ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 4, pp. 307–318. © Allerton Press, Inc., 2019.

An Asymptotically Optimal Transform of Pearson's Correlation Statistic

I. Pinelis[1*]

[1] *Dept. Math. Sci., Michigan Technol. Univ., Houghton, Michigan, USA*

Received July 28, 2019

Abstract—It is shown that for any correlation-parametrized model of dependence and any given significance level $\alpha \in (0, 1)$, there is an asymptotically optimal transform of Pearson's correlation statistic R, for which the generally leading error term for the normal approximation vanishes for all values $\rho \in (-1, 1)$ of the correlation coefficient. This general result is then applied to the bivariate normal (BVN) model of dependence and to what is referred to in this paper as the SquareV model. In the BVN model, Pearson's R turns out to be asymptotically optimal for a rather unusual significance level $\alpha \approx 0.240$, whereas Fisher's transform R_F of R is asymptotically optimal for the limit significance level $\alpha = 0$. In the SquareV model, Pearson's R is asymptotically optimal for a still rather high significance level $\alpha \approx 0.159$, whereas Fisher's transform R_F of R is not asymptotically optimal for any $\alpha \in [0, 1]$. Moreover, it is shown that in both the BVN model and the SquareV model, the transform optimal for a given value of α is in fact asymptotically better than R and R_F in wide ranges of values of the significance level, including α itself. Extensive computer simulations for the BVN and SquareV models of dependence suggest that, for sample sizes $n \geq 100$ and significance levels $\alpha \in \{0.01, 0.05\}$, the mentioned asymptotically optimal transform of R generally outperforms both Pearson's R and Fisher's transform R_F of R, the latter appearing generally much inferior to both R and the asymptotically optimal transform of R in the SquareV model.

Keywords: hypothesis testing, Pearson's correlation statistic, Fisher's z transform, asymptotically optimal transform, models of dependence, copulas, bivariate normal distribution.

AMS 2010 Subject Classification: 62E20, 62F03, 62F12.

DOI: 10.3103/S1066530719040057

1. INTRODUCTION

A statistic closely related to Pearson's R is commonly known as the Fisher z transform, defined by the formula

$$R_F := \tanh^{-1}(R) = \frac{1}{2} \ln \frac{1+R}{1-R}. \tag{1.1}$$

An advantage of using R_F (as opposed to R) in making statistical inferences about the true correlation coefficient ρ is usually ascribed to its variance-stabilizing property in normal populations, see, e.g.. Fisher [4], Gayen [5], and Hotelling [6], that is, $n \operatorname{Var} R_F \to 1$ for all $\rho \in (-1, 1)$ as $n \to \infty$ (as opposed to $n \operatorname{Var} R \to (1-\rho^2)^2$) whenever the underlying distribution is bivariate normal. Everywhere here, n denotes the sample size.

In his discussion of Hotelling's paper [6], Kendall provides heuristics suggesting that such variance stabilization of the distribution of a statistic may often result in it being closer to normality. Namely, if an approximate constancy of the variance of a statistic were the same as an approximate constancy of its distribution itself, and if the distribution is close to normality at least for one value of the parameter (say, ρ, as in the present case), then it would be close to normality for all values of ρ. For normal populations and large enough sample sizes, the Fisher z transform indeed brings the distribution of the correlation statistic closer to normality, and it is especially effective for values of ρ far from 0. However, it is well

[*] E-mail: ipinelis@mtu.edu

known (see, e.g., [1, 10]) that the closeness of the distribution of a statistic to normality is usually mainly determined, not by the variance, but by the third moments of the appropriately standardized statistic.

In this paper, we shall see that for a general and most common class of models of dependence, including the bivariate normal (BVN) model, and for each given significance level $\alpha \in (0,1)$ there is a certain transform $\Psi_\alpha(R)$ of Pearson's statistic R that assures the *vanishing* of the generally leading term of the asymptotics of the probability that an approximately standardized version of the statistic $\Psi_\alpha(R)$ exceeds the standard normal critical value

$$z_\alpha := \Phi^{-1}(1-\alpha); \tag{1.2}$$

here, as usual, Φ is the standard normal cumulative distribution function (cdf) and Φ^{-1} is its inverse; unless otherwise specified, all the asymptotics here are for large sample sizes n. Thus, the transform $\Psi_\alpha(R)$ of R is asymptotically optimal: its distribution is asymptotically the closest to normality exactly at the critical value.

Once this optimality result is obtained for the general class of models of dependence, the rest of the paper is devoted to detailed analysis of the optimal transform $\Psi_\alpha(R)$ of R in the BVN model and another specific model of dependence, referred to in this paper as the SquareV model.

We shall see that, in the BVN model, the mentioned family $(\Psi_\alpha(R))_{\alpha\in(0,1)}$ of transforms of Pearson's statistic R includes R itself: namely,

$$R = \Psi_{\alpha_P}(R), \tag{1.3}$$

where

$$\alpha_P := 1 - \Phi(1/\sqrt{2}) \approx 0.240.$$

Thus, Pearson's statistic R is asymptotically optimal for a significance level α of about 24%, but such a significance level is rather unusual in statistical practice.

As for Fisher's transform R_F of R, we shall see that, again in the BVN model,

$$R_F = \Psi_0(R) := \lim_{\alpha\downarrow 0} \Psi_\alpha(R), \tag{1.4}$$

which means that Fisher's transform is asymptotically optimal for the significance level

$$\alpha_F := 0.$$

Now one might explain the fact that for the usually rather small significance levels, such as 0.05 or 0.01, Fisher's statistic R_F is asymptotically closer to normality than Pearson's statistic R by noting that the significance level $\alpha_F = 0$ (for which R_F is asymptotically optimal) is closer to 0.05 and especially to 0.01 than the significance level $\alpha_R \approx 0.240$ (for which R is asymptotically optimal).

As for the SquareV model, there Pearson's R is asymptotically optimal for a still rather high significance level $\alpha \approx 0.159$, whereas Fisher's transform R_F of R is not asymptotically optimal for any $\alpha \in [0,1]$.

It should be noted that in both the BVN model and the SquareV model, the transform optimal for a given value of α is in fact asymptotically better than R and R_F in wide ranges of values of the significance level, including α itself.

We have also conducted extensive computer simulations for the BVN and SquareV models of dependence, which suggest that, for sample sizes $n \geq 100$ and significance levels $\alpha \in \{0.01, 0.05\}$, the mentioned asymptotically optimal transform of R generally outperforms both Pearson's R and Fisher's transform R_F of R, the latter appearing generally much inferior to both R and the asymptotically optimal transform of R in the SquareV model.

The rest of the paper is organized as follows.

In Section 2 we present an asymptotic expansion for statistics that are general smooth nonlinear functions of the sample mean of iid random vectors in $\mathbb{R}^p$. This expansion, which may be viewed as a far-reaching refinement of the delta method, is a special case of results by Bhattacharya and Ghosh [1]. For Berry–Esseen-type bounds for general nonlinear statistics, see, e.g., [2, 10].

In Section 3, the mentioned asymptotic expansion is specialized for the cases of Pearson's correlation statistic R and its smooth enough transforms. A key observation there is that the main term of the asymptotic for such a transform of R differs from the corresponding main term for R itself only by a comparatively simple expression involving the first two derivatives of the transform function ψ. This allows one to obtain, for any correlation-parametrized model of dependence and for any given significance level $\alpha \in (0,1)$, a rather simple second-order ordinary differential equation (ODE) for the optimal transform function ψ that makes the main term of the asymptotic for the asymptotically optimal transform of R vanish for all values $\rho \in (-1,1)$ of the correlation coefficient. This ODE can be explicitly solved for a number of models of dependence, including the important BVN model and models with a linear dependence of the correlation parameter. The mentioned SquareV model is a model with such a linear dependence.

The BVN model is considered in detail in Section 4.

A similar treatment of the SquareV model is given in Section 5.

Section 6 is a summary of the results of this paper.

2. ASYMPTOTIC EXPANSIONS FOR SMOOTH NONLINEAR STATISTICS

Let

$$V, V_1, V_2, \dots$$

be independent identically distributed (iid) zero-mean random vectors in $\mathbb{R}^p$ with $\mathsf{E}\,\|V\|^3 < \infty$, where p is a natural number and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^p$, which latter will be identified, as usual, with the space of all $p \times 1$ column matrices. Assume also that the Cramér-type condition $\limsup_{\|t\|\to\infty} |\,\mathsf{E}\exp(it^T V)| < 1$ is satisfied, where i is the imaginary unit and ^{T} denotes the transposition, in this case of a column matrix $t \in \mathbb{R}^p$; for this Cramér-type condition to hold, it is enough that, for some natural k, the k-fold convolution of the distribution of V have a nonzero absolutely continuous component. Let Σ stand for the covariance matrix of V:

$$\Sigma := \mathsf{E}\, V V^T. \tag{2.1}$$

Let $f\colon \mathbb{R}^p \to \mathbb{R}$ be a function which is twice continuously differentiable in a neighborhood of $0 \in \mathbb{R}^p$ and such that $f(0) = 0$. Let L and H denote, respectively, the gradient vector and the Hessian matrix of the function f at 0, so that

$$f'(0)(v) = L^T v \quad \text{and} \quad f''(0)(v,v) = v^T H v \tag{2.2}$$

for all $v \in \mathbb{R}^p$. Since V is assumed to be zero-mean, one has $\mathsf{E}\, L^T V = 0$. Introduce now

$$\sigma := \sqrt{\mathsf{E}(L^T V)^2}, \tag{2.3}$$

which will be assumed to be nonzero, so that

$$\Lambda := \frac{L^T V}{\sigma}$$

is a well-defined r.v., with zero mean and unit variance. Consider the r.v.

$$T_n := \frac{\sqrt{n}}{\sigma} f(\bar{V}),$$

where of course $\bar{V} := \frac{1}{n}\sum_1^n V_i$. Then, by Theorem 2 of the paper [1] by Bhattacharya and Ghosh,

$$\sup_{z\in\mathbb{R}} |\,\mathsf{P}(T_n \le z) - \Psi_{3,n}(z)| = o\Big(\frac{1}{\sqrt{n}}\Big), \tag{2.4}$$

where

$$\Psi_{3,n}(z) := \Phi(z) + \frac{\Delta(z)}{\sqrt{n}}, \tag{2.5}$$

$$\Delta(z) := -\Big[\Big(\frac{\mathsf{E}\,\Lambda^3}{6} + a_3\Big)(z^2 - 1) + a_1\Big]\,\varphi(z) \tag{2.6}$$

$$= (Az^2 + B)\,\varphi(z), \tag{2.7}$$

Φ and φ denote, as usual, the distribution and density functions of $N(0,1)$, and a_1, a_2, A, and B are constants depending only on L, H, Σ, σ, and $\mathsf{E}\,\Lambda^3$ (but not on z or n):

$$a_1 := \frac{1}{2\sigma}\ \operatorname{tr} H\Sigma, \tag{2.8}$$

$$a_3 := \frac{1}{4\sigma^3}\,(L^T\Sigma L - \sigma^2)\operatorname{tr}(H\Sigma) + \frac{1}{2\sigma^3}\,L^T\Sigma H\Sigma L, \tag{2.9}$$

$$A := -\Big(\frac{\mathsf{E}\,\Lambda^3}{6} + a_3\Big), \tag{2.10}$$

$$B := -A - a_1, \tag{2.11}$$

with tr denoting the trace of a matrix.

Remark 2.1. If the condition $\mathsf{E}\,\|V\|^3 < \infty$ is strengthened to $\mathsf{E}\,\|V\|^4 < \infty$, then $o\big(\frac{1}{\sqrt{n}}\big)$ in (2.4) can be replaced by $O\big(\frac{1}{n}\big)$.

3. ASYMPTOTICS FOR THE PEARSON STATISTIC AND ITS TRANSFORMS

Let $(Y,Z),(Y_1,Z_1),\dots,(Y_n,Z_n)$ be independent identically distributed random points in $\mathbb{R}^2$ with a correlation coefficient $\rho \in (-1,1)$ and $\mathsf{E}(Y^6 + Z^6) < \infty$. Pearson's sample correlation coefficient based on the observations $(Y_1,Z_1),\dots,(Y_n,Z_n)$ is defined by the formula

$$R := R_n := \frac{\bar{YZ} - \bar{Y}\,\bar{Z}}{\sqrt{\bar{Y^2} - \bar{Y}^2}\sqrt{\bar{Z^2} - \bar{Z}^2}}, \tag{3.1}$$

where $\bar{Y} := \frac{1}{n}\sum_1^n Y_i$, $\bar{Z} := \frac{1}{n}\sum_1^n Z_i$, $\bar{Y^2} := \frac{1}{n}\sum_1^n Y_i^2$, $\bar{Z^2} := \frac{1}{n}\sum_1^n Z_i^2$, and $\bar{YZ} := \frac{1}{n}\sum_1^n Y_iZ_i$; let R take an arbitrarily assigned value in the interval $[-1,1]$ if the denominator of the ratio in (3.1) is 0.

Let us assume that Y and Z are each standardized, that is, zero-mean and unit-variance. This assumption does not diminish generality, because R is invariant with respect to affine transformations $Y_i \mapsto a + b\,Y_i$ and $Z_i \mapsto c + d\,Z_i$ of the Y_i's and Z_i's, for any real constants a,b,c,d such that $b > 0$ and $d > 0$.

Observe that

$$R - \rho = f(\bar{V}), \tag{3.2}$$

where

$$V := (Y, Z, Y^2 - 1, Z^2 - 1, YZ - \rho) \tag{3.3}$$

and

$$f(v) := f_\rho(v) := \frac{\rho + v_5 - v_1v_2}{\sqrt{1 + v_3 - v_1^2}\sqrt{1 + v_4 - v_4^2}} - \rho \tag{3.4}$$

if $v = (v_1,\dots,v_5) \in \mathbb{R}^5$ is such that $1 + v_3 - v_1^2 > 0$ and $1 + v_4 - v_4^2 > 0$; otherwise, let $f(v) := 0$. In this case, $L = (0,0,-\frac{\rho}{2},-\frac{\rho}{2},1)$, whence

$$\sigma = \sqrt{\mathsf{E}\,\big(YZ - \tfrac{\rho}{2}\,(Y^2 + Z^2)\big)^2} \quad \text{and} \quad \Lambda = \frac{YZ - \frac{\rho}{2}\,(Y^2 + Z^2)}{\sigma}. \tag{3.5}$$

As noted in [10], the condition $\sigma = 0$ is equivalent to the following exceptional situation: there exists some $\kappa \in \mathbb{R}$ such that the random point (Y,Z) lies almost surely on the union of the two straight lines through the origin with slopes κ and $1/\kappa$ (for $\kappa = 0$, these two lines should be understood as the two coordinate axes in the plane $\mathbb{R}^2$).

It will be assumed in what follows that the random point (Y,Z) is such that σ is never 0.

Then it is easy to check that all the conditions on f and V stated in Section 2 are satisfied, with $p = 5$; in particular, $f(0) = 0$.

Let now $\Delta_R(z)$ denote $\Delta(z)$ defined by (2.6) with f as in (3.4).

Further, letting

$$\mu_{ij} := \mathsf{E}\, Y^i Z^j, \tag{3.6}$$

one has

$$\sigma = \frac{1}{2}\sqrt{\rho^2\left(\mu_{04} + 2\mu_{22} + \mu_{40}\right) - 4\rho\left(\mu_{13} + \mu_{31}\right) + 4\mu_{22}}, \tag{3.7}$$

$$\begin{aligned}\sigma^3\,\mathsf{E}\,\Lambda^3 = {}& -\frac{\rho^3}{8}\left(\mu_{06} + 3\mu_{24} + 3\mu_{42} + \mu_{60}\right)\\ & + 6\rho^2\left(\mu_{15} + 2\mu_{33} + \mu_{51}\right) - 12\rho\left(\mu_{24} + \mu_{42}\right) + 8\mu_{33},\end{aligned} \tag{3.8}$$

and

$$\begin{aligned}\frac{96\sigma^3}{\varphi(z)}\Delta_R(z) = \tilde\Delta_R(z) := {}& 16\big[(z^2-1)(6\mu_{12}\mu_{21} - \mu_{33}) + 3\sigma^2 z^2(\mu_{13} + \mu_{31})\big]\\ & - 12\rho\big[(z^2-1)(4\mu_{03}\mu_{21} + 4\mu_{12}\mu_{30} + 8\mu_{12}^2 - 2\mu_{13}\mu_{31} + \mu_{13}^2 + 8\mu_{21}^2 - 2\mu_{24} + \mu_{31}^2 - 2\mu_{42})\\ & + \sigma^2\big((2z^2+1)(\mu_{04} + \mu_{40}) + (4z^2-2)\mu_{22}\big)\big]\\ & + 12\rho^2(z^2-1)\big[2\mu_{03}(3\mu_{12} + \mu_{30}) + \mu_{04}(\mu_{13} - \mu_{31}) + 10\mu_{12}\mu_{21} - \mu_{13}\mu_{40} - \mu_{15}\\ & + 6\mu_{21}\mu_{30} + \mu_{31}\mu_{40} - 2\mu_{33} - \mu_{51}\big]\\ & - \rho^3(z^2-1)\big[24\mu_{03}\mu_{21} + 12\mu_{03}^2 - 6\mu_{04}\mu_{40} + 3\mu_{04}^2 - 2\mu_{06}\\ & + 24\mu_{12}\mu_{30} + 12\mu_{12}^2 + 12\mu_{21}^2 - 6\mu_{24} + 12\mu_{30}^2 + 3\mu_{40}^2 - 6\mu_{42} - 2\mu_{60}\big].\end{aligned} \tag{3.9}$$

More generally, let now

$$\psi\colon (-1,1) \to \mathbb{R}$$

be a twice continuously differentiable function whose derivative ψ' does not vanish at any point of the interval $(-1,1)$. Let then

$$g(v) := g_\rho(v) := \frac{\psi\big(f(v)+\rho\big) - \psi(\rho)}{\psi'(\rho)} \tag{3.10}$$

for $v \in \mathbb{R}^5$, with f as defined in (3.4). Then, in view of (3.2),

$$\frac{\psi(R) - \psi(\rho)}{\psi'(\rho)} = g(\bar V), \tag{3.11}$$

Note that $g(0) = f(0) = 0$ and $g'(0) = f'(0)$, so that σ and Λ for the function g are the same as in (3.5) (given there for the function f). Thus, all the conditions stated in Section 2 are satisfied with g in place of f.

Let now $\Delta_{\psi(R)}(z)$ denote $\Delta(z)$ defined by (2.6) with g as in (3.10) in place of f. Then (2.4) will hold with

$$\tau_{\psi,n} := \frac{\psi(R) - \psi(\rho)}{\psi'(\rho)\sigma/\sqrt{n}} \tag{3.12}$$

in place of T_n and $\Delta_{\psi(R)}(z)$ in place of $\Delta(z)$. One may note here that $\tau_{\psi,n}$ may be considered an asymptotically standardized version of $\psi(R)$.

A key observation is that

$$\Delta_{\psi(R)}(z) = \Delta_R(z) - \frac{\psi''(\rho)}{2\psi'(\rho)}\,\sigma z^2\,\varphi(z). \tag{3.13}$$

To begin using this observation, let us refer to any family $(P_\rho)_{\rho\in(-1,1)}$ of distributions of the random pair (Y,Z) in $\mathbb{R}^2$ parametrized by the correlation coefficient ρ of (Y,Z) as a *correlation-parametrized model (CP) of dependence*. The CP condition seems quite natural for parametric models of dependence. Indeed, let a real parameter θ represent the strength of the dependence between Y and Z. Then one should usually expect the correlation coefficient ρ to be a strictly increasing continuous function g of θ: $\rho = g(\theta)$. Replacing then θ by $g^{-1}(\rho)$, one obtains a re-parametrization with ρ as the new parameter.

In this regard, one may recall the formula

$$\mathsf{Cov}(Y,Z) = \iint_{\mathbb{R}^2} a_{Y,Z}(y,z)\,dy\,dz \tag{3.14}$$

for the covariance of Y and Z, where $a_{Y,Z}$ is the association function of r.v.'s Y and Z, given by the formula

$$a_{Y,Z}(y,z) := \mathsf{P}(Y > y, Z > z) - \mathsf{P}(Y > y)\,\mathsf{P}(Z > z)$$

for all $(y,z) \in \mathbb{R}^2$. Since $\mathsf{P}(Y > y) = \mathsf{P}(-Y < -y)$, formula (3.14) can be rewritten as

$$\mathsf{Cov}(Y,Z) = \iint_{\mathbb{R}^2} [F_{Y,Z}(y,z) - F_Y(y)\,F_Z(z)]\,dy\,dz, \tag{3.15}$$

where $F_{Y,Z}$ is the joint cdf of the random pair (Y,Z), and F_Y and F_Z are the corresponding marginal cdf's.

Suppose now that we have a dependence model (P_θ), where θ is a strength of the association/dependence parameter, so that the association function $a_{\theta;Y,Z}$ of the pair (Y,Z) with respect to the probability measure P_θ is increasing in θ on the average in the sense that the integral in (3.14) with $a_{\theta;Y,Z}$ in place of $a_{Y,Z}$ is increasing in θ. Suppose also that the Y- and Z-marginals of the distribution of the pair (Y,Z) with respect to P_θ do not depend on θ. Then the correlation coefficient ρ of (Y,Z) will be an increasing function of θ.

A more specific, but still rather general way to construct a CP model is as follows. By Sklar's theorem (see, e.g., [7], Theorem 2.3.3.),

$$F_{Y,Z}(y,z) = C(F_Y(y), F_Z(z))$$

for some copula C and all $(y,z) \in \mathbb{R}^2$; recall that a copula can be defined as the joint cdf of a random pair with values in the unit square $[0,1]^2$ whose marginals are uniform on the interval $[0,1]$. Let now (C_θ) be any family of copulas increasing in θ; a large number of such families can be found in [7]. Fix the marginal cdf's F_Y and F_Z, and for each value of the parameter θ let $F_{\theta;Y,Z}(y,z) := C_\theta(F_Y(y), F_Z(z))$, again for all $(y,z) \in \mathbb{R}^2$. Then the correlation coefficient ρ corresponding to the joint cdf $F_{\theta;Y,Z}$ will be an increasing function of θ.

In view of (3.6), in any correlation-parametrized model of dependence and for any given real $z \neq 0$, the expressions in (3.7) and (3.9) for σ and $\tilde{\Delta}_R(z)$ will depend on ρ only. Then, by the key observation (3.13), the condition $\Delta_{\psi(R)}(z) = 0$ can be rewritten as the second-order ordinary differential equation (ODE)

$$\frac{\psi''(\rho)}{\psi'(\rho)} = h_z(\rho) \tag{3.16}$$

for the function ψ, where

$$h_z(\rho) := \frac{\tilde{\Delta}_R(z)}{48\sigma^4 z^2}. \tag{3.17}$$

Solving now ODE (3.16) with the natural initial conditions

$$\psi(0) = 0 \quad \text{and} \quad \psi'(0) = 1, \tag{3.18}$$

we have

$$\psi'(\rho) = \exp \int_0^\rho dr\, h_z(r) \tag{3.19}$$

and

$$\psi(\rho) = \psi_z(\rho) := \int_0^\rho dr\, \exp \int_0^r ds\, h_z(s) \tag{3.20}$$

for $\rho \in (-1, 1)$; in (3.19) and (3.20), we use the common convention $\int_0^s := -\int_s^0$ for $s < 0$. Thus, we obtain

Theorem 3.1. *In any correlation-parametrized model of dependence and for any given real $z \neq 0$, the generally leading error term for the normal approximation for $\psi_z(R)$ vanishes:*

$$\Delta_{\psi_z(R)}(z) = 0 \tag{3.21}$$

for all $\rho \in (-1, 1)$.

Letting now

$$\Psi_\alpha := \psi_{z_\alpha}, \tag{3.22}$$

we can rewrite (3.21) as

$$\Delta_{\Psi_\alpha(R)}(z_\alpha) = 0$$

for all $\alpha \in (0, 1)$, with $z_\alpha = \Phi^{-1}(1 - \alpha)$, as defined in (1.2).

One may note here that for a rather large class of models of dependence the functions h_z will be rational, and hence, according to (4.3), ψ'_z will be an elementary, closed-form function. This class of models with rational functions h_z includes the bivariate normal model and models with linear dependence of the joint cdf $F_{\theta;Y,Z}$ on θ. In particular, the class of models with linear dependence of $F_{\theta;Y,Z}$ on θ contains Farlie's model [3].

We shall consider the bivariate normal model and a particular simple model with linear dependence of $F_{\theta;Y,Z}$ on θ in the following sections, to compare the performance of Pearson's R itself, its Fisher transform R_F, and the asymptotically optimal transform $\Psi_\alpha(R)$ of R in non-asymptotic settings, for specific sample sizes.

4. BIVARIATE NORMAL MODEL (BVN)

4.1. Asymptotically Optimal Transform $\Psi_\alpha(R)$ in the BVN Model

Here it is assumed that the random point (Y, Z) has the bivariate normal (BVN) distribution with zero means, unit variances, and an arbitrary correlation coefficient $\rho \in (-1, 1)$. Then the expressions for $\tilde{\Delta}_R$ and σ, and thus for $h_z(\rho)$, in formulas (3.9), (3.7), and (3.17) can be greatly simplified.

Indeed, in this case the pair (Y, Z) equals $(Y, \rho Y + \sqrt{1-\rho^2}\, Y_1)$ in distribution, whence, by (3.6),

$$\mu_{ij} = \sum_{k=0}^{j} \binom{j}{k} \rho^k (1-\rho^2)^{(j-k)/2} m(i+k)\, m(j-k)$$

for all $i, j = 0, 1, \dots$ and

$$m_j := \mathsf{E}\, Y^j,$$

so that $(m_0, \dots, m_6) = (1, 0, 1, 0, 3, 0, 15)$. As the result, ODE (3.16) becomes

$$\frac{\psi''(\rho)}{\psi'(\rho)} = p_z \frac{-2\rho}{1-\rho^2}, \tag{4.1}$$

where

$$p_z := \frac{1}{2z^2} - 1. \tag{4.2}$$

ODE (4.1) is easily solved, yielding

$$\psi'(\rho) = (1-\rho^2)^{p_z} \tag{4.3}$$

and

$$\psi(\rho) = \psi_z(\rho) := \int_0^\rho (1-r^2)^{p_z}\,dr = \rho\,{}_2F_1\big(\tfrac12, -p_z; \tfrac32; \rho^2\big), \tag{4.4}$$

where ${}_2F_1$ is the ordinary hypergeometric function, given by the formula

$${}_2F_1(a,b;c;x) = \sum_{k=0}^\infty \frac{(a)_k(b)_k}{(c)_k}\,\frac{x^k}{k!}$$

for x with $|x|<1$, where $(q)_k := \prod_{i=0}^{k-1}(q+i)$ is the Pochhammer symbol. The last equality in (4.4) can be obtained by expanding the integrand $(1-r^2)^{p_z}$ into the Maclaurin series in powers of r and then integrating the series term-wise.

Thus, recalling (3.22), we see that in the bivariate normal case the transform $\Psi_\alpha(R) = \psi_{z_\alpha}(R)$ of Pearson's R with ψ_z as in (4.4) is asymptotically optimal for any given significance level $\alpha \in (0,1)$.

In particular, choosing

$$z = 1/\sqrt{2} \approx 0.707,$$

we have $p_z = 0$. Hence, in view of the integral expression in (4.4), $\psi_z(\rho) = \rho$, so that we have (1.3), confirming that the family $(\Psi_\alpha(R))_{\alpha\in(0,1)}$ of transforms of Pearson's statistic R includes R itself.

On the other hand, letting $z\to\infty$, we have $p_z \to -1$, so that, using again the integral expression in (4.4) (and, say, the dominated convergence theorem), we see

$$\psi_z(\rho) \underset{z\to\infty}{\longrightarrow} \psi_\infty(\rho) := \frac12 \ln\frac{1+\rho}{1-\rho} \tag{4.5}$$

for all $\rho \in (-1,1)$, thus confirming (1.4).

In view of formula (4.4), the calculation of values of the functions ψ_z or, equivalently, of the functions Ψ_α mainly reduces to the calculation of values of the hypergeometric function ${}_2F_1$. In general, this hypergeometric function is not elementary. However, there are a number of highly efficient ways to compute values of ${}_2F_1$. It takes only about 1.7×10^{-5} sec on an average to compute a value of $\psi_2(\rho)$ (on a standard computer), which may be compared with the corresponding execution time of about 0.45×10^{-5} sec for Fisher's $\psi_\infty(\rho) = \frac12\ln\frac{1+\rho}{1-\rho}$. Therefore and because usually in statistical practice the value of the transform $\Psi_\alpha(R) = \psi_{z_\alpha}(R)$ of the statistic R needs to be computed only once, the use of the hypergeometric function ${}_2F_1$ should not cause any complications.

Also, according to (4.3), the derivarive ψ' of the function $\psi = \psi_z$ is a simple elementary expression, which makes it easy to obtain various analytical properties of ψ_z. For instance, using the special l'Hospital-type rule for monotonicity (see, e.g, [8], Proposition 4.1), we can immediately see that the ratio $\psi_{z_1}(\rho)/\psi_{z_2}(\rho)$ is decreasing in ρ^2 for any real z_1 and z_2 such that $0<|z_1|<|z_2|$. In particular, it follows that the ratio of $\psi_z(\rho)$ to Fisher's $\psi_\infty(\rho) = \frac12\ln\frac{1+\rho}{1-\rho}$ is decreasing in ρ^2 for any real $z\neq0$. One may also note that the values

$$\psi_z(\pm1) = \pm\frac{\sqrt{\pi}\,\Gamma(p_z+1)}{2\,\Gamma(p_z+3/2)}$$

at the endpoints of the interval $[-1,1]$ are finite for all real $z\neq0$, in contrast with Fisher's limit values $\psi_\infty(\pm(1-)) = \pm\infty$; here p_z is as defined in (4.2). It is also clear that $\psi_z(\rho)$ is odd in ρ, for each $z\neq0$.

4.2. The Transform $\Psi_\alpha(R)$ in the BVN model is Asymptotically Better than R and R_F in Wide Ranges of Values of the Significance Level Including α Itself

According to Theorem 3.1, for any given correlation-parametrized model of dependence and any given significance level $\alpha \in (0,1)$, the transform $\Psi_\alpha(R)$ of R is asymptotically optimal for all $\rho \in (-1,1)$ as the sample size n goes to ∞. In particular, for any given significance level $\alpha \in (0,1)$, the transform $\Psi_\alpha(R)$ of R is asymptotically better than both R itself and its Fisher transform R_F. In fact, $\Psi_\alpha(R)$ is asymptotically better than R and R_F for rather wide ranges of values (say β) of the significance level; of course, these ranges include the value α itself. Indeed, one can see that in the BVN model

$$\Delta_{\psi(R)}(z) = \frac{\rho}{2}\,(2z^2-1) - \frac{z^2}{2}\frac{(1-\rho^2)\psi''(\rho)}{\psi'(\rho)}.$$

In particular,

$$\Delta_R(z) = \frac{\rho}{2}\,(2z^2-1),$$

$$\Delta_{R_F}(z) = \Delta_{\psi_\infty(R)}(z) = -\frac{\rho}{2},$$

where ψ_∞ is as defined in (4.5), and

$$\Delta_{\Psi_\alpha(R)}(z) = \Delta_{\psi_{z_\alpha}(R)}(z) = \frac{\rho}{2}\Big(\frac{z^2}{z_\alpha^2} - 1\Big).$$

So, $|\Delta_{\Psi_\alpha(R)}(z_\beta)| < |\Delta_{R_F}(z_\beta)|$ for $\rho \neq 0$ if $0 < z_\beta < z_\alpha\sqrt{2}$. That is, the transform $\Psi_\alpha(R)$, which is asymptotically optimal for the given significance level $\alpha \in (0,1)$, will still be asymptotically better than Fisher's transform R_F for any significance level $\beta \in (0,1)$ such that $0 < z_\beta < z_\alpha\sqrt{2}$. For instance, if $\alpha = 0.05$, then $\Psi_\alpha(R)$ will be asymptotically better than R_F, not just for the significance level $\alpha = 0.05$, but for any significance level $\beta \in (0.01000, 0.5)$ — because $0 < z_\beta < z_{0.05}\sqrt{2}$ for all $\beta \in (0.01000, 0.5)$. Similarly, if $\alpha = 0.01$, then $\Psi_\alpha(R)$ will be asymptotically better than R_F for any significance level $\beta \in (0.00050, 0.5)$.

As for the comparison of the asymptotically optimal transform $\Psi_\alpha(R)$ with R itself, we can similarly see that, for instance, if $\alpha = 0.05$, then $\Psi_\alpha(R)$ will be asymptotically better than R, not just for the significance level $\alpha = 0.05$, but for any significance level $\beta \in (0, 0.17912)$; if $\alpha = 0.01$, then $\Psi_\alpha(R)$ will be asymptotically better than R for any significance level $\beta \in (0, 0.16933)$.

5. SQUAREV MODEL

Here we shall consider the dependence model that is the family $(P_\rho)_{-1<\rho<1}$ of distributions of the random pair (Y,Z) on the vertices of the square $[-1,1]\times[-1,1]$ given by the following formulas:

$$\begin{aligned} \mathsf{P}_\rho\big((Y,Z)=(1,1)\big) = \mathsf{P}_\rho\big((Y,Z)=(-1,-1)\big) = \tfrac{1+\rho}{4}, \\ \mathsf{P}_\rho\big((Y,Z)=(1,-1)\big) = \mathsf{P}_\rho\big((Y,Z)=(-1,1)\big) = \tfrac{1-\rho}{4}. \end{aligned} \tag{5.1}$$

In other words, the distribution $P_{Y,Z} = P_{\rho;Y,Z}$ of (Y,Z) under P_ρ is the mixture

$$\begin{cases} (1-\rho)P_{\varepsilon_1,\varepsilon_2} + \rho P_{\varepsilon_1,\varepsilon_1} & \text{if } \rho \geq 0, \\ (1+\rho)P_{\varepsilon_1,\varepsilon_2} - \rho P_{\varepsilon_1,-\varepsilon_1} & \text{if } \rho < 0, \end{cases} \tag{5.2}$$

where $\varepsilon_1, \varepsilon_2$ are independent Rademacher r.v.'s, with $\mathsf{P}(\varepsilon_j = \pm 1) = 1/2$ for $j = 1,2$. Then $\mathsf{Cov}(Y,Z) = \rho$ under P_ρ, so that the use of the symbol ρ to denote the parameter is consistent.

The just described model of dependence will be referred to as the SquareV model, where "V" stands for "vertices".

In view of the previously mentioned symmetry $\mathsf{Cov}(Y,-Z) = -\mathsf{Cov}(Y,Z)$, negative values of ρ will not be further considered in this section.

5.1. Asymptotically Optimal Transform $\Psi_\alpha(R)$ in the SquareV Model

Using (5.1) and (5.2), we obtain the following expressions for the joint moments of (Y, Z) as defined in (3.6):

$$\mu_{ij} = \frac{1+\rho}{4}\big(1+(-1)^{i+j}\big) + \frac{1-\rho}{4}\big((-1)^i+(-1)^j\big) \tag{5.3}$$

$$= (1-\rho)\,\frac{1+(-1)^i}{2}\,\frac{1+(-1)^j}{2} + \rho\,\frac{1+(-1)^{i+j}}{2} \tag{5.4}$$

for all $i, j = 0, 1, \dots$ and all $\rho \in [0, 1)$.

As the result, ODE (3.16) becomes

$$\frac{\psi''(\rho)}{\psi'(\rho)} = q_z\,\frac{-2\rho}{1-\rho^2}, \tag{5.5}$$

where

$$q_z := \frac{1}{3z^2} - \frac{1}{3}. \tag{5.6}$$

We see that formulas (5.5)–(5.6) are rather similar to (4.1)–(4.2). Hence, quite similarly to (4.3) and (4.4), here we have

$$\psi'(\rho) = (1-\rho^2)^{q_z} \tag{5.7}$$

and

$$\psi(\rho) = \psi_{4;z}(\rho) := \int_0^\rho (1-r^2)^{q_z}\,dr = \rho\,{}_2F_1\big(\tfrac{1}{2}, -q_z; \tfrac{3}{2}; \rho^2\big). \tag{5.8}$$

Here the subscript 4 in $\psi_{4;z}$ refers to the four points of the distribution of the random point (Y, Z) in the SquareV model, currently under consideration; thus, one can distinguish between the function $\psi_{4;z}$ in (5.8) and the function ψ_z in (4.4).

Accordingly, recalling again (3.22), we see that in the SquareV model the transform

$$\Psi_{4;\alpha}(R) = \psi_{4;z_\alpha}(R)$$

of Pearson's R is asymptotically optimal for any given significance level $\alpha \in (0, 1)$.

In particular, choosing

$$z = 1,$$

we have $q_z = 0$. Hence, in view of the integral expression in (5.8), $\psi_{4;z}(\rho) = \rho$, so that the family $(\Psi_{4;\alpha}(R))_{\alpha\in(0,1)}$ of transforms of Pearson's statistic R includes R itself. More specifically,

$$R = \Psi_{4;\alpha}(R) \quad \text{for} \quad \alpha = 1 - \Phi(1) \approx 0.159.$$

However, in order for Fisher's transform R_F of R to belong to the family $(\Psi_{4;\alpha}(R))_{\alpha\in(0,1)}$ of asymptotically optimal transforms of R in the SquareV model, one would have to have $q_z = -1$ for some real z, which is impossible, because, in view of (5.6), q_z is always greater than $-\frac{1}{3}$. So, in contrast with the BVN model $\big($where, according to (1.4), R_F is asymptotically optimal in the limit case with $\alpha = 0$ and $z_\alpha = \infty\big)$, in the SquareV model Fisher's transform R_F is not asymptotically optimal for any significance level $\alpha \in [0, 1]$, even if the endpoints $\alpha = 0$ and $\alpha = 1$ are included as limit cases.

5.2. The Transform $\Psi_\alpha(R)$ in the SquareV Model is Asymptotically Better than R and R_F in Wide Ranges of Values of the Significance Level Including α Itself

This subsection is similar to Subsection 4.4.2. One can see that in the SquareV model

$$\Delta_{\psi(R)}(z) = \frac{\rho}{3\sqrt{1-\rho^2}}\,(z^2-1) - \frac{z^2}{2}\,\frac{\sqrt{1-\rho^2}\,\psi''(\rho)}{\psi'(\rho)}.$$

In particular,

$$\Delta_R(z) = \frac{\rho}{3\sqrt{1-\rho^2}}\,(z^2-1),$$

$$\Delta_{R_F}(z) = \Delta_{\psi_\infty(R)}(z) = -\frac{\rho}{3\sqrt{1-\rho^2}}\,(2z^2+1),$$

where ψ_∞ is as defined in (4.5), and

$$\Delta_{\Psi_{4;\alpha}(R)}(z) = \Delta_{\psi_{4;z_\alpha}(R)}(z) = \frac{\rho}{3\sqrt{1-\rho^2}}\Big(\frac{z^2}{z_\alpha^2}-1\Big).$$

So, we have $|\Delta_{\Psi_{4;\alpha}(R)}(z_\beta)| < |\Delta_{R_F}(z_\beta)|$ for $\rho \neq 0$ whenever $z_\alpha \geq 1/\sqrt{2}$ or, equivalently, $\alpha \in (0, 1-\Phi(1/\sqrt{2}))$, with $1-\Phi(1/\sqrt{2}) = 0.2397\ldots$. Therefore, the transform $\Psi_{4;\alpha}$, which is asymptotically optimal for the given significance level $\alpha \in (0,1)$, will still be asymptotically better than Fisher's transform R_F for any significance level $\beta \in (0, 0.5)$ provided that $\alpha \in (0, 0.2397)$.

As for the comparison of the asymptotically optimal transform $\Psi_\alpha(R)$ with R itself, we can similarly see that, for instance, if $\alpha = 0.05$, then $\Psi_\alpha(R)$ will be asymptotically better than R, not just for the significance level $\alpha = 0.05$, but for any significance level $\beta \in (0, 0.11344)$; if $\alpha = 0.01$, then $\Psi_\alpha(R)$ will be asymptotically better than R for any significance level $\beta \in (0, 0.096927)$.

We see that the advantage of the asymptotically optimal transform of R over R itself is substantially less in the SquareV model than in the BVN model. Vice versa, the advantage of the asymptotically optimal transform of R over the Fisher transform R_F of R is much greater in the SquareV model than in the BVN model.

6. CONCLUSION

The main result of this paper is Theorem 3.1, which shows for any correlation-parametrized model of dependence and for any given significance level $\alpha \in (0,1)$, there is an asymptotically optimal transform of Pearson's correlation statistic R, for which the generally leading error term for the normal approximation vanishes for all values $\rho \in (-1,1)$ of the correlation coefficient.

It is also shown that in the BVN model Pearson's R turns out to be asymptotically optimal for a rather unusual significance level $\alpha \approx 0.240$, whereas Fisher's transform R_F of R is asymptotically optimal for the limit significance level $\alpha = 0$. In the other specific model of dependence considered in this paper — the SquareV model, Pearson's R is asymptotically optimal for a still rather high significance level $\alpha \approx 0.159$, whereas Fisher's transform R_F of R is not asymptotically optimal for any $\alpha \in [0,1]$.

Moreover, we saw that in both the BVN model and the SquareV model, the transform $\Psi_\alpha(R)$, asymptotically optimal for a given value of α, is in fact asymptotically better than R and R_F in wide ranges of values of the significance level, including α itself.

Recall that Fisher's transform R_F of R was designed for the BVN case, with the purpose of making the asymptotic variance constant with respect to the correlation coefficient ρ. That R_F usually turns out to be asymptotically closer to normality than R in the BVN model might now be explained by the observation that the significance level $\alpha = 0$ (for which R_F is asymptotically optimal in the BVN case) is closer to such usual in statistical practice values of the significance level as 0.05 than to the significance level $\alpha_R \approx 0.240$ (for which R is asymptotically optimal in the BVN case).

Extensive computer simulations for the BVN and SquareV models of dependence presented in the detailed, arXiv version of this paper [9] suggest that, for sample sizes $n \geq 100$ and significance levels $\alpha \in \{0.01, 0.05\}$, the mentioned asymptotically optimal transform of R generally outperforms both Pearson's R and Fisher's transform R_F of R, the latter appearing generally much inferior to both R and the asymptotically optimal transform of R in the SquareV model.

REFERENCES

1. R. N. Bhattacharya and J. K. Ghosh, “On the Validity of the formal Edgeworth expansion”, Ann. Statist. **6**, 434–451 (1978).
2. L. H. Y. Chen and Q. M. Shao, “Normal Approximation for Nonlinear Statistics Using a Concentration Inequality Approach”, Bernoulli **13**, 581–599 (2007).
3. D. J. G. Farlie, “The Performance of Some Correlation Coefficients for a General Bivariate Distribution”, Biometrika **47**, 307–323 (1960).
4. R. A. Fisher, “Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population”, Biometrika **10**, 507–521 (1915).
5. A. K. Gayen, “The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Non-Normal Universes”, Biometrika **38**, 219–247 (1951).
6. H. Hotelling, “New Light on the Correlation Coefficient and Its Transforms”, J. Roy. Statist. Soc. Ser. B. **15**, 193–225; discussion, 225–232 (1953).
7. R. B. Nelsen, *An Introduction to Copulas*, in *Springer Series in Statistics* (Springer, New York, 2006), 2nd ed..
8. I. Pinelis, “On l’Hospital-Type Rules for Monotonicity”, J. Inequal. Pure Appl. Math. JIPAM **7**, (2005), Article 40, 19 pp. (electronic), www.emis.de/journals/JIPAM/images/157_05_JIPAM/157_05.pdf.
9. I. Pinelis, “An Asymptotically Optimal Transform of Pearson’s Correlation Statistic”, https://arxiv.org/abs/1907.11579 (20019).
10. I. Pinelis and R. Molzon, “Optimal-Order Bounds on the Rate of Convergence to Normality in the Multivariate Delta Method”, Electron. J. Statist. **10**, 1001–1063 (2016).