 0.625\Pi_{GD} - \Pi_{FD}. \tag{7}$$

Ψ_a and Ψ_c in 7-point Likert scale:

Let us discuss the two formations for $k = 7$. In this case, the weight matrix is of the structure:

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	$Y=4$	$Y=5$
$X=1$	1	δ_1	δ_2	δ_3	δ_4	δ_5	-1
$X=2$	δ_1	1	δ_1	δ_2	δ_3	δ_4	δ_5
$X=3$	δ_2	δ_1	1	δ_1	δ_2	δ_3	δ_4
$X=4$	δ_3	δ_2	δ_1	1	δ_1	δ_2	δ_3
$X=5$	δ_4	δ_3	δ_2	δ_1	1	δ_1	δ_2
$X=6$	δ_5	δ_4	δ_3	δ_2	δ_1	1	δ_1
$X=7$	-1	δ_5	δ_4	δ_3	δ_2	δ_1	1

In the 7-point Likert scale, we use the terms highly-concordant (HC), when $X-Y=1$, mildly-concordant (MC) when $X-Y=2$, neutral when $X-Y=3$. Similarly we consider the pairs to be highly-discordant (HD), when $X-Y=5$, and mildly-discordant (MC) when $X-Y=4$. With this, for variables in 57-point Likert scale, we can express the general form of the measure as:

$$\Psi^7 = \Pi_{FC} + \delta_1\Pi_{HC} + \delta_2\Pi_{MC} + \delta_3\Pi_N + \delta_4\Pi_{MD} + \delta_5\Pi_{HD} - \Pi_{FD},$$

where

$$\Pi_{FC} = \sum_{i=1}^{7} \pi_{ii}$$

denotes the fully-concordant probability, while

$$\Pi_{FD} = \pi_{17} + \pi_{71}$$

denotes the fully-discordant probability.

$$\Pi_{HC} = \Pi_{HC}^{7} = \sum_{\substack{i,j=1 \\ |i-j|=1}}^{7} \pi_{ij}$$

denotes the highly-concordant probability, while

$$\Pi_{HD} = \Pi_{HD}^{7} = \pi_{16} + \pi_{61} + \pi_{27} + \pi_{72}$$

denotes the highly-discordant probability. Similarly,

$$\Pi_{MC} = \Pi_{MC}^{7} = \sum_{\substack{i,j=1 \\ |i-j|=2}}^{7} \pi_{ij}$$

denotes the mildly-concordant probability, while

$$\Pi_{MD} = \Pi_{MD}^{7} = \pi_{15} + \pi_{51} + \pi_{26} + \pi_{62} + \pi_{37} + \pi_{73}$$

denotes the mildly-discordant probability. The neutral probability is:

$$\Pi_N = \Pi_N^7 = \sum_{\substack{i,j=1 \\ |i-j|=3}}^{7} \pi_{ij} = 1 - \Pi_{FC} - \Pi_{HC} - \Pi_{MC} - \Pi_{MD} - \Pi_{HD} - \Pi_{FD}.$$

As before, the superscript in the notations are omitted, when the scale is unambiguous from the context.

The first version measure of the measure with equi-spaced δ_i's turn out to be:

$$\Psi_a^7 = (\Pi_{FC} - \Pi_{FD}) + \frac{2}{3}(\Pi_{HC} - \Pi_{HD}) + \frac{1}{3}(\Pi_{MC} - \Pi_{MD}).$$

As we can extrapolate from the case with lower k, the second version of the measure Ψ_b exists only under very restrictive conditions.

The weights $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$ are to be decided such that

$$-1 < \delta_5 < \delta_4 < \delta_3 < \delta_2 < \delta_1 < 1;$$

$$\Psi_c = 0 \text{ when } \pi_{ij} = \frac{1}{49} \forall i, j = 1, \ldots 7;$$

$$\Leftrightarrow \frac{7-2}{49} + \frac{12\delta_1}{49} + \frac{10\delta_2}{49} + \frac{8\delta_3}{49} + \frac{6\delta_4}{49} + \frac{4\delta_5}{49} = 0.$$

For $k = 7$, ($\eta = \frac{1}{3}$), the solution of this LP (linear programming formulation:

$$Maximize \quad \theta$$

subject to:

$$12\delta_1 + 10\delta_2 + 8\delta_3 + 6\delta_4 + 4\delta_5 = -5,$$

$$\delta_1 + \frac{1}{3}\theta \leq 1,$$

$$\delta_1 - \delta_2 - \frac{1}{3}\theta \geq 0,$$

$$\delta_2 - \delta_3 - \frac{1}{3}\theta \geq 0,$$

$$\delta_3 - \delta_4 - \frac{1}{3}\theta \geq 0,$$

$$\delta_4 - \delta_5 + \frac{1}{3}\theta \geq 0,$$

$$\delta_5 - \frac{1}{3} - \theta \geq -1,$$

$$0 < \theta < 1$$

leads to the same optimal value of $\theta = 0.75$ as in $k = 5$ and values of

$$\delta_1 = 0.25, \quad \delta_2 = 0, \quad \delta_3 = -0.25, \quad \delta_4 = -0.5, \quad \delta_5 = -0.75.$$

Thus, the LP approach leads to the following measure when the variables are in 7-point Likert scale:

$$\Psi_c^7 = \Pi_{FC} + 0.25\Pi_{HC} - 0.25\Pi_N - 0.5\Pi_{MD} - 0.75\Pi_{HD} - \Pi_{FD}. \quad (8)$$

3. Concluding comments

We have proposed new measures of association between variables measured in Likert-scale. The measure has all desirable properties of correlation or associatio, including having natural and attainable boundary values, sign indicating nature of association, following value-validity principle and having a natural sample analogue estimate whose large sample properties can be conveniently studied. The only shortcoming is that it is not necessarily equal to zero when the variables are independent. Given that the variables are only ordinal in nature, this must be quite acceptable. The first version of the measure, Ψ_a, is simpler, but more limited on that front. The third version of the measure, Ψ_c provides an improvement on that front in the sense that it is equal to zero for independent and uniformly distributed variables in the Likert scale and yet the weights for different levels of concordance/discordance are optimally accounted for.

At first look, it may appear bothering that Ψ_c puts unequal weights – weights which are not symmetric from both sides. At closer reflection, we can see that this is justified since getting (completely) discordant response is much more unlikely than getting (completely) concordant observations.

Several interesting conjectures can be arrived at. It appears that the linear and nonlinear optimization framework alternatives would provide equivalent solution in wider contexts. It also appears that at least for odd values of k, possibly for all, the optimal θ=0.75 in linear optimization framework to derive Ψ_c.

The work is extendable when the scale k for X and Y are different. We also plan to apply the methods extensively for real and simulated datesets. Subsequently, this may be used for extensions of factor analysis like methods for Likert scale variables following [Jöreskog and Moustaki (2001)] and [Joreskog and Moustaki (2006)].

References

S. Chatterjee. A new coefficient of correlation. *J. of Amer. Stat. Assoc.,* **116(536)***, 2009 – 2022.*, 2021.

H. Dette, K. F. Siburg, and P. A. Stoimenov. A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.,* **40 (1)***, 21 – 41.*, 2013.

L. Goodman and W. Kruskal. Measures of association for cross classifications. *J. of Amer. Stat. Assoc.,* **49***,764.*, 1954.

K. Joreskog and I. Moustaki. Factor analysis of ordinal variables with full information maximum likelihood. 01 2006.

K. G. Jöreskog and I. Moustaki. Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, 36(3):347–387, 2001. doi: 10.1207/S15327906347-387.

M. G. Kendall. A new measure of rank correlation. *Biometrika,* **30(1/2)***, 81 — 93*, 1938.

T. O. Kvålseth. An alternative measure of ordinal association as a value-validity correction of the goodman–kruskal gamma. *Communications in Statistics -Theory and Methods,* **46:21***, 10582 – 10593.*, 2017.

R. Likert. A technique for measurement of attitudes. *Archives of Psychology, 140, 5-55.*, 1932.

J. Nešlehová. On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, 98:544–567, 03 2007. doi: 10.1016/j.jmva.2005.11.007.

K. Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London,* **58***, 240 – 242.*, 1895.

C. Spearman. The proof and measurement of association between two things. *Amer. J. Psychol.,* **15(1)***, 72 -– 101.*, 1904.