Title
An Empirical Polychoric Correlation Coefficient

Author
Ekström, Joakim

Publication Date
2011-10-25

Peer reviewed

AN EMPIRICAL POLYCHORIC CORRELATION COEFFICIENT

JOAKIM EKSTRÖM

Abstract. A new measure of association for ordinal variables is proposed. The new measure of association, named the empirical polychoric correlation coefficient, builds upon the theoretical framework of the polychoric correlation coefficient, but relaxes its fundamental assumption so that an underlying continuous joint distribution is only assumed to exist, not to be of any specific distributional family. The empirical polychoric correlation has good properties in terms of statistical robustness and asymptotics, and is easy to compute by hand. Moreover, a simulation study indicates that the new measure of association is more stable, in terms of standard deviation, than conventional polychoric correlation coefficients.

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

Financial support from the Jan Wallander and Tom Hedelius Research Foundation, project P2008-0102:1, is gratefully acknowledged.

1. Introduction

The polychoric correlation coefficient is a measure of association for ordinal variables. Originally proposed by Karl Pearson (1900), the measure of association rests upon an assumption of an underlying continuous joint distribution. Consequently, the contingency table of the two ordinal variables is assumed to be the result of a double discretization of the continuous joint distribution. Under a joint normal distribution assumption, the case studied in Pearson (1900), the polychoric correlation coefficient corresponds to 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.

Besides the difficulties in its computation, which have been substantially reduced with the availability of modern computers, the polychoric correlation coefficient has by some been regarded with scepticism because of its elaborate and restrictive fundamental assumptions, notably by Yule (1912). Of course not all bivariate distributions are normal, but 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."

It is not clear how Pearson & Heron (1913) arrived at this conclusion, but a generalization of the definition to all families of continuous bivariate distributions (Ekström, 2008), and subsequent analysis of statistical robustness properties, provides ample evidence that the measure of association is not robust to changes of the underlying distributional assumption. In fact, the conclusions of the association analysis can change profoundly as a consequence of a change of the distributional assumption. Moreover, the polychoric correlation coefficient in general needs to be fitted to a contingency table with respect to a loss function, and the coefficient is not robust to a change of loss function either.

The present article proposes a new measure of association for ordinal variables, named the empirical polychoric correlation coefficient, which is designed to enhance statistical robustness. By utilizing the joint empirical distribution of the ordinal variables, it builds upon the theoretical construction of the polychoric correlation coefficient while at the same time removing the need to specify the underlying continuous joint distribution up to its distributional family.

Theoretical results on the new measure of association is provided; that it is well defined, takes values between -1 and 1, and that it converges almost surely to the theoretical population value of the polychoric correlation in a certain sense. As a consequence of the latter, in particular, the empirical polychoric correlation coefficient has better theoretical properties than the polychoric correlation coefficient. The rate of convergence is studied in a simulation study, and the results indicate that the empirical polychoric correlation coefficient is better than the polychoric correlation coefficient in terms of statistical robustness and is, in terms of standard deviation, more stable. Another advantage is, moreover, that the empirical polychoric correlation coefficient requires neither fitting nor optimization. In fact it is easily computed by hand, thereby mitigating the difficulties in the polychoric correlation coefficient's computation.

In Section 2, the conventional polychoric correlation coefficient is presented with its implicit assumptions. In Section 3, the empirical polychoric correlation coefficient is defined and properties are derived. Section 4 contains the results of a simulation study and a discussion of the fixed sample size properties of the empirical polychoric correlation coefficient in comparison to conventional polychoric correlation coefficients computed under five distributional assumptions. And lastly, the article is concluded with Section 5.

2. The polychoric correlation coefficient

The fundamental idea of the polychoric correlation coefficient is to assume that the two ordinal variables are, into r and s ordered categories respectively, discretized random variables with a continuous joint distribution belonging to some family of bivariate distributions. The discretization cuts the domain of the bivariate density function into rectangles corresponding to the cells of the contingency table, see Figure 1 for an illustration. Ideally, the hypothesized probability masses of the rectangles, i.e. the volumes of the rectangles, should equal the corresponding joint probabilities of the two ordinal variables. The fundamental assumption is formalized below.

Assumption A1. *The two ordinal variables are, into r and s ordered categories respectively, discretized random variables with a continuous joint distribution belonging to the family of bivariate distributions $\{H_\theta\}_{\theta\in\Theta}$.*

Pearson (1900) studied the case assuming a bivariate standard normal distribution. Under a standard normal distribution assumption, the polychoric correlation coefficient is precisely the parameter value for which the volumes of the rectangles equal the joint probabilities of the two ordinal variables. For the bivariate standard normal distribution, that parameter also corresponds to the linear correlation of the two such distributed random variables.

For a bivariate probability distribution H and a rectangle $A = [a, b] \times [c, d]$, the volume of the rectangle equals $H(A) = H(b, d) - H(b, c) - H(a, d) + H(a, c)$. If the distribution function is absolutely continuous, i.e. has a density function, then the volume $H(A)$ equals the integral of the density function over the rectangle A. This illustrates the fact that if Z is a bivariate random variable with distribution function H, then $P(Z \in A) = H(A)$.

Let F and G be the marginal distribution functions of H and denote by $F^{(-1)}$ the quasi-inverse of a continuous distribution function, $F^{(-1)}(0) = \max\{x : F(x) = 0\}$, and $F^{(-1)}(y) = \min\{x : F(x) = y\}$ for $y > 0$. If the set $\{x : F(x) = 0\}$ is empty then $F^{(-1)}(0)$ is set to $-\infty$, and if $\{x : F(x) = 1\}$ is empty then $F^{(-1)}(1)$ is set

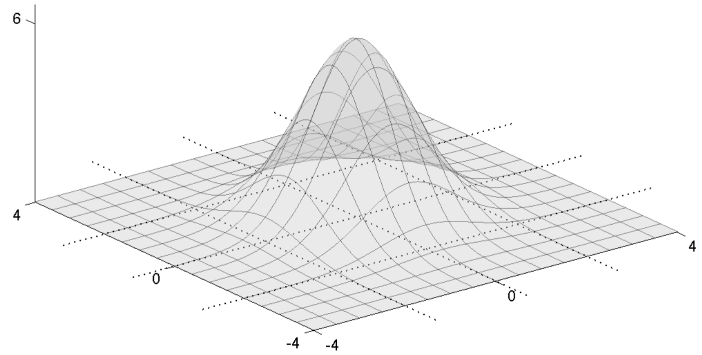

FIGURE 1. Illustration of the domain of the standard normal density function being discretized by the dotted lines into a 4×4 contingency table.

to ∞. Furthermore, let $u_0, \ldots, u_r$ and $v_0, \ldots, v_s$ be the cumulative marginal probabilities of the two ordinal variables, respectively, i.e. u_0 and v_0 are zero and u_i and v_j are the probabilities of observing values with order less than or equal to i and j, respectively, of each ordinal variable. For all $i = 1, \ldots, r$ and $j = 1, \ldots, s$, create rectangles $[F^{(-1)}(u_{i-1}), F^{(-1)}(u_i)] \times [G^{(-1)}(v_{j-1}), G^{(-1)}(v_j)]$, enumerate them and denote them $A_1, \ldots, A_{rs}$. The rectangles $A_1, \ldots, A_{rs}$ are interpreted as the result of the discretization of the domain of the bivariate distribution function, cf. Figure 1. Moreover, let $p_1, \ldots, p_{rs}$ denote the joint probabilities of the ordinal variables corresponding to rectangles $A_1, \ldots, A_{rs}$, respectively.

Under Assumption A1, it should, ideally, hold that the volumes of the rectangles equal the joint probabilities of the two ordinal variables. Hence it should hold that

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

For the solution θ to the above equation, the polychoric correlation coefficient is defined as

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

where ρ_S denotes the Spearman grade correlation. If all points of cumulative marginal probabilities (u_i, v_j) are elements of the boundary of the unit square, ∂I^2, then any parameter θ will satisfy Equation (1), but in this case the polychoric correlation coefficient is defined to be zero, in part because of a reasoning of presuming independence until evidence of association is found. Moreover, note that if the family of bivariate standard normal distributions, with correlation parameter ρ, is assumed then the function $\rho \mapsto 2\mathrm{sin}(\rho_S(\Phi_\rho)\pi/6)$ equals identity, see Pearson (1907), and thus the definition above agrees with the definition in Pearson (1900). A detailed proof can be found in Ekström (2008).

If the numbers of categories, r and s, both equal 2 then a unique polychoric correlation coefficient exists for each contingency table under some general conditions on the bivariate family. If one of r and s is greater than 2 and the other is greater than or equal to 2, then a solution to Equation (1) does in general not exist. Given Assumption A1, a solution may not exist due to, e.g., fixed sample sizes and/or noisy observations. In that case it is standard statistical procedure to look for a best fit of the parameter θ with respect to some loss function. Any distance function between the two vectors of Equation (1) works, and the usual ℓ^p-norm, $||x||_p = (\sum |x_i|^p)^{1/p}$, is in many ways natural.

Denote the left hand side of Equation (1) by $\vec{H}_\theta$ and the right hand side by $\vec{p}$. The polychoric correlation coefficient fitted with respect to the ℓ^p-norm is $r_{pc}^{(p)} = 2\sin(\rho_S(H_{\hat{\theta}^{(p)}})\pi/6)$, where $\hat{\theta}^{(p)}$ is given by

$$\hat{\theta}^{(p)} = \arg\min_{\theta \in \Theta} ||\vec{H}_\theta - \vec{p}||_p. \tag{2}$$

The loss function is zero if and only if θ is a solution to Equation (1). The ℓ^2-norm, in particular, can be interpreted as the Euclidean distance and the minimum also corresponds to the method of least squares.

Martinson & Hamdan (1971) suggested

$$\hat{\theta}^{(MH)} = \arg\min_{\theta \in \Theta} - \prod_{k=1}^{rs} (H_\theta(A_k))^{p_k}\,, \tag{3}$$

based on a likelihood argument. The Martinson-Hamdan loss function does not necessarily have a unique global minimum, but it is bounded from below and continuous. However, examples in Ekström (2008) show that the Martinson-Hamdan loss function is inherently unstable and fails under a variety of circumstances. For additional loss function suggestions, see, e.g., Martinson & Hamdan (1971). The minimum of a loss function can be found by method of numerical optimization.

It is both mathematically and practically convenient to use the copula corresponding to a bivariate distribution instead of the bivariate distribution function itself. As a consequence of Sklar's theorem, for each continuous bivariate distribution function H with (continuous) marginal distribution functions F and G and each rectangle of the form $A = [F^{(-1)}(a), F^{(-1)}(b)] \times [G^{(-1)}(c), G^{(-1)}(d)]$, there exists a unique copula C such that $H(A) = C(B)$, where $B = [a, b] \times [c, d]$. Hence, the use of copulas eliminates the need for computing quasi-inverses of the marginal distribution functions. In the standardization to copulas all moments of the marginal distributions are lost, but since the Spearman grade correlation is invariant under strictly increasing transformations of the marginal distribution functions it comes at no cost in this setting.

If X and Y are continuous random variables with copula C, then the Spearman grade correlation can be expressed as

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

where I is the unit interval, $[0, 1]$, and λ is the Lebesgue measure (Nelsen, 2006). Consequently, the polychoric correlation coefficient can be expressed as a function of the copula, C, something which will be exploited in the following section.

3. An empirical version

The aim of this section is to develop a new measure of association for ordinal variables, named the empirical polychoric correlation coefficient and denoted r_{epc}, which is based on the empirical distribution of the contingency table. The need for the new measure of association arises out of the following set of problems.

Because of Assumption A1, it is for the polychoric correlation coefficient necessary to specify a specific family of distributions, $\{H_\theta\}_{\theta\in\Theta}$. In most applications, however, there is no natural candidate for a choice of distributional family. A distributional family could be chosen by random, or by habit, but an unsubstantiated assumption would open up to criticism, and rightfully so because the polychoric correlation coefficient is not statistically robust to changes of the distributional assumption. The analysis could be performed under a number of distributional assumptions, and a distribution then chosen based on performance in a goodness-of-fit test. However, it could be that no tested distribution fits the contingency table. Moreover, in general a loss function must be chosen for the fitting process, and the polychoric correlation coefficient is not robust to changes of loss function either.

Pearson's idea of an underlying distribution is in many ways attractive, but Assumption A1 is too restrictive in that the underlying distribution must be specified up to its distributional family. With the empirical polychoric correlation coefficient, on the other hand, the fundamental assumption is relaxed so that the distributional family does not need to be specified; it only needs to be assumed that an underlying continuous distribution exists. In the statistical sense, therefore, the empirical polychoric correlation coefficient is non-parametric. 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 continuous joint distribution.*

The assumption of a continuous distribution function is mostly a matter of convenience. Every distribution function can be arbitrarily well approximated by a continuous distribution function. But also, the idea of continuous underlying marginal distributions was central to Karl Pearson's idea for the measure of association. Because Assumption A2 is necessary but not sufficient for Assumption A1, Assumption A2 is strictly weaker than Assumption A1. Therefore, the empirical polychoric correlation coefficient is theoretically more statistically robust than the polychoric correlation coefficient.

3.1. Construction and definition.

Let the two ordinal variables be denoted X and Y, respectively, and assume Assumption A2. Moreover, let the theoretical non-discretized random variables be denoted $\tilde{X}$ and $\tilde{Y}$. By Sklar's theorem, $\tilde{X}$ and $\tilde{Y}$ have a copula which is denoted C. For convenience, it is assumed throughout this section that $\tilde{X}$ and

$\tilde{Y}$ have standard uniform marginal distributions, i.e. that C is the bivariate distribution function of $\tilde{X}$ and $\tilde{Y}$. Consequently, X and Y are without loss of generality assumed to be discretized numerical random variables with support on the unit interval. The aim is to find an estimate of C, and use it to define the empirical polychoric correlation coefficient.

Let $u_0, \dots, u_r$ and $v_0, \dots, v_s$ be the cumulative marginal probabilities of X and Y respectively, $u_0 = v_0 = 0$ and $u_r = v_s = 1$, and create rectangles $A_1, \dots, A_{rs}$ as in Section 2. In accordance with Assumption A2, X and Y can be expressed as functions of the non-discretized random variables $\tilde{X}$ and $\tilde{Y}$ by $X = \sum_{i=1}^{r} u_i \mathbb{1}_{(u_{i-1},u_i]}(\tilde{X})$ and $Y = \sum_{j=1}^{s} v_j \mathbb{1}_{(v_{j-1},v_j]}(\tilde{Y})$, where $\mathbb{1}_A(x)$ is the indicator function of the set A. Let $(x_k, y_k)_{k=1}^n$ and $(\tilde{x}_k, \tilde{y}_k)_{k=1}^n$ be the observations of (X, Y) and $(\tilde{X}, \tilde{Y})$, respectively. The observations (x_k, y_k) can be expressed as discretizations of the theoretical non-discretized observations $(\tilde{x}_k, \tilde{y}_k)$ in the same way as (X, Y) can be expressed as discretizations of $(\tilde{X}, \tilde{Y})$.

The empirical polychoric correlation coefficient uses the empirical copula, which is defined

$$\hat{C}_n(u_i, v_j) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[0,u_i]\times[0,v_j]}(x_k, y_k).$$

The empirical copula is only defined on the points of cumulative marginal probabilities (u_i, v_j), $i = 0, \dots, r$ and $j = 0, \dots, s$. Where it is defined, the empirical copula is an unbiased estimator of C, and converges almost surely to C by the strong law of large numbers. For the use of Expression (4), however, the estimate of C must be defined on all of I^2.

The empirical measure of C is

$$\hat{D}_n(u, v) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[0,u]\times[0,v]}(\tilde{x}_k, \tilde{y}_k).$$

The empirical measure is an unbiased estimator of C which converges almost surely to C for all $(u, v) \in I^2$. Consequently, the empirical measure has good properties which a measure of association ideally could leverage. However, because $\hat{D}_n$ is a function of the theoretical non-discretized sample $(\tilde{x}_k, \tilde{y}_k)_{k=1}^n$, it is not available in practice but only as a theoretical construct. From the expression of (x_k, y_k) as discretizations of $(\tilde{x}_k, \tilde{y}_k)$, though, it follows that the empirical measure is identical to the empirical copula where the latter is defined.

The empirical measure $\hat{D}_n$ is approximated by a simple function on $A_1, \dots, A_{rs}$, i.e. a function of type $f = \sum_{k=1}^{rs} a_k \mathbb{1}_{A_k}$. If the values a_k are chosen to be the empirical copula values of the lower vertices of the rectangles A_k, i.e. $A_k = [u_{i-1}, u_i) \times [v_{j-1}, v_j)$ yielding $a_k = \hat{C}_n(u_{i-1}, v_{j-1})$, then the simple function is less than or equal to $\hat{D}_n$ on every rectangle, and is hence an underestimation of $\hat{D}_n$. If the values a_k are chosen to be the empirical copula values of the upper vertices of the rectangles A_k, then the simple function is greater than or equal to $\hat{D}_n$, and hence an overestimation of the same.

Because all $\hat{D}_n$ values of A_k lie between these two mentioned choices of a_k-values, a better approximation is yielded for a middle choice.

By choosing the a_k values to be the means of the empirical copula values of all four vertices of the rectangles A_k, a simple function which is neither an underestimation nor an overestimation, in the sense that the resulting measure of association is unbiased for independent ordinal variables, is yielded. The such obtained simple function is denoted $\hat{E}_n$. Consequently,

$$\hat{E}_n(u,v) = \sum_{k=1}^{rs} a_k \mathbb{1}_{A_k}(u,v),$$

where if $A_k = [u_{i-1}, u_i) \times [v_{j-1}, v_j)$,

$$a_k = \frac{1}{4}\left(\hat{C}_n(u_i,v_j) + \hat{C}_n(u_i,v_{j-1}) + \hat{C}_n(u_{i-1},v_j) + \hat{C}_n(u_{i-1},v_{j-1})\right).$$

The empirical polychoric correlation coefficient is then, in analogy with the conventional polychoric correlation coefficient, defined as $r_{epc} = 2\sin(\rho_S(\hat{E}_n)\pi/6)$.

Because $\hat{E}_n$ is a simple function, $\rho_S(\hat{E}_n)$, given by Expression (4), is easy to compute by hand. More precisely, the expression reduces to

$$\rho_S(\hat{E}_n) = 12\sum_{k=1}^{rs} a_k \lambda(A_k) - 3.$$

Therefore the empirical polychoric correlation coefficient is easily computed by hand, in contrast with the conventional polychoric correlation coefficient which in practice demands computer assisted numerical optimization. The transformation $f(x) = 2\sin(x\pi/6)$, while insignificant from the practical perspective, is kept for the purpose of theoretical aspects such as asymptotic properties.

3.2. **Properties.** If all points of cumulative marginal probabilities (u_i, v_j) are elements of the boundary of the unit square, ∂I^2, then it follows that $\hat{E}_n \equiv 0.25$ and hence that $r_{epc} = 0$. As a consequence, a special definition for this case is not needed, unlike for the polychoric correlation coefficient. In fact, for all contingency tables there exists a unique empirical polychoric correlation coefficient. If a unique correlation coefficient exists for every contingency table, then the coefficient is said to be well defined.

Theorem 1. *The empirical polychoric correlation coefficient is well defined and takes values on the interval* $[-1, 1]$.

Proof. The simple function $\hat{E}_n$ exists and is uniquely determined for every contingency table. Because $2\sin(\rho_S(\hat{E}_n)\pi/6)$ is a well defined function, the empirical polychoric correlation coefficient, r_{epc}, exists and is unique.

To show that $-1 \le r_{epc} \le 1$, it is first noted that everywhere on I^2, $W \le \hat{C}_n \le M$, where $W(u,v) = \max(u+v-1, 0)$ is the minimum copula and $M(u,v) = \min(u,v)$ is the maximum copula. Let $\delta_M(t) = M(t,t)$ and $\delta_W(t) = W(t, 1-t)$ be the diagonal sections of M and W respectively. Assume first that $\hat{C}_n = W$. For all rectangles A that do not contain any segment of the diagonal δ_W, the graph $(u, v, W(u, v))$ is a plane, and it is easily seen that $\int_A \hat{E}_n d\lambda = \int_A W d\lambda$. For rectangles A that contain a segment of the diagonal, assume that the diagonal passes through two vertices of A. If the diagonal does not pass through two vertices of A, then it is possible to subdivide the rectangle into one rectangle for which the diagonal passes though two vertices and other rectangles that do not contain any segment of the diagonal. It is easily verified using simple geometry that the integrals over the rectangle are $\int_A W d\lambda = \lambda(A)^{3/2}/6$ and $\int_A \hat{E}_n d\lambda = \lambda(A)^{3/2}/4$. So for all rectangles $A \subset I^2$, $\int_A \hat{E}_n d\lambda \geq \int_A W d\lambda$, and thus $\rho_S(\hat{E}_n) \geq \rho_S(W) = -1$.

For the other extreme case, when $\hat{C}_n = M$, the integrals are also equal whenever the rectangle does not contain any segment of the diagonal δ_M. For rectangles A such that the diagonal passes through two vertices of A the integrals are $\int_A M d\lambda = \lambda(A)^{3/2}/3 + c$ and $\int_A \hat{E}_n d\lambda = \lambda(A)^{3/2}/4 + c$, where c is a constant dependent on the position of A. So for all rectangles $A \subset I^2$, $\int_A \hat{E}_n d\lambda \leq \int_A M d\lambda$, which implies $\rho_S(\hat{E}_n) \leq \rho_S(M) = 1$. And because the function $2\sin(x\pi/6)$ maps $[-1, 1]$ to $[-1, 1]$ it follows that $-1 \leq r_{epc} \leq 1$. □

It is also easy to check that if $\hat{C}_n = \Pi$, where $\Pi(u, v) = uv$ is the product copula, then for all rectangles $A_1, \ldots, A_{rs}$, $\int_A \hat{E}_n d\lambda = \int_A \Pi d\lambda$. Hence, the empirical polychoric correlation coefficient is unbiased whenever the ordinal variables are statistically independent.

The next theorem is an asymptotic result by which the empirical polychoric correlation coefficient under general conditions converges almost surely to the theoretical population analogue, $\rho_{pc} = 2\sin(\rho_S(C)\pi/6)$.

Theorem 2. *For a given underlying joint distribution, if the numbers of categories, r and s, increase such that the maximal difference of cumulative marginal probabilities goes to zero as $r, s \to \infty$, then*

$$\lim_{n\to\infty} \lim_{r,s\to\infty} r_{epc} = \rho_{pc} \quad \textit{almost surely.}$$

For the proof of Theorem 2, the following two results are needed. Note that the following lemma is based on a standard argument, and the lack of bibliographical citation should not be taken as a claim of originality.

Lemma 3. *Let $f : X \times \Omega \to \mathbb{R}$ be a random variable and $g : X \to \mathbb{R}$ a measurable function. If f and g are both λ-almost everywhere continuous in $x \in X$, and agree almost surely on each element of a countable dense subset of X, then $\int f d\lambda = \int g d\lambda$ almost surely.*

Proof. Denote the countable dense subset $D \subset X$ and let $\{q_n\}_{n=1}^{\infty}$ be an enumeration of D. For a fixed $\omega \in \Omega$, the notation $f_\omega(x) = f(x, \omega)$ is used; $f_\omega : X \to \mathbb{R}$. Because continuous functions are completely determined by their values on dense subsets, it follows that if two functions f and g are both continuous λ-almost everywhere then they agree on a dense subset if and only if they agree λ-almost everywhere. Thus, it follows that $P\left(\{\omega : \int f_\omega d\lambda \neq \int g d\lambda\}\right) \leq P\left(\cup_{n=1}^{\infty}\{\omega : f_\omega(q_n) \neq g(q_n)\}\right)$ which by countable subadditivity is less than or equal to $\sum_{n=1}^{\infty} P\left(\{\omega : f_\omega(q_n) \neq g(q_n)\}\right) = 0$. This proves the lemma. $\square$

Theorem 4. *Let $C : I^2 \to I$ be a copula, $X_1, X_2, X_3, \ldots$ independent random vectors with distribution function C, and let $\hat{D}_n(u,v) = n^{-1}\sum_{k=1}^{n} \mathbb{1}_{[0,u]\times[0,v]}(X_k)$. Then $\int \lim_{n\to\infty} \hat{D}_n d\lambda = \int C d\lambda$ almost surely.*

Proof. It is clear that $\hat{D}_n$ is λ-almost everywhere continuous for all $n \in \mathbb{N}$, and every copula is continuous. By the strong law of large numbers, for all $(u,v) \in I^2$, $\lim_{n\to\infty} \hat{D}_n(u,v) = C(u,v)$ almost surely. The statement then follows by Lemma 3. $\square$

Proof of Theorem 2. Note first that $\lim_n \lim_{r,s} r_{epc} = 2\sin(\pi/6 \lim_n \lim_{r,s} \rho_S(\hat{E}_n))$ since the sine function is continuous, and by the Lebesgue dominated convergence theorem, $\lim_n \lim_{r,s} \rho_S(\hat{E}_n) = 12 \int_{I^2} \lim_n \lim_{r,s} \hat{E}_n d\lambda - 3$.

Since r and s increase such that the maximal difference of cumulative marginal probabilities goes to zero as $r, s \to \infty$, the maximal side length of the rectangles $A_1, A_2, \ldots$ goes to zero as $r, s \to \infty$. Note also that where $\hat{C}_n$ is defined, $\hat{C}_n = \hat{D}_n$. Thus, $\hat{E}_n$ can be expressed as a function of $\hat{D}_n$. It is clear that as the maximal side length of $A_1, A_2, \ldots$ goes to zero, $\hat{E}_n$ converges to $\hat{D}_n$ at every point where $\hat{D}_n$ is continuous. And since $\hat{D}_n$ is λ-almost everywhere continuous for all $n \in \mathbb{N}$, $\lim_{r,s} \hat{E}_n = \hat{D}_n$ λ-almost everywhere. The identity $\int_{I^2} \lim_n \hat{D}_n d\lambda = \int_{I^2} C d\lambda$ almost surely then follows by Theorem 4. Thus, $\lim_n \lim_{r,s} r_{epc} = 2\sin(\rho_S(C)\pi/6) = \rho_{pc}$ almost surely. $\square$

A result corresponding to Theorem 2 does not hold for the conventional polychoric correlation coefficient under Assumption A2. Thus, the empirical polychoric correlation coefficient is theoretically more robust to changes of the distributional assumption than the polychoric correlation coefficient. The rate of convergence is studied by means of simulation in Section 4.

Clearly, because the empirical polychoric correlation coefficient is based on the empirical distribution, a goodness-of-fit test is not relevant. The empirical polychoric correlation coefficient fits every contingency table perfectly every time; for the empirical distribution, Equation (1) always holds. A solution to Equation (1) need not be found because the empirical distribution is a guaranteed solution to Equation (1).

A confidence interval for the empirical polychoric correlation coefficient can be found by means of simulation under the null hypothesis, i.e. under the empirical distribution. The simulated critical values converge almost surely to the correct theoretical critical values by the strong law of large numbers. Simulation under the empirical distribution is sometimes referred to as non-parametric bootstrap. Do note, however, that there are many good tests for independence that do not require even Assumption A2, such as, e.g., the Pearson chi-square test.

4. Simulation setup and results

To gain an understanding of the size of the approximation error under fixed sample sizes and numbers of categories, a simulation study was conducted. The numbers of categories were chosen to both equal 3, 5 and 7 respectively, yielding square tables, and the sample sizes were chosen to 100 and 500. These numbers can be considered relevant for, e.g., survey applications. The marginal probabilities were set to uniform over the categories throughout.

Contingency tables were simulated from the volumes of five families of bivariate distributions, with five parameter values each. So in all, contingency tables were simulated from 25 bivariate distributions. The parameter values for the different distributional families were chosen so that the theoretical population polychoric correlations equal -0.67, -0.33, 0, 0.33 and 0.67, respectively, covering with equal spacing a range of common values in many applications.

The families of bivariate distributions were chosen such that all have the minimum and the maximum copulas as limits, corresponding to perfect positive and negative association between the random variables, respectively, are continuous in the parameter and ordered. The Gaussian, Frank, Clayton, Nelsen-(2) and the Genest-Ghoudi families, are quite common and have a mix of dependency structures suitable for the simulation study. The Gaussian and Frank families have monotonic dependency structures. The Clayton family has an asymmetric left tail dependency structure, while the Nelsen-(2) and Genest-Ghoudi families have asymmetric right tail dependency structures.

For every simulated contingency table, six measures of association were computed. Five of them are polychoric correlation coefficients computed under the five distributional assumptions discussed in the preceding paragraph. All polychoric correlation coefficients were fitted with respect to the ℓ^2-norm, because this loss function worked best in Ekström (2008). The sixth measure of association is the empirical polychoric correlation coefficient. The simulation was replicated 1000 times, generating for each set of numbers of categories, each distribution and each measure of association 1000 computed coefficients.

In Tables 1 to 6, the results of the simulation are presented. Under the different numbers of categories and sample sizes the broad patterns of the simulation are consistent, and therefore the results will be not discussed table by table.

In the tables, it is seen that the empirical polychoric correlation coefficient is the winner in terms of stability as measured by the standard deviation. The empirical polychoric correlation coefficient has a consistently low standard deviation of less than or equal to 0.06 at sample size 500, and less than or equal to 0.12 at sample size 100. The other measures of association have higher standard deviations in general, with several occurrences of standard deviations of 0.20 and higher. Moreover, for the conventional polychoric correlation coefficients the standard deviations seem to be increasing with higher numbers of categories, while the standard deviation of the empirical polychoric correlation coefficient seems to be constant over the numbers of categories 3, 5 and 7.

Based on this simulation, the empirical polychoric correlation coefficient can be deemed statistically robust under different distributional families. For 5×5 contingency tables, for example, the maximum difference in the empirical polychoric correlation coefficient under different distributional families is 0.03, compared to 0.20 and higher for the other polychoric correlation coefficients. Moreover, for the empirical polychoric correlation coefficient this lack-of-robustness error is decreasing with an increasing number of categories, while it is not for the other polychoric correlation coefficients. So in terms of statistical robustness the empirical polychoric correlation coefficient seems to be decidedly better than the conventional polychoric correlation coefficients.

The empirical polychoric correlation coefficient is biased so that its absolute value is smaller than that of the theoretical polychoric correlation. When the ordinal variables are independent, however, the empirical polychoric correlation coefficient is unbiased, as was shown in Section 3.2. Based on this simulation, for 3×3 contingency tables the empirical polychoric correlation coefficient is consistently 20% too small. For 5×5 contingency tables, the empirical polychoric correlation coefficient is 8% too small, and for 7×7 contingency tables the empirical polychoric correlation coefficient is 4% too small. The bias seems to be stable under different sample sizes. The reason for the bias was hinted at in the proof of Theorem 1.

It is not hard to think of a bias correction for the empirical polychoric correlation coefficient. Based on this simulation, each coefficient could simply be multiplied with a bias correcting factor. However, the bias is likely dependent on more parameters than the numbers of categories, so for an extensive bias correction simulation could be used with the actual numbers of categories and the marginal probabilities of the contingency table at hand.

However, in many applications it is of minor importance whether the correlation coefficient is, say, 0.61 or 0.67. The general conclusion is the same, that there is a strong positive association between the two ordinal variables. A more problematic error would be if the correlation coefficient would be for example 0.2 instead of -0.1, because then the conclusion of the association analysis would change appreciably. The empirical polychoric correlation coefficient is unbiased at zero, and it is the measure of association with least standard deviation, also at zero. If the analyst is aware of the fact that the empirical polychoric correlation coefficient is a conservative estimate of the theoretical population polychoric correlation, then a bias correction is for most purposes not necessary. Based on this simulation, for 5×5 contingency tables the empirical polychoric correlation coefficient is 8% too small, a difference which is for most applications negligible.

5. Conclusions

The empirical polychoric correlation coefficient is a statistically robust and, in terms of standard deviation, stable measure of association for ordinal variables. While unbiased at zero, the absolute value of the empirical polychoric correlation coefficient, based on a simulation study, is some 4 to 20 percent too small, depending on the number of categories. The empirical polychoric correlation coefficient can in that sense be therefore be considered a conservative estimate of the theoretical population polychoric correlation. The empirical polychoric correlation coefficient also has good asymptotic properties. As a consequence of Theorem 2, studies of association of ordinal variables should be designed to have the largest number of categories feasible.

Statistical robustness is an important property because statistical conclusions should in principle follow from data, not the assumptions that were made at the outset of the statistical analysis. If there is uncertainty about which specific distributional family that satisfies the statement of Assumption A1, then from the perspective of statistical robustness it is preferable to assume the weaker Assumption A2 and use the empirical polychoric correlation coefficient. The cost of the additional statistical robustness seems rather low, since the statistically robust method has better asymptotic properties and is more stable in simulations.

For practical purposes, the empirical polychoric correlation coefficient also has the advantages compared to a conventional polychoric correlation coefficient that neither a distributional family nor a loss function need to be chosen. And consequently, such an assumption does not need to be substantiated, something which can often be quite difficult. Furthermore, since the empirical polychoric correlation coefficient is based on the assumption of an underlying continuous joint distribution it retains the essence of Karl Pearson's original idea.

References

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

Ekström, J. (2008). A generalized definition of the polychoric correlation coefficient. In *Contributions to the Theory of Measures of Association for Ordinal Variables*. Ph.D. thesis, Uppsala: Acta Universitatis Upsaliensis.

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.

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.

Yule, G. U. (1912). On the methods of measuring the association between two attributes. *J. Roy. Statist. Soc.*, *75*, 579–652.

TABLE 1. Polychoric correlation coefficients with standard deviations for 3×3 contingency tables, sample size 100.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.08)	−.33 (.12)	−.00 (.13)	.33 (.12)	.67 (.09)
	Frank	−.64 (.08)	−.31 (.11)	−.00 (.12)	.31 (.11)	.63 (.09)
	Clayton	−.66 (.10)	−.30 (.11)	−.00 (.12)	.30 (.12)	.64 (.10)
	Nelsen-(2)	−.59 (.12)	−.22 (.28)	.22 (.21)	.47 (.11)	.68 (.08)
	Genest-G.	−.63 (.09)	−.28 (.19)	.09 (.14)	.38 (.11)	.66 (.09)
	Empirical	−.51 (.06)	−.25 (.09)	−.00 (.09)	.25 (.09)	.51 (.07)
Frank	Gaussian	−.70 (.08)	−.36 (.12)	−.00 (.13)	.35 (.12)	.70 (.08)
	Frank	−.67 (.08)	−.34 (.11)	−.00 (.12)	.33 (.11)	.67 (.08)
	Clayton	−.69 (.09)	−.32 (.12)	−.00 (.12)	.33 (.11)	.67 (.09)
	Nelsen-(2)	−.61 (.11)	−.26 (.28)	.21 (.22)	.49 (.11)	.71 (.08)
	Genest-G.	−.64 (.09)	−.32 (.19)	.10 (.14)	.40 (.11)	.70 (.08)
	Empirical	−.54 (.06)	−.26 (.09)	−.00 (.09)	.26 (.09)	.53 (.06)
Clayton	Gaussian	−.67 (.09)	−.33 (.12)	−.00 (.13)	.34 (.12)	.69 (.09)
	Frank	−.64 (.09)	−.31 (.11)	−.00 (.12)	.32 (.11)	.65 (.09)
	Clayton	−.66 (.07)	−.34 (.11)	−.00 (.12)	.34 (.11)	.67 (.09)
	Nelsen-(2)	−.54 (.10)	−.28 (.23)	.22 (.21)	.51 (.11)	.74 (.07)
	Genest-G.	−.59 (.09)	−.29 (.16)	.10 (.14)	.41 (.11)	.71 (.07)
	Empirical	−.52 (.06)	−.24 (.09)	−.00 (.09)	.25 (.09)	.52 (.07)
Nelsen-(2)	Gaussian	−.84 (.08)	−.36 (.15)	−.01 (.14)	.37 (.12)	.73 (.08)
	Frank	−.81 (.08)	−.33 (.14)	−.01 (.14)	.35 (.12)	.70 (.09)
	Clayton	−.79 (.06)	−.50 (.10)	−.26 (.14)	.21 (.16)	.70 (.12)
	Nelsen-(2)	−.67 (.11)	−.33 (.13)	−.02 (.14)	.34 (.12)	.67 (.10)
	Genest-G.	−.71 (.08)	−.36 (.12)	−.11 (.13)	.28 (.12)	.68 (.10)
	Empirical	−.54 (.07)	−.26 (.09)	−.01 (.10)	.30 (.09)	.58 (.06)
Genest-Ghoudi	Gaussian	−.77 (.09)	−.35 (.12)	.01 (.13)	.36 (.12)	.70 (.09)
	Frank	−.74 (.09)	−.33 (.12)	.01 (.13)	.34 (.12)	.67 (.09)
	Clayton	−.74 (.06)	−.48 (.09)	−.13 (.13)	.25 (.14)	.66 (.11)
	Nelsen-(2)	−.62 (.10)	−.31 (.12)	.09 (.18)	.41 (.12)	.67 (.09)
	Genest-G.	−.66 (.08)	−.34 (.11)	.00 (.12)	.33 (.12)	.67 (.09)
	Empirical	−.53 (.07)	−.28 (.08)	.01 (.09)	.27 (.09)	.53 (.07)

Note. Simulation of 3×3 contingency tables of sample size $n = 100$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.

Table 2. Polychoric correlation coefficients with standard deviations for 5×5 contingency tables, sample size 100.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.08)	−.33 (.11)	−.01 (.12)	.33 (.11)	.67 (.07)
	Frank	−.65 (.08)	−.32 (.11)	−.01 (.12)	.32 (.10)	.65 (.08)
	Clayton	−.55 (.10)	−.27 (.10)	−.01 (.11)	.28 (.10)	.61 (.09)
	Nelsen-(2)	−.38 (.37)	.09 (.29)	.35 (.20)	.58 (.16)	.77 (.07)
	Genest-G.	−.44 (.18)	−.11 (.17)	.14 (.14)	.43 (.11)	.69 (.07)
	Empirical	−.60 (.06)	−.29 (.09)	−.01 (.10)	.30 (.09)	.60 (.06)
Frank	Gaussian	−.67 (.08)	−.34 (.10)	−.00 (.12)	.34 (.11)	.67 (.07)
	Frank	−.67 (.08)	−.33 (.10)	−.00 (.12)	.33 (.10)	.66 (.07)
	Clayton	−.55 (.10)	−.28 (.10)	−.00 (.11)	.28 (.10)	.62 (.09)
	Nelsen-(2)	−.40 (.37)	.08 (.28)	.35 (.20)	.58 (.14)	.78 (.07)
	Genest-G.	−.45 (.18)	−.12 (.16)	.15 (.14)	.44 (.10)	.69 (.06)
	Empirical	−.61 (.06)	−.30 (.09)	.00 (.10)	.30 (.09)	.61 (.06)
Clayton	Gaussian	−.69 (.08)	−.32 (.10)	−.00 (.11)	.34 (.12)	.69 (.08)
	Frank	−.67 (.08)	−.31 (.10)	−.00 (.11)	.33 (.11)	.68 (.09)
	Clayton	−.68 (.06)	−.33 (.09)	−.00 (.10)	.33 (.10)	.67 (.08)
	Nelsen-(2)	−.49 (.23)	.03 (.24)	.35 (.19)	.65 (.15)	.83 (.07)
	Genest-G.	−.55 (.11)	−.14 (.15)	.15 (.13)	.48 (.12)	.75 (.06)
	Empirical	−.61 (.07)	−.29 (.09)	−.00 (.10)	.29 (.09)	.60 (.06)
Nelsen-(2)	Gaussian	−.86 (.07)	−.44 (.21)	.02 (.19)	.33 (.16)	.64 (.11)
	Frank	−.84 (.08)	−.43 (.21)	.02 (.19)	.33 (.17)	.65 (.11)
	Clayton	−.80 (.05)	−.68 (.08)	−.38 (.17)	.02 (.19)	.43 (.20)
	Nelsen-(2)	−.72 (.18)	−.29 (.27)	.03 (.16)	.33 (.12)	.66 (.08)
	Genest-G.	−.72 (.07)	−.50 (.09)	−.14 (.11)	.16 (.15)	.58 (.10)
	Empirical	−.62 (.08)	−.32 (.11)	.01 (.11)	.32 (.10)	.63 (.06)
Genest-Ghoudi	Gaussian	−.79 (.08)	−.38 (.15)	.00 (.15)	.33 (.13)	.69 (.08)
	Frank	−.77 (.09)	−.36 (.15)	.01 (.15)	.32 (.13)	.68 (.09)
	Clayton	−.76 (.05)	−.55 (.07)	−.23 (.13)	.11 (.13)	.57 (.14)
	Nelsen-(2)	−.61 (.24)	−.17 (.15)	.15 (.17)	.46 (.12)	.74 (.07)
	Genest-G.	−.66 (.08)	−.33 (.10)	−.01 (.11)	.32 (.11)	.67 (.08)
	Empirical	−.62 (.08)	−.31 (.10)	.00 (.10)	.30 (.09)	.61 (.07)

Note. Simulation of 5×5 contingency tables of sample size $n = 100$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.

Table 3. Polychoric correlation coefficients with standard deviations for 7×7 contingency tables, sample size 100.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.08)	−.33 (.11)	.00 (.11)	.33 (.10)	.66 (.07)
	Frank	−.66 (.08)	−.32 (.11)	.00 (.11)	.33 (.11)	.66 (.08)
	Clayton	−.51 (.09)	−.26 (.10)	.00 (.10)	.26 (.10)	.58 (.09)
	Nelsen-(2)	−.19 (.41)	.22 (.26)	.45 (.19)	.61 (.15)	.78 (.10)
	Genest-G.	−.33 (.18)	−.03 (.15)	.20 (.13)	.44 (.12)	.69 (.07)
	Empirical	−.63 (.07)	−.31 (.09)	.00 (.10)	.31 (.09)	.63 (.06)
Frank	Gaussian	−.65 (.08)	−.33 (.10)	.00 (.11)	.33 (.10)	.65 (.08)
	Frank	−.66 (.08)	−.33 (.10)	.00 (.11)	.33 (.10)	.66 (.08)
	Clayton	−.50 (.09)	−.26 (.09)	−.00 (.10)	.26 (.10)	.57 (.10)
	Nelsen-(2)	−.20 (.41)	.21 (.27)	.44 (.20)	.61 (.14)	.77 (.10)
	Genest-G.	−.33 (.17)	−.04 (.15)	.20 (.14)	.43 (.11)	.68 (.07)
	Empirical	−.63 (.06)	−.32 (.09)	.00 (.10)	.32 (.09)	.63 (.06)
Clayton	Gaussian	−.69 (.08)	−.32 (.10)	.00 (.11)	.35 (.12)	.69 (.08)
	Frank	−.67 (.08)	−.31 (.10)	.00 (.11)	.34 (.12)	.69 (.09)
	Clayton	−.66 (.05)	−.33 (.08)	.00 (.09)	.34 (.10)	.67 (.07)
	Nelsen-(2)	−.33 (.32)	.16 (.22)	.43 (.19)	.65 (.16)	.84 (.11)
	Genest-G.	−.49 (.08)	−.06 (.13)	.19 (.13)	.48 (.15)	.75 (.08)
	Empirical	−.63 (.07)	−.31 (.09)	.00 (.10)	.31 (.09)	.63 (.06)
Nelsen-(2)	Gaussian	−.87 (.07)	−.44 (.21)	.01 (.23)	.35 (.19)	.64 (.13)
	Frank	−.85 (.08)	−.43 (.21)	.00 (.22)	.34 (.19)	.65 (.13)
	Clayton	−.81 (.04)	−.66 (.05)	−.48 (.14)	−.05 (.25)	.35 (.25)
	Nelsen-(2)	−.57 (.52)	−.15 (.39)	.04 (.20)	.33 (.12)	.66 (.08)
	Genest-G.	−.72 (.07)	−.47 (.09)	−.23 (.10)	.06 (.12)	.56 (.11)
	Empirical	−.65 (.09)	−.32 (.11)	.00 (.12)	.32 (.11)	.65 (.07)
Genest-Ghoudi	Gaussian	−.82 (.08)	−.39 (.16)	.01 (.17)	.33 (.14)	.68 (.09)
	Frank	−.80 (.08)	−.37 (.16)	.01 (.16)	.33 (.15)	.68 (.09)
	Clayton	−.79 (.04)	−.58 (.05)	−.28 (.14)	.07 (.13)	.51 (.15)
	Nelsen-(2)	−.47 (.50)	−.10 (.25)	.23 (.15)	.55 (.13)	.74 (.09)
	Genest-G.	−.67 (.07)	−.33 (.09)	−.00 (.10)	.33 (.10)	.66 (.07)
	Empirical	−.64 (.08)	−.32 (.11)	.01 (.11)	.32 (.10)	.64 (.06)

Note. Simulation of 7×7 contingency tables of sample size $n = 100$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.

TABLE 4. Polychoric correlation coefficients with standard deviations for 3×3 contingency tables, sample size 500.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.04)	−.33 (.05)	−.00 (.06)	.33 (.05)	.67 (.04)
	Frank	−.63 (.04)	−.31 (.05)	−.00 (.05)	.31 (.05)	.63 (.04)
	Clayton	−.68 (.05)	−.29 (.05)	−.00 (.05)	.31 (.05)	.64 (.04)
	Nelsen-(2)	−.59 (.04)	−.32 (.20)	.26 (.09)	.47 (.05)	.68 (.04)
	Genest-G.	−.63 (.04)	−.32 (.16)	.10 (.05)	.38 (.05)	.67 (.04)
	Empirical	−.51 (.03)	−.25 (.04)	−.00 (.04)	.25 (.04)	.51 (.03)
Frank	Gaussian	−.70 (.04)	−.35 (.05)	−.00 (.06)	.36 (.05)	.70 (.04)
	Frank	−.67 (.04)	−.33 (.05)	−.00 (.05)	.33 (.05)	.67 (.04)
	Clayton	−.71 (.03)	−.31 (.05)	−.00 (.06)	.33 (.05)	.67 (.04)
	Nelsen-(2)	−.61 (.04)	−.36 (.18)	.26 (.10)	.49 (.05)	.71 (.03)
	Genest-G.	−.65 (.04)	−.36 (.15)	.10 (.05)	.40 (.05)	.70 (.03)
	Empirical	−.54 (.03)	−.26 (.04)	−.00 (.04)	.26 (.04)	.54 (.03)
Clayton	Gaussian	−.67 (.04)	−.32 (.05)	.00 (.06)	.33 (.05)	.69 (.04)
	Frank	−.64 (.04)	−.30 (.05)	−.00 (.05)	.31 (.05)	.65 (.04)
	Clayton	−.67 (.03)	−.33 (.05)	−.00 (.06)	.33 (.05)	.67 (.04)
	Nelsen-(2)	−.55 (.04)	−.34 (.14)	.26 (.09)	.51 (.05)	.73 (.03)
	Genest-G.	−.59 (.04)	−.32 (.12)	.10 (.05)	.41 (.05)	.71 (.03)
	Empirical	−.52 (.03)	−.24 (.04)	.00 (.04)	.25 (.04)	.52 (.03)
Nelsen-(2)	Gaussian	−.85 (.03)	−.36 (.06)	−.02 (.06)	.37 (.05)	.73 (.04)
	Frank	−.82 (.03)	−.33 (.06)	−.01 (.06)	.35 (.05)	.70 (.04)
	Clayton	−.79 (.03)	−.49 (.04)	−.27 (.06)	.21 (.07)	.70 (.05)
	Nelsen-(2)	−.67 (.04)	−.33 (.06)	−.01 (.06)	.33 (.05)	.67 (.04)
	Genest-G.	−.71 (.04)	−.36 (.06)	−.11 (.06)	.28 (.06)	.68 (.04)
	Empirical	−.55 (.03)	−.27 (.04)	−.01 (.04)	.30 (.04)	.58 (.03)
Genest-Ghoudi	Gaussian	−.78 (.04)	−.35 (.06)	.01 (.06)	.36 (.05)	.70 (.04)
	Frank	−.74 (.04)	−.33 (.05)	.01 (.05)	.34 (.05)	.67 (.04)
	Clayton	−.75 (.03)	−.47 (.04)	−.14 (.06)	.26 (.06)	.66 (.05)
	Nelsen-(2)	−.62 (.04)	−.31 (.06)	.10 (.12)	.41 (.05)	.67 (.04)
	Genest-G.	−.67 (.04)	−.34 (.05)	−.00 (.05)	.33 (.05)	.67 (.04)
	Empirical	−.53 (.03)	−.28 (.04)	.01 (.04)	.27 (.04)	.53 (.03)

Note. Simulation of 3×3 contingency tables of sample size $n = 500$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.

TABLE 5. Polychoric correlation coefficients with standard deviations for 5×5 contingency tables, sample size 500.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.03)	−.33 (.05)	−.00 (.05)	.33 (.05)	.67 (.03)
	Frank	−.66 (.03)	−.32 (.05)	−.00 (.05)	.32 (.05)	.66 (.03)
	Clayton	−.56 (.04)	−.27 (.04)	−.00 (.04)	.28 (.05)	.60 (.04)
	Nelsen-(2)	−.12 (.29)	.22 (.21)	.37 (.09)	.55 (.13)	.78 (.02)
	Genest-G.	−.40 (.11)	−.10 (.13)	.13 (.06)	.45 (.04)	.69 (.03)
	Empirical	−.61 (.03)	−.30 (.04)	−.00 (.04)	.30 (.04)	.61 (.03)
Frank	Gaussian	−.67 (.03)	−.34 (.05)	.00 (.05)	.34 (.04)	.67 (.03)
	Frank	−.67 (.03)	−.34 (.05)	.00 (.05)	.33 (.04)	.67 (.03)
	Clayton	−.56 (.04)	−.28 (.05)	.00 (.05)	.27 (.04)	.62 (.04)
	Nelsen-(2)	−.11 (.30)	.20 (.22)	.37 (.08)	.54 (.13)	.79 (.02)
	Genest-G.	−.41 (.12)	−.11 (.13)	.13 (.06)	.45 (.04)	.70 (.03)
	Empirical	−.62 (.03)	−.31 (.04)	.00 (.04)	.31 (.04)	.62 (.03)
Clayton	Gaussian	−.69 (.03)	−.32 (.05)	−.00 (.05)	.34 (.05)	.69 (.04)
	Frank	−.67 (.03)	−.31 (.05)	−.00 (.05)	.33 (.05)	.68 (.04)
	Clayton	−.67 (.02)	−.33 (.04)	−.00 (.04)	.33 (.04)	.67 (.03)
	Nelsen-(2)	−.45 (.21)	.13 (.21)	.37 (.10)	.67 (.12)	.83 (.02)
	Genest-G.	−.56 (.09)	−.11 (.11)	.13 (.06)	.50 (.04)	.75 (.03)
	Empirical	−.61 (.03)	−.30 (.04)	−.00 (.05)	.30 (.04)	.61 (.03)
Nelsen-(2)	Gaussian	−.86 (.03)	−.44 (.09)	.01 (.09)	.32 (.07)	.64 (.05)
	Frank	−.85 (.03)	−.43 (.09)	.00 (.09)	.32 (.08)	.64 (.05)
	Clayton	−.80 (.02)	−.69 (.03)	−.43 (.06)	−.01 (.07)	.37 (.08)
	Nelsen-(2)	−.31 (.47)	−.30 (.15)	−.00 (.06)	.33 (.05)	.67 (.04)
	Genest-G.	−.72 (.03)	−.51 (.04)	−.15 (.05)	.14 (.08)	.57 (.04)
	Empirical	−.62 (.04)	−.32 (.05)	.00 (.05)	.32 (.04)	.64 (.03)
Genest-Ghoudi	Gaussian	−.80 (.03)	−.38 (.07)	.00 (.07)	.33 (.06)	.69 (.04)
	Frank	−.78 (.03)	−.37 (.07)	.00 (.07)	.33 (.06)	.68 (.04)
	Clayton	−.75 (.02)	−.55 (.03)	−.23 (.06)	.10 (.05)	.56 (.06)
	Nelsen-(2)	−.33 (.42)	−.16 (.05)	.17 (.15)	.44 (.05)	.76 (.03)
	Genest-G.	−.67 (.03)	−.34 (.04)	.00 (.04)	.33 (.04)	.67 (.03)
	Empirical	−.62 (.03)	−.31 (.05)	.00 (.05)	.31 (.04)	.62 (.03)

Note. Simulation of 5×5 contingency tables of sample size $n = 500$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.

TABLE 6. Polychoric correlation coefficients with standard deviations for 7×7 contingency tables, sample size 500.

True	Assumed	$\rho_{pc} = -.67$	$\rho_{pc} = -.33$	$\rho_{pc} = 0$	$\rho_{pc} = .33$	$\rho_{pc} = .67$
Gaussian	Gaussian	−.67 (.03)	−.33 (.05)	−.00 (.05)	.33 (.05)	.67 (.03)
	Frank	−.67 (.03)	−.33 (.05)	−.00 (.05)	.33 (.05)	.67 (.03)
	Clayton	−.51 (.04)	−.26 (.04)	−.00 (.04)	.26 (.04)	.58 (.04)
	Nelsen-(2)	−.14 (.26)	.22 (.17)	.42 (.11)	.58 (.11)	.77 (.08)
	Genest-G.	−.32 (.12)	−.02 (.09)	.21 (.07)	.44 (.06)	.70 (.03)
	Empirical	−.63 (.03)	−.32 (.04)	−.00 (.05)	.31 (.04)	.64 (.03)
Frank	Gaussian	−.65 (.03)	−.33 (.05)	.00 (.05)	.33 (.04)	.65 (.03)
	Frank	−.66 (.03)	−.33 (.05)	−.00 (.05)	.33 (.04)	.67 (.03)
	Clayton	−.49 (.05)	−.26 (.04)	−.00 (.04)	.25 (.04)	.57 (.04)
	Nelsen-(2)	−.14 (.27)	.19 (.18)	.41 (.12)	.59 (.11)	.76 (.07)
	Genest-G.	−.30 (.12)	−.03 (.09)	.21 (.08)	.44 (.06)	.69 (.03)
	Empirical	−.64 (.03)	−.32 (.04)	−.00 (.05)	.32 (.04)	.64 (.03)
Clayton	Gaussian	−.69 (.03)	−.32 (.05)	.00 (.05)	.35 (.05)	.69 (.04)
	Frank	−.68 (.04)	−.32 (.05)	−.00 (.05)	.34 (.06)	.69 (.04)
	Clayton	−.67 (.02)	−.33 (.03)	−.00 (.04)	.33 (.05)	.66 (.03)
	Nelsen-(2)	−.35 (.08)	.15 (.15)	.41 (.11)	.59 (.12)	.87 (.05)
	Genest-G.	−.50 (.03)	−.03 (.08)	.21 (.07)	.50 (.08)	.76 (.03)
	Empirical	−.64 (.03)	−.32 (.04)	.00 (.05)	.31 (.05)	.64 (.03)
Nelsen-(2)	Gaussian	−.88 (.03)	−.47 (.09)	.01 (.11)	.35 (.09)	.64 (.06)
	Frank	−.86 (.03)	−.45 (.10)	.00 (.10)	.35 (.09)	.66 (.06)
	Clayton	−.81 (.02)	−.66 (.02)	−.51 (.04)	−.07 (.09)	.27 (.09)
	Nelsen-(2)	−.17 (.51)	−.31 (.14)	−.00 (.06)	.33 (.05)	.67 (.04)
	Genest-G.	−.72 (.03)	−.48 (.04)	−.24 (.04)	.04 (.05)	.56 (.05)
	Empirical	−.65 (.04)	−.33 (.05)	.00 (.05)	.33 (.04)	.65 (.03)
Genest-Ghoudi	Gaussian	−.82 (.03)	−.40 (.07)	.00 (.08)	.33 (.06)	.68 (.04)
	Frank	−.80 (.03)	−.39 (.07)	.00 (.08)	.33 (.07)	.69 (.04)
	Clayton	−.79 (.02)	−.59 (.02)	−.29 (.06)	.05 (.05)	.48 (.07)
	Nelsen-(2)	−.44 (.37)	−.14 (.08)	.21 (.10)	.52 (.10)	.73 (.05)
	Genest-G.	−.67 (.03)	−.33 (.04)	.00 (.04)	.34 (.04)	.67 (.03)
	Empirical	−.65 (.04)	−.33 (.05)	.00 (.05)	.32 (.04)	.64 (.03)

Note. Simulation of 7×7 contingency tables of sample size $n = 500$ from different distributions with theoretical population polychoric correlation equal to −.67, −.33, 0, .33 and .67. 1000 contingency tables were simulated and for each, polychoric correlation coefficients were calculated. Figures in the table are mean correlation coefficients and in parentheses the standard deviations.