A GENERALIZED DEFINITION OF THE POLYCHORIC CORRELATION COEFFICIENT

JOAKIM EKSTRÖM

Abstract. The polychoric correlation coefficient is a measure of association for ordinal variables which rests upon an assumption of an underlying joint continuous distribution. More specifically, in Karl Pearson's original definition an underlying joint normal distribution is assumed. In this article, the definition of the polychoric correlation coefficient is generalized so that it allows for other distributional assumptions than the joint normal distribution. The generalized definition is analogous to Pearson's definition, and the two definitions agree under bivariate normal distributions. Moreover, the polychoric correlation coefficient is put into a framework of copulas which is both mathematically and practically convenient. The theory is illustrated with examples which, among other things, show that the measure of association suffers from lack of statistical robustness.

Key words and phrases. Polychoric Correlation Coefficient, Contingency Table, Measure of Association, Ordinal Variables.

1. Introduction

The polychoric correlation coefficient is a measure of association for ordinal variables. It was first proposed by Karl Pearson in year 1900, and throughout his career, Pearson was an advocate of the statistical method.

A measure of association is, loosely, a function which maps a pair of random variables to a subset of the real line. The first and likely most well-known measure of association is the ordinary linear correlation. The linear correlation was first envisioned in year 1888 by Francis Galton, who originally named it *co-relation*. Galton (1888) claimed that the proposed measure of association was particularly useful for the study of social problems, such as the relationship between poverty and crime. More than a hundred years later, though, it is safe to say that the study of association between variables indeed is of fundamental interest in nearly every scientific discipline.

Ordinal variables, also called ordered category variables, are variables whose values are ordered, but cannot be added or multiplied. For example, the extent to which a person experiences an emotion, values a product, or agrees to an opinion are usually thought of as inherently ordinal variables. Ordinal variables can also occur as a consequence of difficulties in measurement. Numerical variables are sometimes considered as merely ordinal if the measurements are blurred to such an extent that it only is meaningful to compare values in terms of order. Ordinal variables are common in, for example, medicine and the social sciences.

Data for pairs of ordinal variables are often presented in the form of contingency tables. Each value of an ordinal variable represent either a row or a column, and frequencies or relative frequencies are shown in corresponding cells. Sometimes the marginal frequencies are printed in a separate row and column. The historically prominent smallpox data set, from Pearson (1900), is shown in Table 1.

Measures of association for ordinal variables is a subject that has been studied from the very infancy of modern statistics. In the 7th article in the seminal series *Mathematical contributions to the theory of evolution*, Pearson (1900) proposed what later became

Table 1. Karl Pearson's smallpox recovery data.

	Recovery	Death	
Vaccinated	1562	42	1604
Unvaccinated	383	94	477
	1945	136	2081

Pearson's chi-square test for independence

$\chi^2_{obs} = 176$ p-value < 0.0001

Source: Metropolitan Asylums Board: Small-pox epidemic 1893. (Pearson, 1900)

known as the polychoric correlation coefficient. The fundamental idea is to think of the two ordinal variables as having an underlying joint normal distribution, and that the contingency table consequently is a result of a double discretization of the joint distribution, see Figure 1 for an illustration. The polychoric correlation coefficient is the linear correlation of the postulated joint normal distribution.

According to Pearson's colleague Burton H. Camp (1933), Pearson considered the polychoric correlation coefficient as being one of his most important contributions to the theory of statistics. However, the polychoric correlation coefficient suffered in popularity because of the difficulty in its computation. Throughout his career, Pearson published statistical tables aimed at reducing that difficulty (Camp, 1933), reflecting an interest in promoting a wider adoption of the polychoric correlation coefficient among practitioners.

There are several theories why Pearson for the purpose of the definition of the polychoric correlation coefficient chose the family of bivariate normal distributions. Central to Pearson's thinking was the idea of the ordinal variables having continuous underlying distributions, and at the time the normal distribution was prevalent. In fact, according to Pearson & Heron (1913) there were no other continuous bivariate distribution that up until the time had been discussed effectively. Furthermore, Pearson (1900) was primarily interested in applications in the fields of evolution and natural selection, which is evident from the article's title, and such variables were generally assumed to be normally distributed. Pearson's mentor Francis Galton even had a philosophical argument why all variables in nature ought to be normally distributed. Also, the parameter of the parametric family of bivariate normal distributions happens to be a measure of association, and this in combination with other nice properties makes the choice of the bivariate normal distribution most convenient.

Of course, not all continuous bivariate distributions are normal. Pearson & Heron (1913) addresses this apparent weakness, and claims that for the purpose of the polychoric correlation coefficient, divergence between the actual joint distribution and the normal distribution is hardly ever of practical importance. It is not mentioned how Pearson & Heron arrived at this conclusion. In the present article, it will be shown that the distributional assumption in fact has a profound impact on the polychoric correlation coefficient. For example, for Pearson's smallpox data set, Table 1, the polychoric correlation coefficient ranges from 0.9, to 0, to -0.2, only because of changes of the distributional assumption. Correspondingly, the conclusions of the association analysis ranges from near perfect positive association, to statistical independence, to negative association between the two ordinal variables only as a consequence of changes of the distributional assumption. So contrary to Pearson's intuition, it will be seen that the distributional assumption is indeed of profound importance for the purpose of the polychoric correlation coefficient.

1.1. **Bibliographical summary.** Originally, Pearson (1900) studied the special case of association between dichotomous variables, i.e. 2×2 contingency tables. The computation of the coefficient amounts to the solving of an integral equation involving the

bivariate standard normal density function. So-called tetrachoric series were used as a means to solve the integral equation, and the use of such is, in all likelihood, the reason why the measure of association, for 2×2 contingency tables, became known as the tetrachoric correlation coefficient.

Ritchie-Scott (1918) extended the tetrachoric correlation coefficient to general ordinal variables. The extension is not trivial because for general $r \times s$ contingency tables, a solution to an analogous integral equation does in general not exist. Ritchie-Scott's suggestion was to dichotomize the ordinal variables in all possible ways, calculate a tetrachoric correlation coefficient for each dichotomization, and then to compute a weighted average of those so obtained tetrachoric correlation coefficients. The weighted average was subsequently called the polychoric correlation coefficient.

Tallis (1962) suggested that a polychoric correlation coefficient could be fitted to the contingency table with respect to a multiplicative loss function referred to as a likelihood. Martinson & Hamdan (1971) merged the idea of Tallis (1962) with the works of Pearson and Ritchie-Scott, along with some computational simplifications. Martinson & Hamdan (1971) also provided some additional suggestions of loss functions. Moreover, Olsson (1979) suggested a slightly modified approach, allowing for reclassifications. All of the above was done under the assumption of an underlying joint normal distribution.

Quiroga (1992) and Roscino & Pollice (2006) suggested a polychoric correlation coefficient with a mixture of an independent bivariate skew-normal distribution and a bivariate normal distribution, and a bivariate skew-normal distribution, respectively, as underlying distributional assumptions. However, the theories are not fully developed, e.g. they provide neither results on existence, nor that the definitions agree for bivariate normal distributions.

1.2. **Outline of the present article.** The aim of the present article is to generalize the definition so that the polychoric correlation coefficient can be computed under other distributional assumptions than the normal distribution. The generalization makes it possible to explore the extent to which the polychoric correlation coefficient depends on the distributional assumption. Using the generalized definition, statistical robustness properties will be evaluated.

The generalization is made in analogy with Pearson's original definition, and the generalized definition is shown to agree with Pearson's definition under a joint normal distribution assumption. Extensive results on existence and uniqueness of a polychoric correlation coefficient in various circumstances are provided. Moreover, the generalized polychoric correlation coefficient is put into the framework of copulas, a framework which is both mathematically suitable as well as convenient in practice.

It is briefly discussed how a goodness-of-fit analysis can yield further information on the association between the ordinal variables. When in doubt, different distributional assumptions can be tested. Furthermore, consideration of goodness-of-fit can enrich the association analysis with an analysis of, for example, possible tail dependence.

In Section 2, the framework of ordinal variables is discussed, and the polychoric correlation coefficient is presented with its assumptions formalized. In Section 3, the generalized definition is introduced along with some results on existence and uniqueness of a polychoric correlation coefficient for a given family of bivariate distributions. Furthermore, it is shown that the generalized definition agrees with the conventional definition. Section 4 contains suggestions for goodness-of-fit tests, and Section 5 contains examples to illustrate the use of the generalized definition. And finally, the article is concluded with Section 6.

2. The polychoric correlation coefficient

2.1. **Ordinal variables.** Ordinal variables are variables whose values are ordered but cannot in general be added, multiplied, or otherwise acted on by any binary operator save projection. In the framework of Kolmogorov's definition of random variables, an ordinal variable is a measurable function from a probability space Ω to some sample space, $\mathcal{C}$. The sample space $\mathcal{C} = \{c_1, c_2, \dots\}$ is totally ordered, i.e. for any c_i and c_j it holds that either $c_i \preceq c_j$, $c_i \succeq c_j$, or both. But characteristically, the sample space is not equipped with any binary operation. The equality notation $c_i = c_j$ is shorthand for $c_i \preceq c_j$ and $c_i \succeq c_j$, and the strict notation $c_i \prec c_j$ is shorthand for $c_i \preceq c_j$ and $c_i \not\succeq c_j$.

In the present context, what the elements of $\mathcal{C}$ represent is not of interest. The elements may represent colors, opinions, species, or anything else, but the only characteristic that is of relevance in this statistical context is the ordering. Therefore, all elements that have the same order are considered equal. Let $[c]_{\mathcal{C}}$ denote the equivalence class $\{x \in \mathcal{C} : x = c\}$, and let $\lfloor c \rfloor_{\mathcal{C}}$ denote the lower half-space $\{x \in \mathcal{C} : x \preceq c\}$. The index $\mathcal{C}$ is sometimes omitted when the ordered set is clear from the context.

Since the concern of the analysis is the equivalence classes $[c_i]$, for an ordinal variable $X : \Omega \to \mathcal{C}$ it is assumed without loss of generality that the strict inequalities $c_1 \prec c_2 \prec c_3 \prec \cdots$ hold. That this can be assumed is clear when considering that one always can map each equivalence class to any element of the class, relabel them if necessary, and then get a totally ordered set for which the strict inequalities hold. The values of an ordinal variable are sometimes referred to as *categories*, the ordinal variable as an *ordered categorical variable*, and the cardinality of the sample space as the *number of categories*.

Let X and Y be the two ordinal variables whose association is to be studied, and denote their numbers of categories r and s, respectively. Let the cumulative marginal probabilities be denoted $u_0, u_1, \dots, u_r$ for X, i.e. $u_0 = 0$, $u_r = 1$ and $u_i = P(X \preceq c_i)$, and $v_0, v_1, \dots, v_s$ for Y. The marginal probabilities are denoted ∇u_i and ∇v_j, respectively. The symbol ∇ can be interpreted as a difference operator, yielding $\nabla u_i = u_i - u_{i-1} = P(X = c_i)$.

The joint distribution of X and Y is often illustrated with an $r \times s$ contingency table, where the values of X and Y are labeling, in correct order, the columns and rows and the joint probabilities are printed in the corresponding cells. The joint probabilities are

sometimes denoted with a double index, each referring to a value of one of the ordinal variables. In the present article, however, the joint probabilities will be denoted with single index, $p_1, \ldots, p_{rs}$, each index referring to a specific cell of the contingency table. The way the in which the cells of the contingency table is enumerated is unimportant. For example, the cells could be enumerated column-wise, row-wise, or via Cantor's diagonal method.

As always, Kolmogorov's axioms imply that the joint probabilities are elements of the unit interval, $I = [0, 1]$, and that they sum to one. It is sometimes necessary to separate the two cases where the cumulative marginal probabilities, (u_i, v_j), are elements of the boundary and the interior of the unit square, I^2, respectively. Because it is closed, the unit square is the disjoint union of its boundary and its interior, $I^2 = \partial I^2 \cup \mathrm{Int}(I^2)$.

If X and Y are statistically independent, then the joint probabilities are the products of the marginal probabilities. Given a sample, independence can be tested with, e.g., the Pearson chi-square test. An example of such a test is shown in Table 1. If X and Y are found to be statistically dependent, then it is often of interest to estimate some measure of association. The polychoric correlation coefficient is one such measure of association, especially defined for ordinal variables.

2.2. **Pearson's definition.** The fundamental idea of the polychoric correlation coefficient, as presented in Pearson (1900), is to think of the ordinal variables as discretized random variables with a joint normal distribution. Pearson likely visualized the contingency table with the bell-shaped bivariate normal density function standing on top of it. The discretization cuts the domain of the bivariate normal density function into rectangles corresponding to the cells of the contingency table, see Figure 1 for an illustration. Since the ordinal variables are both scale and origin free, and the family of normal distributions is closed under linear transformations, the normal distribution can without loss of generality be set to standard normal. Changing the parameter value of the bivariate standard normal distribution will change the shape of the bell-shaped bivariate normal density function, and hence the probability masses over the rectangles that results from the discretization. The polychoric correlation coefficient is the parameter value for which the volumes of the discretized bivariate standard normal distribution equal the joint probabilities of the contingency table, i.e. the parameter value for which the probability measures, as induced by the bivariate standard normal distribution, of the rectangles resulting from the discretization equal the joint probabilities of the contingency table. The parameter value of the bivariate standard normal distribution equals, of course, the linear correlation of the two jointly normally distributed random variables.

The fundamental assumption is formalized as follows.

Assumption A1. *The two ordinal variables are, into r and s ordered categories respectively, discretized random variables with a joint normal distribution.*

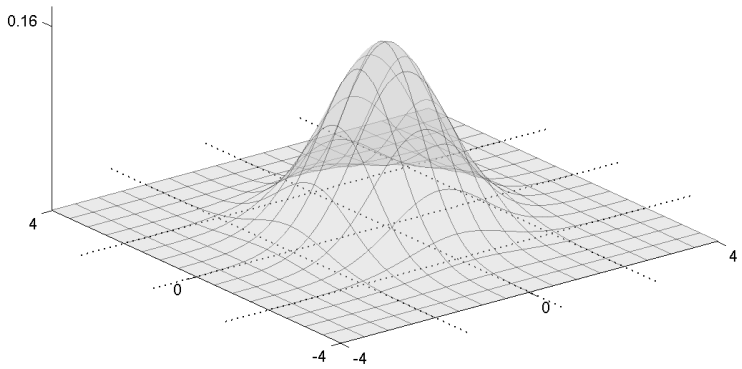

FIGURE 1. Illustration of the domain of the standard normal density function being discretized by the dotted lines into a 4×4 contingency table.

It follows from Assumption A1 that the joint normal distribution must be discretized such that the marginal probabilities of the discretized bivariate standard normal distribution equal the marginal probabilities of the contingency table. Thus, the rectangles resulting from the discretization of the joint normal distribution are all of of the form $[\Phi^{-1}(u_{i-1}), \Phi^{-1}(u_i)] \times [\Phi^{-1}(v_{j-1}), \Phi^{-1}(v_j)]$, where Φ^{-1} is the inverse of the univariate standard normal distribution function. Create such rectangles for all $i = 1, \ldots, r$ and $j = 1, \ldots, s$, enumerate them in the same way as the joint probabilities $p_1, \ldots, p_{rs}$, and denote them $A_1, \ldots, A_{rs}$.

Under Assumption A1 it should also, ideally, hold that the joint probabilities of the discretized joint normal distribution equal the joint probabilities of the ordinal variables. Hence, the equation

$$\int_{A_k} \phi_\rho \, d\lambda = p_k,$$

where ϕ_ρ is the bivariate standard normal density function with parameter ρ and λ is the Lebesgue measure, should hold for all $k = 1, \ldots, rs$. Note that the left hand side of the equation above is the probability measure of the rectangle A_k, as induced by the postulated underlying distribution. The left hand side is often called the volume of the rectangle A_k and is here denoted $\Phi_\rho(A_k)$. Using vector notation, the equalities above can be written

$$(\Phi_\rho(A_1), \ldots, \Phi_\rho(A_{rs})) = (p_1, \ldots, p_{rs}) \,. \tag{1}$$

The polychoric correlation coefficient, r_{pc}, is defined as the solution ρ to Equation (1). If all cumulative marginal probabilities (u_i, v_j) are elements of the boundary of the unit square, ∂I^2, then any value of ρ will satisfy Equation (1). In this case, however, the the polychoric correlation coefficient is defined to be zero, in part because of a reasoning of presuming independence until evidence of association is found.

Originally, Pearson (1900) studied the case of dichotomous variables, i.e. 2×2 contingency tables. In the 2×2 case, a unique solution of Equation (1) always exists. In other cases a solution does in general not exist.

Proposition 1. *For every 2×2 contingency table, a unique solution to Equation (1) exists, and the polychoric correlation coefficient is consequently well defined for all 2×2 contingency tables.*

Proof. See Ekström (2008). □

Proposition 2. *If one of the numbers of categories is greater than 2 and the other is greater or equal to 2, then a solution to Equation (1) does in general not exist.*

Proof. Assume that one of the numbers of categories, r, say, is greater than 2 and the other, s, is greater than or equal to 2. Take the contingency table for which the joint probability corresponding to the second category of both ordinal variables is zero and the other joint probabilities are equal to $(rs-1)^{-1}$. Because the normal distribution is elliptic, no value of ρ can then satisfy Equation (1). Thus, the statement is proved by counter-example. □

2.3. **Methods of fitting a coefficient.** If a solution to Equation (1), the defining relation of the polychoric correlation coefficient, does not exist then it can still be argued that Assumption A1 basically is valid, for example based on a reasoning of fixed sample sizes or a reasoning of noisy observations. One could also argue that Assumption A1 is approximately true, i.e. that the underlying joint distribution of the ordinal variables is approximately normal in some sense.

From this point of view it is natural to look for some best compromise value of the parameter ρ of Equation (1), or more specifically a best fit with respect to a loss function. A distance between the vectors of (1) is suitable because if a unique solution exists such a loss function has a unique global minimum which corresponds to the solution. Thus, a best fit with respect to such a loss function will yield the solution of Equation (1) whenever a solution exists. Moreover, a distance is non-negative, geometrically interpretable, and hence suits the intuitive notion of a loss function.

The usual L^p norm, $||\mathbf{x}||_p = (\sum x_i^p)^{1/p}$, is in many ways natural. Denote the vector on the left hand side of Equation (1) by $\mathbf{\Phi}_\rho$ and the vector on the right hand side by $\mathbf{p}$. The fitted polychoric correlation coefficient with respect to the L^p norm is then defined as

$$r_{pc}^{(p)} = \underset{\rho \in [-1,1]}{\arg\min} ||\mathbf{\Phi}_\rho - \mathbf{p}||_p \,. \tag{2}$$

The L^2 norm, in particular, is interpretable as the Euclidean distance and its minimum, $r_{pc}^{(2)}$, corresponds to the method of least squares.

Martinson & Hamdan (1971) suggested the multiplicative loss function

$$r_{pc}^{(MH)} = \underset{\rho \in [-1,1]}{\arg\min} \; -\prod_{k=1}^{rs} (\Phi_\rho(A_k))^{p_k} \,, \tag{3}$$

based on a likelihood argument. Olsson (1979) suggested a similar approach, but one which allows for reclassifications. While the loss function of Equation (3) is not non-negative, it is bounded from below and continuous. However, values of the loss function of (3) are not geometrically interpretable. For more loss function suggestions see, e.g., Martinson & Hamdan (1971).

If the global minimum of a loss function is not unique, then from a view of presuming statistical independence until evidence of association is found the global minimum with least absolute value should be chosen as fitted polychoric correlation coefficient. If two distinct global minima share the least absolute value $|a|$, say, then the fitted polychoric correlation coefficient can be expressed as $\pm a$.

The loss functions of Equations (2) and (3) are both differentiable, bounded from below, and since the bivariate normal distribution is differentiable in the parameter the minimum can be found by method of differential calculus.

3. A generalized polychoric correlation coefficient

The aim of this section is to define a generalized polychoric correlation coefficient analogous to Pearson's definition, hence based on Equation (1), that allows for other distributional assumptions than the joint normal distribution. First, a few aspects need to be considered.

3.1. **Preliminaries.** The family of bivariate standard normal distributions is parameterized by the linear correlation coefficient ρ. But for other families of bivariate distributions, the parameter may not be a measure of association by and of itself. Furthermore, other families of bivariate distributions may not have a linear dependency structure, something which compromises the suitability of the linear correlation. The Spearman grade correlation is a measure of association which can measure all types of monotonic association.

The Spearman grade correlation is the linear correlation of the grades of the two random variables, i.e. the percentiles of the two random variables. More precisely, for random variables U and V with marginal distribution functions F and G, the Spearman grade correlation ρ_S is the linear correlation between $F(U)$ and $G(V)$. The Spearman grade correlation is the population analog of Spearman's rank correlation coefficient, which was proposed by Charles Spearman (1904). The latter fact is also the reason for the grade correlation's name. However, according to Pearson (1907) the idea of correlation of percentiles was first introduced by Francis Galton. The Spearman grade correlation has good properties in the sense that it satisfies Scarsini's axioms for measures of concordance (Scarsini, 1984). For example, it is invariant under strictly increasing transformations of each of the two random variables U and V.

In the present setup, copulas are most fitting. A copula is a function that couples a joint probability distribution function to its standard uniform marginal distribution functions. More formally, a copula $C : I^2 \to I$ is a function which has non-negative volume for all rectangles, and satisfies boundary conditions $C(u, 0) = C(0, v) = 0$ and

$C(u,1) = u$, $C(1,v) = v$. The fact that a copula couples a joint probability distribution to its standard uniform marginal distribution functions is the statement of Sklar's theorem (see e.g. Nelsen, 2006). Because the Spearman grade correlation is invariant under strictly increasing transformations of the random variables, it can be expressed as a function of their copula. More precisely, if the two random variables are continuous and have copula C, then the Spearman grade correlation can be expressed as

$$\rho_S(C) = 12 \int_{I^2} C d\lambda - 3. \tag{4}$$

Among all copulas, three especially noteworthy ones are $W(u,v) = \max(u+v-1, 0)$, $\Pi(u,v) = uv$, and $M(u,v) = \min(u,v)$. These copulas correspond to perfect negative association, independence, and perfect positive association between the two random variables, respectively, and are called the minimum, the product, and the maximum copula. For all $(u,v) \in I^2$, it holds that $W(u,v) \leq \Pi(u,v) \leq M(u,v)$, i.e. the inequalities hold everywhere. The relation $C_1 \leq C_2$ everywhere implies an ordering between the two copulas C_1 and C_2. Many families of copulas are totally ordered, i.e. for any two members C_1 and C_2 it either holds that $C_1 \leq C_2$ everywhere or $C_1 \geq C_2$ everywhere. Moreover, if it for every $\alpha \leq \beta$ holds that $C_\alpha \leq C_\beta$ everywhere or $C_\alpha \geq C_\beta$ everywhere, then the family of copulas is called totally ordered and directed.

For a continuous univariate distribution function F, the inverse may not exist. However, since the function is continuous the preimage of each one-point set in the range I is a non-empty closed set in the domain. And with the equivalence relation $x \sim y$ if $F(x) = F(y)$, the preimages constitute equivalence classes. By choosing an element from each equivalence class, a function that is sufficiently similar to an inverse, for the purposes of this article, can be constructed. Therefore, let the quasi-inverse of a continuous univariate distribution function be defined by $F^{(-1)}(0) = \max F^{-1}(\{0\})$, and $F^{(-1)}(y) = \min F^{-1}(\{y\})$ for $y > 0$. The fact that all preimages has a minimum and maximum element follows since each one-point set is closed and F is continuous. For every continuous random variable U with distribution function F, $F^{(-1)}F(U)$ is defined and equals U with probability one.

3.2. **The generalized definition.** The generalized polychoric correlation coefficient is defined analogously to Pearson's definition, but with the exception that the assumption of a joint underlying normal distribution is relaxed. Instead of being a member of the family of normal distributions, the postulated underlying distribution is allowed to be a member of any family of continuous bivariate distributions. The assumption is formalized as follows.

Assumption A2. *The two ordinal variables are, into r and s ordered categories respectively, discretized random variables with a joint distribution belonging to the family of continuous bivariate distributions $\{H_\theta\}_{\theta \in \Theta}$.*

Appropriately, Assumption A2 reduces to Assumption A1 if the family of bivariate normal distributions is assumed. That the underlying joint distribution is assumed to be continuous is mostly a matter of convenience. Any distribution function can be arbitrarily well approximated by a continuous distribution function. But also, the idea of an underlying continuous distribution was central to Karl Pearson's view of the measure of association for ordinal variables.

Let F and G denote the marginal distribution functions of the joint distribution H. Like in Section 2.2, rectangles $[F^{(-1)}(u_{i-1}), F^{(-1)}(u_i)] \times [G^{(-1)}(v_{j-1}), G^{(-1)}(v_j)]$ are created for all $i = 1, \ldots, r$ and $j = 1, \ldots, s$, enumerated in the same way as the joint probabilities $p_1, \ldots, p_{rs}$, and denoted $A_1, \ldots, A_{rs}$.

For a general bivariate distribution function H, the probability of a such distributed random vector being an element of the rectangle $A = [a, b] \times [c, d]$ is equal to $H(a, c) - H(a, d) - H(b, c) + H(b, d)$. That probability is called the volume of the rectangle A, and is here denoted $H(A)$. In analogy with the reasoning in Section 2.2, it should under Assumption A2 ideally hold that the joint probabilities of the discretized bivariate distribution equal the joint probabilities of the ordinal variables. Hence it should hold that

$$(H_\theta(A_1), \ldots, H_\theta(A_{rs})) = (p_1, \ldots, p_{rs})\,. \tag{5}$$

Note that Equation (5) also appropriately reduces to Equation (1) if the family of bivariate standard normal distributions is assumed.

Since the solution θ of Equation (5) in general is not a measures of association, the value θ is by and of itself not of much interest. Instead the generalized polychoric correlation coefficient is based on the Spearman grade correlation of H_θ. And in order to make the generalized and the conventional definitions agree for the family of bivariate normal distributions, a result from Pearson (1907) must be utilized. Consequently, the generalized polychoric correlation coefficient is defined as

$$r_{pc} = 2\sin(\rho_S(H_\theta)\pi/6),$$

for the solution θ to Equation (5). If all cumulative marginal probabilities (u_i, v_j) are elements of the boundary of the unit square, ∂I^2, then the generalized polychoric correlation coefficient is defined to be zero, in agreement with the reasoning in Section 2.2.

By Proposition 2, a solution θ to Equation (5) does in general not exist. Then, again in analogy with Section 2.3, a parameter can be fitted with respect to some loss function. For example, let $\hat{\theta}^{(p)}$ be the parameter fitted with respect to the L^p distance, cf. Equation (2), then the such fitted polychoric correlation coefficient is defined as $r_{pc}^{(p)} = 2\sin(\rho_S(H_{\hat{\theta}^{(p)}})\pi/6)$.

3.3. **Agreement between the two definitions.** A generalization must agree with what is generalized wherever the latter is defined. In this subsection a generalized polychoric correlation coefficient is first shown to exist, and then shown to agree with Pearson's definition under a joint normal distribution assumption.

Proposition 3. *For any contingency table for which at least one point, (u_i, v_j), of the cumulative marginal probabilities is not an element of the boundary of the unit square, ∂I^2, and any family of continuous bivariate distributions $\{H_\theta\}_{\theta\in\Theta}$, a fitted polychoric correlation coefficient exists if and only if the family $\{H_\theta\}_{\theta\in\Theta}$ is non-empty.*

Proof. Assume that at least one point (u_i, v_j), for $i = 1, \ldots, r$ and $j = 1, \ldots, s$, is not an element of ∂I^2 and that a fitted polychoric correlation coefficient exists. Then, there is some best fit $\hat{\theta}$ to (5) with respect to some loss function. Thus, the family of continuous bivariate distributions $\{H_\theta\}_{\theta\in\Theta}$ has at least one element $H_{\hat{\theta}}$ and is therefore non-empty.

Conversely, assume that the family of continuous bivariate distributions $\{H_\theta\}_{\theta\in\Theta}$ is non-empty. Then, because all volumes and joint probabilities are finite, the loss function is defined for at least one parameter-value. Hence some global minimum of the loss function exists and therefore a fitted polychoric correlation coefficient exists. □

The following corollary is needed for Theorem 5.

Corollary 4. *Under a joint normal distribution assumption, a fitted polychoric correlation coefficient exists for every contingency table.*

By the next theorem, the generalized and the conventional definitions agree wherever the conventional definition is defined, and therefore the generalized polychoric correlation coefficient is a generalization in the true sense of the word.

Theorem 5. *Under a joint normal distribution assumption, the generalized and the conventional definitions of the polychoric correlation coefficient agree.*

Proof. Let temporarily the conventional polychoric correlation coefficient be denoted r_{cpc} and the generalized polychoric correlation coefficient be denoted r_{gpc}. If for all $i = 1, \ldots, r$ and $j = 1, \ldots, s$, the points (u_i, v_j) are elements of ∂I^2, then by both definitions the polychoric correlation coefficient is zero, so here the definitions clearly agree.

If at least one point (u_i, v_j) is not an element of ∂I^2, then the conventional polychoric correlation coefficient r_{cpc} is the best fit to (1) with respect to some loss function. For the generalized version under a normal distribution assumption, Equation (5) reduces to Equation (1). Thus, r_{cpc} is also the best fit to (5) with respect to the same loss function. But for the bivariate normal distribution it holds that $\rho_S(\Phi_\rho) = 6/\pi \arcsin(\rho/2)$ (Pearson, 1907). Therefore if follows after inserting the arcsine expression into the definition of the generalized polychoric correlation coefficient that $r_{gpc} = r_{cpc}$. So, by construction one may say, the definitions agree here as well. □

3.4. **The copula formulation.** If the marginal distribution functions of the postulated underlying continuous joint distribution are standardized to standard uniform via the quasi-inverses $F^{(-1)}$ and $G^{(-1)}$, the standardized joint distribution is a copula. The cost of the standardization is that all moments of the marginal distribution functions are lost, but since the Spearman grade correlation is invariant under strictly increasing

transformations it will not affect the measure of association. Therefore, for this purpose the standardization comes at no cost. The gains are substantial, though, since the need for computing quasi-inverses is eliminated.

By the following proposition, for continuous bivariate distributions the left-hand side of Equation (5) can be expressed as volumes of a copula. And since the inverse of the marginal distribution functions of a copula is the identity operator, the rectangles are written without quasi-inverses of the marginal distribution functions.

Proposition 6. *Suppose the bivariate distribution function H has continuous marginal distribution functions F and G, $0 \leq a \leq b \leq 1$ and $0 \leq c \leq d \leq 1$, then for each rectangle $A = [F^{(-1)}(a), F^{(-1)}(b)] \times [G^{(-1)}(c), G^{(-1)}(d)]$ there is a unique copula C such that*

$$H(A) = C(B),$$

where B is the rectangle $[a, b] \times [c, d]$.

Proof. Assume that H has continuous marginal distribution functions F and G. By Sklar's theorem, there exists a unique copula C such that $H(x, y) = C(F(x), G(y))$ for all $(x, y) \in I^2$. Because the quasi-inverses are well defined, change of variables yield $H(F^{(-1)}(u), G^{(-1)}(v)) = C(F^{(-1)}F(u), G^{(-1)}G(v)) = C(u, v)$. The statement then follows by the definition of the volume of a rectangle. □

Instead of considering the bivariate distribution $\{H_\theta\}$ that was assumed in Assumption A2, it is often more convenient to consider its corresponding family of copulas $\{C_\lambda\}$. In the copula framework, the rectangles $A_1, \ldots, A_{rs}$ are of the more simple form $[u_{i-1}, u_i] \times [v_{j-1}, v_j]$. With the change of notation, Equation (5) is written

$$(C_\lambda(A_1), \ldots, C_\lambda(A_{rs})) = (p_1, \ldots, p_{rs}),$$

and the polychoric correlation coefficient is written $r_{pc} = 2\sin(\rho_S(C_\lambda)\pi/6)$ for the solution λ to the equation above.

3.5. **Existence and uniqueness for 2×2 tables.** Originally, Pearson (1900) studied association for dichotomous variables, i.e 2×2 contingency tables. As mentioned, the polychoric correlation coefficient for this special case is sometimes called the tetrachoric correlation coefficient. For the generalized version, there are some existence and uniqueness results corresponding to Proposition 1, i.e. existence and uniqueness of a solution to Equation (5) for every contingency table, and hence existence of a unique polychoric correlation coefficient.

Because a 2×2 contingency table has four elements which sum to one, every such table is fully determined by the triple (u_1, v_1, p_1), where p_1 is the joint probability corresponding to the first categories of both dichotomous variables. Moreover, as a consequence of Kolmogorov's axioms the inequalities $\max(u_1 + v_1 - 1, 0) \leq p_1 \leq \min(u_1, v_1)$ hold. Throughout this subsection, results will be stated and proved for copulas. But as a consequence of Proposition 6, the results hold for all bivariate distributions that correspond

to a copula, which includes all families possible under Assumption A2. In this sense, the word *copula* can here be considered equivalent to *continuous bivariate distribution*.

The following is a necessary and sufficient condition for existence and uniqueness of a solution to Equation (5) for every contingency table, and hence for the polychoric correlation coefficient to be well defined.

Theorem 7. *The polychoric correlation coefficient is well defined for the family of copulas* $\{C_\lambda\}_{\lambda\in\Lambda}$ *if and only if for every contingency table* (u_1, v_1, p_1) *there exists a copula* $C \in \{C_\lambda\}$ *such that Equation (5) holds, and if* C_α *and* C_β *are two copulas such that Equation (5) holds then* $\rho_S(C_\alpha) = \rho_S(C_\beta)$.

Proof. Assume that for every contingency table (u_1, v_1, p_1) with $(u_1, v_1) \in \mathrm{Int}(I^2)$ there exists a copula $C \in \{C_\lambda\}_{\lambda\in\Lambda}$ such that (5) holds, and if C_α and C_β are two copulas such that (5) holds then $\rho_S(C_\alpha) = \rho_S(C_\beta)$. It will be shown that for any contingency table the set of polychoric correlation coefficients R is a one-point set, which implies that for every contingency table a unique polychoric correlation coefficient exists and, hence, is well defined.

Take any contingency table (u_1, v_1, p_1) and let $g : \Lambda \to [\max(u_1+v_1-1, 0), \min(u_1, v_1)]$ be the function defined by $g(\lambda) = C_\lambda(u_1, v_1)$. Furthermore, let $h : \Lambda \to [-1, 1]$ be defined $h(\lambda) = \rho_S(C_\lambda)$ and let $f : [-1, 1] \to [-1, 1]$ be defined $f(x) = 2\sin(x\pi/6)$. Then $r_{pc} \in R = f \circ h \circ g^{-1}(\{p_1\})$. By assumption, $g^{-1}(\{p_1\})$ is non-empty and $h \circ g^{-1}(\{p_1\})$ is a one-point set. And since f is a homeomorphism, $f \circ h \circ g^{-1}(\{p_a\}) = R$ is also a one-point set. Thus, the polychoric correlation coefficient exists and is unique. For a contingency table with marginal probabilities $(u_1, v_1) \in \partial I^2$, the polychoric correlation coefficient is identically zero, so in this case a unique coefficient always exists. Hence, for every contingency table the polychoric correlation coefficient exists and is unique, and thus it is well defined.

Conversely, assume that the polychoric correlation coefficient is well defined for the parametric family of copulas $\{C_\lambda\}_{\lambda\in\Lambda}$. Then for every contingency table, R is a one-point set. Take any contingency table (u_1, v_1, p_1) with $(u_1, v_1) \in \mathrm{Int}(I^2)$, and consider the construction $R = f \circ h \circ g^{-1}(\{p_1\})$. Since f is a homeomorphism, $f^{-1}(R)$ is a one-point set. By construction $f^{-1}(R) = h \circ g^{-1}(\{p_1\})$. Thus, $g^{-1}(\{p_1\})$ is non-empty, so for every contingency table there exists a copula $C \in \{C_\lambda\}$ such that Equation (5) holds. Also, since $h \circ g^{-1}(\{p_1\})$ is a one-point set it must hold that if C_α and C_β are two copulas such that Equation (5) holds then $\rho_S(C_\alpha) = \rho_S(C_\beta)$. □

The following result is a sufficient condition which is convenient in many situations. A family of copulas $\{C_\lambda\}_{\lambda\in\Lambda}$ is defined to be strictly ordered and directed if for every $\alpha < \beta$ it holds that $C_\alpha < C_\beta$ or $C_\alpha > C_\beta$ everywhere on $\mathrm{Int}(I^2)$.

Theorem 8. *If the assumed family of copulas has limits* W *and* M*, is continuous in the parameter, and is strictly ordered and directed, then the polychoric correlation coefficient is well defined for all* 2×2 *contingency tables.*

Proof. Because $\{C_\lambda\}_{\lambda\in\Lambda}$ has limits W and M, for any $(u_1, v_1) \in I^2$ there are parameters λ_1 and λ_2 in $\bar{\Lambda}$ such that $C_{\lambda_1}(u_1, v_1) = \max(u_1+v_1-1, 0)$ and $C_{\lambda_2}(u_1, v_1) = \min(u_1, v_1)$. Since $\{C_\lambda\}$ is continuous in λ, by the intermediate value theorem there exists a copula C_λ such that $C_\lambda(u_1, v_1) = p_1$ for every $p_1 \in [\max(u_1 + v_1 - 1, 0), \min(u_1, v_1)]$. Because $\{C_\lambda\}$ is strictly ordered and directed, the solution is unique. Thus, by Theorem 7, the polychoric correlation coefficient is well defined. □

Corollary 9. *Under a joint normal distribution assumption, the polychoric correlation coefficient is well defined for all* 2×2 *contingency tables.*

Many commonly used copula families, for instance the Clayton family, are totally ordered and directed but not strictly ordered and directed. For a family of copulas that has limits W and M, is continuous in the parameter, and is totally ordered but not strictly ordered, a polychoric correlation coefficient exists but is in general not unique. However, the set of polychoric correlation coefficients is a closed interval and, moreover, for any other contingency table with the same marginal probabilities, the polychoric correlation coefficient is not an element of that interval. Thus, the sets of polychoric correlation coefficients constitute equivalence classes.

By mapping each equivalence class to an element of the same class, the polychoric correlation coefficient is made well defined for the totally ordered and directed families of copulas. From the perspective of presuming statistical independence until evidence of association is found, it is natural to map the equivalence class to the element with least absolute value. The following theorem is key.

Theorem 10. *If the assumed family of copulas has limits* W *and* M*, is continuous in the parameter, and is totally ordered and directed, then for all* 2×2 *contingency tables the set of polychoric correlation coefficients is a non-empty closed interval. Moreover, the polychoric correlation coefficient for any other contingency table with the same marginal probabilities is not an element of that interval.*

Proof. For contingency tables with marginal probabilities $(u_1, v_1) \in \partial I^2$, the polychoric correlation coefficient is identically zero. So here the set of polychoric correlation coefficients is clearly a non-empty closed interval. Moreover, for such marginal probabilities it holds that $\max(u_1+v_1-1, 0) = \min(u_1, v_1)$, so there exists no other contingency table with the same marginal probabilities.

Take any contingency table (u_1, v_1, p_1) with $(u_1, v_1) \in \mathrm{Int}(I^2)$. Assume that the parametric family of copulas $\{C_\lambda\}_{\lambda\in\Lambda}$ is continuous in λ, totally ordered and directed, and has limits W and M. Since $\{C_\lambda\}$ is continuous in λ and has limits W and M, by the intermediate value theorem there exists a solution to (5). Thus, the set of polychoric correlation coefficients R is non-empty.

To show that R is a closed interval, let $g : \Lambda \to [\max(u_1 + v_1 - 1, 0), \min(u_1, v_1)]$ be the function defined by $g(\lambda) = C_\lambda(u_1, v_1)$, let $h : \Lambda \to [-1, 1]$ be defined $h(\lambda) = \rho_S(C_\lambda)$ and let $f : [-1, 1] \to [-1, 1]$ be defined $f(x) = 2\sin(x\pi/6)$. Then $R = f \circ h \circ g^{-1}(\{p_1\})$. Because $\{C_\lambda\}$ is continuous in λ and totally ordered and directed, g is continuous and

monotonic. Since the one-point set $\{p_1\} \subset \mathbb{R}$ is closed, the preimage $g^{-1}(\{p_1\})$ is closed and since $\{p_1\}$ is connected and g is monotonic, $g^{-1}(\{p_1\})$ is connected. It is clear from the relation $\alpha, \beta \in \Lambda$, $\alpha \geq \beta$, in the definition of a totally ordered and directed copula family that Λ is a subset of $\mathbb{R}$. Thus, $g^{-1}(\{p_1\})$ is a closed interval. Because C_λ is continuous and monotonic in λ and $\int_{I^2} C_\lambda d\lambda$ is continuous and monotonic in C_λ, h is continuous and monotonic. And since f is a homeomorphism, the set of polychoric correlation coefficients R is also a closed, non-empty interval.

To complete the proof, suppose that $(u_1, v_1, \tilde{p}_1)$ is another contingency table with the same marginal probabilities, i.e. $\tilde{p}_1 \neq p_1$. By the fact that Equation (5) reduces to $C(u_1, v_1) = p_1$ for 2×2 contingency tables, it follows that $C(u_1, v_1) \neq \tilde{C}(u_1, v_1)$, where $\tilde{C}$ is the copula that solves Equation (5) for the contingency table $(u_1, v_1, \tilde{p}_1)$. Because the copula family is totally ordered, and all copulas are continuous, $\int_{I^2} C d\lambda \neq \int_{I^2} \tilde{C} d\lambda$, and hence the sets $h \circ g^{-1}(\{\tilde{p}_1\})$ and $h \circ g^{-1}(\{p_1\})$ do not meet. And because f is a homeomorphism, the polychoric correlation coefficient for $(u_1, v_1, \tilde{p}_1)$ is not contained in $R = f \circ h \circ g^{-1}(\{p_1\})$. □

Under the hypotheses of Theorem 10, the polychoric correlation coefficient is chosen to be the unique element of the closed interval with least absolute value.

4. Goodness-of-fit

As is stated in Proposition 2, a solution to Equation (5) may not exist. While this for many purposes should be considered a downside, it opens up the possibility of comparing different distributional assumptions based on goodness of fit, with respect to some loss function. This provides the possibility of testing different distributional assumptions and use the results for the enhancement of the association analysis.

While a comprehensive study of goodness-of-fit tests for ordinal variables is beyond the scope of this article, examples in Section 5 will illustrate how an analysis of goodness of fit can enhance the association analysis. Under the null hypothesis, as stated in Assumption A2, an acceptance region for a goodness-of-fit test statistic can be created. The acceptance region should contain the value of the test statistic under a perfect fit and the test should have as high statistical power as possible, given some fixed type-I error probability α. The exact type-I error probability is in practice often difficult to compute, and is therefore often approximated either by means of an asymptotical approximation or by simulation.

The Pearson chi-square test statistic,

$$Q = n \sum_{k=1}^{rs} \frac{\left(H_{\hat{\theta}}(A_k) - p_k\right)^2}{H_{\hat{\theta}}(A_k)}, \tag{6}$$

where n is the sample size, is asymptotically χ^2_{rs-1-k}-distributed, where k is the number of parameters estimated, under the null hypothesis as stated in Assumption A2. However, the approximation is often rather poor for small sample sizes, in the sense that the actual type-I error probability is appreciably higher than intended. Making matters

worse, the approximation error also depends on the vector $\mathbf{H}_{\hat{\theta}}$. One textbook rule of thumb for an acceptable approximation is that the sample size, n, should be greater than $10/\min\mathbf{H}_{\hat{\theta}}$.

By method of simulation, acceptance regions can be created such that the difference between intended and actual type-I error probabilities can be made arbitrarily small. In the present setup, the computational demands of simulation of acceptance regions are modest. Therefore, approximation of acceptance regions by simulation is generally to recommend since it yields better control of the actual type-I error probability.

While the Pearson chi-square test statistic is asymptotically χ^2-distributed, it weighs rectangles by the inverse of the probability under the null hypothesis. Consequently, rectangles with near zero probability under the null hypothesis are given near infinite weight. In practice, therefore, the test commonly rejects or accepts the null hypothesis exclusively on the basis of whether the observed joint probabilities of the ordinal variables corresponding to such rectangles is zero or non-zero. As an effect, the statistical power is often low.

In the examples of Section 5, the L^2-norm has been chosen as test statistic (cf. Equation (2)),

$$T = ||\mathbf{H}_{\hat{\theta}} - \mathbf{p}||_2.$$

The L^2-norm is similar to Pearson's chi-square test statistic with the difference that all rectangles are weighted equally. While the L^2-norm may not have the best possible statistical power, it apparently does have some power based on the examples of Section 5. Because it is a norm, the test statistic is zero if and only if there is a perfect fit, and as a distance it suits the intuitive notion of a test statistic based on a loss function. The acceptance region is the interval $[0, c]$, with the critical value c, for a given sample size, type-I error probability, and $\mathbf{H}_{\hat{\theta}}$ vector, being found by simulation under the null hypothesis. The simulated critical value, c, converges to the true critical value with probability one as the simulation size goes to infinity by the strong law of large numbers.

5. Examples

In this section, examples are used to illustrate how the distributional assumption can impact the conclusion of the association analysis. Moreover, goodness-of-fit p-values are used to enhance the association analysis with an analysis of the dependency structure of the ordinal variables' postulated underlying joint distribution.

5.1. **Distributional assumptions.** All families of continuous bivariate distributions used in this section are copula families. Listed in Table 4 are six copula families that have limits W and M, are continuous in the parameter, and are ordered and directed. The copula families are quite common and have a mix of properties suitable for the examples of this section.

The family of Gaussian copulas is the family of copulas corresponding to bivariate normal distributions. It is strictly ordered, and has a monotonic dependency structure,

i.e. the association is similar conditioning on any value of either variable. The family of Student copulas is based on the multivariate t-distribution and is also strictly ordered. The Student copula has a symmetric tail dependency structure, i.e. the association is stronger in the tails than in the center of each marginal distribution function. The family of Frank copulas is a strictly ordered copula family which has a monotonic dependency structure.

The family of Clayton copulas is a totally ordered copula family. The Clayton copula is asymmetric and has a left tail dependency structure, i.e. the association is strongest in the left tails of the marginal distribution functions. The family of copulas denoted Nelsen-(2) is the copula denoted (4.2.2) in Nelsen (2006), though there is neither a name for this copula family nor any reference on it, therefore the notation. The Nelsen-(2) copula is asymmetric and has a right tail dependency structure. Notably, the product copula, Π, which implies independence of the random variables, is not a member of this family. Lastly, the Genest-Ghoudi family of copulas is also totally ordered, and has an asymmetric right-tail dependency structure.

5.2. **Examples of 2×2 contingency tables.** In Table 2, polychoric correlation coefficients have been calculated for eight 2×2 contingency tables. The contingency tables have been chosen so as to illustrate differences in the coefficient caused by changes of distributional assumptions. Both contingency tables in the first row of Table 2 have marginal probabilities $(0.5, 0.5)$. The right contingency table has a solution M, which is a member of all four copula families, thus the polychoric correlation coefficients represent perfect positive association under all four distributional assumptions. The left contingency table has a solution Π, but Π is not a member of the Nelsen-(2) copula family. So for this contingency table the polychoric correlation coefficients differ. By Theorem 10, there is a solution for the Nelsen-(2) copula family, and that solution gives a slightly negative polychoric correlation coefficient. However, for this row of contingency tables, the choice of distributional assumption seems to have a modest impact on the polychoric correlation coefficient.

In the second row of Table 2, there are two contingency tables with marginal probabilities $(0.8, 0.8)$ and $(0.8, 0.2)$, respectively, that both have Π as a solution. Since Π is not a member of the Nelsen-(2) copula family, this copula family has another solution. Noteworthy here is the fact that, because this copula family has an asymmetric dependency structure, the polychoric correlation coefficients have different absolute values. Under the Nelsen-(2) distributional assumption, the polychoric correlation coefficient of the left contingency table represents strong positive association while the polychoric correlation coefficient of the right contingency table represents near independence.

In the third row of Table 2, the two contingency tables also have marginal probabilities $(0.8, 0.8)$ and $(0.8, 0.2)$, respectively. In this row, however, both contingency tables have a zero element. Under the monotonically dependent Gaussian and Frank distributional assumptions, these contingency tables have solutions W and M respectively, implying

TABLE 2. Contingency tables with generalized polychoric correlation coefficients under four distributional assumptions.

0.25	0.25	Gaussian:	0.00	0.5	0
		Frank:	0.00		
0.25	0.25	Clayton:	0.00	0	0.5
		Nelsen-(2):	−0.07		

Gaussian:	1.00
Frank:	1.00
Clayton:	1.00
Nelsen-(2):	1.00

0.64	0.16	Gaussian:	0.00	0.16	0.64
		Frank:	0.00		
0.16	0.04	Clayton:	0.00	0.04	0.16
		Nelsen-(2):	0.70		

Gaussian:	0.00
Frank:	0.00
Clayton:	0.00
Nelsen-(2):	−0.04

0.6	0.2	Gaussian:	−1.00	0.2	0.6
		Frank:	−1.00		
0.2	0	Clayton:	−0.41	0	0.2
		Nelsen-(2):	0.00		

Gaussian:	1.00
Frank:	1.00
Clayton:	0.88
Nelsen-(2):	0.87

0.08	0.12	Gaussian:	0.44	0.68	0.12
		Frank:	0.46		
0.12	0.68	Clayton:	0.65	0.12	0.08
		Nelsen-(2):	−0.30		

Gaussian:	0.44
Frank:	0.46
Clayton:	0.29
Nelsen-(2):	0.81

perfect association. But under the asymmetrically tail dependent Clayton and Nelsen-(2) distributional assumptions, the polychoric correlation coefficients do not represent perfect association. In fact, under the Nelsen-(2) distributional assumption, the left contingency table has polychoric correlation coefficient zero, representing independence. Under the Clayton distributional assumption, the polychoric correlation coefficient of the left contingency table represents weak negative association. In order to understand of the difference, recall that the Clayton copula family has a left tail dependency structure while the Nelsen-(2) copula family has a right tail dependency structure. The polychoric correlation coefficients of the right contingency table in the third row represent strong positive association under both Clayton and Nelsen-(2) distributional assumptions.

Lastly, in the fourth row of Table 2 the two contingency tables have marginal probabilities $(0.2, 0.2)$ and $(0.8, 0.8)$, respectively. Note here that the contingency tables are symmetric. Under the Gaussian and Frank distributional assumptions, these contingency tables have the same solutions, since the copula families have symmetric dependency structures. However, under the asymmetric Clayton and Nelsen-(2) distributional assumptions, the contingency tables have different solutions. Because the Clayton copula family has a left tail dependency structure and the Nelsen-(2) copula family has a

right tail dependency structure, the polychoric correlation coefficient of the left contingency table represents strong association under the Clayton distributional assumption, but weak association under the Nelsen-(2) distributional assumption. And for the right contingency table vice versa. Remarkable here is also the fact that the polychoric correlation coefficient of the left contingency table represents negative association under the Nelsen-(2) distributional assumption, but positive association under the Clayton distributional assumption.

The conclusion of the discussion of Table 2 is that the distributional assumption can have a profound impact on the polychoric correlation coefficient, and thus, that the polychoric correlation coefficient is not robust to changes of the distributional assumption. In fact, the polychoric correlation coefficient seems to be quite the opposite of robust. The polychoric correlation coefficient can represent perfect association or independence, positive or negative association, or anything in between, only as a consequence of a change of distributional assumption. Moreover, another conclusion is that large differences tend to occur for distributions that have different dependency structures. The Gaussian and Frank copula families both have a monotonic dependency structure, and under these two distributional assumptions differences tend to be small.

5.3. **Examples of $r \times s$ contingency tables.** In Table 3, a 5×5 contingency table with survey data is shown. Also in the table are p-values for the L^2-norm goodness-of-fit test described in Section 4. The survey was conducted on statistics students and they had to consider a number of statements and answer whether they agreed strongly, agreed, neither agreed nor disagreed, disagreed, or disagreed strongly. In Table 3 statement X is *I feel that I must perform well in statistics*, and statement Y is *I do not like mathematical formulae*. Note that statement Y is written with negation. Here, again, the Martinson-Hamdan loss function fails. For all distributional assumptions the polychoric correlation coefficients, fitted with the L^2-norm loss function, are slightly negative, suggesting that on average, students that do not like mathematical formulae are a bit more anxious about their performance. However, because of the following reasons, that conclusion is actually erroneous.

From the goodness-of-fit p-values, the bivariate normal distribution does not fit the contingency table at the 5% level, whereas the Clayton copula does. This suggests that the underlying joint distribution is asymmetrically left tail dependent. Thus, on average, students that disagree to either statement tend to agree to the other statement, while on average, students that agree to either statement relate to the other statement just as the average student. Assuming the bivariate normal distribution, the analyst will erroneously conclude that students that do not like mathematical formulae do not feel that they have to perform as well in statistics. Assuming the left tail dependent Clayton copula, on the other hand, the conclusion is that the joint distribution is weakly associated in its left tail, but nearly independent in its right tail. Hence, students that like mathematical formulae tend to, on average, feel that they must perform well in statistics, while for students that do not like mathematical formulae there is no particular

TABLE 3. Survey of undergraduate statistics students.

		Y				
		Disagree			Agree	
X	Disagree	0	7	0	0	3
		3	10	25	10	3
		18	84	80	47	7
		40	54	65	43	10
	Agree	43	29	29	14	10

Copula	$r_{pc}^{(MH)}$	p-value	$r_{pc}^{(2)}$	p-value
Gaussian	−.18	.03	−.23	.01
Frank	−.18	.01	−.16	.00
Clayton	−.16	.12	−.31	.14
Nelsen-(2)	.82	.00	.20	.00
Genest-G.	.39	.00	−.08	.00

Note. Contingency table with polychoric correlation coefficients and goodness-of-fit p-values. Statement X is *I feel that I must perform well in statistics* , and statement Y is *I do not like mathematical formulae*.

relationship. This example illustrates how the analyst gets additional information from the goodness-of-fit p-values.

6. Conclusions

The definition of the polychoric correlation coefficient has been generalized so that, in addition to the bivariate normal distribution, a large class of continuous bivariate distributions can be assumed as the underlying joint distribution. The generalized definition is analogous to Karl Pearson's original definition, and the two definitions agree under a joint normal distribution assumption.

The generalized definition makes it possible to evaluate the statistical robustness of the polychoric correlation coefficient. Pearson & Heron (1913) claimed that for the purpose of the polychoric correlation coefficient, divergence between the actual joint distribution and the normal distribution is hardly ever of practical importance, i.e. claiming ideal statistical robustness properties. However, in the article there are no details on how they were able to come to that conclusion. Examples in Section 5 point towards a different conclusion, that the polychoric correlation coefficient in fact has poor statistical robustness properties.

The lack of statistical robustness must be considered a weakness of the polychoric correlation coefficient. For ordinal variables it is without prior knowledge difficult make

an assertion about the specific distribution function of a postulated underlying distribution. A practical solution in this case is to do the association analysis under a number of distributional assumptions, and then choose a distribution based on an analysis of goodness of fit. For 2×2 contingency tables, however, a large class of families of bivariate distributions have guaranteed perfect fit, and therefore the goodness-of-fit circumvention does not work.

On the other hand, because of its assumption of an underlying joint distribution, the generalized polychoric correlation coefficient makes it possible to analyze the dependency structure of the postulated underlying distribution. An analysis of goodness of fit can provide valuable information on possible deviation from the monotonic dependency structure, such as for example symmetric or asymmetric tail dependence. All in all, the polychoric correlation coefficient has both strengths and weaknesses. Consequently, the discussion on the optimal measure of association for ordinal variables will continue.

Acknowledgements

This article was prepared during a visit to UCLA Department of Statistics, and the author is grateful for the generosity and hospitality of all department faculty and staff, and in particular Distinguished Professor Jan de Leeuw. The author is also grateful Katrin Kraus for the data of Table 3.

References

Camp, B. H. (1933). Karl Pearson and Mathematical Statistics. *J. Amer. Statist. Assoc.*, *28*, 395–401.

Ekström, J. (2008). On the relation between the phi-coefficient and the tetrachoric correlation coefficient. In *Contributions to the Theory of Measures of Association for Ordinal Variables*. Ph.D. thesis, Uppsala: Acta Universitatis Upsaliensis.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proc. Roy. Soc. London*, *45*, 135–145.

Martinson, E. O., & Hamdan, M. A. (1971). Maximum likelihood and some other asymptotically efficient estimators of correlation in two way contingency tables. *J. Stat. Comput. Simul.*, *1*, 45–54.

Nelsen, R. B. (2006). *An Introduction to Copulas, 2nd ed*. New York: Springer.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, *195*, 1–47.

Pearson, K. (1907). *Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation*, vol. 4 of *Drapers' Company Research Memoirs, Biometric series*. London: Cambridge University Press.

Pearson, K., & Heron, D. (1913). On theories of association. *Biometrika*, *9*, 159–315.

Quiroga, A. M. (1992). *Studies of the polychoric correlation and other correlation measures for ordinal variables*. Ph.D. thesis, Uppsala University.

Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biometrika*, *12*, 93–133.

Roscino, A., & Pollice, A. (2006). A generalization of the polychoric correlation coefficient. In *Data analysis, classification and the forward search*, (pp. 135–142). Classification and Data Analysis Group of the Italian Statistical Society, Berlin: Springer.

Scarsini, M. (1984). On measures of concordance. *Stochastica*, *8*, 201–218.

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

Tallis, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, *18*, 342–353.

UCLA Department of Statistics, 8125 Mathematical Sciences Building, Box 951554, Los Angeles CA, 90095-1554

E-mail address: joakim.ekstrom@stat.ucla.edu

TABLE 4. Some continuous, ordered and directed copula families with limits W and M.

Family	$C_\lambda(u,v)$	ordered	Λ	Limits and special cases
Gaussian	$\Phi_\lambda(\Phi^{-1}(u),\Phi^{-1}(v))$	strictly	$[-1,1]$	$C_{-1}=W$, $C_0=\Pi$, $C_1=M$
Student	$T_\lambda(t_\nu^{-1}(u),t_\nu^{-1}(v),\nu)$	strictly	$[-1,1]$	$C_{-1}=W$, $C_0=\Pi$, $C_1=M$
Frank	$-\frac{1}{\lambda}\ln\left(1+\frac{(\exp\{-\lambda u\}-1)(\exp\{-\lambda v\}-1)}{(\exp\{-\lambda\}-1)}\right)$	strictly	$\mathbf{R}-\{0\}$	$C_{-\infty}=W$, $C_0=\Pi$, $C_\infty=M$
Clayton	$\left[\max\left(u^{-\lambda}+v^{-\lambda}-1,0\right)\right]^{-1/\lambda}$	totally	$[-1,\infty)-\{0\}$	$C_{-1}=W$, $C_0=\Pi$, $C_\infty=M$
Nelsen-(2)	$\max\left(1-\left((1-u)^\lambda+(1-v)^\lambda\right)^{1/\lambda},0\right)$	totally	$[1,\infty)$	$C_1=W$, $C_\infty=M$
Genest-Ghoudi	$\left[\max\left(1-\left[(1-u^{1/\lambda})^\lambda+(1-v^{1/\lambda})^\lambda\right]^{1/\lambda},0\right)\right]^\lambda$	totally	$[1,\infty)$	$C_1=W$, $C_\infty=M$