Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

On estimating a transformation correlation coefficient

Kelly H. Zou ;W. J. Hall

To cite this Article Zou, Kelly H. andHall, W. J.(2002) 'On estimating a transformation correlation coefficient', Journal of Applied Statistics, 29: 5, 745 — 760
To link to this Article: DOI: 10.1080/02664760120098801
URL: http://dx.doi.org/10.1080/02664760120098801

Journal of Applied Statistics, Vol. 29, No. 5, 2002, 745–760

On estimating a transformation correlation coefficient

KELLY H. ZOU[1,2] **& W. J. HALL**[3], [1]*Department of Health Care Policy, Harvard Medical School, Boston, USA,* [2]*Department of Radiology, Brigham and Women's Hospital, Boston, USA and* [3]*Department of Biostatistics, University of Rochester, USA*

ABSTRACT *We consider a semiparametric and a parametric transformation-to-normality model for bivariate data. After an unstructured or structured monotone transformation of the measurement scales, the measurements are assumed to have a bivariate normal distribution with correlation coefficient ρ, here termed the 'transformation correlation coefficient'. Under the semiparametric model with unstructured transformation, the principle of invariance leads to basing inference on the marginal ranks. The resulting rank-based likelihood function of ρ is maximized via a Monte Carlo procedure. Under the parametric model, we consider Box–Cox type transformations and maximize the likelihood of ρ along with the nuisance parameters. Efficiencies of competing methods are reported, both theoretically and by simulations. The methods are illustrated on a real-data example.*

1 Introduction

Let (X_0, Y_0) represent a bivariate random variable with an absolutely continuous distribution, and $(X_{0,i}, Y_{0,i})$ $(i = 1, \ldots, n)$ a random sample from this distribution. We assume that there exist two monotone increasing transformations ψ_1 and ψ_2 from the measurement scales of X_0 and Y_0, respectively, for which $X \equiv \psi_1(X_0)$ and $Y \equiv \psi_2(Y_0)$ jointly have a standard bivariate normal distribution, with 0 means, unit variances, and correlation coefficient ρ. (If the marginal distribution functions of X_0 and Y_0 are F_1 and F_2, then $\psi_i = \Phi^{-1} \circ F_i$ $(i = 1, 2)$ with Φ the standard normal distribution function, but the F_is are unknown). The parameter ρ, which we call the 'transformation correlation coefficient', serves as a measure of the relationship between the variables X_0 and Y_0 and, of course, differs from the product-moment correlation coefficient unless only linear transformations are required. We consider

Correspondence: K. H. Zou, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, Massachusetts 02115, USA. E-mail: zou@hcp.med.harvard.edu.

ISSN 0266-4763 print; 1360-0532 online/02/050745-16
DOI: 10.1080/02664760120098801

here estimation of ρ, first in this semiparametric transformation model without further assumptions about the required transformations, and secondly in a Box-Cox type parametric transformation model. We compare our estimates with others derived from various rank correlation coefficients.

Many statistical works have examined the linear or monotone relationships between paired-data samples (see, for example, Kendall & Gibbons, 1990 for a review; also Daniels, 1944; Hall, 1970; Klaassen & Wellner, 1997, and references therein). The former (linear) is commonly measured by the sample product-moment correlation coefficient and is invariant to only linear transformations of the marginal data. The latter (monotone) is measured using rank correlation coefficients such as (1) *di erence sign* (or 'Kendall's tau'; see Esscher, 1924; Kendall, 1938; Hoeffding, 1947; Kruskal, 1958; Hettmansperger, 1991), (2) *rank* (or 'Spearman's rho'; see Spearman, 1904; Fieller *et al.*, 1957; Kruskal, 1958; David & Mallows, 1961; Fieller & Pearson, 1961), and (3) *normal-scores* (or 'Fisher-Yates'; see Fisher & Yates, 1948; Hájek & Šidák, 1967; van der Waerden, 1952; Klaassen & Wellner, 1997). Rank correlation coefficients are invariant to any monotone transformations of the measurement scales. When using rank data, it is immaterial whether the actual joint distribution is bivariate normal or whether it is bivariate normal only after unspecified monotone transformations. The transformation correlation coefficient ρ can be estimated from these various rank correlation coefficients (see Appendix 4).

Doksum (1987) considered a general class of transformation models, and proposed a Monte Carlo 'likelihood sampler' method to estimate parameters in them. Our semiparametric estimation method may be considered a modification of his likelihood sampler, with greater efficiency in the Monte Carlo step.

Klaassen & Wellner (1997) also considered such a transformation-to-bivariate-normality model, which they termed a 'normal copula model' and termed ρ the 'normal correlation coefficient'. They proved that the normal scores correlation coefficient (see Appendix 4) is an asymptotically efficient estimate of ρ. Our semiparametric model estimate provides an alternative asymptotically efficient estimate. Although not as easily computed, it has some intuitive appeal and may have somewhat differing behaviour in finite samples; moreover, it has potential for generalization in some settings where a normal scores approach does not; see Section 7.

This article is organized as follows. In Sections 2 and 3, we develop the semiparametric and parametric transformation estimates for ρ. Under the semiparametric model, we base inference about ρ only on the ranks of each coordinate, with the transformations left unspecified. We construct and maximize the likelihood function using an extension of Hoeffding's (1951) Theorem to produce an estimate of ρ. Under the parametric model, the structured transformations ψ_1 and ψ_2, along with other nuisance parameters, are first estimated by maximizing the likelihood function, followed by estimating ρ using the product-moment correlation coefficient after these estimated transformations. In Section 4, we describe construction of confidence intervals for ρ, using Fisher's (1915) z-transformation method, and report and compare the asymptotic (large-sample) variances and efficiencies associated with the various methods of estimating ρ. In Section 5, we illustrate six competing methods on a published data set. A Monte Carlo efficiency study is summarized in Section 6. Finally, Section 7 provides discussion, concluding remarks and some potential extensions.

Programs in S-PLUS, for computation of the semiparametric and parametric model estimates, are available from the first author.

2 A semiparametric transformation model

In order to estimate ρ in this model, defined in Section 1, we base inference on the ranks of the data in each of the two coordinates. The rank-based likelihood can be derived by the following theorem, an extension of Hoeffding's (1951) Theorem to this bivariate case.

Theorem 1

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from an absolutely continuous bivariate distribution with joint density function $f(x,y)$ and marginal density functions $f(x)$ and $g(y)$. Let $\underline{R} = (R_1, \ldots, R_n)$ and $\underline{S} = (S_1, \ldots, S_n)$ denote the marginal ranks of Xs and Ys. Consider two absolutely continuous univariate distributions with densities h_1 and h_2 such that $f(x,y) = 0$ whenever $h_1(x)h_2(y) = 0$, and define $L_f(z_1, z_2) = f(z_1, z_2)/\{h_1(z_1)h_2(z_2)\}$. Further, for each $k = 1, 2$, let $Z_{k,(1)} < \ldots < Z_{k,(n)}$ denote the (unobserved) order statistics of a random sample of size n from h_k. Then

$$\Pr\{\underline{R} = \underline{r}, \underline{S} = \underline{s}\} = \frac{1}{(n!)^2} E_{h_1,h_2}\left\{\prod_{i=1}^{n} L_f(Z_{1,(r_i)}, Z_{2,(s_i)})\right\} \tag{1}$$

(see Appendix 1 for proof).

In our application, we take $f(x,y) = \phi_\rho(x,y)$ (standard bivariate normal) and $f(x) = g(x) = \phi(x)$ (standard normal) and $h_1 = h_2 = \phi$. The resulting distribution of the ranks in (1) depends only on ρ. This leads to the following likelihood function of ρ for the observed ranks $\underline{r}$ and $\underline{s}$ of the $\underline{X}_0$ and $\underline{Y}_0$ samples:

$$L(\rho \mid \underline{r}, \underline{s}) \propto k_1(\rho) E_{Z_1,Z_2}\left[\exp\left\{k_2(\rho)\sum_{i=1}^{n}\left(\rho Z_{1,(r_i)}^2 - 2Z_{1,(r_i)}Z_{2,(s_i)} + \rho Z_{2,(s_i)}^2\right)\right\}\right] \tag{2}$$

where $k_1(\rho) \equiv (1-\rho^2)^{-n/2}$ and $k_2(\rho) \equiv -\rho/\{2(1-\rho^2)\}$. Here, $Z_{k,(1)} < \ldots < Z_{k,(n)}$ are the order statistics of a sample (unobserved) of size n from independent $N(0,1)$ distributions $(k = 1, 2)$; we associate the $Z_{1,(r_i)}$s with the sample of X_0s and the $Z_{2,(s_i)}$s with the sample of Y_0s.

Denote residuals from the marginal mean $\bar{Z}_1 \equiv \sum_{i=1}^{n} Z_{1,i}/n$ for the X_0-sample by $D_{1,i} \equiv Z_{1,i} - \bar{Z}_1$, and similarly from $\bar{Z}_2 \equiv \sum_{i=1}^{n} Z_{2,i}/n$ for the Y_0-sample by $D_{2,i} \equiv Z_{2,i} - \bar{Z}_2$, and define $A \equiv \sum_{i=1}^{n}(D_{1,i}^2 + D_{2,i}^2)/2$ and $B \equiv \sum_{i=1}^{n}(D_{1,i}D_{2,i})$, the latter being a function of the observed rank vectors $\underline{r}$ and $\underline{s}$. Then the expectation in (2) becomes

$$E_{A,B,\bar{Z}_1,\bar{Z}_2}[\exp\{2k_2(\rho)\ (\rho A - B) + nk_2(\rho)\ (\rho\bar{Z}_1^2 + \rho\bar{Z}_2^2 - 2\bar{Z}_1\bar{Z}_2)\}] \tag{3}$$

By a conditioning argument, (3) becomes

$$E_{A,B}[\exp\{U(\rho, A, B)\}E_{\bar{Z}_1,\bar{Z}_2|A,B}\exp(V)] \tag{4}$$

where $U(\rho, A, B) \equiv 2k_2(\rho)\ (\rho A - B) = -(\rho/1-\rho^2)\ (\rho A - B)$, $V \equiv c(\rho\bar{Z}_1^2 + \rho\bar{Z}_2^2 - 2\bar{Z}_1\bar{Z}_2)$ and $c \equiv nk_2(\rho)$. The inner conditional expectation in (4), $E_{\bar{Z}_1,\bar{Z}_2|A,B}\{\exp(V)\}$, may be obtained mathematically, yielding $(1-\rho^2)^{1/2}$ (see Appendix 2). The following proposition summarizes these results.

Proposition 1

The likelihood function for ρ, given observed ranks $\underline{r}$ and $\underline{s}$ of the marginal samples of X_0s and Y_0s, is proportional to

$$(1-\rho^2)^{(1-n)/2} \cdot E_{A,B}\left[\exp\left\{-\frac{\rho}{1-\rho^2}(\rho A - B)\right\}\right] \tag{5}$$

where

$$A = \sum_{i=1}^{n} \{(Z_{1,(r_i)} - \bar{Z}_1)^2 + (Z_{2,(s_i)} - \bar{Z}_2)^2\}/2 = \sum_{i=1}^{n} \{(Z_{1i} - \bar{Z}_1)^2 + (Z_{2i} - \bar{Z}_2)^2\}/2,$$

$$B = \sum_{i=1}^{n} (Z_{1,(r_i)} - \bar{Z}_1)(Z_{2,(s_i)} - \bar{Z}_2),$$

and $Z_{k,(1)} < \ldots < Z_{k,(n)}$ are unobserved order statistics from two independent samples of size n from $N(0, 1)$ $(k = 1, 2)$.

Now A is proportional to a chi-square random variable, but A and B are dependent, so no further evaluation in (5) appears possible. The expectation w.r.t. the As and Bs in (5) may be obtained by a Monte Carlo procedure (see the algorithm in Appendix 3). Doksum's (1987) 'likelihood sampler' adapted for this problem would evaluate the expectation in (2) by Monte Carlo whereas we have done part of the computation mathematically, and thereby reduced the Monte Carlo variability.

The root of the resulting score equation is determined numerically to obtain the MLE $\hat{\rho}_S$ of ρ in this semiparametric model.

3 A parametric transformation model

We consider a Box & Cox (1964) parametric class of power transformations instead of unstructured ψ_1 and ψ_2. (We assume that the bivariate measurements take on positive values ($\mathscr{R}^+$); if not, first transform them to $\mathscr{R}^+$ either by exponentiating or by using a logit type of transformation.)

We assume that the data may be transformed to bivariate normality in two steps, from (X_0, Y_0) to (X', Y') to (X, Y): let ψ_1' and ψ_2' be Box-Cox power transformations with parameters (often called 'power coefficients') λ_1 and λ_2:

$$X' \equiv \psi'(X_0) \equiv (X_0^{\lambda_1} - 1)/\lambda_1, \qquad Y' \equiv \psi_2'(Y_0) \equiv (Y_0^{\lambda_2} - 1)/\lambda_2$$

(and $\equiv \log X_0$ or $\log Y_0$ if λ_1 or λ_2 is 0), and let

$$X \equiv \psi(X_0) \equiv (X' - \mu)/\sigma, \qquad Y \equiv \psi(Y_0) \equiv (Y' - \nu)/\tau$$

We assume a bivariate normal distribution for (X', Y'), with means (μ, ν), variances (σ^2, τ^2), and correlation coefficient ρ, and hence (X, Y) has a standard bivariate normal distribution with correlation coefficient ρ. (It should be noted that exact normality is only possible when both power coefficients are zero (log normal transformations); otherwise the range after transformation is not the whole real line. However, approximate normality is adequate for most purposes.)

There are seven parameters in this model: $(\rho, \lambda_1, \lambda_2, \mu, \nu, \sigma, \tau)$ in which the transformation correlation ρ is the parameter of interest and the rest are nuisance

parameters. The log likelihood associated with observed sample values x_0s and y_0s is analogous to that for the univariate case (Hernandez & Johnson, 1980):

$$l(\rho, \lambda_1, \lambda_2, \mu, \nu, \sigma, \tau | \underset{\sim}{x}_0, \underset{\sim}{y}_0) = -\frac{n}{2} \log(1-\rho^2) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} (x_i^2 - 2\rho x_i y_i + y_i^2) + (\lambda_1 - 1) \sum_{i=1}^{n} \log x_{0,i} + (\lambda_2 - 1) \sum_{i=1}^{n} \log y_{0,i} - n \log(2\pi\sigma\tau) \quad (6)$$

Note that x_i is a function of $(x_{0,i}, \lambda_1, \mu, \sigma)$ and y_i a function of $(y_{0,i}, \lambda_2, \nu, \tau)$.

To carry out the maximization, we proceed in two steps: for a grid of potential λ_1- and λ_2-values, maximize equation (6) w.r.t. the five bivariate normal parameters by substituting corresponding sample statistics of the $x_i'(\lambda_1)$s and $y_i'(\lambda_2)$s—that is, sample means, standard deviations and correlation coefficient. Then equation (6) is calculated at each point on the grid, locating the maximizing grid point. Both steps are repeated, using a finer grid in this neighbourhood, etc. It should be noted that this simultaneous determination of both λs is different from the less efficient separate marginal determination of each λ; see Section 5.

Users may prefer to limit the power coefficients to a small set of values, perhaps a few integers and half-integers; if so, only these need be used in the grid at step 1.

A large-sample variance for $\hat{\rho}_P$ could be derived by obtaining the 7×7 sample information matrix and inverting it. However, we argue in Section 4 that its large-sample variance is the same as if the power coefficients were known. This is in sharp contrast to findings of Bickel & Doksum (1981) regarding estimation of some other parameters in an analogous Box-Cox model.

The validity of the parametric transformation to bivariate normality may be examined. We propose using the Z_2-test of Mudholkar *et al.* (1992) for bivariate normality, and the Z-test of Lin & Mudholkar (1980) for testing univariate normality.

4 Some competitors, the z-transformation, asymptotic variances and efficiencies

We compare our semiparametric transformation estimate $\hat{\rho}_S$ and parametric Box-Cox transformation estimate $\hat{\rho}_P$ with three rank-based estimates and the sample correlation coefficient r. The other rank-based methods are derived from the difference-sign correlation coefficient of Kendall, the rank correlation coefficient of Spearman, and the Fisher-Yates or van der Waerden normal scores correlation coefficient; see Appendix 4 and references listed in Section 1.

For inference about ρ from an estimate $\hat{\rho}$, we may use Fisher's (1915) z-transformation, defined as

$$z = z(\hat{\rho}) \equiv \frac{1}{2} \log\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) = \tanh^{-1}(\hat{\rho})$$

and $\zeta = z(\rho)$. For product-moment $r = \hat{\rho}$ and ρ, this is a variance-stabilizing and skewness-reducing transformation (e.g. Mudholkar, 1983): z has an asymptotically normal distribution around ζ, with a ρ-free asymptotic variance $1/n$ and with a skewness coefficient of order $1/n^{3/2}$ (compared with order $1/n^{1/2}$ for r). To infer about ρ, one instead infers about ζ and then back-transforms to make inference

about ρ. We propose doing the same when inferring about the transformation correlation ρ, whatever the estimate $\hat{\rho}$. We can expect similar, although perhaps not as dramatic, advantages.

Specifically, by the delta method, if $\hat{\rho}$ has asymptotic variance v_n, then $z(\hat{\rho})$ has asymptotic variance $w_n = v_n/(1-\rho^2)^2$. Hence, a confidence interval for ρ, with asymptotic confidence coefficient 95%, is

$$z^{-1}\{z(\hat{\rho}) \pm 1.96\sqrt{\hat{w}_n}\} \tag{7}$$

with $z^{-1}(\zeta) \equiv \tanh(\zeta) = \{\exp(2\zeta) - 1\}/\{\exp(2\zeta) + 1\}$.

We now give values of v_n corresponding to various $\hat{\rho}$s, from which the corresponding w_ns are readily obtained; $w_n(\hat{\rho})$ provides a needed estimate $\hat{w}_n$.

If no transformation is needed to achieve bivariate normality, the sample product-moment r is well-known to be efficient, with $n \cdot v_n \sim (1-\rho^2)^2$ and hence $n \cdot w_n \sim 1$; otherwise, r estimates the product-moment ρ rather than the transformation ρ. This variance can be attained if the transformations were known. Hence, asymptotic relative efficiencies of other methods are given simply by $1/(nw_n)$. (Asymptotic efficiencies for inference about ρ are identical to those for inference about ζ.)

The normal scores $\hat{\rho}_N$ (see Appendix 4) is an asymptotically efficient estimate of ρ and hence has the same v_n and w_n as does r, that is, $n \cdot w_n \sim 1$.

Our $\hat{\rho}_S$, being the rank-based MLE, should be fully efficient among rank-based methods (although no formal proof is given here); since the rank-based $\hat{\rho}_N$ is fully efficient, then so must be $\hat{\rho}_S$, with $n \cdot w_n \sim 1$.

Our $\hat{\rho}_P$ is the MLE within the power-transformation model, and hence is efficient there. This model is sandwiched between the semiparametric model of Section 2 and a bivariate normal model—that is, a Box-Cox model with linear transformations. Hence, the large-sample variance of $\hat{\rho}_P$ must be sandwiched between that for $\hat{\rho}_S$ (or $\hat{\rho}_N$) and that of r. Since the two extreme models have identical large-sample variances, we conclude that the large-sample variance of $\hat{\rho}_P$ is also $(1-\rho^2)^2/n$, and $n \cdot w_n \sim 1$. However, if the Box-Cox model is incorrect, the estimate is not even consistent. Likewise, unless ψ_1 and ψ_2 are linear, r is not consistent.

Hence, r, $\hat{\rho}_N$, $\hat{\rho}_S$ and $\hat{\rho}_P$ each have asymptotic variances for their z-transforms equal to $1/n$ (when the corresponding models are correct).

In Appendix 4, we derive two other rank-based estimates of ρ. One, denoted $\hat{\rho}_D$, is based on Kendall's difference-sign correlation coefficient, and the other, denoted $\hat{\rho}_R$, on Spearman's rank correlation coefficient; their large-sample variances are also given there. Both estimates have larger v_ns and w_ns than do the estimates above. Their asymptotic efficiencies when estimating the transformation correlation ρ (or when estimating a bivariate normal ρ) are given in Table 1. For small ρ, about 9% efficiency is lost, with greater losses elsewhere.

Finite-sample efficiency has been investigated in a simulation study (Section 6).

TABLE 1. Asymptotic relative efficiencies of the difference-sign-based and rank-correlation-based estimates of ρ

ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
D	0.912	0.911	0.909	0.905	0.899	0.892	0.882	0.872	0.859
R	0.912	0.910	0.905	0.896	0.884	0.867	0.846	0.820	0.791

D = Difference sign; *R* = Rank.

5 An example

We illustrate this with the following fish-stock example (Ricker & Smith, 1975; Carroll & Ruppert, 1988), where the bivariate data consist of the numbers of spawners and recruits in the 28 years for the Skeena River sockeye salmon stock. A rockslide occurred in 1951, resulting in a drastically reduced number of recruits during that year. Hence, we drop this observation and have sample size $n = 27$. We estimate the transformation correlation coefficient ρ based on six correlation coefficients, namely difference sign (D), rank (R), normal scores (N), semiparametric transformation (S), parametric transformation (P) and sample (r), the latter two requiring additional assumptions for validity.

Our parametric transformation program yields the estimated power coefficients $\hat{\lambda}_1 = 0.473$ and $\hat{\lambda}_2 = 0.007$. These transformations reduce the skewness coefficient of the y sample from 0.807 to -0.051. However, the transformations make the x-sample slightly more skewed with coefficients 0.186 (before) and -0.344 (after). The Z_2-test for bivariate normality gives a p-value of 0.09 after the transformations, in contrast to a prior p-value of 0.01. The resulting correlation coefficient is labelled $\hat{\rho}_P$. Hence, in this example, square-root ($\lambda_1 = 0.5$) and log ($\lambda_2 = 0$) transformations make the data roughly bivariate normal.

If we instead ignore the dependency structure and maximize the marginal likelihood functions for each of the samples, we find the estimated power coefficients to be $\hat{\lambda}_1 = 0.696$ and $\hat{\lambda}_2 = 0.140$, slightly different from those obtained by our bivariate method, and asymptotically inefficient. Moreover, the p-value from the Z_2-test for bivariate normality is now 0.04, although the data are marginally normal after these transformations (with p-values from the Z-test for normality of 0.78 and 0.96).

In our example, the difference sign (Kendall's tau) and rank (Spearman's rho) correlation coefficients (see Appendix 4) are $t = 0.368$ and $r_S = 0.503$, which are used to obtain the estimates $\hat{\rho}_D$ and $\hat{\rho}_R$, respectively. We chose a Monte Carlo size of 50 000 in our semiparametric estimation method, resulting in $\hat{\rho}_S$. We also obtained the normal scores correlation coefficient $\hat{\rho}_N$.

We then z-transformed each of the above correlation coefficients. The standard errors of the resulting $\hat{\zeta}$s were estimated ($se = \sqrt{\hat{w}_n}$); then confidence intervals for ζ, and hence ρ, were constructed from (7).

The estimates $\hat{\rho}$s and $\hat{\zeta}$s with their standard errors, along with 95% confidence intervals for ρ, appear in Table 2.

All of the methods yield quite similar results for this example, especially the (N), (S) and (P) methods; that method (P) agrees with the others is a reflection of the

TABLE 2. Estimates $\hat{\rho}$s and $\hat{\zeta}$s (with s.e.s) and 95% confidence intervals (CIs) for the fish-stock example

Method	$\hat{\rho}$	$\hat{\zeta}$	CI for ρ
D	0.546 (0.152)	0.613 (0.217)	(0.186, 0.777)
R	0.521 (0.151)	0.578 (0.207)	(0.170, 0.755)
N	0.563 (0.131)	0.637 (0.192)	(0.254, 0.768)
S	0.565 (0.131)	0.640 (0.192)	(0.257, 0.769)
P	0.557 (0.133)	0.628 (0.192)	(0.246, 0.764)
r	0.512 (0.142)	0.565 (0.192)	(0.186, 0.736)

D = Difference sign; R = Rank; N = Normal scores; S = Semiparametric transformation; P = Parametric Box-Cox transformation.

possible adequacy ($p = 0.09$) of the fit of the Box-Cox model. Note that the sample version r is slightly smaller than the others, but it is estimating the transformation correlation coefficient ρ only if the data are bivariate normal without transformation, for which we found $p = 0.01$. All of the confidence intervals are wide due to the small sample size $n = 27$.

6 A simulation study

We conducted a Monte Carlo simulation study to investigate and compare the adequacy of the large-sample approximations for the various estimates, for several sample sizes and several ρ-values. A summary is presented here; fuller results are available in Zou (1997).

For each of three sample sizes $n = 20$, 50 and 100, we generated 500 replications of $2n$ $N(0, 1)$ random variables in S-PLUS. We labelled n of them as Xs and the remaining n as Zs. For each of the following true parameter values: $\rho = 0$, 0.5, 0.7, 0.8 or 0.9, we defined $Y_i = \rho X_i + \sqrt{1 - \rho^2} Z_i$ $(i = 1, \ldots, n)$, so that the correlation coefficient between X_i and Y_i is ρ. Hence, we have a total of 15 combinations of sample size and parameter values.

We estimated ρ semiparametrically (S) (with Monte Carlo size of 1000) and parametrically (P), along with difference sign (D), rank (R), and normal scores (N) methods, and the sample correlation coefficient (r). For the parametric method (P), we exponentiated the bivariate data prior to it being input into the computation program, as if the data were from a bivariate log-normal distribution.

Hence, for each of the 15 combinations, we recorded 500 $\hat{\rho}$s by each of the six methods. The mean, standard deviation, skewness coefficient and excess kurtosis coefficient of the 500 values were computed for each method, and designated Monte Carlo summary statistics. Then we z-transformed the $\hat{\rho}$s to $\hat{\zeta}$s, and again calculated Monte Carlo summary statistics. Subtracting true ρs or ζs from Monte Carlo means yielded bias estimates of the $\hat{\rho}$s and $\hat{\zeta}$s. Comparison of square roots of asymptotic variance formulas with Monte Carlo standard deviations provided evaluations of the accuracy of asymptotic formulas for standard errors compared with finite-n computations. In addition, for each method, we recorded Monte Carlo coverage probabilities of 95% confidence intervals for ρ, obtained by the back-transform method of Section 4.

Selected simulation results are presented in Tables 3 to 5, all for the case with $n = 50$, for $\rho = 0$, 0.5, 0.7 and 0.9, and for each of the six methods. Table 3 presents summary statistics for $\hat{\rho}$ and Table 4 for $\hat{\xi}$. Table 5 is for the Box-Cox method only; means of the estimated power coefficients $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are themselves bias estimates since the true λs are both 0 (a log transformation on each coordinate); Monte Carlo standard deviations are also reported.

Some of the conclusions to be drawn are as follows.

(i) There is little bias in $\hat{\rho}$ or $\hat{\zeta}$ for any of the methods or ρ-values. The z-transform succeeds in reducing the skewness for all methods and ρs, and there is little excess kurtosis.

(ii) Comparing standard errors computed from large-sample formulas (se) with the Monte Carlo standard deviations (sd_M) in Table 4, a general conclusion is that the large-sample formulas slightly underestimate the true variability. The underestimation is more pronounced for the rank-based methods at $\rho = 0.9$, but standard errors for parametric methods (P) and (r) are quite

TABLE 3. A comparison of six methods for estimating ρ, without the z-transformation

ρ	Method	$\hat{\rho} - \rho$	$\mathrm{sd}_M(\hat{\rho})$	$\mathrm{SK}_M(\hat{\rho})$
0	D	−0.006	0.153	+0.017
	R	−0.006	0.151	+0.025
	N	−0.005	0.145	−0.012
	S	−0.009	0.153	−0.003
	P	−0.005	0.144	+0.001
	r	−0.005	0.142	+0.001
0.5	D	−0.010	0.119	−0.537
	R	−0.015	0.117	−0.538
	N	−0.018	0.112	−0.493
	S	−0.008	0.113	−0.489
	P	−0.006	0.110	−0.482
	r	−0.008	0.109	−0.470
0.7	D	−0.008	0.084	−0.778
	R	−0.015	0.084	−0.817
	N	−0.020	0.078	−0.695
	S	−0.007	0.080	−0.763
	P	−0.005	0.076	−0.652
	r	−0.007	0.076	−0.624
0.9	D	−0.004	0.035	−1.035
	R	−0.012	0.036	−1.029
	N	−0.016	0.033	−0.773
	S	−0.007	0.032	−0.792
	P	−0.002	0.028	−0.724
	r	−0.003	0.028	−0.713

D = Difference sign; R = Rank; N = Normal scores; S = Semiparametric transformation; P = Parametric Box-Cox transformation; r = sample correlation.

good even then. This underestimation results in overstating coverage probabilities; the nominal value (95%) is slightly too high whenever the standard error is slightly too small. (The standard error of these estimated coverage probabilities, based on 500 simulations, is 0.010.)

(iii) Similar tables for $n = 20$ (not shown) lead to similar conclusions, with only slightly greater underestimation of standard errors and overstatement of coverage probabilities. For the rank-based methods, true coverage probabilities were approximately 92.5% when $\rho = 0.7$ and approximately 90% when $\rho = 0.9$. The parametric methods remained fully valid. Results for $n = 100$ were essentially indistinguishable from those with $n = 50$.

(iv) The finite-sample performances of (N) and of (S) were essentially indistinguishable; (P) gave a more reliable performance, and was found to be quite reliable for all ρ even at $n = 20$.

7 Conclusions

We have considered estimating the transformation correlation coefficient in both semiparametric and parametric transformation-to-bivariate-normality models. Maximum likelihood algorithms have been created, and the methods compared with three existing rank-based methods and the sample correlation.

TABLE 4. A comparison of six methods for estimating ρ, after the z-transformation

ρ	Method	$\hat{\zeta}-\zeta$	$\mathrm{sd}_M(\hat{\zeta})$	$\mathrm{SK}_M(\hat{\zeta})$	$\mathrm{KUR}_M(\hat{\zeta})$	$n\mathrm{Var}(\hat{\zeta})$	$se(\hat{\zeta})$	covprob
0	*D*	−0.006	0.157	+0.013	−0.245	1.048	0.145	0.936
	R	−0.006	0.154	+0.022	−0.243	1.049	0.145	0.944
	N	−0.005	0.148	−0.021	−0.257	1.000	0.141	0.940
	S	−0.004	0.150	−0.020	−0.258	1.000	0.141	0.948
	P	−0.005	0.146	−0.004	−0.243	1.000	0.141	0.950
	r	−0.005	0.145	−0.005	−0.247	1.000	0.141	0.954
0.5	*D*	−0.002	0.157	−0.121	−0.046	1.059	0.146	0.942
	R	−0.009	0.154	−0.122	+0.013	1.074	0.147	0.944
	N	−0.014	0.146	−0.127	−0.135	1.000	0.141	0.940
	S	−0.002	0.150	−0.125	−0.030	1.000	0.141	0.946
	P	+0.002	0.146	−0.085	−0.042	1.000	0.141	0.948
	r	−0.001	0.145	−0.086	−0.095	1.000	0.141	0.952
0.7	*D*	+0.002	0.160	−0.102	+0.069	1.071	0.146	0.938
	R	−0.012	0.157	−0.147	+0.120	1.103	0.149	0.946
	N	−0.025	0.144	−0.161	−0.129	1.000	0.141	0.948
	S	−0.002	0.155	−0.132	−0.070	1.000	0.141	0.946
	P	+0.006	0.145	−0.081	−0.042	1.000	0.141	0.948
	r	+0.001	0.145	−0.072	−0.098	1.000	0.141	0.944
0.9	*D*	+0.006	0.172	−0.025	−0.038	1.088	0.148	0.934
	R	−0.033	0.169	−0.061	−0.024	1.093	0.148	0.910
	N	−0.059	0.150	−0.040	−0.313	1.000	0.141	0.924
	S	−0.020	0.161	−0.035	−0.034	1.000	0.141	0.934
	P	+0.010	0.144	+0.014	−0.097	1.000	0.141	0.948
	r	+0.004	0.144	+0.032	−0.084	1.000	0.141	0.950

D = Difference sign; *R* = Rank; *N* = Normal scores; *S* = Semiparametric transformation; *P* = Parametric Box-Cox transformation; *r* = sample correlation.

TABLE 5. Estimated power coefficients and Monte Carlo standard errors, by the parametric Box-Cox transformation method for estimating ρ

ρ	$\hat{\lambda}_1$	$\mathrm{sd}_M(\hat{\lambda}_1)$	$\hat{\lambda}_2$	$\mathrm{sd}_M(\hat{\lambda}_2)$
0	−0.003	0.121	−0.001	0.128
0.5	−0.005	0.117	−0.001	0.122
0.7	−0.006	0.113	−0.001	0.114
0.9	−0.006	0.103	−0.002	0.100

Both new methods are asymptotically efficient (when the model is correct), and estimate ρ as well as if the transformations taking the bivariate data to bivariate normality were known. A normal scores correlation coefficient is likewise efficient, and simpler to compute. Transforming either Kendall's or Spearman's rank correlation coefficient to estimate ρ has no advantage, and efficiency is lost relative to the other methods.

Application of Fisher's z-transform method for inferring about ρ works very well—even in samples of size 20 for the Box-Cox method (P), whatever the true ρ. For moderate ρ, say within ± 0.5, normal scores (N) and our rank method (S) also work well for samples of this size, and work quite well for large ρ in samples of size 50 or more.

An advantage of the Box-Cox method is that we can measure goodness-of-fit of

the model. Moreover, it is reliable for quite small sample sizes. Although there are semiparametric rank methods ((*N*) and (*S*)) that have the same large-sample efficiency, our Monte Carlo study shows that, with small to moderate sample sizes, there is some cost for allowing unstructured transformations. Hence, whenever the fit is adequate, the Box-Cox method is the preferred choice. When the fit is not adequate, the normal scores method must be considered the method of choice—being simpler than our semiparametric method and apparently with essentially the same performance.

One variation on the Box-Cox method is as follows. When both variables are of the same type, a common power coefficient may be appropriate. Our method may be easily adapted to this. However, there is no way to adapt the rank-based methods—ours or any of the others—to deal with any assumed common features of the two transformations. Still, (*N*) and (*S*) remain fully efficient, as would the modified (*P*).

Several kinds of extensions may be possible. First, for measurement data of dimension $d > 2$, a transformation-to-multivariate-normality model may be considered. Both of our methods may be extended to this case, estimating the $(d/2)$ pairwise transformation correlation coefficients. However, algorithms for their computation, not presented here, would become increasingly complicated as the dimensionality d grows. By contrast, the normal scores method remains quite simple.

Other extensions may not be amenable to a normal scores solution, whereas our two methods may be. In particular, consider a repeated-measures problem in which a single transformation takes correlated 'before' and 'after' measurements on individuals into bivariate normality with a change in the 'after' mean (see Section 6.9 of Zou, 1997). Another extension is to the problem of comparing two sets of pairs of correlated measurements, say diagnostic test measurements on two occasions, or with two diagnostic tests, on samples of healthy and diseased subjects (Zou & Hall, 2001). In each of these extensions, both semiparametric and parametric (Box-Cox) methods have been developed, but normal scores cannot be adapted. Indeed, it was problems such as these that motivated development of our methods presented here.

Acknowledgement

This research was supported in part by a grant from the Agency for Healthcare Research and Quality (AHRQ, USA).

REFERENCES

BICKEL, P. J. & DOKSUM, K. A. (1981) An analysis of transformations revisited, *Journal of the American Statistical Association*, 76, pp. 296-311.

BOX, G. E. P. & COX, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 42, pp. 71-78.

CARROLL, R. J. & RUPPERT, D. (1988) *Transformation and Weighting in Regression* (New York, Chapman and Hall).

DANIELS, H. E. (1944) The relation between measures of correlation in the universe of sample permutation, *Biometrika*, 44, pp. 129-135.

DAVID, F. N. & MALLOWS, C. L. (1961) The variance of Spearman's rho in normal samples, *Biometrika*, 48, pp. 19-28.

DOKSUM, K. A. (1987) An extension of partial likelihood methods from proportional hazard models to general transformation models, *Annals of Statistics*, 15, pp. 325-345.

ESSCHER, F. (1924) On a method of determining correlation from the ranks of variates, *Skandinavisk Aktuarietidskrift*, 7, pp. 201-219.
FIELLER, E. C., HARTLEY, H. O. & PEARSON, E. S. (1957) Tests for rank correlation coefficients: I, *Biometrika*, 44, pp. 470-481.
FIELLER, E. C. & PEARSON, E. S. (1961) Tests for rank correlation coefficients: II, *Biometrika*, 48, pp. 29-40.
FISHER, R. A. (1915) Frequency distributions of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika*, 10, pp. 507-521.
FISHER, R. A. & YATES, F. (1948) *Statistical Tables for Biological, Agricultural, and Medical Research* (New York, Hafner).
HÁJEK, J. & ŠIDÁK, Z. (1967) *Theory of Rank Tests* (New York, Academic Press).
HALL, W. J. (1970) On characterizing dependence in joint distributions. Chapter 17 in: R. C. BOSE *et al.* (eds), *Essays in Probability and Statistics* (Chapel Hill, University of North Carolina Press), pp. 339-376.
HERNANDEZ, F. & JOHNSON, R. A. (1980) The large-sample behavior of transformations to normality, *Journal of American Statistical Association*, 75, pp. 855-861.
HETTMANSPERGER, T. P. (1991) *Statistical Inference Based on Ranks* (Malabar, FL, Krieger).
HOEFFDING, W. (1947) On the distribution of the rank correlation coefficient τ when the variates are not independent, *Biometrika*, 34, pp. 183-196.
HOEFFDING, W. (1951) 'Optimum' nonparametric tests, *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 83-92.
KENDALL, M. G. (1938) A new measure of rank correlation, *Biometrika*, 30, pp. 81-93.
KENDALL, M. & GIBBONS, J. D. (1990) *Rank Correlation Methods* (New York, Oxford University Press).
KLAASSEN, C. A. J. & WELLNER, J. A. (1997) Efficient estimation in the bivariate normal copula model: normal margins are least favourable, *Bernoulli*, 3, pp. 55-77.
KRUSKAL, W. H. (1958) Oridinal measures of association, *Journal of the American Statistical Association*, 53, pp. 814-861.
LIN, C. C. & MUDHOLKAR, G. S. (1980) A simple test for normality against asymmetric alternatives, *Biometrika*, 67, pp. 455-461.
MUDHOLKAR, G. S. (1983) Fisher's z-transformation. In: S. KOTZ & N. L. JOHNSON (eds), *Encyclopedia of Statistical Sciences*, 3, pp. 130-135.
MUDHOLKAR, G. S., MCDERMOTT, M. & SRIVASTAVA, D. K. (1992) A test of p-variate normality, *Biometrika*, 79, pp. 850-854.
RICKER, W. E. & SMITH, H. D. (1975) A revised interpretation of the history of the Skeena River sockeye salmon, *Journal of the Fisheries Research Board of Canada*, 32, pp. 1369-1381.
SPEARMAN, C. (1904) The proof and measurement of association between two things, *American Journal of Psychology*, 15, pp. 72-101.
VAN DER WAERDEN, B. L. (1952) Order tests for the two sample problem and their power, I, II, III, *Indagationes Mathematicae*, 14, pp. 453-458.
ZOU, K. H. (1997) Analysis of some transformation models for the two-sample problem with special reference to receiver operating characteristic curves. PhD dissertation, University of Rochester, Rochester, NY.
ZOU, K. H. & HALL, W. J. (2002) Semiparametric and parametric transformation models for comparing diagnostic markers with paired design, *Journal of Applied Statistics*, accepted for publication.

Appendix 1

Proof of Theorem 1

We provide a proof quite different from that of Hoeffding (1951).

For given rank values $\underline{r}$ and $\underline{s}$, each a permutation of $(1, \ldots, n)$, let $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ be their respective anti-ranks. Then

$$P \equiv \Pr\{\underline{R} = \underline{r}, \underline{S} = \underline{s}\} = \Pr\{X_{u_1} < \ldots < X_{u_n}, Y_{v_1} < \ldots < Y_{v_n}\}$$

$$= \int \ldots \int_{\mathcal{R}^{2n}} 1\{x_{u_1} < \ldots < x_{u_n}\} 1\{y_{v_1} < \ldots < y_{v_n}\} \cdot \prod_{i=1}^{n} \{f(x_i, y_i)\, \mathrm{d}x_i\, \mathrm{d}y_i\}$$

where $1\{A\}$ is the indicator of the event A.

Since $f(x,y) = g(x,y) = 0$ whenever $h_1(x)h_2(y) = 0$, we have

$$P = \int \cdots \int_{\mathcal{R}^{2n}} \prod_{i=1}^{n} L_f(x_i, y_i) \cdot 1\{x_{u_1} < \cdots < x_{u_n}\} 1\{y_{v_1} < \cdots < y_{v_n}\}$$

$$\cdot \prod_{l=1}^{n} \{h_1(x_l) h_2(y_l)\, \mathrm{d}x_l\, \mathrm{d}y_l\}$$

Now let $z_{1i} = x_{u_i}$ and $z_{2i} = y_{v_i}$ so that $x_i = z_{1r_i}$ and $y_i = z_{2s_i}$. Hence,

$$P = \int \cdots \int_{\mathcal{R}^{2n}} \frac{1}{(n!)^2} \prod_{i=1}^{n} L_f(z_{1r_i}, z_{2s_i}) \cdot 1\{z_{11} < \cdots < z_{1n}\} n! \prod_{i=1}^{n} h_1(z_{1i})\, \mathrm{d}z_{1i}$$

$$\cdot 1\{z_{21} < \cdots < z_{2n}\} n! \prod_{i=1}^{n} h_2(z_{2i})\, \mathrm{d}z_{2i}$$

$$= E_{h_1,h_2} \left\{ \frac{1}{(n!)^2} \prod_{i=1}^{n} L_f(Z_{1(r_i)}, Z_{2(s_i)}) \right\}$$

Appendix 2

Computation of the inner integral in equation (4) To evaluate equation (4) in Section 2, because of the product term $\bar{Z}_1\bar{Z}_2$ in V, a conditioning argument is needed. We have

$$E_{\bar{Z}_1,\bar{Z}_2|A,B}\{\exp(V)\} = E_{\bar{Z}_1|A,B}[\exp(c\rho\bar{Z}_1^2) \cdot E_{\bar{Z}_2|\bar{Z}_1,A,B}\{\exp(c\rho\bar{Z}_2^2 - 2c\bar{Z}_1\bar{Z}_2)\}] \qquad (8)$$

The following lemma is needed:

Lemma 1

If $T \sim N(0, \sigma^2)$, then, for $a < 1/(2\sigma^2)$,

$$E_T \exp(aT^2 + bT) = \left(\frac{1}{1 - 2a\sigma^2}\right) \exp\left(\frac{b^2\sigma^2}{2 - 4a\sigma^2}\right) \qquad (9)$$

Proof

Writing $c = 1/(2\sigma^2) - a > 0$, the left-hand side in equation (9) equals

$$\frac{1}{\sigma} \int \phi\left(\frac{t}{\sigma}\right) \exp(at^2 + bt)\, \mathrm{d}t = \frac{1}{\sigma} \int \frac{1}{\sqrt{2\pi}} \exp\left\{-c\left(t^2 - \frac{b}{c}t + \frac{b^2}{4c^2}\right) + \frac{b^2}{4c}\right\} \mathrm{d}t$$

$$= \exp\left(\frac{b^2}{4c}\right) \frac{1}{\sigma} \int \frac{1}{\sqrt{2\pi}} \exp\left\{-c\left(t - \frac{b}{2c}\right)^2\right\} \mathrm{d}t$$

Realize that the last integral is proportional to the integral of the p.d.f. of a normal distribution $N(b/(2c), 1/(2c))$ over $\mathcal{R}$, and hence equals $1/\sqrt{2c}$. Therefore, the expectation becomes $(1/\sigma\sqrt{2c}) \exp(b^2/4c)$, which reduces to the right-hand side in (9).

By Lemma 1, the inner conditional expectation in (8) becomes

$$\exp\left\{\frac{(-2c\bar{Z}_1)^2}{2(n-2c\rho)}\right\}\left(\frac{n}{n-2c\rho}\right)^{1/2}$$

Hence, again by Lemma 1, (8) may be reduced to $\{n/(n-2c\rho)\}^{1/2} = (1-\rho^2)^{1/2}$.

Appendix 3

Algorithm for semiparametric maximum likelihood estimate of ρ

Step 1. Determine the marginal ranks $\underline{r}$ and $\underline{s}$ of the observed x_0s and y_0s. Randomly break any ties resulting from rounding.

Step 2. Start with an initial estimate of ρ by the sample correlation coefficient of the x_0s and the y_0s, respectively (possibly first applying Box-Cox transformations, formally or informally).

Step 3. Repeat the following procedure M times. (1) Generate a random sample of size $2n$ from $N(0, 1)$. (2) Label the first half of the sample as z_1s and the second half as z_2s. (3) Identify subscripts on the marginal z_1s and z_2s so that the ranks are $\underline{r}$ and $\underline{s}$, respectively. Record only the A and B sums, say a_m and b_m, at the mth step of the M iterations.

Step 4. At the end of the M runs, re-centre the a_ms and b_ms into $e_m = a_m - \bar{a}$ and $g_m = b_m - \bar{b}$ (in order to avoid potential overflow problems).

For large M, equation (5) may be approximated by

$$(1-\rho^2)^{(1-n)/2}\exp\{-U(\rho, \bar{a}, \bar{b})\}\cdot\frac{1}{M}\sum_{m=1}^{M}\exp\{U(\rho, e_m, g_m)\}$$

where the function U is given after (4). Hence, the log likelihood function of ρ is

$$-\frac{n-1}{2}\log(1-\rho^2) - U(\rho, \bar{a}, \bar{b}) + \log\left[\frac{1}{M}\sum_{m=1}^{M}\exp\{U(\rho, e_m, g_m)\}\right] + c$$

Differentiating w.r.t. ρ yields the score

$$S(\rho) \approx -\frac{2\rho\bar{a}_m}{(1-\rho^2)^2} + \frac{(1+\rho^2)\bar{b}_m}{(1-\rho^2)^2} + \frac{(n-1)\rho}{1-\rho^2} + \left[\sum_{m=1}^{M}\left\{-\frac{2\rho e_m}{(1-\rho^2)^2} + \frac{(1+\rho^2)g_m}{(1-\rho^2)^2}\right\}\exp\{U(\rho, e_m, g_m)\}\right] \Bigg/ \sum_{m=1}^{M}\exp\{U(\rho, e_m, g_m)\}$$

Step 5. Equate the score above to zero and solve iteratively for $\rho \equiv \hat{\rho}_S$, using an initial value from Step 2.

Appendix 4

Estimating ρ from three rank correlation coefficients

We consider a bivariate normal population of (X, Y) values with correlation coefficient ρ, and a random sample of size n. Here, we deal only with the marginal ranks from this sample, and these are the same whether sampling directly from (X, Y) or from (X_0, Y_0), which can be transformed to (X, Y) as in Section 1; that is, this bivariate normal ρ is the transformation correlation coefficient.

We show how each of three rank correlation coefficients can be used to estimate the transformation correlation ρ. (For references to these rank correlations see Section 1.)

(1) The sample *difference sign correlation coefficient* (Kendall's tau) is

$$t = \frac{1}{\binom{n}{2}} \sum_{i<j} \sum \operatorname{sgn}\{(x_j - x_i)\ (y_j - y_i)\}$$

where $\operatorname{sgn}(z) = 1, 0, -1$ for $z > 0, = 0, < 0$. It depends only on orderings within the xs and within the ys, and is therefore a rank statistic. The population version is $\tau \equiv 2\omega_1 - 1$ where ω_1 is the probability that two pairs of observations, say (X_1, Y_1) and (X_2, Y_2), are type I concordant—that is, $\omega_1 \equiv \Pr\{(X_1 - X_2)\ (Y_1 - Y_2) > 0\}$.

The estimate t is unbiased for τ and is asymptotically normal with large-sample variance

$$\operatorname{Var}(t) \approx \frac{4}{9n}\left[1 - \left\{\frac{6}{\pi}\sin^{-1}\left(\frac{\rho}{2}\right)\right\}^2\right]$$

Esscher (1924) noted, in the bivariate normal case, $\tau = (2/\pi)\sin^{-1}\rho$ and inversely $\rho = \sin(\frac{1}{2}\pi\tau)$. Hence $\hat{\rho}_D \equiv \sin(\frac{1}{2}\pi t)$ estimates ρ. The delta method yields $\operatorname{Var}(\hat{\rho}_D) \approx \frac{1}{4}\pi^2(1 - \rho^2)\operatorname{Var}(t)$.

(2) The sample *rank correlation coefficient* (Spearman's rho) r_S is the product-moment correlation coefficient of the paired marginal rank data $\underline{r}$ and $\underline{s}$. The population version is $\rho_S \equiv 6\omega_2 - 3$ where ω_2 is the probability that, among three pairs of observations, say (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3), at least one is concordant with the other two—namely, $\Pr\{(X_1 - X_2)\ (Y_1 - Y_3) > 0\}$.

The estimate r_S of ρ_S has bias of order $1/n$, and is asymptotically normal with large-sample variance given by the series expansion:

$$\operatorname{Var}(r_S) \approx \frac{1}{n}\left(1 - 1.5635\rho^2 + 0.3047\rho^4 + 0.1553\rho^6 + 0.0616\rho^8 + 0.0242\rho^{10} + \ldots\right)$$

(Additional terms may be needed for $\rho > 0.8$.)

In the bivariate normal case, $\rho_S = (6/\pi)\sin^{-1}(\rho/2)$ so that, inversely, $\rho = 2\sin((\pi/6)\rho_S)$ and $\hat{\rho}_R \equiv 2\sin((\pi/6)r_S)$ estimates ρ. The delta method yields $\operatorname{Var}(\hat{\rho}_R) \approx \frac{1}{9}\pi^2(1 - \frac{1}{4}\rho^2)\operatorname{Var}(r_S)$.

(3) The sample *normal scores correlation coefficient* (Fisher-Yates) r_N is the product-moment correlation coefficient of the n pairs of normal scores (r_i', s_i'). The normal score for x_i is defined as $r_i' = E(Z_{(r_i)})$ where $Z_{(1)} < \ldots < Z_{(n)}$ are order statistics generated from $N(0, 1)$ and r_i is the rank of x_i among $(x_1, \ldots, x_n)$; similarly for the normal score s_i' for y_i. However, the van der Waerden (1952) versions are more commonly used today, namely $r_i' = \Phi^{-1}(r_i/(n+1))$, and the resulting van der Waerden and Fisher-Yates normal scores correlation coefficients are asymptotically equivalent (see Hàjek & Sĭdàk, 1967).

The population version ρ_N of the van der Waerden form is the product-moment correlation coefficient between $(\Phi^{-1} \circ F)(X)$ and $(\Phi^{-1} \circ G)(Y)$ with F and G the marginal distribution functions of X and Y. Hence, in the bivariate normal case, $\rho_N = \rho$ (Klaassen and Wellner, 1997). They term ρ_N the 'normal correlation coefficient' (rather than transformation correlation coefficient), and refer to the transformation model as a 'normal copula model'.

Klaassen and Wellner go on to prove that $\hat{\rho}_N \equiv r_N$ is an asymptotically efficient estimate of ρ with large-sample variance $\mathrm{Var}(\hat{\rho}_N) \approx (1 - \rho^2)^2/n$; that is, in large samples, estimating ρ by $\hat{\rho}_N$ is just as efficient as if the transformations ψ_1 and ψ_2 of Section 1 were known, applied to the data, and the product-moment correlation coefficient r then computed.