Statistics and Probability Letters 178 (2021) 109183

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Measure of asymmetric association for ordinal contingency tables via the bilinear extension copula

Zheng Wei [a], Daeyoung Kim [b,*]

[a] Department of Mathematics and Statistics, University of Maine, Orono, ME, 04469-5752, USA
[b] Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA, 01003-9305, USA

ARTICLE INFO

Article history:
Received 9 May 2020
Received in revised form 23 May 2021
Accepted 7 June 2021
Available online 23 June 2021

Keywords:
Asymmetric association
Bilinear extension copula
Ordinal variable

ABSTRACT

This paper proposes a new correlation ratio based on the bilinear extension copula regression to describe asymmetric association for ordinal contingency tables. Theoretical and finite sample properties of the proposed measure and its estimator are studied.

1. Introduction

For two variables X and Y that are asymmetrically dependent, the association of Y on X is not the same as that of X on Y. The analysis of asymmetric dependence is important in many areas such as gene regulatory network reconstruction (Wang and Huang, 2014) and risk assessments in stock market (Okimoto, 2008).

Several asymmetric dependence measures were proposed in the literatures (see Kruskal and Goodman, 1954; Dabrowska, 1981; Sungur, 2005; Trutschnig, 2011; Dette et al., 2013; Shan et al., 2015; Wei and Kim, 2017; Chatterjee, 2020; Junker et al., 2021, and references therein). Among these measures, only a few of them are mainly developed for discrete random variables: Goodman–Kruskal tau τ_{GK} (Kruskal and Goodman, 1954), MCD (mutually completely dependent) measure μ_t (Shan et al., 2015) and subcopula correlation ratio ρ_S^2 (Wei and Kim, 2017).

It is well-known that τ_{GK} is sensitive to the number of categories in the dependent variable and thus its magnitude cannot be interpreted as the degree of asymmetric association (Beh and Lombardo, 2014). To overcome the issue above, the subcopula correlation ratio ρ_S^2 through the subcopula regression was proposed. The MCD measure μ_t, defined through subcopula, is a normalized expectation of weighted distance between the marginal distribution of one variable and its conditional distribution given the other variable.

The well-known Sklar's theorem (Sklar, 1959) states that, for any two random variables, a subcopula is a function that uniquely links the marginal distributions to their joint distribution on the ranges of its marginal distributions. If the random variables involved are continuous, a subcopula is known as a copula on the unit square. For the discrete random variables, any subcopula can be extended to a copula in a non-unique way. A common practice to construct a valid copula extension is to employ the distributional transform of the discrete variables and obtain the bilinear extension copula, also called checkerboard copula (Schweizer and Sklar, 1974; Rüschendorf, 2009). Many researchers have demonstrated that the bilinear extension copula can well capture the dependence of discrete variables (Genest and Nešlehová, 2007; Nešlehová, 2007; Faugeras, 2015).

* Corresponding author.
E-mail addresses: zheng.wei@maine.edu (Z. Wei), daeyoung@math.umass.edu (D. Kim).

https://doi.org/10.1016/j.spl.2021.109183
0167-7152/

In this paper we propose a new copula correlation ratio, named as *Bilinear Extension Copula based Correlation Ratio (BECCR)*, for the asymmetric association in two-way ordinal contingency tables. The BECCR is a measure of regression association utilizing the bilinear extension copula to quantify the average proportion of the variance associated with one variable explained by the other variable. We also propose a data-dependent nonparametric estimator of the BECCR using the empirical bilinear extension copula and prove its strong consistency.

The rest of the article is organized as follows. Section 2 reviews the bilinear extension copula for a two-way ordinal contingency table and proposes the bilinear extension copula regression. Section 3 proposes a new copula correlation ratio, *BECCR*, for the asymmetric dependence between two ordinal variables, and its theoretical properties are investigated. Section 4 proposes a strongly consistent nonparametric estimator of the BECCR. A conclusion is given in Section 5. Additional results for the paper including a running example, simulation and data analysis are provided in Sections S1 and S2 of the Supplementary Material.

2. Bilinear extension copula regression

Consider an $I \times J$ contingency table for ordinal variables X and Y with I ordered categories $\{x_1 < \cdots < x_I\}$ and J ordered categories $\{y_1 < \cdots < y_J\}$, respectively, and the joint probability mass function (p.m.f.) $P = \{p_{ij}\}$ where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$. The ith row and the jth column marginal p.m.f.s are denoted as $p_{i\cdot} = \sum_{j=1}^{J} p_{ij}$ and $p_{\cdot j} = \sum_{i=1}^{I} p_{ij}$, respectively. The conditional p.m.f.s of Y given X and X given Y are respectively denoted as $p_{j|i} = p_{ij}/p_{i\cdot}$ if $p_{i\cdot} \neq 0$ and zero otherwise, and $p_{i|j} = p_{ij}/p_{\cdot j}$ if $p_{\cdot j} \neq 0$ and zero otherwise.

Denote the ranges of the marginal distributions of X and Y by $D_0 = \{u_0, \ldots, u_i, \ldots, u_I\}$ and $D_1 = \{v_0, \ldots, v_j, \ldots, v_J\}$, where $u_0 = v_0 = 0$, $u_I = v_J = 1$, $u_i = \sum_{s=1}^{i} p_{s\cdot}$ and $v_j = \sum_{t=1}^{j} p_{\cdot t}$. Then, according to Sklar's theorem (Sklar, 1959), there exists a unique subcopula C^s associated with X and Y on $D_0 \times D_1$ such that

$$H(x_i, y_j) = C^s(F(x_i), G(y_j)) = C^s(u_i, v_j) = \sum_{s \le i} \sum_{t \le j} c^s(u_s, v_t), \tag{1}$$

where $H(x, y)$, $F(x)$ and $G(y)$ are the joint distribution function for X and Y and its marginal distributions, and $c^s(u_s, v_t) = p_{st}$ is the p.m.f. of $C^s(u_s, v_t)$.

The subcopula C^s in Eq. (1) can be extended to a copula on $I^2 = [0, 1]^2$ via the bilinear extension (Schweizer and Sklar, 1974), called *bilinear extension copula* (Nelsen, 2007), as shown below. For any $(u, v) \in I^2$, let u_1 and u_2 be respectively, the greatest and least elements of $\overline{D_0}$ satisfying $u_1 \le u \le u_2$; and v_1 and v_2 be respectively, the greatest and least elements of $\overline{D_1}$ satisfying $v_1 \le v \le v_2$, where $\overline{D}$ is the closure of set D. Then the bilinear extension copula $C^+(u, v)$ is defined to be

$$C^+(u, v) = (1\text{-}\lambda)(1\text{-}\mu)C^s(u_1, v_1) + (1\text{-}\lambda)\mu C^s(u_1, v_2) + \lambda(1\text{-}\mu)C^s(u_2, v_1) + \lambda\mu C^s(u_2, v_2), \tag{2}$$

where $\lambda = \frac{u - u_1}{u_2 - u_1}$ and $\mu = \frac{v - v_1}{v_2 - v_1}$. The density of $C^+(u, v)$ is defined to be

$$c^+(u, v) = p_{ij}\big/p_{i\cdot}p_{\cdot j}, \quad \text{if} \quad u_{i-1} < u \le u_i,\ v_{j-1} < v \le v_j. \tag{3}$$

The stochastic representation of $C^+(u, v)$ in Eq. (2) (Nešlehová, 2007; Rüschendorf, 2009) is that $C^+(u, v)$ is the joint distribution of two standard uniform variables U and V, the distributional transform of the ordinal variables X and Y :

$$U = F(X-) + [F(X) - F(X-)]W_1, \quad V = G(Y-) + [G(Y) - G(Y-)]W_2, \tag{4}$$

where $F(x-)$ refers to the left limit of F and W_1, W_2 denote the independent uniform random variables on $[0, 1]$ independent of X and Y.

Using the bilinear extension copula, we define an important tool to identify the asymmetric regression association for a two-way ordinal contingency table.

Definition 1. Let U and V be standard uniform variables with $c^+(u, v)$ in Eq. (3) associated with X and Y. The **bilinear extension copula regression** of V on U is defined as follows : for $u_{i-1} < u \le u_i$,

$$r_{V|U}(u) \equiv E_{c^+}(V|U = u) = \int_0^1 v c^+(v|u) dv = \sum_{j=1}^{J} p_{j|i} \left(\frac{v_j + v_{j-1}}{2} \right), \tag{5}$$

where $c^+(v|u)$ is the conditional density of V given U. The bilinear extension copula regression of U on V for $v_{j-1} < v \le v_j$, denoted as $r_{U|V}(v)$, can be defined similarly.

3. Asymmetric association measure via bilinear extension copula regression

This section proposes a new measure of asymmetric association for the ordinal contingency table and studies its theoretical properties.

Definition 2. For the ordinal variables X and Y in an $I \times J$ table, the **Bilinear Extension Copula based Correlation Ratio (BECCR) of** Y **on** X is defined to be

$$\rho^2_{(X\to Y)} \equiv \frac{Var(r_{V|U}(U))}{Var(V)} = \frac{E[\left(r_{V|U}(U) - 1/2\right)^2]}{1/12} = 3\sum_{i=1}^{I}\left(\sum_{j=1}^{J} p_{j|i}\left(v_j + v_{j-1}\right) - 1\right)^2 p_{i\cdot}\ . \tag{6}$$

The BECCR of X on Y, denoted as $\rho^2_{(Y\to X)}$, can be defined similarly.

The proposition below investigates the properties of the proposed measures.

Proposition 1. *(a)* $0 \le \rho^2_{(X\to Y)} \le 3\sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}$ *and* $0 \le \rho^2_{(Y\to X)} \le 3\sum_{i=1}^{I} u_{i-1}u_i p_{i\cdot}$.

(b) (i) If X *and* Y *are independent, then* $\rho^2_{(X\to Y)} = \rho^2_{(Y\to X)} = 0$*, (ii) If* $\rho^2_{(X\to Y)} = 0$ *or* $\rho^2_{(Y\to X)} = 0$*, then* $r_{V|U}(U) = E(V) = 1/2$ *or* $r_{U|V}(V) = E(U) = 1/2$*, and* $corr(U, V) = 0$.

(c) $\rho^2_{(X\to Y)} = 3\sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}$ *if and only if* Y *is a function of* X *almost surely. Similarly,* $\rho^2_{(Y\to X)} = 3\sum_{i=1}^{I} u_{i-1}u_i p_{i\cdot}$ *if and only if* X *is a function of* Y *almost surely.*

(d) If $V = g(U)+\epsilon$ *where* ϵ*, being independent of* U*, is a random variable with finite second moments and* g *is a measurable function, then* $\rho^2_{(X\to Y)} = \frac{Var(g(U))}{[Var(g(U))+Var(\epsilon)]}$.

(e) $\rho_{(Y\to X)} = corr(U, r_{U|V}(V))$ *and* $\rho_{(X\to Y)} = corr(V, r_{V|U}(U))$.

Proof. See Appendix A.

Proposition 1(a) and (c) indicate that the BECCR reaches at its upper bound if one variable is uniquely determined by the other variable. From Proposition 1(d) and (e), we see that the BECCR represents the average proportion of the variance associated with the (tentative) response variable explained by the bilinear extension copula regression of the (tentative) predictor and its square root describes the predictive power of the corresponding regression. Note that the BECCR is asymmetric in the sense that in general, $\rho^2_{(X\to Y)}$ is not equal to $\rho^2_{(Y\to X)}$.

Proposition 1(a) indicates that the upper bounds of $\rho^2_{(X\to Y)}$ and $\rho^2_{(Y\to X)}$ depend on the marginals of Y and X, respectively. Therefore, it is natural to consider the standardized versions of $\rho^2_{(X\to Y)}$ and $\rho^2_{(Y\to X)}$.

Definition 3. For ordinal variables X and Y in an $I \times J$ contingency table, the **standardized BECCR** of Y on X and of X on Y are defined as follows, respectively:

$$\rho^{*2}_{(X\to Y)} \equiv \frac{\rho^2_{(X\to Y)}}{3\sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}} \quad \text{and} \quad \rho^{*2}_{(Y\to X)} \equiv \frac{\rho^2_{(Y\to X)}}{3\sum_{i=1}^{I} u_{i-1}u_i p_{i\cdot}}. \tag{7}$$

The standardized BECCRs are both between 0 and 1, and so they can be compared on the same scale when the question of interest is to explore the asymmetric association. Note that Proposition 1(b)–(c) hold for the standardized BECCR.

The standardized BECCRs also have similar properties as Proposition 1(d) and (e). Suppose that, using the ranges D_0 and D_1 of the marginal distributions of X and Y, we assign the ordered scores $\{(u_{i-1} + u_i)/2\}$ and $\{(v_{j-1} + v_j)/2\}$ to the ith and jth categories of X and Y, respectively. Denote X^* and Y^* to be the ordinal random variables equipped with the ordered scores defined above and the same joint p.m.f $P = \{p_{ij}\}$ as that of X and Y. Then it can be shown that $Var(X^*) = \sum_{i=1}^{I} u_{i-1}u_i p_{i\cdot}/4$ and $Var(Y^*) = \sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}/4$, and thus, $\rho^{*2}_{(X\to Y)}$ is equal to $Var(r_{V|U}(U))/Var(Y^*)$ and $\rho^{*2}_{(Y\to X)}$ is equal to $Var(r_{U|V}(V))/Var(X^*)$. We can also show that $\rho^*_{(X\to Y)} = corr(Y^*, E[Y^*|X^*])$ and $\rho^*_{(Y\to X)} = corr(X^*, E[X^*|Y^*])$.

Remark 1. As a reviewer pointed out, Proposition 1(b)-(ii) indicates that the (standardized) BECCR is not able to detect independence. However, it has easy and meaningful interpretation in terms of regression association. That is, $\rho^2_{(X\to Y)} = 0$ and $\rho^{*2}_{(X\to Y)} = 0$ are equivalent to $r_{V|U}(U) = E(V) = 1/2$ and, due to Proposition 1(d), this indicates that X has no contribution in explaining the variance associated with Y.

Remark 2. Section S3 in the Supplementary Material highlights the similarities and differences between the proposed BECCR in Eq. (7) and two existing asymmetric association measures constructed through subcopulas, μ_t of Shan et al. (2015) and ρ^2_S of Wei and Kim (2017).

4. Estimation

In order to estimate the proposed asymmetric association measures, we use the observed cell count $\boldsymbol{n} = \{n_{ij}\}$ in an $I \times J$ contingency table for two ordinal variables X and Y. First, the estimators for p_{ij}, $p_{i\cdot}$, $p_{\cdot j}$, u_i, v_j, $p_{i|j}$, and $p_{j|i}$ are given by $\hat{p}_{ij} = n_{ij}/n$, $\hat{p}_{i\cdot} = n_{i\cdot}/n$, $\hat{p}_{\cdot j} = n_{\cdot j}/n$, $\hat{u}_i = \sum_{s=1}^{i} \hat{p}_{s\cdot}$, $\hat{v}_j = \sum_{t=1}^{j} \hat{p}_{\cdot t}$, $\hat{p}_{i|j} = \hat{p}_{ij}/\hat{p}_{\cdot j}$ if $\hat{p}_{\cdot j} \neq 0$ and zero otherwise, and $\hat{p}_{j|i} = \hat{p}_{ij}/\hat{p}_{i\cdot}$ if $\hat{p}_{i\cdot} \neq 0$ and zero otherwise, where $n = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$.

We can then obtain the empirical distribution H_n of (X, Y), and the empirical counterparts of the subcopula in Eq. (1) and the bilinear extension copula in Eq. (2), the *empirical subcopula* and the *empirical bilinear extension copula* denoted as C_n^s and C_n^+, respectively. Finally, the estimators of the (standardized) BECCRs in Definitions 2 and 3 can be obtained by plugging in the above estimators. We denote the obtained estimators by $\hat{\rho}^2_{(X\to Y)}$, $\hat{\rho}^2_{(Y\to X)}$, $\hat{\rho}^{*2}_{(X\to Y)}$ and $\hat{\rho}^{*2}_{(Y\to X)}$.

Using the strong consistency of the empirical subcopula given in Rachasingho and Tasena (2020), the corollary below presents almost sure consistency of the proposed estimator of the BECCR.

Corollary 1. *With probability one,*

$$\hat{\rho}^2_{X\to Y} \to \rho^2_{X\to Y}, \quad \textit{and} \quad \hat{\rho}^{*2}_{X\to Y} \to \rho^{*2}_{X\to Y}, \quad \textit{as } n \to \infty. \tag{8}$$

Proof. See Appendix B.

To evaluate the uncertainty in the estimates of the proposed measures $\rho^2_{(X\to Y)}$, $\rho^2_{(Y\to X)}$, $\rho^{*2}_{(X\to Y)}$, $\rho^{*2}_{(Y\to X)}$ and their pairwise difference $\rho^{*2}_{(X\to Y)} - \rho^{*2}_{(Y\to X)}$, we suggest using the bootstrap confidence interval method (Efron and Tibshirani, 1994).

5. Conclusion

In this paper, we proposed a bilinear extension copula correlation ratio for asymmetric association in a two-way ordinal contingency table and a strongly consistent nonparametric estimator for the proposed measure.

To illustrate the underlying idea of the proposed measure, we include a running example in Section S1 of the Supplementary Material. We also provide additional results for simulations and real data analysis in Sections S2.1 and S2.2 of the Supplementary Material to examine the finite-sample performance of the proposed nonparametric estimator and its utility.

A valuable extension of this research would be to investigate the strong consistency of the estimators of the BECCR when the ordinal random variables X and Y take values in a countable set.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We appreciate the Editor, Associate Editor and two anonymous referees for the insightful comments that substantially improved the article.

Appendix A. The proof of Proposition 1

(a) Using $Var(V) = E[Var(V|U)] + Var(r^C_{V|U}(U))$, we can rewrite $\rho^2_{(X\to Y)}$ as $\rho^2_{(X\to Y)} = 1 - E[Var(V|U)]/Var(V) = 1 - E[(V - E[V|U])^2]/Var(V)$. The following inequality holds for the numerator of $\rho^2_{(X\to Y)}$:

$$Var\left(r^C_{V|U}(U)\right) = E\left[r^C_{V|U}(U) - 1/2\right]^2 = \sum_{i=1}^{I}\left(\sum_{j=1}^{J} p_{j|i}\frac{v_j + v_{j-1}}{2} - 1/2\right)^2 p_{i\cdot}$$
$$= \sum_{i=1}^{I}\left(\sum_{j=1}^{J} p_{j|i}\frac{v_j + v_{j-1}}{2} - \sum_{j=1}^{J} p_{\cdot j}\frac{v_j + v_{j-1}}{2}\right)^2 p_{i\cdot} = \sum_{j=1}^{J}\left(\frac{v_j + v_{j-1}}{2} - \frac{1}{2}\right)^2 p_{\cdot j}$$
$$- \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{v_j + v_{j-1}}{2} - \sum_{j=1}^{J}\left(\frac{v_j + v_{j-1}}{2}\right)p_{j|i}\right)^2 p_{ij} \le \sum_{j=1}^{J}\left(\frac{v_j + v_{j-1}}{2} - \frac{1}{2}\right)^2 p_{\cdot j} = \frac{1}{4}\sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}$$

(b) If X and Y are independent, the bilinear extension copula $C^+(u, v)$ is also independent and then $r_{V|U}(u) = E[V]$. On the other hand, if $\rho^2_{(X\to Y)} = 0$, we have $E[(r_{V|U}(U) - E(V))^2] = 0$ which in turn implies $r_{V|U}(U) = E(V) = 1/2$ almost everywhere and $\int_0^1 vc^+(u, v)dv = \int_0^1 vdv$. This means $E(UV) = E[U]E[V]$.

(c) By (a), we know $\rho^2_{(X\to Y)} = 3\sum_{j=1}^{J} v_{j-1}v_j p_{\cdot j}$ if and only if $\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\frac{u_i+u_{i-1}}{2} - \sum_{i=1}^{I}\left(\frac{u_i+u_{i-1}}{2}\right) p_{i|j}\right)^2 p_{ij} = 0$. This is equivalent to Y is a function of X almost surely.

(d) If $V = g(U) + \epsilon$, then $r_{V|U}(u) = g(u) + E[\epsilon]$, and $Var(r_{V|U}(U)) = Var(g(U))$. Therefore, we have $\rho^2_{(X\to Y)} = Var(g(U))/(Var(g(U)) + Var(\epsilon))$.

(e) Since $Var(r_{V|U}(U)) = E[VE[V \mid U]] - E[V]E[E[V \mid U]] = cov(V, E[V \mid U])$,

$$cor(V, E[V \mid U]) = \frac{cov(V, E[V \mid U])}{\sqrt{Var(V)}\sqrt{Var(E[V \mid U])}} = \sqrt{\frac{Var(r_{V|U}(U))}{Var(V)}} = \sqrt{\rho^2_{(X\to Y)}}.$$

Appendix B. The proof of Corollary 1

Let U and V be the uniform random variables in Eq. (4) transformed from the ordinal variables (X, Y) with the joint distribution H and its marginals F and G in Eq. (1). Then the bilinear extension copula C^+ in Eq. (2) is the joint distribution of (U, V) (Nešlehová, 2007). We denote H_n, F_n and G_n to be the joint empirical distribution of (X, Y) and its marginals, respectively. For (X_n^*, Y_n^*) generated from H_n, define U_n^* and V_n^* (empirical counterpart of U and V in Eq (4)) by $U_n^* = F_n(X_n^*-) + [F_n(X_n^*) - F_n(X_n^*-)]W_1$, $V_n^* = G_n(Y_n^*-) + [G_n(Y_n^*) - G_n(Y_n^*-)]W_2$, where W_1 and W_2 are independent uniform random variables. Then the joint distribution of (U_n^*, V_n^*) is the empirical bilinear extension copula associated with H_n, denoted as C_n^+, which is the empirical counterpart of C^+.

The bilinear extension copula regression in Eq. (5) and the proposed association measure in Eq. (6) can be written as a function of $\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv$:

$$r_{V|U}(u) = 1 - \int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv,$$
$$\rho^2_{X\to Y} = 12\left[1 - 2E_U\left(\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv\right) + E_U\left(\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv\right)^2\right] - 3.$$

By the definition of C^+ in Eq. (2) we have that, for $u_{i-1} < u < u_i$ and $v_{j-1} < v < v_j$,

$$\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv$$
$$= \sum_{i=1}^{I} 1_{(u_{i-1},u_i)} \sum_{j=1}^{J} \frac{v_j - v_{j-1}}{2(u_i - u_{i-1})}\left(-C^S(u_{i-1}, v_{j-1}) - C^S(u_{i-1}, v_j) + C^S(u_i, v_{j-1}) + C^S(u_i, v_j)\right).$$

Let $\hat{u}_i$ and $\hat{v}_j$ be the empirical counterpart of $u_i = \sum_{s=1}^{i} p_{s\cdot}$ and $v_j = \sum_{t=1}^{j} p_{\cdot t}$. By the law of large number, $\hat{u}_i \to u_i$ a.s. and $\hat{v}_j \to v_j$ a.s. for all i and j. It follows from the condition (C2) and Corollary 5.3 in Rachasingho and Tasena (2020) that the empirical subcopula $C_n^S(\hat{u}_i, \hat{v}_j)$ converges to the true subcopula $C^S(u_i, v_j)$ for all i and j. Thus, $\int_0^1 \frac{\partial}{\partial u} C_n^+(u, v)dv$ (the empirical counterpart of $\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv$ expressed as a linear combination of $C_n^S(\hat{u}_i, \hat{v}_j)$ for all i, j) converges to $\int_0^1 \frac{\partial}{\partial u} C^+(u, v)dv$ a.s., and then $\hat{\rho}^2_{X_n^*\to Y_n^*}$ converges to $\rho^2_{X\to Y}$ a.s. by the bounded convergence theorem.

Furthermore, let Y^* to be the ordinal random variable equipped with the ordered scores $\{(v_{j-1} + v_j)/2\}$ with v_j in D_1 and the same marginal distribution G as that of Y. Then the plug-in estimator for $Var(Y^*)$ used in the denominator of $\rho^{*2}_{X\to Y}$ can be rewritten in terms of the empirical distribution function G_n of Y:

$$\widehat{Var(Y^*)} = \int \left(\frac{G_n(y) + G_n(y-)}{2} - \frac{1}{2}\right)^2 dG_n(y).$$

Then we have the a.s. convergence of $\widehat{Var(Y^*)}$ to $Var(Y^*)$ by Proposition 4 in Genest et al. (2013). This proves the strong consistency of $\hat{\rho}^{*2}_{X_n^*\to Y_n^*}$.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2021.109183.

References

Beh, E.J., Lombardo, R., 2014. Correspondence Analysis: Theory, Practice and New Strategies. John Wiley & Sons.
Chatterjee, S., 2020. A new coefficient of correlation. J. Amer. Statist. Assoc. 1–21. http://dx.doi.org/10.1080/01621459.2020.1758115.
Dabrowska, D., 1981. Regression-based orderings and measures of stochastic dependence. Ser. Stat. 12, 317–325.
Dette, H., Siburg, K., Stoimenov, P., 2013. A copula-based non-parametric measure of regression dependence. Scand. J. Stat. 40, 21–41.
Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press.

Faugeras, O., 2015. Maximal coupling of empirical copulas for discrete vectors. J. Multivariate Anal. 137, 179–186.
Genest, C., Nešlehová, J.G., 2007. A primer on copulas for count data. Astin Bull. 37, 475–515.
Genest, C., Nešlehová, J.G., Rémillard, B., 2013. On the estimation of Spearman's rho and related tests of independence for possibly discontinuous multivariate data. J. Multivariate Anal. 117, 214–228.
Junker, R., Griessenberger, F., Trutschnig, W., 2021. Estimating scale-invariant directed dependence of bivariate distributions. Comput. Statist. Data Anal. 153, 107058. http://dx.doi.org/10.1016/j.csda.2020.107058.
Kruskal, W.H., Goodman, L., 1954. Measures of association for cross classifications. J. Amer. Statist. Assoc. 49 (268), 732–764.
Nelsen, R.B., 2007. An Introduction to Copulas. Springer, New York.
Nešlehová, J., 2007. On rank correlation measures for non-continuous random variables. J. Multivariate Anal. 98 (3), 544–567.
Okimoto, T., 2008. New evidence of asymmetric dependence structures in international equity markets. J. Financ. Quant. Anal. 43, 787–815.
Rachasingho, J., Tasena, S., 2020. A metric space of subcopulas – An approach via Hausdorff distance. Fuzzy Sets and Systems 378 (1), 144–156.
Rüschendorf, L., 2009. On the distributional transform, sklar's theorem, and the empirical copula process. J. Statist. Plann. Inference 139 (11), 3921–3927.
Schweizer, B., Sklar, A., 1974. Operation on distribution functions not derivable from operations on random variables. Studia Math. 52, 43–52.
Shan, Q., Wongyang, T., Wang, T., Tasena, S., 2015. A measure of mutual complete dependence in discrete variables through subcopula. Internat. J. Approx. Reason. 65, 11–23.
Sklar, A., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.
Sungur, E., 2005. A note on directional dependence in regression setting. Comm. Statist. Theory Methods 34, 1957–1965.
Trutschnig, W., 2011. On a strong metric on the space of copulas and its induced dependence measure. J. Math. Anal. Appl. 384, 690–705.
Wang, Y.X., Huang, H., 2014. Review on statistical methods for gene network reconstruction using expression data. J. Theoret. Biol. 362, 53–61.
Wei, Z., Kim, D., 2017. Subcopula-based measure of asymmetric association for contingency tables. Stat. Med. 36 (24), 3875–3894.