Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Maximal point-polyserial correlation for non-normal random distributions

Abstract

We consider the problem of determining the maximal value of the point-polyserial correlation for a bivariate random vector, when the continuous distributions of both components are not necessarily normal and one component is discretized into an ordinal random variable with k categories, which are assigned the first k natural values $1, 2, \ldots, k$, and arbitrary probabilities p_i. For different parametric distributions and number of categories k, we derive the formula of the maximal point-polyserial correlation as a function of the p_i and of the distribution's parameters; we devise an algorithm for obtaining its maximum value numerically for a given k. These maximum values and the features of the corresponding k-point discrete rvs are discussed with respect to the underlying continuous distribution. We also focus our attention on the case when discretization is implemented through equal probabilities: a simple expression for the maximal point-polyserial correlation is derivable and its limiting behavior is investigated as k tends to ∞. An application to real data exemplifies the main findings. A comparison between the discretization leading to the maximum point-polyserial correlation and those based on quantization or moment matching, is sketched.

KEYWORDS

attainable correlations, biserial correlation, discretization, latent variable, non-normal distribution

1 | INTRODUCTION

In behavioral, educational, and psychological studies, the observed variables are frequently measured using ordinal scales. For example, Likert scale is widely used to measure responses in surveys, allowing respondents to express how much they agree or disagree with a particular statement or the level of satisfaction they show towards a product they bought or a service they experienced, in a (typically) five- or seven-point scale (e.g., 1="Completely disagree" or "Completely unsatisfied", ..., 5="Completely agree" or "Completely satisfied"). These categorical or ordinal variables can be treated as being discretized from an underlying continuous variable for degree of agreement on the statement or level of satisfaction (see, e.g., Bartholomew, 1980). There are also many examples of quantitative variables that are discretized explicitly in social science studies: for instance, when asking questions about sensitive or personal quantitative attributes (e.g., income, alcohol consumption, time spent on social media, etc.), the non-response rate may often be reduced by simply asking the respondent to select one of two very broad categories (under 50K/over 50K, etc.). When analyzing this kind of data, a common approach is to assign integer values to each category and proceed in the analysis as if the data had been measured on an interval scale with desired distributional properties (Norman, 2010); 'Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of "coming to the wrong conclusion".' The most common choice for the distribution of the latent variables is the (multivariate) normal distribution, because the dependence structure among them can be fully captured by the variance-covariance matrix and each of its elements can be estimated using a bivariate normal distribution separately (McNeil, Frey, & Embrechts, 2015).

Let X_2 be an observed ordinal variable that depends on an underlying latent continuous random variable Z_2 and let Z_1 represent another observed continuous variable. It is assumed that the joint distribution of Z_1 and Z_2 is bivariate normal. The product moment correlation between Z_1 and X_2 is called the point-polyserial correlation, while the correlation between Z_1 and Z_2 is called the polyserial correlation. As a particular case, if X_1 is a dichotomous random variable, we refer to them as point-biserial and biserial correlations. The problem of estimating the polyserial correlation based on a bivariate sample has been studied by Cox (1974), who derived the MLE; Olsson, Drasgow, and Dorans (1982) derived the relationship between the polyserial and the point-polyserial correlation and compared the MLE of polyserial correlation with a two-step estimator and with a computationally convenient ad hoc estimator. Bedrick (1995) studied the attenuation of the correlation coefficient (the polyserial correlation) when one of the continuous variables is categorized. The attenuation is shown to depend critically on the distribution of the underlying latent variable, and on the scores assigned to the categories. It is observed that the reduction in correlation can be substantially greater under exponential, double exponential, and *t* distributions than is expected assuming normality. However, attenuation becomes less severe as the number of categories increases, provided the category scores are carefully selected. In particular, equally-spaced scores (e.g., $1, 2, \dots, k$) give reasonable protection against gross attenuation across a variety of distributions. On the problem of assigning scores to ordered categories, one can refer to Ivanova and Berger (2001) and Fernández, Liu, Costilla, and Gu (2020).

Demirtas and Hedeker (2016) and later, Demirtas and Vardar-Acar (2017) studied the relationship between the biserial and the point-biserial correlations by devising an algorithm working for any underlying distribution other than the (bivariate) normal for the bivariate vector (Z_1, Z_2) . The authors state that "it works for ordinal-continuous data combinations, and so one can compute the polyserial correlation given the point–polyserial correlation (or vice versa) when the relative proportions of the ordinal categories are specified." The algorithm is based on the generation of a huge sample (of size, say, $N = 100,000$) from a bivariate random vector (Z_1, Z_2) with assigned marginal distributions and dependence structure, implicitly induced by the method of Fleishman polynomials (Fleishman, 1978) for the construction of bivariate random vectors (Foldnes & Grønneberg, 2015). Although the numerical experiments carried out in Demirtas and Hedeker (2016) are shown to produce negligible errors (when an analytical solution is also available), nevertheless the sampling error naturally induced by random simulation can hardly be controlled, especially if one is interested in determining the maximal point-polyserial correlation and its asymptotic value for $k \to \infty$, which is expected to be close to 1 if not exactly 1. Cheng and Liu (2016) derived the maximal point-biserial correlation under several non-normal distributions, namely, the uniform, Student's *t*, exponential, and a mixture of two normal distributions. They showed that the maximal point-biserial correlation, depending on the non-normal continuous distribution, may not be a function of the probability p that the dichotomous variable takes the value 1; may be symmetric or non-symmetric around $p = 0.5$. The relatively easy analytical derivation of (maximal) point-biserial correlation relies on the (availability of expression for) moments of truncated continuous distributions.

The aim of this paper is to derive the maximal point-polyserial correlation, i.e., the maximum linear correlation between two continuous, not necessarily normal, random variables after one of them is ordinalized. We will start from the general case (an ordinal random variable with support values $1, 2, \dots, k$, and corresponding probabilities p_i, $i = 1, 2, \dots, k$) and then move to the case of equal-probability discretization ($p_i = 1/k \; \forall i = 1, 2, \dots, k$), which is particularly suitable if one wants to study the limit behavior of the maximal point-polyserial correlation. Along with the normal, several widely used non-normal distributions are considered, namely the uniform, the exponential, the Pareto, the logistic, and the power distributions. Theoretical arguments imply that the maximal point-polyserial correlation is always smaller than 1, which is confirmed by the numerical experiments conducted under the `R` statistical environment.

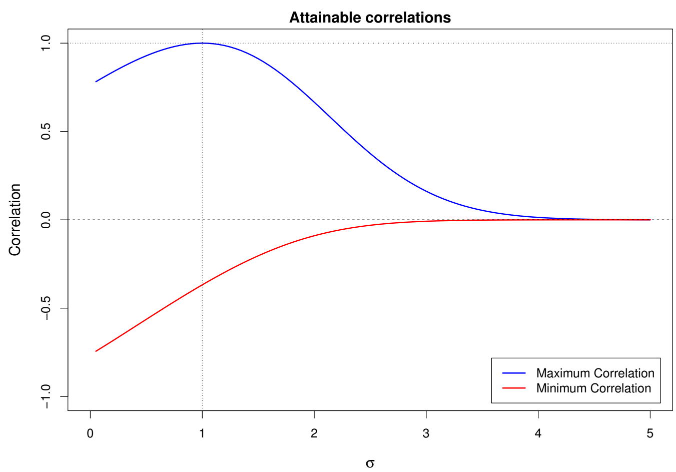
The paper is structured as follows. Section 2 recalls some results about attainable correlations between two random variables with assigned margins. Section 3 synthesizes and integrates the main findings about the maximal point-biserial and point-polyserial correlation under bivariate normality. Section 3 investigates the behaviour of the maximal point-polyserial correlation under several continuous non-normal distributions. Section 4 illustrates the main findings through a real data set. Section 5 hints at a possible application of the results on maximal point-polyserial correlations in finding an optimal *k*-point approximation of a continuous distribution. Section 6 concludes the paper with some final remarks.

2 | ATTAINABLE CORRELATIONS

Although Pearson's correlation ρ between two random variables (rvs) *X* and *Y* can theoretically take on any value between –1 and +1, however, when the marginal distributions of *X* and *Y* are assigned, it may generally not span the entire $[-1, +1]$ interval and achieve either its natural lower or upper bounds; the constraint on the marginal distributions typically reduces the range of Pearson's correlation to a narrower interval. In more detail (Hoeffding, 1940; Fréchet, 1951), the minimal and maximal attainable correlations that Pearson's ρ can achieve form a closed interval $[\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} < 0 < \rho_{\max}$. The minimum correlation $\rho_{\min}$ is attained if and only if *X* and *Y* are countermonotonic; the maximum correlation $\rho_{\max}$ is attained if and only if *X* and *Y* are comonotonic. Moreover, $\rho_{\min} = -1$ if and only if *X* and $-Y$ are of the same type, and $\rho_{\max} = 1$ if and only if *X* and *Y* are of the same type. We recall that two rvs X and Y (or their random distributions) are said of the same type if there exist two constants $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$ such that $X \stackrel{d}{=} a + bY$; in other terms, X and Y are rvs of the same type if they are a location-scale transformation of each other. The bounds for ρ are computed as $\rho_{\text{min}} = \rho(F_1^{-1}(U), F_2^{-1}(1 - U))$ and $\rho_{\text{max}} = \rho(F_1^{-1}(U), F_2^{-1}(U))$, where U is a standard uniform rv and F_1 and F_2 are the marginal distributions of rvs X and Y, respectively. It is often possible to determine analytically the minimum and maximum correlations by using the two formulas above; otherwise, they can be computed numerically by resorting to the algorithm in Demirtas and Hedeker (2011). A correlation value ρ is said "feasible" given the assigned margins F_1 and F_2 if it falls within $[\rho_{\text{min}}, \rho_{\text{max}}]$.

This feature of Pearson's correlation, which is well known in the quantitative risk management field (Embrechts, McNeil, & Straumann, 2002), but is often overlooked in other applied areas, represents a drawback and can lead to misinterpretations of its observed sample values; a typical example concerns two lognormal distributions with parameters $\mu_1 = 0, \sigma_1 = 1$ and $\mu_2 = 0, \sigma_2 > 0$. The two distributions are not of the same type unless $\sigma_2 = \sigma_1$; the value of the minimal correlation is given by $\rho_{\text{min}} = \frac{e^{-\sigma_2}-1}{\sqrt{(e-1)(e^{\sigma_2^2}-1)}}$, the value of the maximal correlation is $\rho_{\text{max}} = \frac{e^{\sigma_2}-1}{\sqrt{(e-1)(e^{\sigma_2^2}-1)}}$. Therefore, if $\sigma_2 = \sigma_1 = 1$, $\rho_{\text{max}} = 1$ and $\rho_{\text{min}} \approx -0.368$: X_1 and X_2 are of the same type, but X_1 and $-X_2$ are not, since the lognormal distribution is supported on $\mathbb{R}^+$ and is consequently asymmetric; for any $\sigma_2 \neq \sigma_1$, X_1 and X_2 are not rvs of the same type, and the interval $[\rho_{\text{min}}, \rho_{\text{max}}]$ tends to get narrower as σ_2 increases. For example, if $\sigma_2 = 2$ we have that $\rho_{\text{max}} = 0.666$ and $\rho_{\text{min}} \approx -0.090$; if $\sigma_2 = 4$, $\rho_{\text{max}} \approx 0.014$ and $\rho_{\text{min}} \approx 0.000$: these latter values can lead the inadvertent researcher to claim that the two rvs are nearly uncorrelated, whereas the two rvs are indeed perfectly (positively/negatively) correlated! Figure 1 displays the maximum and minimum attainable correlations for the two lognormal rvs as functions of σ_2.

FIGURE 1 Attainable correlations between two lognormal rvs, $X \sim \mathcal{LN}(\mu_1 = 0, \sigma_1 = 1)$ and $Y \sim \mathcal{LN}(\mu_2 = 0, \sigma_2)$



From what explained above, it is thus clear that if we consider a first rv with a continuous random distribution and a second rv whose distribution is obtained by discretizing the former, then the maximum and minimum correlations can never be $+1$ and -1, since a discrete distribution can never be of the same type of a continuous distribution, just for the fact that the latter has a non-countable support and the former is defined over a finite or countable set. The extreme values -1 and $+1$ can be potentially obtained only as limits when the cardinality of the support of the discrete rv increases and resembles a continuous one, or when the continuous rv converges to a discrete rv when one of its parameters tends to a limiting value, as can occur in the case of a mixture of two normal distributions with the same variance (Cheng & Liu, 2016).

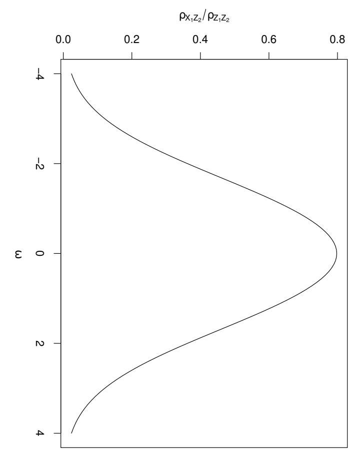
3 | POINT-POLYSERIAL CORRELATION UNDER NORMALITY

Let (Z_1, Z_2) be a bivariate standard normal rv, and let X_2 be a dichotomy of Z_2, with the point of dichotomy ω; thus X_2 is a rv which takes the value 1 when $Z_2 \geq \omega$ and the value 0 when $Z_2 < \omega$. By denoting with $\varphi(\cdot)$ the pdf of a standard normal rv and letting $P(X_2 = 1) = \int_{\omega}^{\infty} \varphi(y)dy = p(\omega)$ and $P(X_2 = 0) = q(\omega) = 1 - p(\omega)$, the relationship between $\rho(Z_1, Z_2)$ (the biserial correlation) and $\rho(Z_1, X_2)$ (the point biserial correlation) is due to Karl Pearson (Pearson, 1909) and reported also in MacCallum, Zhang, Preacher, and Rucker (2002), where the consequences of dichotomization for measurement and statistical analyses are illustrated and discussed in a more general context:

$$\frac{\rho(Z_1, X_2)}{\rho(Z_1, Z_2)} = \frac{\phi(\omega)}{\sqrt{pq}}. \tag{1}$$

It is interesting to take a look at the plot of this function, displayed in Figure 2, and to note that it is symmetrical and presents its unique maximum (equal to $2\varphi(0) = .7979$) in $\omega = 0$, which corresponds to the "equal-probability" dichotomization ($p = q = 1/2$). Please note that changing the two values of the support of the discrete rv X_2, by default set at 0 and 1, does not affect the value of the biserial correlation coefficient (this is due the well-known invariance of Pearson's ρ under any positive linear transformation).

FIGURE 2 Maximal point-biserial correlation (i.e., ratio between point-biserial and biserial correlations) as a function of the cut-point ω for a bivariate normal rv - Equation (1); the maximum, equal to $2\phi(0)$, is attained at $\omega = 0$



A generalization of Pearson's point-biserial correlation to the case of discretization into a k point-scale distribution (supported on $\{1, 2, \ldots, k\}$) is easily provided, again starting from a bivariate normal rv. Let then X_2 be the discrete rv obtained by discretizing the component Z_2. Recalling that the following relationship holds for the probability density function (pdf) of a standard normal rv:

$$\int x\varphi(x)dx = -\varphi(x) + \text{constant},$$

it can be proved that the resulting Pearson's correlation coefficient between Z_1 and X_2, i.e., the point-polyserial correlation coefficient, is:

$$\rho_{PP} = \rho(Z_1, X_2) = \rho(Z_1, Z_2) \sum_{i=1}^{k} \varphi\left[\Phi^{-1}(F_i)\right] \Bigg/ \sqrt{\sum_{i=1}^{k} i^2 p_i - \left(\sum_{i=1}^{k} i p_i\right)^2}, \tag{2}$$

where p_i and F_i are the probability and cumulative probability of the value i, respectively. Eq. (2) indicates that there is a linear relationship between the polyserial and the point-polyserial correlations, at least when working with a bivariate normal rv. The ratio between the point-polyserial correlation and the (polyserial) correlation of the bivariate normal distribution is therefore constant once the p_i's are assigned and is equal to

$$\rho_{PP}/\rho = \sum_{i=1}^{k} \varphi\left[\Phi^{-1}(F_i)\right] \Bigg/ \sqrt{\left[\sum_{i=1}^{k} i^2 p_i - \left(\sum_{i=1}^{k} i p_i\right)^2\right]} = \sum_{i=1}^{k} \varphi\left[\Phi^{-1}\left(\sum_{j=1}^{i} p_j\right)\right] \Bigg/ \sqrt{\left[\sum_{i=1}^{k} i^2 p_i - \left(\sum_{i=1}^{k} i p_i\right)^2\right]}, \tag{3}$$

which consequently corresponds to the maximal point-polyserial correlation, which is obtained by letting $\rho_{Z_1 Z_2} = 1$.

We can particularize the formulas above in the case of discretization into k equal-probability categories ($p_i = 1/k$ for each $i = 1, \ldots, k$), i.e., if the rv X_2 is defined as

$$X_2 = \begin{cases} 1 & \text{if } Z_2 < \Phi^{-1}(1/k) \\ i & \text{if } \Phi^{-1}\left(\frac{i-1}{k}\right) \leq Z_2 < \Phi^{-1}\left(\frac{i}{k}\right), 1 < i < k \\ k & \text{if } Z_2 \geq \Phi^{-1}\left(\frac{k-1}{k}\right) \end{cases}$$

Then, specializing (3), we obtain

$$\rho_{PP}/\rho = \sum_{i=1}^{k-1} \varphi(\Phi^{-1}(i/k))/\sqrt{(k^2-1)/12}, \tag{4}$$

since $\phi(\Phi^{-1}(1)) = 0$, and for a discrete uniform rv X_2, $\mathbb{E}(X_2) = (1+k)/2$ and $\text{Var}(X_2) = \sum_{i=1}^{k} i^2/k - [(1+k)/2]^2 = (k+1)(2k+1)/6 - (k+1)^2/4 = (k^2-1)/12$.

Recalling the symmetry of φ about 0, for k odd $\mathbb{E}(Z_1X_2) = 2\rho\sum_{i=1}^{(k-1)/2} \varphi(\Phi^{-1}(i/k))$, for k even $\mathbb{E}(Z_1X_2) = \rho\left\{\varphi(0) + 2\sum_{i=1}^{k/2-1} \varphi[\Phi^{-1}(i/k)]\right\}$; therefore Equation (4) can be rewritten as

$$\rho_{PP}/\rho = \begin{cases} \dfrac{\varphi(0) + 2\sum_{i=1}^{k/2-1} \varphi\left[\Phi^{-1}(i/k)\right]}{\sqrt{(k^2-1)/12}} & \text{if } k \text{ even} \\ \dfrac{2\sum_{i=1}^{(k-1)/2} \varphi\left[\Phi^{-1}(i/k)\right]}{\sqrt{(k^2-1)/12}} & \text{if } k \text{ odd,} \end{cases} \tag{5}$$

implying that for $k = 2$ the sum at the numerator vanishes and thus ρ_{PP}/ρ reduces to $2\varphi(0) = \sqrt{2/\pi}$, which is consistent with the expression of the biserial correlation coefficient with cut-point $\omega = 0$.

For $k \to \infty$, the ratio in (5), and then the maximal point-polyserial correlation, tends asymptotically to $\sqrt{3/\pi}$. In fact, we can write:

$$\int_0^1 \varphi[\Phi^{-1}(x)]\mathrm{d}x = \int_{-\infty}^{+\infty} \varphi(u)\Phi'(u)\mathrm{d}u = \int_{-\infty}^{+\infty} \varphi^2(u)\mathrm{d}u = \int_{-\infty}^{+\infty} \frac{1}{2\pi}e^{-z^2}\mathrm{d}u = \frac{1}{2\pi}\sqrt{\pi} = \frac{1}{2\sqrt{\pi}}, \tag{6}$$

but the integral on the left side of (6) is related to the finite sum above through

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \varphi[\Phi^{-1}(i/k)] = \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k-1} \varphi[\Phi^{-1}(i/k)] = \int_0^1 \varphi[\Phi^{-1}(x)]\mathrm{d}x = \frac{1}{2\sqrt{\pi}},$$

and then it easily follows that

$$\lim_{k\to\infty} \rho_{PP,\max} = \frac{2\sqrt{3}}{2\sqrt{\pi}} = \sqrt{3/\pi} \approx 0.977205.$$

This is an important theoretical result: starting from a bivariate standard normal distribution with correlation coefficient ρ and discretizing one of its component through an equal-probability discretization process, the resulting correlation coefficient between the unaltered component and the new discrete one, letting k go to ∞, tends to a value strictly smaller than ρ. This result is not unexpected since discretizing a normal distribution although through “many” equal-probability categories produces a distribution that cannot resemble the unimodal normal pdf (Barbiero & Hitaj, 2023).

TABLE 1 Ratio between point-polyserial correlation ρ_{PP} and polyserial correlation ρ as a function of k, number of equal-probability categories

k	2	3	4	5	6	7	8	9	10	20	50	100	1000
ρ'/ρ	0.7979	0.8906	0.9253	0.9423	0.9520	0.9581	0.9622	0.9650	0.9672	0.9744	0.9767	0.9771	0.9772

Moving back to the general case of unequal p_i’s, for an assigned $k \geq 2$, the ratio in (3) can be maximized with respect to the p_i’s satisfying the customary contraints $p_i \geq 0$, $i = 1, 2, \ldots, k$, and $\sum_{i=1}^{k} p_i = 1$. Figure 3 displays for $k = 2, \ldots, 10$, the maximal point-polyserial correlation that can be achieved by ordinalizing/discretizing into k categories a continuous (standard) normal distribution. For each k, a barplot is drawn that represents the k probabilites leading to the maximal point-polyserial correlation. We notice that all these ordinalized distributions maximizing the maximal point-polyserial correlation are symmetrical, as one could have expected, with a unique mode – the central category – if k is odd, with two modes – the central categories – is k

is even; therefore they inherit or, better, mirror the two main features of the continuous Gaussian distribution, symmetry and unimodality. For illustrative purposes, we report here the R code used to determine the value of the maximal point-polyserial correlation for $k = 5$.

```
library(Rsolnp)
k <- 5
p <- rep(1/k,k)
fn1 <- function(p){
F <- head(cumsum(p),-1)
i <- 1:k
-sum(dnorm(qnorm(F)))/sqrt(sum(p*i^2)-(sum(i*p))^2)
}
fnB <- function(p){sum(p)}

sol <- solnp(pars=p, fun=fn1, eqfun = fnB,  eqB=1, LB=rep(0,k), UB=rep(1,k))
sum(sol$pars)
print(sol$pars) # prints the probabilities
print(qnorm(head(cumsum(sol$pars),-1))) # prints the thresholds
print(tail(-sol$values,1)) # print the maximal point-polyserial correlation
```

which produces the following output:

```
sol <- solnp(pars=p, fun=fn1, eqfun = fnB,  eqB=1, LB=rep(0,k), UB=rep(1,k))

Iter: 1 fn: -0.9580      Pars:  0.10302 0.23367 0.32660 0.23367 0.10303
Iter: 2 fn: -0.9580      Pars:  0.10303 0.23367 0.32661 0.23367 0.10303
solnp--> Completed in 2 iterations
> sum(sol$pars)
[1] 1
> print(sol$pars) # prints the probabilities
[1] 0.1030275 0.2336679 0.3266053 0.2336698 0.1030294
> print(qnorm(head(cumsum(sol$pars),-1))) # prints the thresholds
[1] -1.2644876 -0.4214987  0.4214884  1.2644770
> print(tail(-sol$values,1)) # print the maximal point-polyserial correlation
[1] 0.9580304
```

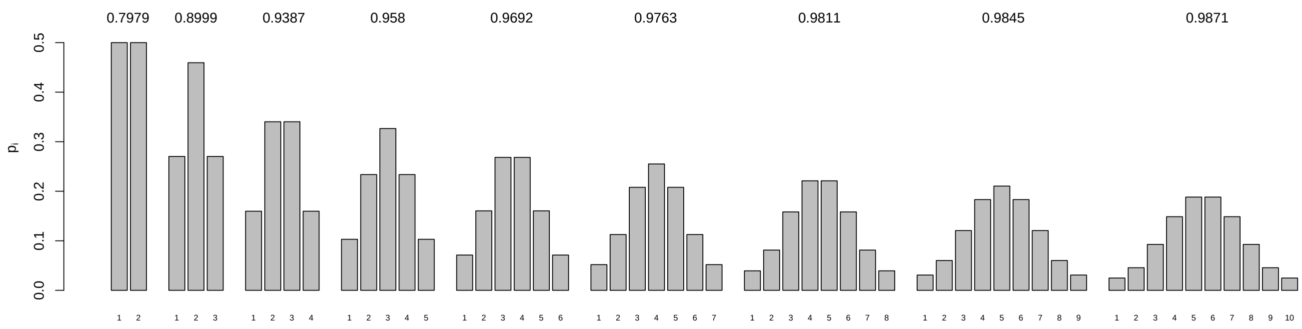
We used the `solnp` function included in the `Rsolnp` package (Ghalanos & Theussl, 2015; Ye, 1987) for solving our non-linear maximization problem, which is actually converted into a minimization problem by just changing the sign to the expression of the point-polyserial correlation for an underlying normal distribution (3). The constraints on the p_i's are provided through the arguments `eqfun` and `eqB` (through which we impose that $\sum_{i=1}^{k} p_i = 1$), `LB` (lower bounds for the p_i), and `UB` (upper bounds for the p_i's).

From the R output, one can notice an important feature of the p_i's that solve the optimization problem: for any $k \geq 4$, the corresponding thresholds constitute a set of equally-spaced values.

For a standard normal distribution, the problem of maximizing the point-polyserial correlation can be written in the following terms:

$$\max_{p_1,\ldots,p_k} \sum_{i=1}^{k} \varphi\left[\Phi^{-1}\left(\sum_{j=1}^{i} p_j\right)\right] \Bigg/ \sqrt{\left[\sum_{i=1}^{k} i^2 p_i - \left(\sum_{i=1}^{k} i p_i\right)^2\right]}$$

FIGURE 3 Maximal point-polyserial correlations and corresponding configurations $p_1, \dots, p_k$ for a number of categories $k = 2, \dots, 10$, when the continuous distribution is normal.



subject to the constraints $p_i \geq 0$ for $i = 1, \dots, k$ and $\sum_{i=1}^k p_i = 1$. We can rewrite the problem as a non-linear optimization problem by using Lagrange multipliers:

$$\min_{p_1,\dots,p_k} \sum_{i=1}^k \varphi\left[\Phi^{-1}\left(\sum_{j=1}^i p_j\right)\right] \Big/ \sqrt{\left[\sum_{i=1}^k i^2 p_i - \left(\sum_{i=1}^k i p_i\right)^2\right]} + \lambda\left(\sum_{i=1}^k p_i - 1\right)$$

from which, recalling that $\phi'(x) = -x\phi(x)$, and computing the first-order derivatives with respect to p_i's and to λ and equating them to zero:

$$\begin{cases} \dfrac{-\sum_{j=i}^{k-1}\Phi^{-1}(\sum_{h=1}^j p_h)\cdot\left[\sum_{i=1}^k i^2 p_i - (\sum_{i=1}^k ip_i)^2\right] - \frac{1}{2}\sum_{i=1}^{k-1}\phi(\Phi^{-1}(\sum_{j=1}^i p_j))\cdot(i^2 - 2i\sum_{i=1}^k ip_i)}{\left[\sum_{i=1}^k i^2 p_i - (\sum_{i=1}^k ip_i)^2\right]^{3/2}} + \lambda & = 0 \quad i = 1, \dots, k \\ \sum_{i=1}^k p_i - 1 & = 0 \end{cases}$$

It is not possible to find the analytical expression of the p_i's solution to the above optimization problem. However, by evaluating the first equation of the system above for two consecutive values of i, we obtain

$$\Phi^{-1}\left(\sum_{h=1}^i p_h\right) \cdot V - \frac{1}{2}A\left(2i + 1 - 2\sum_{i=1}^k ip_i\right) = 0,$$

where $V = \sum_{i=1}^k i^2 p_i - (\sum_{i=1}^k ip_i)^2$ and $A = \sum_{i=1}^{k-1}\phi\left(\Phi^{-1}(\sum_{j=1}^i p_j)\right)$, from which, for all $i = 1, \dots, k-2$,

$$\Phi^{-1}\left(\sum_{j=1}^{i+1} p_i\right) - \Phi^{-1}\left(\sum_{j=1}^{i} p_i\right) = \frac{A}{V} = \text{const},$$

which means that the discrete distribution maximizing the point-polyserial correlation has cumulative probabilities F_i whose corresponding standard normal quantiles are, for $k \geq 4$, equally-spaced. It should be also expected that the k-point distribution maximizing the point-polyserial correlation is symmetrical, i.e., $p_j = p_{k+1-j}, j = 1, \dots, k$; hence the thresholds are symmetrical around zero.

4 | POINT-POLYSERIAL CORRELATION UNDER NON-NORMALITY

If we consider a bivariate continuous rv (Z_1, Z_2) that is not bivariate normal, then Formula (2) does not hold and then one cannot claim there exists a linear relationship between the linear correlation coefficient before and after the discretization of Z_2. This means that for fixed k and p_i's, the ratio between the correlations after and before dicretization is not constant, but depends on the value of the latter, although an approximately linear relationship can be assumed, as done in Demirtas and Vardar-Acar (2017).

In the following subsections, we want to assess the maximum value that the point-polyserial correlation can assume when we consider two rvs X_1 and X_2 with the same continuous, not necessarily normal, distributions. Discretizing the second component, we obtain a discrete rv, which we call X_d; the first component, which remains unaltered, will be named simply X. We will review several continuous parametric families widely used in many fields of statistics, such as the uniform, the exponential, the Pareto, the logistic, and the power distributions. For each family, we will study the maximal value of the point-polyserial coefficient as a function of the number of categories of the ordinalized distribution, by providing an algorithm that returns the maximum value of the point-polyserial correlation within the class of all possible k-point ordinal distributions, supported on the first k natural numbers, and the features of the ordinal random distribution producing the maximum point-polyserial correlation.

4.1 | Uniform

Let X be a uniform rv in $(0, 1)$, then $\mathbb{E}(X) = 1/2$ and $\mathrm{Var}(X) = 1/12$; and let X_d be a k-point discrete rv with values $1, 2, \ldots, k$, and corresponding probabilities p_i, $i = 1, 2, \ldots, k$. Then the point-polyserial correlation is

$$\rho_{PP}(XX_d) = \frac{\mathbb{E}(XX_d) - \mathbb{E}(X)\mathbb{E}(X_d)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(X_d)}},$$

which is maximized, for any given feasible $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, when the mixed moment $\mathbb{E}(XX_d)$ is maximized. This occurs if X and X_d are comonotonic and in this case $\mathbb{E}(XX_d)$ becomes

$$\max \mathbb{E}(XX_d) = 1 \cdot \int_0^{F_1} x\mathrm{d}x + 2 \cdot \int_{F_1}^{F_2} x\mathrm{d}x + \cdots + k \cdot \int_{F_{k-1}}^{1} x\mathrm{d}x = \sum_{i=1}^{k} i \cdot \frac{F_i^2 - F_{i-1}^2}{2} = \frac{1}{2}\left(k - \sum_{i=1}^{k-1} F_i^2\right).$$

In order to find the maximum (among all the probability vectors $\boldsymbol{p}$) of the maximal value of ρ_{PP}, the optimization problem to be solved is

$$\max(\mathbb{E}(XX_d) - \mathbb{E}(X)\mathbb{E}(X_d)) \Bigg/ \sqrt{\frac{1}{12}\left(\sum_{i=1}^{k} i^2 p_i - (\sum_{i=1}^{k} ip_i)^2\right)} = \max\left[\frac{1}{2}\left(k - \sum_{i=1}^{k-1} F_i^2\right) - \frac{1}{2}\sum_{i=1}^{k} ip_i\right] \Bigg/ \sqrt{\sum_{i=1}^{k} i^2 p_i - (\sum_{i=1}^{k} ip_i)^2}$$
$$= \max \sum_{i=1}^{k}\left(\frac{1}{2}F_i(1-F_i)\right) \Bigg/ \sqrt{\sum_{i=1}^{k} 2i(1-F_i)}$$

subject to the usual constraints on the vector $\boldsymbol{p}$. By denoting with V the variance of X_d, with E its expectation, and with N the covariance between X and X_d (E, V, and N all depend on the p_i's, but for the sake of simplicity of notation we omitted this dependence), the first order derivative of the Lagrangian function with respect to p_i is

$$\frac{1}{2}\left[-\left(2\sum_{j=i}^{k-1}(k-j+1)p_j + i\right)V - 0.5\left(k - \sum_{i=1}^{k-1}(\sum_{j=1}^{i} p_i)^2 - \sum ip_i\right)(i^2 - 2i\sum_{i=1}^{k} ip_i)\right] \Bigg/ V^{3/2} + \lambda = 0, i = 1, \ldots, k-1$$

By subtracting the second ($i = 2$) from the first ($i = 1$) equation, one obtains

$$-\frac{1}{2}(1 - 2p_1)V - \frac{1}{2}N(3 - 2E) = 0,$$

and then

$$(1 - 2p_1)V + N(3 - 2E) = 0,$$

from which

$$p_1 = \frac{V - N(2E - 3)}{2V} = \frac{1}{2} - \frac{N(2E - 3)}{2V}.$$

By subtracting the third and the second equation,

$$-\frac{1}{2}(1-2p_1-2p_2)V-\frac{1}{2}N(5-2E)=0,$$

from which, recalling the previous expression obtained for p_1,

$$\left[1-1+\frac{N}{V}(2E-3)-2p_2\right]V+N(5-2E)=0,$$

from which one derives $p_2=N/V$, and in a similar manner $p_3=\cdots=p_{k-1}=N/V$; finally, $p_k=\frac{1}{2}-\frac{N}{V}(k-1/2-E)$.

Although the optimization problem above cannot be solved analytically, it can be numerically proved that for any $k\geq 2$, the discrete uniform distribution, which assigns each category a constant probability $p_i=1/k$ to all $i=1,\ldots,k$, is the one, among all the k-point discrete distributions, that maximizes the maximal point-polyserial correlation. In fact, letting $p_i=1/k$ for all $i=1,\ldots,$ we obtain $E=(k+1)/2$, $V=(k^2-1)/12$, $N=\mathbb{E}(XX_d)-\mathbb{E}(X)\mathbb{E}(X_d)=\frac{1}{2}\left(k-\sum_{i=1}^{k-1}(i/k)^2\right)-(k+1)/4=\frac{1}{2}\left(k-\frac{1}{k^2}\frac{(k-1)k(2(k-1)+1)}{6}\right)-(k+1)/4=(k^2-1)/(12k)$, and all the equations obtained by setting equal to zero the derivatives of the Lagrangian function are satisfied. Below is the R code that can be employed for determining the maximal point-polyserial correlation with a number of categories from 2 to 10:

```
library(Rsolnp)
maxrho <- numeric(9)
for(k in 2:10)
{
p   <- rep(1/k,k)
fn1 <- function(p){
F   <- cumsum(p)[-length(p)]
i   <- 1:k
-(1/2*(k-sum(F^2))-1/2*sum(i*p))/sqrt(1/12)/sqrt(sum(p*i^2)-(sum(i*p))^2)
}
fnB <- function(p){sum(p)}
sol <- solnp(pars=p, fun=fn1, eqfun = fnB,  eqB=1, LB=rep(0,k), UB=rep(1,k))
sum(sol$pars)
print(sol$pars) # prints the probabilities
print(tail(-sol$values,1)) # print the maximal polserial correlation
maxrho[k-1] <- -tail(sol$values,1)
}
```

Table 2 displays for several values of k the maximal point-polyserial correlation, for which an analytical expression is readily obtained. For the discrete uniform case, in fact, the maximal point-polyserial correlation becomes

$$\rho_{PP,\max}=\frac{\sqrt{k^2-1}}{k}=\sqrt{1-1/k^2}.$$

We thus observe that $\lim_{k\to\infty}\rho_{PP,\max}=1$: (under the equal-probability setting), the maximal point-polyserial correlation tends to 1, the natural upper bound of Pearson's correlation.

TABLE 2 Values of the maximal polyserial correlation between a uniformly distributed rv and a discrete rv for several values of k.

k	2	3	4	5	6	7	8	9	10	20	50	100	200
max $\rho^{PP\max}$	0.8660	0.9428	0.9682	0.9798	0.9860	0.9897	0.9922	0.9938	0.9950	0.9987	0.9998	0.9999	1.0000

It is important to remark that although it is quite easy to derive the expression of the maximal point-polyserial correlation, starting from the bivariate continuous distribution, finding the point-polyserial correlation is more challenging or, better, it requires some additional information: while for the former it is sufficient to fully specify the univariate non-normal continuous distribution, for the latter, it is necessary to specify the joint random distribution of (Z_1, Z_2), or equivalently, the two marginal distributions of Z_1 and Z_2 and the copula $C(u_1, u_2)$ linking them into the joint distribution. To better understand this point, we carried out the following numerical experiment. We considered four different parametric copulas $C(u_1, u_2; \theta)$ (Gauss, Frank, Clayton, and Gumbel), whose marginal distributions are by definition standard uniform. For each copula and for different values of the linear correlation ρ (the biserial/polyserial correlation), properly induced by the copula parameter θ, we computed the point-biserial/polyserial correlation, and the corresponding ratio, by considering for the sake of simplicity $k = 2$ and $k = 3$ equal-probability categories for the discretized random variable. The results indicate that the ratio between point-polyserial and polyserial correlations is not constant with ρ (although it can be considered as nearly constant), confirming the fact that a constant ratio characterizes the (bivariate) normal distribution only. The range of values that the ratio can span, though narrow, sensibly varies depending on the copula selected. We considered only positive values of ρ, since whereas the Frank and the Gauss copulas are comprehensive copulas (i.e., they are able to model the entire range of dependence, from countermonotonicity to comonotonicity, passing through independence), and then they are able to induce all the values of ρ in $[-1, +1]$), the Gumbel and the Clayton copulas can only model positive dependence and thus induce only positive values of linear correlation. The point-biserial (point-polyserial) correlation can be computed as usual as

$$\rho_{PP} = \frac{\mathbb{E}(U_1 U_{2d}) - \mathbb{E}(U_1)\mathbb{E}(U_{2d})}{\sqrt{\mathrm{Var}(U_1)\mathrm{Var}(U_{2d})}},$$

where the maximum value of the mixed moment can be expressed, in case of 2 equal-probability categories for U_{2d}, as

$$\max \mathbb{E}(U_1 U_{2d}) = 1 \int_0^1 \mathrm{d}u_1 \int_0^{1/2} u_1 c(u_1, u_2; \theta)\mathrm{d}u_2 + 2 \int_0^1 \mathrm{d}u_1 \int_{1/2}^1 u_1 c(u_1, u_2; \theta)\mathrm{d}u_2, \tag{7}$$

where $c(u_1, u_2; \theta)$ is the copula density; with $\mathbb{E}(U_1) = 1/2$, $\mathbb{E}(U_{2d}) = 3/2$, $\mathrm{Var}(U_1) = 1/12$, $\mathrm{Var}(U_{2d}) = 1/4$. In case of 3 equal-probability categories for U_{2d}, the maximum value of the mixed moment takes on the expression

$$\max \mathbb{E}(U_1 U_{2d}) = 1 \int_0^1 \mathrm{d}u_1 \int_0^{1/3} u_1 c(u_1, u_2; \theta)\mathrm{d}u_2 + 2 \int_0^1 \mathrm{d}u_1 \int_{1/3}^{2/3} u_1 c(u_1, u_2; \theta)\mathrm{d}u_2 + 3 \int_0^1 \mathrm{d}u_1 \int_{2/3}^1 u_1 c(u_1, u_2; \theta)\mathrm{d}u_2 \tag{8}$$

and it is easy to check that $\mathbb{E}(U_{2d}) = 2$ and $\mathrm{Var}(U_{2d}) = 2/3$. The point-polyserial correlation is readily computed once the quantities in (7) and (8) are evaluated: to this aim, one can resort to the package `cubature` (Narasimhan, Johnson, Hahn, Bouvier, & Kiêu, 2023) in `R`, which implements adaptive multivariate integration over hypercubes. The function `iRho`, provided by the package `copula` (Hofert, Kojadinovic, Maechler, & Jun, 2023), determines ("calibrate") the copula parameter θ given the value of Spearman's ρ, which coincides with Pearson's ρ for a bivariate copula.

Figure 4 displays, for each copula examined, the values of the ratio between point-biserial and biserial correlations for different values of the latter (from 0.05 to 0.95 with steps of 0.05). One can note the values of the ratio are all around the value 0.8660, which is reported in Table 2 as the maximum value of point-biserial correlation for the uniform distribution. Analogously, Figure 5 displays, for each copula examined, the values of the ratio between point-polyserial and polyserial ($k = 3$) correlations for different values of the latter (the same grid adopted for $k = 2$). One can note the values of the ratio are all around the value 0.9428, which is reported in Table 2 as the maximum value of point-biserial correlation for the uniform distribution for $k = 3$.

4.2 | Exponential

Let X be an exponential rv with pdf $f(x) = \lambda e^{-\lambda x}$ and cdf $F(x) = 1 - e^{-\lambda x}$, $x > 0$, $\lambda > 0$. It is well known that $\mathbb{E}(X) = 1/\lambda$ and $\mathrm{Var}(X) = 1/\lambda^2$. The quantile of level $0 < u < 1$ is $x_u = -\log(1 - u)/\lambda$. Let X_d be a discrete rv taking on the value i with probability p_i, $i = 1, \ldots, k$.

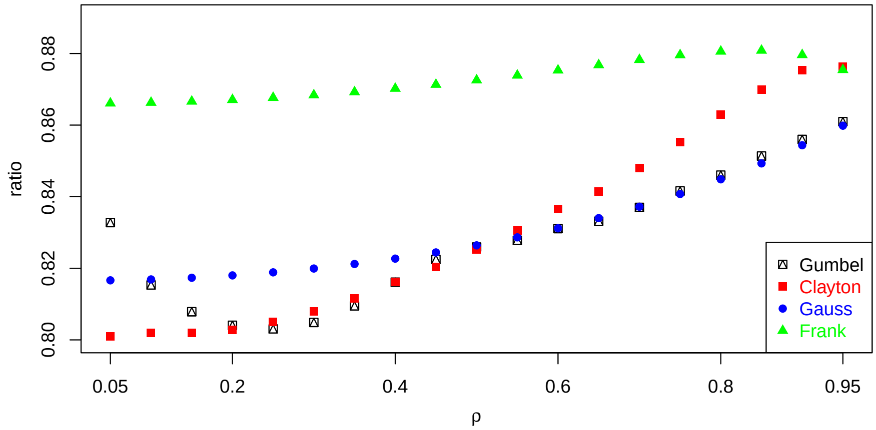


FIGURE 4 Graph of the ratio between point-biserial correlation and biserial correlation for several copulas and values of ρ

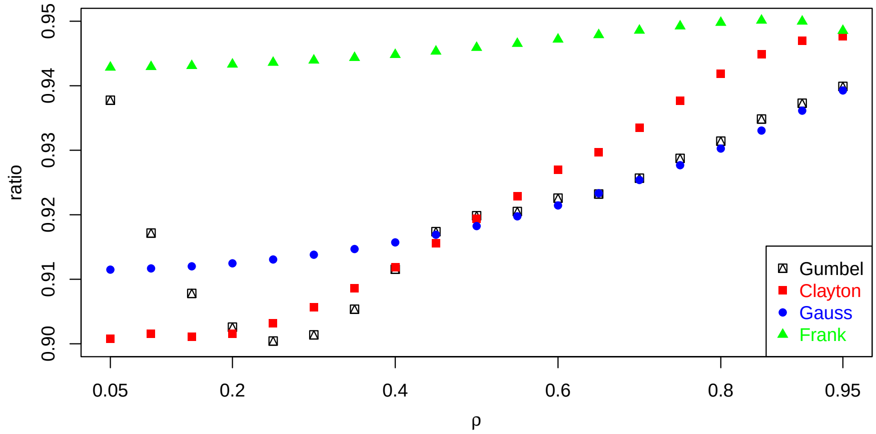


FIGURE 5 Graph of the ratio between point-polyserial correlation and biserial correlation for several copulas and values of ρ

The maximal point-polyserial correlation can be computed by noting that the maximum value of the mixed moment $\mathbb{E}(XX_d)$ is obtained (again) when X and X_d are comonotonic and can be written as

$$\max \mathbb{E}(XX_d) = \sum_{i=1}^{k} i \int_{F^{-1}(F_{i-1})}^{F^{-1}(F_i)} x\lambda e^{-\lambda x} \mathrm{d}x = \sum_{i=1}^{k} i \left[-(x+1/\lambda)e^{-\lambda x}\right]_{-\log(1-F_{i-1})/\lambda}^{-\log(1-F_i)/\lambda}$$
$$= \frac{1}{\lambda}\sum_{i=1}^{k} i\{[\log(1-F_i)-1](1-F_i)-[\log(1-F_{i-1})-1](1-F_{i-1})\} = \frac{1}{\lambda}\left[1-\sum_{i=1}^{k-1}[\log(1-F_i)-1](1-F_i)\right];$$

since

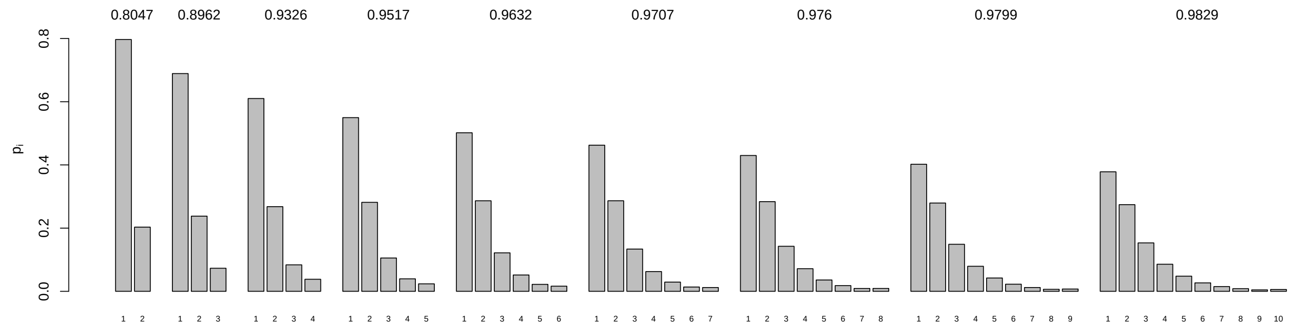
$$\int_a^b x\lambda e^{-\lambda x}\mathrm{d}x = \left[-\left(x+\frac{1}{\lambda}\right)e^{-\lambda x}\right]_a^b .$$

Therefore, the corresponding maximal point-polyserial correlation is equal to

$$\rho_{PP,\max} = \frac{1-\sum_{i=1}^{k-1}[\log(1-F_i)-1](1-F_i)-\sum_{i=1}^{k} ip_i}{\sqrt{\sum_{i=1}^{k} i^2 p_i - (\sum_{i=1}^{k} ip_i)^2}}. \tag{9}$$

Maximizing the function in Eq. (9), for a fixed k, with respect to the vector $\boldsymbol{p}$, does not return a closed form solution for $\boldsymbol{p}$ and for the maximum value of $\rho_{PP,\max}$; one needs to resort to numerical optimization as already done for the normal and uniform distributions. The k-point distribution maximizing the maximal point-polyserial correlation is empirically proved to have decreasing probabilities p_i's for $k \leq 7$, thus resembling the trend of the exponential pdf; for $k \geq 8$ the probabilities are decreasing till the second to last category, but the last category has a larger, though very small, probability than the former ($p_k > p_{k-1}$); one can empirically ascertain this by looking at the three most to the far-right graphs of Figure 6, where the k-point discrete distributions maximizing $\rho_{PP,\max}$ are displayed for $k = 2, 3, \ldots, 10$. We note that, for any k there examined, the values of $\rho_{PP,\max}$ for the exponential distribution are not very different from the analogue values for the normal distribution, reported in Figure 3, and then are a bit smaller than those obtained for the uniform distribution. Despite being strongly asymmetrical, the exponential distribution is still able to assure high values of point-polyserial correlation. This is a consequence of the fact that the exponential distribution is a scale family of distributions.

FIGURE 6 Maximal point-polyserial correlation for an exponential distribution as a function of the number of categories k



If we restrict our attention to a uniform discrete rv, then $F_i = i/k$ and the i/k-order quantile is $x_{i/k} = \frac{\log k - \log(k-i)}{\lambda}$, and then one obtains, by specializing Eq. (9), after some algebraic steps, the following expression for the maximum mixed moment:

$$\max \mathbb{E}(XX_d)^{(\text{eq})} = \frac{1}{\lambda k}\left[\frac{k(k+1)}{2}(1 + \log k) - \sum_{i=2}^{k} i \log i\right]$$

and for the maximal point-polyserial correlation:

$$\rho_{PP,\max}^{(\text{eq})} = \frac{\frac{1}{k}\left[\frac{k(k+1)}{2}(1+\log k) - \sum_{i=2}^{k} i\log i\right] - \frac{k+1}{2}}{\sqrt{\frac{k^2-1}{12}}},$$

which tends to $\sqrt{3}/2 \approx 0.866$ as k tends to infinity. In fact, since

$$\int_1^k x \log x \mathrm{d}x = \left[\frac{1}{4}x^2(2\log x - 1)\right]_1^k = \frac{1}{2}k^2 \log k - \frac{1}{4}(k^2 - 1)$$

and the sum appearing at the numerator of $\rho_{PP,\max}^{(\text{eq})}$ can be approximated for large k as

$$\sum_{i=2}^{k} i \log i = \sum_{i=1}^{k} i \log i \approx \frac{1}{2}k^2 \log k - \frac{1}{4}(k^2 - 1),$$

it is immediate to prove the asymptotic result.

TABLE 3 Maximal point-polyserial correlation between an exponentially distributed rv and an ordinal rv with k equal-probability categories

k	2	3	4	5	6	7	8	9	10	20	50	100	1000
ρ^P_{max}	0.6931	0.7796	0.8130	0.8297	0.8395	0.8456	0.8498	0.8528	0.8550	0.8628	0.8654	0.8658	0.8660

Table 3 reports the values of the maximal point-polyserial correlation under the equal-probability setting for different values of k. By comparing them to the values of the maximal point-polyserial correlations displayed in Figure 6, we can conclude that properly diversifying the probabilities of the k categories significantly increases the maximal value of point-polyserial correlation even when k becomes larger: for $k = 10$, the increase in maximal correlation is more than 13%, and this is ascribable to the highly non-uniform and asymmetrical nature of the exponential pdf.

Moreover, one can note that the limiting value of the point-polyserial correlation for the exponential distribution under the equal-probability setting is quite smaller than its analogue resulting for the normal rv ($\sqrt{3}/2 < \sqrt{3/\pi}$); this clearly descends from the asymmetrical nature of the exponential distribution, which mismatches with the equal probabilities characterizing the k-point discrete uniform rv considered in the limit case.

4.3 Pareto (Lomax)

The one-parameter Pareto distribution is characterized by the pdf $f(x) = \alpha/x^{\alpha+1}$ and the cdf $F(x) = 1 - 1/x^{\alpha}$ for $x > 1$, with $\alpha > 0$; its expectation is $\mathbb{E}(X) = \alpha/(\alpha - 1)$ for $\alpha > 1$; its variance is $\mathrm{Var}(X) = \alpha/[(\alpha - 1)^2(\alpha - 2)]$ for $\alpha > 2$. The quantile function is $x_u = F^{-1}(u) = 1/(1 - u)^{1/\alpha}$, $0 < u < 1$.

In order to obtain the maximal point-polyserial correlation between a Pareto rv X with parameter α and a k-point discrete rv X_d, one can follow the lines of the previous subsections. It is easy to find the expression of the maximum value of the mixed moment, obtained when the two rvs X and X_d are comonotonic, which is equal to

$$\max \mathbb{E}(XX_d) = \sum_{i=1}^{k} i \int_{F^{-1}(F_{i-1})}^{F^{-1}(F_i)} \frac{\alpha x}{x^{\alpha+1}} \mathrm{d}x = \sum_{i=1}^{k} i \frac{\alpha}{1-\alpha} \left[x^{1-\alpha}\right]_{(1-F_{i-1})^{1/\alpha}}^{(1-F_i)^{1/\alpha}} = \frac{\alpha}{\alpha - 1}\left[1 + \sum_{i=1}^{k-1} \frac{1}{(1-F_i)^{\frac{1-\alpha}{\alpha}}}\right], \tag{10}$$

and then the corresponding point-polyserial correlation, whose maximum value, for a given k, can be obtained as the solution of a numerical optimization with respect to the p_i's. Here, in Table 4, we report the maximum value of the point-polyserial correlation for several combinations of the parameter α and of the number of categories k. Figure 7 displays for $k = 2, \ldots, 10$, the discrete distributions maximizing the maximal point-polyserial correlation when $\alpha = 3$. In general, for an assigned k, it can be shown numerically that the discrete distribution maximizing the point-polyserial correlation has most of the probability concentrated at one category (which depends on the values of k and α: for higher values of k and α, the mode tends to move towards higher integers), whereas much smaller probabilities compete to the others. Moreover, a change in the mode of the discrete distribution maximizing $\rho_{PP,\max}$ between two consecutive values of k is accompanied by a preservation of the value of $\rho_{PP,\max}$ itself (see again Figure 7, for $k = 5$ and $k = 6$). In this case, we observe that the "old" probabilities of value i ($i = 1, \ldots, k$) remain the same but are assigned to $i + 1$, whereas the "new" probability of 1 is zero. This represents a very interesting feature, especially if the Pareto is compared to the exponential distribution, for which, as seen in the previous section, the mode of the discrete distribution maximizing the maximal point-polyserial correlation always remains equal to the smallest value of the support, 1. We note also that for the same number of categories k, the maximum value of the point-polyserial correlation for the Pareto distribution is smaller for any value of α than for the exponential distribution.

FIGURE 7 Maximal point-polyserial correlations and corresponding configurations $p_1, \ldots, p_k$ for a number of categories $k = 2, \ldots, 10$, when the underlying continuous distribution is Pareto with $\alpha = 3$.

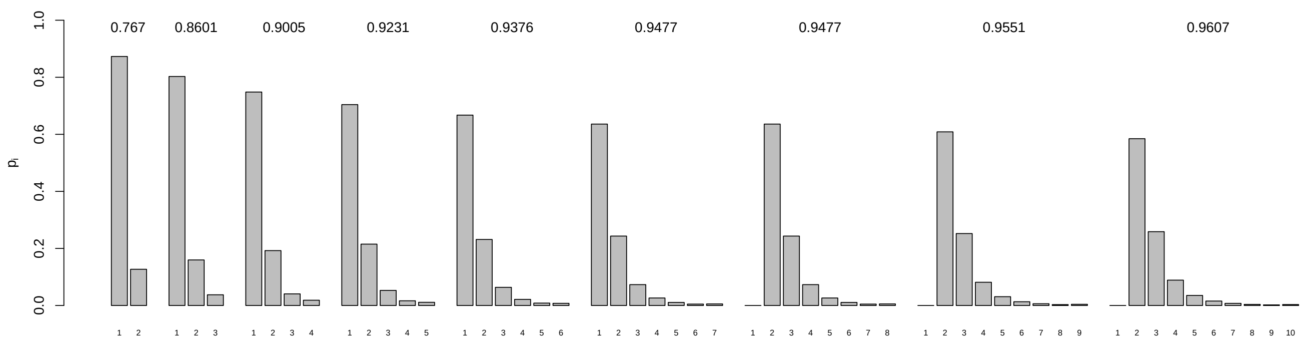


TABLE 4 Maximal polyserial correlation between a Pareto distributed rv and a discrete rv with k equal-probability categories

α, k	2	3	4	5	6	7	8	9	10	20	50	100	1000
3	0.6813	0.7716	0.8148	0.8413	0.8413	0.8596	0.8731	0.8837	0.8922	0.9280	0.9593	0.9734	0.9894
4	0.7345	0.8274	0.8695	0.8942	0.8942	0.9106	0.9106	0.9313	0.9382	0.9653	0.9852	0.9923	0.9988
5	0.7556	0.8488	0.8899	0.9134	0.9286	0.9286	0.9393	0.9473	0.9534	0.9762	0.9911	0.9958	0.9995

Focusing on the equal-probability case, one can derive the expression of the quantile of level i/k as $x_{i/k} = \left(\frac{k}{k-i}\right)^{1/\alpha}$ and then compute the maximum mixed moment arising when the two rvs are comonotonic, specializing the general expression in (10):

$$\max \mathbb{E}(XX_d)^{(eq)} = \sum_{i=1}^{k} i \int_{x_{(i-1)/k}}^{x_{i/k}} x \cdot \frac{\alpha}{x^{\alpha+1}} \mathrm{d}x = \sum_{i=1}^{k} i \left[\frac{\alpha}{1-\alpha} x^{1-\alpha}\right]_{\left(\frac{k}{k-i}\right)^{1/\alpha}}^{\left(\frac{k}{k-i+1}\right)^{1/\alpha}} = \frac{\alpha}{\alpha-1}\left[1 + \sum_{i=1}^{k-1}\left(\frac{k}{k-i}\right)^{(1-\alpha)/\alpha}\right]$$

and therefore the expression of the maximum polyserial correlation becomes:

$$\rho_{PP,\max}^{(\mathrm{eq})} = \frac{\frac{\alpha}{\alpha-1}\left[1 + \sum_{i=1}^{k-1}\left(\frac{k}{k-i}\right)^{(1-\alpha)/\alpha} - (k+1)/2\right]}{\sqrt{\frac{\alpha}{(\alpha-1)^2(\alpha-2)}\frac{k^2-1}{12}}}.$$

Since we have that $\sum_{i=1}^{k}(k/(k-i))^{(1-\alpha)/\alpha} \approx \int_1^k (k/(k-x))^{(1-\alpha)/\alpha}\mathrm{d}x = \left[\frac{\alpha k(k/(k-x))^{1/\alpha-2}}{1-2\alpha}\right]_1^k = \alpha k\left(\frac{k}{k-1}\right)^{1/\alpha-2}/(2\alpha-1)$, then for $k \to \infty$, provided that $\alpha > 2$,

$$\lim_{k\to\infty} \rho_{PP,\max}^{(\mathrm{eq})} = \frac{\sqrt{3\alpha(\alpha-2)}}{2\alpha-1}.$$

For $\alpha = 3$ we have $\rho_{pp,\max}^{(\mathrm{eq})} = 0.6$; for $\alpha = 5$ we have $\rho_{pp,\max}^{(\mathrm{eq})} = 0.745356$; for $\alpha \to \infty$, the maximum of the point-polyserial correlation, under the equal-probability setting, tends to $\sqrt{3}/2$, i.e., the same value as for the exponential distribution.

4.4 | Logistic

The logistic distribution, in its standard version, has pdf $f(x) = \frac{e^x}{(1+e^x)^2}$ and cdf $F(x) = \frac{e^x}{1+e^x}$. The quantile function is $x_u = \ln(u/(1-u))$, $0 < u < 1$; moreover, $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = \pi^2/3$. Since $\int_a^b \frac{xe^x}{(1+e^x)^2}\mathrm{d}x = \left[\frac{xe^x}{1+e^x} - \ln(1+e^x)\right]_a^b$, it is easy to derive the expression of the maximum value of the mixed moment between a logistic rv X and a discrete rv X_d:

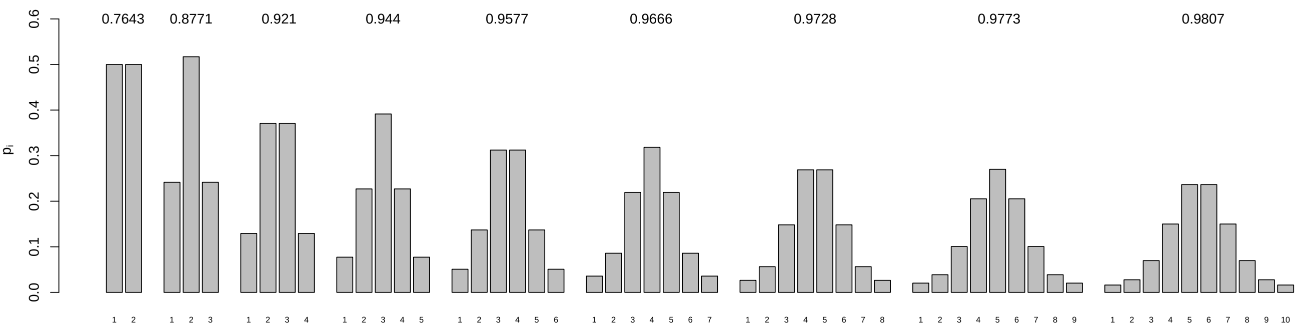
$$\max \mathbb{E}(XX_d) = -\sum_{i=1}^{k-1}[F_i \cdot \log(F_i/(1-F_i)) + \log(1-F_i)],$$

and the maximal point-polyserial correlation for a given k-dimensional vector $\boldsymbol{p}$ is then

$$\rho_{PP,\max} = -\sum_{i=1}^{k-1}[F_i \cdot \log(F_i/(1-F_i)) + \log(1-F_i)] \Big/ \sqrt{\frac{\pi^2}{3}\left[\sum_{i=1}^{k} i^2 p_i - \left(\sum_{i=1}^{k} i p_i\right)^2\right]},$$

which can maximized with respect to $\boldsymbol{p}$ for any k, by resorting to the same optimization routines used in the previous subsections. Figure 8 displays the k-point discrete distributions ($k = 2, \ldots, 10$) that maximize the maximal point-polyserial correlation, whose value is also shown above each graph. We notice that, as an expected consequence of the symmetry of the logistic distribution, they are all symmetrical around the mid-value $(k+1)/2$ and unimodal (for k odd) or bimodal (with k even) with the mode(s) coinciding with the central value(s). It is the same situation that occurs with the normal distribution; the only differences are observed in the magnitude of the probabilities p_i and of the maximal point-polyserial correlation; in particular, for any value k examined here, the maximum of $\rho_{PP,\max}$ for the logistic distribution is smaller than for the normal distribution.

FIGURE 8 Maximal polyserial correlation and corresponding configurations $p_1, \ldots, p_k$ for a number of categories $k = 2, \ldots, 10$, when the underlying distribution is logistic.



Let us study the asymptotic behaviour of $\rho_{PP,\max}$ with k in case of equal-probability categories; in this case the maximum value of the mixed moment between X and X_d takes on the following expression:

$$\begin{aligned}\max \mathbb{E}(XX_d)^{(\mathrm{eq})} &= \sum_{i=1}^{k} i \left[\frac{\frac{i}{k-i} \ln \frac{i}{k-i}}{1 + \frac{i}{k-i}} - \ln\left(1 + \frac{i}{k-i}\right) - \frac{\frac{i-1}{k-i+1} \ln \frac{i-1}{k-i+1}}{1 + \frac{i-1}{k-i+1}} - \ln\left(1 + \frac{i-1}{k-i+1}\right) \right] \\ &= \sum_{i=1}^{k-1} \ln\left(1 + \frac{i}{k-i}\right) - \frac{\frac{i}{k-i} \ln \frac{i}{k-i}}{1 + i/(k-i)} = \sum_{i=1}^{k-1} \ln \frac{k}{k-i} - \frac{i}{k} \ln \frac{i}{k-i} \\ &= \sum_{i=1}^{k-1} \ln k - \ln(k-i) - \frac{i}{k} \ln i + \frac{i}{k} \ln(k-i) = (k-1)\ln k - \sum_{i=1}^{k-1} \frac{i}{k} \ln i + (1 - i/k)\ln(k-i);\end{aligned}$$

therefore, the maximal polyserial-point correlation is given by

$$\rho_{PP,\max}^{(\mathrm{eq})} = \frac{(k-1)\ln k - \sum_{i=1}^{k-1} \left[\frac{i}{k} \ln i + (1 - i/k)\ln(k-i)\right]}{\sqrt{\frac{\pi^2}{3} \frac{k^2-1}{12}}}.$$

Now, for large k, the sum $\sum_{i=1}^{k} \frac{i}{k} \ln i + (1 - i/k)\ln(k-i)$ can be approximated by $\frac{1}{k} \int_0^k x \ln x + (k-x)\ln(k-x)\mathrm{d}x = \left[(-(k-x)^2 \ln(k-x) + x(-k + x\ln x))/(2k)\right]_0^k = k(2\ln k - 1)/2$, from which $\rho_{PP,\max}^{(\mathrm{eq})}$ can be approximated by $\frac{(k-1)\ln k - k \ln k + k/2}{\sqrt{\frac{\pi^2 (k^2-1)}{36}}}$; therefore its limiting value is $\lim_{k\to\infty} \rho_{PP,\max}^{(\mathrm{eq})} = 3\sqrt{2/\pi^2} \approx 0.9549$.

4.5 | Power distribution

The cdf and the pdf of a power rv, which is a particular case of the Beta rv, with the second shape parameter β equal to 1, are $F(x) = x^\alpha$ and $f(x) = \alpha x^{\alpha-1}, 0 < x < 1, \alpha > 0$. When $\alpha = 1$, it reduces to the uniform distribution, see Section 4.1. The quantile of level u is $x_u = u^{1/\alpha}$. Recalling the expressions for the expectation and the variance of a Beta rv, we have $\mathbb{E}(X) = \alpha/(\alpha+1)$ and $\mathrm{var}(X) = \alpha/(\alpha+1)^2/(\alpha+2)$.

It is then easy to derive the expression of the maximum attainable value of the mixed moment between a power rv of parameter α and a k-point discrete rv, which is given by

$$\max \mathbb{E}(XX_d) = \sum_{i=1}^{k} i \int_{F_{i-1}^{1/\alpha}}^{F_i^{1/\alpha}} x \alpha x^{\alpha-1} \mathrm{d}x = \alpha \sum_{i=1}^{k} i \int_{F_{i-1}^{1/\alpha}}^{F_i^{1/\alpha}} x^\alpha \mathrm{d}x = \frac{\alpha}{\alpha+1} \sum_{i=1}^{k} i \left[F_i^{(\alpha+1)/\alpha} - F_{i-1}^{(\alpha+1)/\alpha}\right] = \frac{\alpha}{\alpha+1} \left[k - \sum_{i=1}^{k-1} i F_i^{(\alpha+1)/\alpha}\right].$$

Figure 9 displays the results of the maximization of $\rho_{PP,\max}$ for $\alpha = 2$. For each of the values of k examined, the discrete distribution has increasing probabilities, thus mimicking the increasingness of the pdf of the power rv. $\rho_{PP,\max}$ converges to 1 quite quickly; when $k = 10$, it is equal to 0.9934, a value just slight smaller than the corresponding value 0.9950 obtained for a uniform rv for the same k (see Table 2).

FIGURE 9 Maximum point-polyserial correlation for the power distribution with parameter $\alpha = 2$

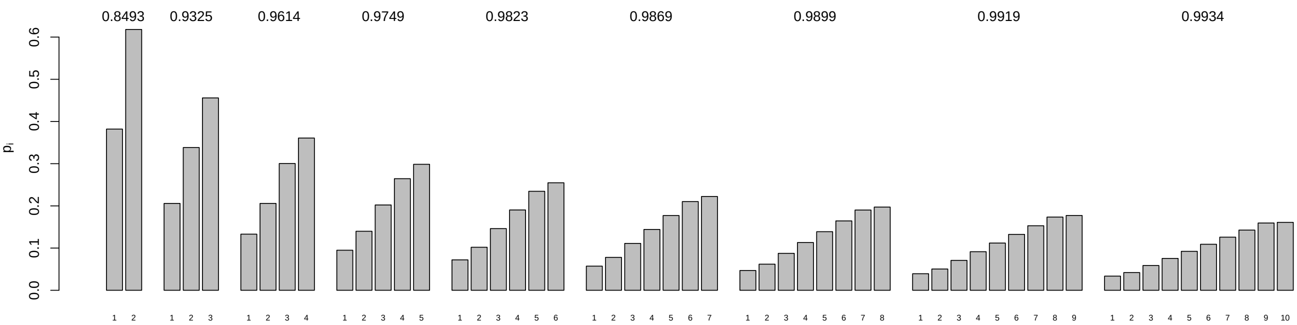


TABLE 5 Limits of the maximum point-polyserial correlation in case of equal-probability categories for the *k*-point scale rv as *k* tends to $+\infty$.

distribution	$\lim_{k\to\infty} \rho_{PP}^{(eq)}(k)$
uniform	1
normal	$\sqrt{3/\pi}$
exponential	$\sqrt{3}/2$
Pareto	$\frac{\sqrt{3\alpha(\alpha-2)}}{2\alpha-1}$
logistic	$3\sqrt{2/\pi^2}$
power	$\sqrt{3\alpha(\alpha+2)}/(2\alpha+1)$

Under the equal-probability setting, the maximum attainable value of the mixed moment is

$$\max \mathbb{E}(XX_d) = \frac{\alpha}{\alpha+1}\left[k - \sum_{i=1}^{k-1} i\left(\frac{i}{k}\right)^{(\alpha+1)/\alpha}\right],$$

and the maximum point-polyserial correlation is

$$\max \rho_{PP}^{(\text{eq})} = \frac{\frac{\alpha}{\alpha+1}\left[k - \frac{k+1}{2} - \sum_{i=1}^{k-1}\left(\frac{i}{k}\right)^{(\alpha+1)/\alpha}\right]}{\sqrt{\frac{k^2-1}{12}\frac{\alpha}{(\alpha+1)^2(\alpha+2)}}} = \frac{\left[k/2 - 1/2 - \sum_{i=1}^{k-1}\left(\frac{i}{k}\right)^{(\alpha+1)/\alpha}\right]}{\sqrt{\frac{k^2-1}{12\alpha(\alpha+2)}}}.$$

In order to evaluate the limit of $\max \rho_{PP}^{(\text{eq})}$ for k tending to infinity, we can approximate the finite sum at the numerator with $\int_0^1 \left(\frac{x}{k}\right)^{(\alpha+1)/\alpha} \mathrm{d}x = \alpha k/(2\alpha+1)$. Then the limiting value can be calculated as

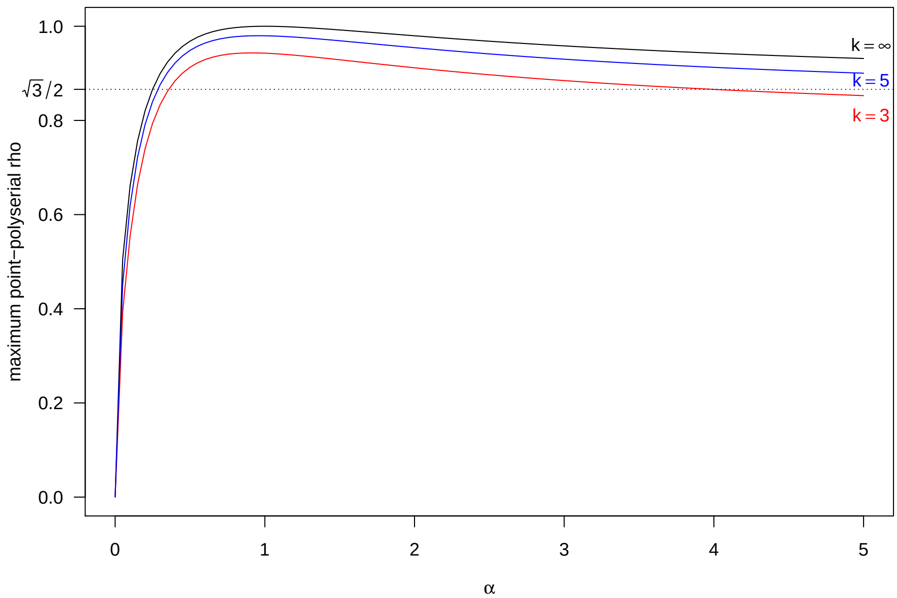
$$\lim_{k\to\infty} \rho_{PP,\max}^{(\text{eq})} = \sqrt{3\alpha(\alpha+2)}/(2\alpha+1).$$

We notice that the limiting value is equal to 1 if and only if $\alpha = 1$, i.e., if we consider a uniform distribution (see also Table 5 for a distributions' summary). For all the other positive values of α, the limit of the maximum point-polyserial correlation - in case of equal probabilities - is strictly smaller than 1. Figure 10 displays the maximum point-polyserial correlation as a function of α for $k = 3; 5; \infty$. As expected, for a fixed α, the point-polyserial correlation increase with k. For a given k, the maximum point-polyserial correlation is attained at $\alpha = 1$ (when the power distributions boils down to a standard uniform).

5 | AN EXAMPLE WITH REAL DATA

Quinn (2004) considered measuring the (latent) political-economic risk of 62 countries for the year 1987. The political-economic risk is defined as the country's risk in manipulating economic rules for its own and its constituents' advantage. Quinn (2004) used five mixed-type variables, namely, the black-market premium in each country (continuous, used as a proxy for illegal economic activity), productivity as measured by the natural logarithm of the real gross domestic product per worker at 1985 international prices (`gdpw2`, continuous), the independence of the national judiciary (dichotomous; 1 if the judiciary is judged

FIGURE 10 Maximum point-polyserial correlation for a power rv as a function of its parameter α, for $k = 3; 5; \infty$, in case of equal-probability discretization. The dotted horizontal line indicates the limit, for k and α both tending to ∞, of the maximum point-polyserial correlation.



to be independent and 0 otherwise), and two ordinal variables (both with levels $0 < 1 < 2 < 3 < 4 < 5$) measuring the lack of expropriation risk (`prsexp`) and lack of corruption (`prscorr`). The data set and a complete description thereof can be found in Quinn (2004) or in the `R` package `MCMCpack` (Martin, Quinn, & Park, 2011). Kadhem and Nikoloulopoulos (2021) applied on this dataset a factor model with bivariate copulas that link the latent variable (which can be interpreted as "political-economic certainty") to each of the observed variables. Here, we just want to apply the results on maximal point-polyserial correlation to (a sample drawn from) a bivariate continuous-ordinal rv; we will consider `gdpw2` as the continuous component and `prsexp` and `prscorr` as two possible ordinal components, which can be assumed to be the result of ordinalization/discretization of some latent continuous variable. Computations show that point-polyserial correlation between `gdpw2` and lack of expropriation risk is 0.4804; the point-polyserial correlation between `gdpw2` and lack of corruption is 0.7250.

Plotting and looking at the histogram and boxplot of the empirical distribution of `gdpw2` and examining its summary statistics, it turns out that it is slighty left-skewed and platykurtic. One can consider fitting a normal and a uniform distribution to these data. Implementing the Kolmogorov-Smirnov test for assessing normality/uniformity for the continuous variable, by adopting the Lilliefors correction in order to take into account the fact that the parameters have to be estimated (Lilliefors, 1967; Novack-Gottshall & Wang, 2019), we obtain a p-value equal to 0.2066 and 0.034, respectively, which means that the distribution of the continuous variable can be hardly assumed to be uniform, but can be more plausibly assumed to be normal.

Taking the two continuous and marginal distributions as assigned, one can compute the maximal (sample) point-polyserial correlation by simply computing the correlation between the two samples sorted in ascending order (Demirtas & Hedeker, 2011), see also Figure 11; we obtain 0.9704 and 0.9531. These values are quite close to the maximum value obtained between a normal rv and a discrete rv with 6 categories, which is 0.9692 (see Figure 3); they are slightly smaller than the maximum point-polyserial correlation between a uniform rv and a discrete rv with 6 categories, which is 0.9860 (see Table 2).

6 | MAXIMAL POINT-POLYSERIAL CORRELATION AS A BASIS FOR DEFINING A *K*-DISCRETE APPROXIMATION OF A CONTINUOUS RANDOM DISTRIBUTION

The oldest and most popular criterion for constructing a k-point approximation of an absolutely continuous rv X, with pdf $f(x)$ and cdf $F(x)$, is based on moment-matching, i.e., matching as many moments as possible of the continuous rv (provided they exist and are finite). Through a discrete rv sitting on k points, it is possible to match $2k-1$ moments; the algorithm that can be used for determining the discrete distribution satisfying this matching is described for example in Golub and Welsch (1969).

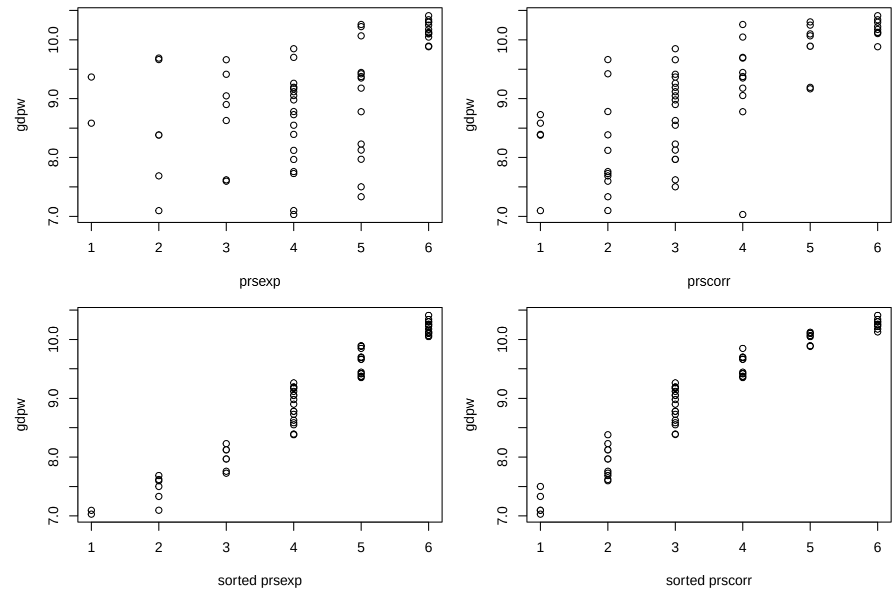


FIGURE 11 Analysis of real data: scatterplots between the continuous and the two ordinal variables before (top panel) and after (bottom panel) reordering

Another way of constructing a k-point discrete approximation is quantization (Lloyd, 1982), which is based on the minimization of the expected squared distance between X and the closest of the k values. Given k values $x_1 < x_2 < \cdots < x_k$, we define the expected squared distance as

$$\mathcal{L}(x_1, x_2, \ldots, x_k) = \mathbb{E} \min_{x_1,\ldots,x_k \in \mathbb{R}} (x - x_i)^2 = \int_{\mathbb{R}} \min_{x_1,\ldots,x_k \in \mathbb{R}} (x - x_i)^2 f(x) \mathrm{d}x.$$

The quantizers $\tilde{x}_1 < \cdots < \tilde{x}_k$ are the values minimizing $\mathcal{L}(x_1, \ldots, x_k)$ (those points minimizing the expected squared distance from X to the closest point) and can be obtained by rewriting the expected squared distance after introducing $k+1$ thresholds or cut-points c_i, $i = 0, 1, \ldots, k$:

$$\mathcal{L}(x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k} \int_{c_{i-1}}^{c_i} (x - x_i)^2 f(x) \mathrm{d}x,$$

where the cut-point c_i is the midpoint between x_i and x_{i+1}, $c_i = (x_i + x_{i+1})/2$ for $i = 1, \ldots, k-1$, and $c_0 = -\infty$, $c_k = +\infty$. The k quantizers are also known as “principal points” (Flury, 1990). To each optimal $\tilde{x}_i$, the probability $\int_{c_{i-1}}^{c_i} f(x)\mathrm{d}x$ remains naturally associated. One can refer to the recent work by Chakraborty, Roychowdhury, and Sifuentes (2021) where the k principal points ($k = 2, \ldots, 8$) of several random distributions have been computed with high numerical precision.

Barbiero and Hitaj (2023) proposed constructing the optimal k-point approximation to a continuous random distribution as the discrete distribution sitting on k distinct values that minimizes a discrepancy measure (the Cramér, Cramér-von Mises, or Anderson-Darling distance) between the two cumulative cdf; their work is based on that of Kennan (2006), where the author distinguishes the case where the approximating points are assigned a priori, and one needs only to compute the optimal probabilities, from the case were the approximating values are not assigned a priori but have to be determined jointly with their probabilities. Barbiero and Hitaj (2021) proposed a similar criterion for constructing a discrete analogue, which is supported over a lattice: $\mathbb{Z}$ if the continuous rv is real or $\mathbb{N}$ if it is positive.

A further alternative for constructing a k-point approximation to a continuous rv X consists of considering the discrete distribution sitting on the first k natural values that maximizes the maximal point-polyserial correlation with X, which we discussed in this work. However, rather than considering $1, 2, \ldots, k$ as the support values, one can instead compute the conditional moments $x_i^* = \int_{F^{-1}(F_{i-1})}^{F^{1}(F_i)} xf(x)\mathrm{d}x/p_i$, $i = 1, \ldots, k$ as the optimal values of the discretization/approximation. A more logical and refined criterion would consist in jointly determine the values x_i and the probabilities p_i that concur to define the discrete rv showing the maximum correlation with the underlying continuous rv. Following the former approach, in Table 6, just as a first comparison, for a standard normal rv, we report the $k = 7$ optimal values and corresponding probabilities of the k-point scale rv maximizing the point-polyserial correlation, of the rv obtained by minimizing the expected square distance (the optimal values are directly taken from Chakraborty et al. (2021), Table 1, A.9), and of the discrete rv obtained by moment matching (Golub & Welsch, 1969), which preserves the $2 \cdot k - 1 = 13$ moments of the parent distribution. Analogously, for an exponential rv with unit rate parameter, we report the 7 optimal values and probabilities calculated according to the three different approaches (for quantization, the optimal values are directly taken from Chakraborty et al. (2021), Table 2, A.9). For both continuous distributions, differences across values and probabilities can be easily appreciated and after all were expected, since the criteria by which we obtained the optimal approximations are based on different rationales. In particular, moment matching produces discrete rvs with a larger range and tends to assign small probabilities to extreme values: one can just notice the values in the last column of Table 6.

TABLE 6 Optimal *k*-point discrete approximations of a standard normal rv and of an exponential with unit parameter

standard normal						exponential with unit parameter					
max. point-polyserial		quantization		moment matching		max. point-polyserial		quantization		moment matching	
values	probabilities	values	probabilites	values	probabilities	values	probabilites	values	probabilites	values	probabilites
–2.0473	0.0519	–2.0334	0.0536	–3.7504	0.0005	0.2785	0.4625	0.1986	0.3479	0.1930	0.4093
–1.2568	0.1126	–1.1881	0.1373	–2.3668	0.0308	0.9537	0.2865	0.6565	0.2563	1.0267	0.4218
–0.6282	0.2080	–0.5606	0.1987	–1.1544	0.2401	1.7153	0.1338	1.1972	0.1787	2.5679	0.1471
0	0.2551	0	0.2207	0	0.4571	2.4768	0.0625	1.8574	0.1150	4.9004	0.0206
0.6282	0.2080	0.5606	0.1987	1.1544	0.2401	3.2383	0.0292	2.7053	0.0652	8.1822	0.0011
1.2568	0.1126	1.1881	0.1373	2.3668	0.0308	3.9998	0.0136	3.8925	0.0294	12.7342	0.0000
2.0473	0.0519	2.0334	0.0536	3.7504	0.0005	5.4284	0.0119	5.8925	0.0075	19.3957	0.0000

7 | CONCLUSION

The object of this work was studying the range of the point-polyserial correlation for several (non-normal) bivariate distributions and, in particular, determining the maximal attainable value as a function of the distribution parameters and of the number k of the ordered categories into which one of the continuous distributions is discretized. Finding the expression of the maximal point-polyserial correlation is often possible (its derivation is related to the availability of closed-form expressions for partial moments of the continuous distribution) and it can be easily evaluated under any mathematical and statistical software like R; just as easily, one can find the maximum value of the maximal point-polyserial correlation for a given k numerically, by using standard constrained optimization routines. Several examples concerning well-known parametric continuous distributions are detailed and indicate that the maximum point-polyserial correlation, computed over all the k-point discrete distributions sitting on $\{1, 2, \ldots, k\}$, is attained at a distribution whose probability values are strictly connected to the continuous random distribution examined: if the continuous distribution is unimodal and symmetrical (e.g., normal and logistic distribution), then the corresponding discrete distribution is unimodal and symmatrical, too; if the continuous distribution is uniform, then the corresponding discrete distribution is a discrete uniform; in case of a decreasing pdf (exponential, Pareto, power), then the k probabilities (under some circumstances) are decreasing as well. From the numerical experiments, it turns out that whatever the continuous distribution is, the maximal point-polyserial correlation always tends to 1 as the number of categories tends to infinity. We also focused on the equal-probability setting and determined the limiting value of the maximal point-polyserial correlation as the number of categories tends to infinity: we find out that in all cases, except – as expected – for the uniform distribution, this limiting value is strictly smaller than 1. We remark that in our analysis we have always assumed that the k ordered categories of the ordinalized rv are assigned the first k positive integers, which seems to be a natural choice, as ordinal variables are standardly handled in this way when it comes to implement any statistical analysis. This can be questionable, however, and one can think of letting the scores of the k categories be unknown and treat them as additional variables to be included into the optimization problem; however, this would introduce further complexity into the theoretical framework and would not allow us to derive some nice results as those discussed here.

With this in mind, future research will investigate the properties of the k-point discrete distribution that maximizes the (maximal) point-polyserial correlation: are there any cases for which the probabilities of this discrete distribution can be determined analytically and not just numerically? Can these probabilities be determined analytically as $k \to \infty$? Can this discrete distribution be regarded as a valid k-point approximation of the parent continuous distribution? What are the main differences with other k-point approximations available in the literature, such as those obtained through quantization, or moment matching, or based on the minimization of some discrepancy measure?

Another future direction of this contribution will consider the determination of the minimum attainable point-polyserial correlation, following the same lines of investigation as in Sections 3 and 4; such complemented work could be helpful for random generation routines involving mixed-type data, by providing a lower and an upper bound to the correlation between ordinal and continuous variables, which can be required when constructing a huge array of artificial scenarios for the assessment of some mixed-type data analysis technique. Being aware of the bounds of point-polyserial correlation is also obviously useful when interpreting its sample value on a real data set.

SUPPORTING INFORMATION

Relevant `R` code implementing the numerical evaluations of the maximum point-polyserial correlation of Sections 3 and 4, and the real data analysis of Section 5 is available as supplementary material.

AUTHOR CONTRIBUTIONS

The authors equally contributed to the manuscript and have agreed to the submitted version.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

References

Barbiero, A., & Hitaj, A. (2021). A new method for building a discrete analogue to a continuous random variable based on minimization of a distance between distribution functions. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 338–341).

Barbiero, A., & Hitaj, A. (2023). Discrete approximations of continuous probability distributions obtained by minimizing Cramér-von Mises-type distances. *Statistical Papers*, *64*(5), 1669-1697.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *42*(3), 293–312.

Bedrick, E. J. (1995). A note on the attenuation of correlation. *British Journal of Mathematical and Statistical Psychology*, *48*(2), 271–280.

Chakraborty, S., Roychowdhury, M. K., & Sifuentes, J. (2021). High precision numerical computation of principal points for univariate distributions. *Sankhya B*, *83*, 558–584.

Cheng, Y., & Liu, H. (2016). A short note on the maximal point-biserial correlation under non-normality. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 344–351.

Cox, N. (1974). Estimation of the correlation between a continuous and a discrete variable. *Biometrics*, 171–178.

Demirtas, H., & Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. *The American Statistician*, *65*(2), 104–109.

Demirtas, H., & Hedeker, D. (2016). Computing the point-biserial correlation under any underlying continuous distribution. *Communications in Statistics-Simulation and Computation*, *45*(8), 2744–2751.

Demirtas, H., & Vardar-Acar, C. (2017). Anatomy of correlational magnitude transformations in latency and discretization contexts in Monte-Carlo studies. *Monte-Carlo Simulation-Based Statistical Modeling*, 59–84.

Embrechts, P., McNeil, A. J., & Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In M. A. H. Dempster (Ed.), *Risk management: Value at risk and beyond* (p. 176–223). Cambridge University Press. doi: 10.1017/CBO9780511615337.008

Fernández, D., Liu, I., Costilla, R., & Gu, P. Y. (2020). Assigning scores for ordered categorical responses. *Journal of Applied Statistics*, *47*(7), 1261–1281.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(4), 521–532.

Flury, B. A. (1990). Principal points. *Biometrika*, *77*(1), 33–41.

Foldnes, N., & Grønneberg, S. (2015). How general is the Vale–Maurelli simulation approach? *Psychometrika*, *80*, 1066–1083.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3^ e serie, Sciences, Sect. A*, *14*, 53–77.

Ghalanos, A., & Theussl, S. (2015). Rsolnp: General non-linear optimization using augmented lagrange multiplier method [Computer software manual]. R package version 1.16..

Golub, G. H., & Welsch, J. H. (1969). Calculation of gauss quadrature rules. *Mathematics of computation*, *23*(106), 221–230.

Hoeffding, W. (1940). Masstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin*, *5*, 181–233.

Hofert, M., Kojadinovic, I., Maechler, M., & Jun, Y. (2023). copula: Multivariate dependence with copulas [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=copula/` R package version 1.1-2.

Ivanova, A., & Berger, V. W. (2001). Drawbacks to integer scoring for ordered categorical data. *Biometrics*, *57*(2), 567–570.

Kadhem, S. H., & Nikoloulopoulos, A. K. (2021). Factor copula models for mixed data. *British Journal of Mathematical and Statistical Psychology*, *74*(3), 365–403.

Kennan, J. (2006). A note on discrete approximations of continuous distributions. *University of Wisconsin-Madison*.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, *28*(2), 129–137.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1), 19.

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, *42*(9), 22. doi: 10.18637/jss.v042.i09

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools*. Princeton University Press.

Narasimhan, B., Johnson, S. G., Hahn, T., Bouvier, A., & Kiêu, K. (2023). cubature: Adaptive multivariate integration over hypercubes [Computer software manual]. Retrieved from `https://bnaras.github.io/cubature/` R package version 2.1.0.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, *15*, 625–632.

Novack-Gottshall, P., & Wang, S. C. (2019). KScorrect: Lilliefors-corrected Kolmogorov-Smirnov goodness-of-fit tests [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=KScorrect` R package version 1.4.0.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–347.

Pearson, K. (1909). On a new method of determining correlation between a measured character A, and a character B, of which only the percentage of cases wherein B exceeds (or falls short of) a given intensity is recorded for each grade of A. *Biometrika*, *7*(1/2), 96–105.

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis*, *12*(4), 338–353.

Ye, Y. (1987). *Interior algorithms for linear, quadratic, and linearly constrained non-linear programming* (Unpublished doctoral dissertation). Department of Engineering-Economic Systems, Stanford University.