Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Multivariate Analysis 95 (2005) 107 – 118
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Efficiency of test for independence after Box–Cox transformation

Jade Freeman[a,b], Reza Modarres[b,*]

[a] *Office of Water, U.S. Environmental Protection Agency, 1200 Pennsylvania Ave., N.W., Washington, DC 20052, USA*

[b] *Department of Statistics, George Washington University, 2201 G St., N.W., Washington, DC 20052, USA*

Received 18 August 2003
Available online 21 September 2004

Abstract

We consider the efficiency and the power of the normal theory test for independence after a Box–Cox transformation. We obtain an expression for the correlation between the variates after a Box–Cox transformation in terms of the correlation on the normal scale. We discuss the efficiency of test of independence after a Box–Cox transformation and show that for the family considered it is always more efficient to conduct the test of independence based on Pearson correlation coefficient after transformation to normality. Power of test of independence before and after a Box–Cox transformation is studied for a finite sample size using Monte Carlo simulation. Our results show that we can increase the power of the normal-theory test for independence after estimating the transformation parameter from the data. The procedure has application for generating non-negative random variables with prescribed correlation.

AMS subject classification: 62E20; 62F03; 62H10; 62H15; 62H20

Keywords: Bivariate non-normal variables; Box–Cox transformation; Pitman's efficiency; Power; Independence

1. Introduction

We are interested in testing independence of two non-negative random variables. In reliability investigations, for example, X_1 and X_2 represent the lifetimes of components in

* Corresponding author. Fax: +1-202-9946917.
E-mail address: reza@gwu.edu (R. Modarres).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2004.08.005

a two-component system. Bivariate exponential distributions have been used to model the lifetimes. Moran [16] gives examples in point processes where it is desirable to test independence of successive intervals between points. The lengths of such intervals are unimodal and right skewed of the gamma type. Moran [16] develops a bivariate negative exponential distribution to model such data. In many environmental applications, the distribution of risk factors such as body weight, total skin area, concentration, inhalation, digestion, and consumption rates are positive and skewed to the right. Bivariate lognormal distribution is used to model the joint occurrence of the risk factors. In the preliminary examination of bivariate samples from (X_1, X_2), we would like to test the hypothesis of independence of the variables. To test for independence between pairs of variates, which are non-negative and non-symmetric, Moran [16] uses the Pearson correlation coefficient. Al-Saadi et al. [1] study the properties and compare the performance of several tests, including Pearson correlation, for independence of exponential variates. More generally, one may be interested in testing independence of two quadratic forms, whose marginal distributions are of gamma type (see [13, Chapter 48]).

Let $(y_{11}, y_{21}),\ldots,(y_{1n}, y_{2n})$ represent a sample from a bivariate normal distribution. The test of $H_0 : \rho_y = 0$ is a test of independence of Y_1 and Y_2. Let r_y be the sample correlation coefficient of Y_1 and Y_2. Fisher [5] first discovered the sampling distribution of r for samples from a bivariate normal distribution. When $\rho_y = 0$, Anderson [2, Chapter 4] shows $r_y\sqrt{n-2}/\sqrt{1-r_y^2}$ has the t-distribution with $n-2$ degrees of freedom. More generally, one may obtain the exact permutation distribution of the correlation coefficient by enumeration of the $n!$ possibilities. Pitman [18] shows that r_x, the sample correlation between X_1 and X_2, has zero expectation and variance $1/(n-1)$ when X_1 and X_2 are independent. Kowalski and Tarter [12] study the use of normalizing transformations as a prelude to applying normal-theory techniques. Given non-normal bivariate random variables (X_1, X_2), the method consists of making co-ordinate transformations $Y_i = \Phi^{-1}(\hat{F}_i(X_i))$ for $i = 1, 2$ where Φ^{-1} is the quantile function of the standard normal distribution and $\hat{F}_i$ are the Fourier estimators of the marginal distribution functions. Normal theory test of independence is then applied to (y_{1i}, y_{2i}) for $i = 1, \ldots, n$. Using Monte Carlo simulation, Kowalski and Tarter [12] show that the normal-theory test of independence is generally more powerful if they are based on r_y than r_x.

Box–Cox transformation to normality [4] is often used in practice to obtain nearly normal variates. The Box–Cox transformation parameter λ is defined (see next section) on $(-\infty, \infty)$. The literature on the Box–Cox transformation assumes that there exists a value λ such that Y has a normal distribution. However, $X > 0$ implies $Y > -1/\lambda$ for $\lambda > 0$ and $Y < -1/\lambda$ for $\lambda < 0$. Thus, for $\lambda \neq 0$, the domain of Y is not the entire real line. Researchers have generally assumed that Y has an approximate normal distribution. For example, Moore [15] sidesteps the issue by assuming that μ is large and the coefficient of variation $\kappa = \sigma/\mu$ is small for $\lambda > 0$ so that $P(Y < -1/\lambda) = P[Z < -(\frac{1}{\kappa} + 1/(\lambda\sigma))] < \varepsilon$, where ε is a small value (e.g., $\varepsilon \leqslant 10^{-6}$). Another strategy is to transform $X + c$ instead of X where c is a sufficiently large constant [6, p. 143]. Hernandez and Johnson [7] use Kullback–Leibler information number to provide benchmarks for maximum amount of improvement to normality after a Box–Cox transformation. It is evident from Hernandez and Johnson [7, p. 859, Eq. (4.2)] that any procedure based on multivariate normality of the observations should benefit from transformations to normality of the marginal distributions. Box–Cox transformation to normality provides an alternative to the Fourier-based estimator. We consider the efficiency and the power of the normal theory test for independence after a Box–Cox transformation. The next section obtains an expression for the correlation between the variates after a Box–Cox transformation in terms of the correlation on the normal scale. In Section 3, we discuss the efficiency of test of independence after a Box–Cox transformation. The last section considers a simulation study and a method for generating non-negative random variables with prescribed correlation.

2. Correlation of X_1 and X_2

Suppose (X_{11}, X_{21}),..., (X_{1n}, X_{2n}) represents n i.i.d. bivariate vectors from (X_1,X_2) with a joint distribution function F and a density function f. We will assume that all observations are non-negative and consider the bivariate Box–Cox transformation

$$Y_j = p_{\lambda_j}(X_j) \begin{cases} \frac{X_j^{\lambda_j}-1}{\lambda_j}, & \lambda_j \neq 0, \\ \ln(X_j), & \lambda_j = 0, \end{cases} \quad j = 1, 2.$$

We assume Y_1 and Y_2 have a bivariate normal distribution with mean $\vec{\mu} = (\mu_1, \mu_2)$, covariance $\Sigma = (\sigma_{ij})$ for some value of the transformation parameter $\vec{\lambda} = (\lambda_1, \lambda_2)$. The p.d.f. of $\vec{X} = (X_1, X_2)$ is

$$f(\vec{X} \mid \vec{\mu}, \Sigma, \vec{\lambda}) = \frac{1}{(2\pi) \mid \Sigma \mid^{1/2}} \prod_{j=1}^{2} x_j^{\lambda_j - 1} \exp\left[-\tfrac{1}{2}\left(p_{\vec{\lambda}}(\vec{X}) - \vec{\mu}\right)' \Sigma^{-1} \left(p_{\vec{\lambda}}(\vec{X}) - \vec{\mu}\right)\right].$$

Let ρ_x denote the correlation coefficient on the original scale. We will show that $\rho_x = h(\rho_y)$, where the form of the function h depends on $\vec{\lambda}$, $\vec{\mu}$, and Σ. Lancaster [14] uses Chebyshev–Hermite polynomial to obtain the correlation coefficient of transformed bivariate random vectors. We will obtain the form of the correlation after a Box–Cox transformation. Let $\phi^{(r)}(z)$ denote the rth derivative of the standard normal density function $\phi(z)$ with respect to z. Chebyshev–Hermite polynomial, $H_r(z)$, is defined by the identity $(-1)^r \frac{d^r}{dz^r}\phi(z) = H_r(z)\phi(z)$. It follows that $H_0(z) = 1$, $H_1(z) = z$, $H_2(z) = z^2 - 1$, $H_3(z) = z^3 - 3z$, and so on. For $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, we have $X_j = (\lambda_j Y_j + 1)^{1/\lambda_j} = \sum_{i=0}^{\infty} \binom{1/\lambda_j}{i} \lambda_j^i (\lambda_j \mu_j + 1)^{(1/\lambda_j - i)} (Y_j - \mu_j)^i$ for $j = 1, 2$. The last expression can also be written as $\sum_{i=0}^{\infty} a_{ij} Z_j^i = \sum_{i=0}^{\infty} b_{ij} H_i(Z_j)$ for some b_{ij} where $H_i(Z_j)$ is the ith Chebyshev–Hermit polynomial evaluated at $Z_j = (Y_j - \mu_j)/\sigma_j$, where $\sigma_j^2 = \sigma_{jj}$. The sum is finite and extends to $m_j = 1/\lambda_j$ when m_j is an integer. By the following orthogonal property of the polynomials, $\int_{-\infty}^{\infty} H_r(x) H_s(x) \phi(x)\, dx = \begin{cases} 0, & i \neq r, \\ r!, & i = r, \end{cases}$ we have $E(H_i(Z_j)) = \int_{-\infty}^{\infty} H_i(z_j)\phi(z_j) dz_j = 1$ at $i = 0$ and zero for $i > 0$. Therefore, we have $E(X_j) = b_{0j}$, where b_{0j} is a function of λ_j, μ_j, and σ_j^2 for $j = 1, 2$. Similarly, $E(X_j^2) = E\left(\sum_{i=0}^{\infty} b_{ij} H_i(Z_j) \sum_{k=0}^{\infty} b_{kj} H_k(Z_j)\right) = \sum_{i=0}^{\infty} b_{ij}^2 i!$ and we obtain $Var(X_j) =$

$\sum_{i=1}^{\infty} b_{ij}^2 i!$. Furthermore, $X_1X_2 = (\lambda_1 Y_1 + 1)^{1/\lambda_1}(\lambda_2 Y_2 + 1)^{1/\lambda_2}$, which can be written as $\sum_{i=0}^{\infty}\sum_{k=0}^{\infty} b_{i1}b_{k2}H_i(Z_1)H_k(Z_2)$. When both m_1 and m_2 are integers, the sum is finite and extends to $m = min(m_1, m_2)$. Table 1 contains the means, variances and covariances of (X_1, X_2) for specified $\vec{\lambda}$.

Suppose (Z_1, Z_2) is a bivariate standard normal random variable with correlation ρ_z. Then, $E(H_i(Z_1)H_k(Z_2)) = \rho_z^i i!$ for $i = k$ and zero for $i \neq k$. It follows that $E(X_1X_2) = \sum_{i=0}^{\infty} b_{i1}b_{i2}\rho_y^i i!$. Hence, the correlation coefficient ρ_x is

$$\rho_x = h(\rho_y) = \frac{\sum_{i=1}^{\infty} b_{i1}b_{i2}\rho_y^i i!}{\sqrt{(\sum_{i=1}^{\infty} b_{i1}^2 i!)(\sum_{k=1}^{\infty} b_{k2}^2 k!)}}$$

for $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$. The covariance of a bivariate lognormal distribution; i.e., $\lambda_1 = \lambda_2 = 0$, can be obtained using Chevyshev–Hermit Polynomials as follows. By Taylor's expansion of $X_j = \exp(Y_j)$ about μ_j, we have $\exp(Y_j) = \sum_{i=0}^{\infty}\exp(\mu_j)\frac{\sigma_j^i}{i!}z_j^i = \sum_{i=0}^{\infty} b_{ij}H_i(z_j)$. Note $Var(X_j) = \exp(2\mu_j + \sigma_j^2)(\exp(\sigma_j^2) - 1) = \sum_{i=1}^{\infty}[\exp(\mu_j + \frac{\sigma_j^2}{2})\frac{\sigma_j^i}{i!}]^2 i!$ and for $i \geqslant 1$ and $j = 1, 2$. Thus, $b_{ij} = \exp(\mu_j + \frac{\sigma_j^2}{2})\frac{\sigma_j^i}{i!}$. By the fact that $E(H_i(Z_1), H_j(Z_2)) = \begin{cases} 0, & i \neq j, \\ \rho^i i!, & i = j, \end{cases}$ [15], $E(X_1X_2) = \sum_{i=0}^{\infty} b_{1i}b_{2i}\rho^i i!$. Consequently, the covariance of X_1 and X_2 simplifies to $Cov(X_1, X_2) = \sum_{i=1}^{\infty}\exp(\mu_1 + \sigma_1^2/2)\sigma_1^i\exp(\mu_2 + \sigma_2^2/2)\sigma_2^i/(i!)^2\rho_y^i i!$. The last expression equals the well-known form $\exp\left(\mu_1 + \mu_2 + \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2}\right)(\exp(\rho_y\sigma_1\sigma_2) - 1)$. When $\lambda_1 = 0$ and $\lambda_2 \neq 0$, we have $E(X_1X_2) = \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right)\sum_{i=0}^{\infty} b_{2i}\frac{\sigma_1^i}{i!}\rho^i i!$. Hence, $Cov(X_1, X_2) = \exp(\mu_1 + \sigma_1^2/2)\sum_{i=1}^{\infty} b_{2i}\sigma_1\rho_y^2$. The form of $Cov(X_1, X_2)$ follows by symmetry when $\lambda_1 \neq 0$ and $\lambda_2 = 0$. The values of $b_{0i} = E(X_i)$, b_{1i}, and b_{2i} for specified $\vec{\lambda}$ appear in Tables 1 and 2. It follows from the form of the joint density of (X_1, X_2) and $h(\rho_y)$ that if $\rho_y = 0$, then $\rho_x = h(\rho_y) = 0$. Further, ρ_y is zero when $h(\rho_y) = 0$ by the transformation property of functions of independent random variables [9].

3. Efficiency of test for independence

Let $r_y = \sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)/S_1(\lambda_1)S_2(\lambda_2)$, where $S_j(\lambda_j) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{ji} - \bar{Y}_j)^2$, $j = 1, 2$, $i = 1, ..., n$. Let $r_x = \sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)/S_{X_1}S_{X_2}$. For testing the hypothesis $H_0 : \rho_x = 0$ against $H_1 : \rho_x = \tau/\sqrt{n}$ for some $\tau > 0$, we will compare the asymptotic efficiencies of the test statistics based on r_x and r_y.

Let X_i, X_j, X_k, and X_h be four random variables describing a multivariate distribution with finite fourth-order moments. Define $\mu_i = E(X_i)$, $\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$, $\sigma_{ijkh} = E(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_h - \mu_h)$, $\rho_{ij} = \sigma_{ij}(\sigma_{ii}\sigma_{jj})^{-1/2}$, and $\rho_{ijkh} = \sigma_{ijkh}(\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{hh})^{-1/2}$. Let $\vec{\rho} = (\rho_{ij}, \rho_{kh})$ and its sample correlations $\vec{r} = (r_{ij}, r_{kh})$. Then, $\sqrt{n}(\vec{r} - \vec{\rho})$ has an asymptotic normal distribution with mean vector zero and covariance matrix Γ where elements $\rho_{ij,kh}$ of Γ are given by $n \cdot Cov(r_{ij}, r_{kh})$, which equals

Table 1
Means, variances, and covariances of (X_1, X_2)

	$\vec{\lambda} = (0, 0)$
$E(X_1)$	$\exp(\mu_1 + \frac{1}{2}\sigma_1^2)$
$E(X_2)$	$\exp(\mu_2 + \frac{1}{2}\sigma_2^2)$
$Var(X_1)$	$\exp(2\mu_1 + \sigma_1^2)(\exp(\sigma_1^2) - 1)$
$Var(X_2)$	$\exp(2\mu_2 + \sigma_2^2)(\exp(\sigma_2^2) - 1)$
$Cov(X_1, X_2)$	$\exp(\mu_1 + \mu_2 + \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2)(\exp(\rho\sigma_1\sigma_2) - 1)$
	$\vec{\lambda} = (0, \frac{1}{2})$
$E(X_1)$	$\exp(\mu_1 + \frac{1}{2}\sigma_1^2)$
$E(X_2)$	$(\frac{1}{2}\mu_2 + 1)^2 + \frac{1}{4}\sigma_2^2$
$Var(X_1)$	$\exp(2\mu_1 + \sigma_1^2)(\exp(\sigma_1^2) - 1)$
$Var(X_2)$	$\sigma_2^2(\frac{1}{8}\sigma_2^2 + \frac{1}{4}\mu_2^2 + \mu_2 + 1)$
$Cov(X_1, X_2)$	$\rho\sigma_1\sigma_2 \exp(\mu_1 + \frac{1}{2}\sigma_1^2)(\frac{1}{4}\rho\sigma_1\sigma_2 + \frac{1}{2}\mu_2 + 1)$
	$\vec{\lambda} = (0, 1)$
$E(X_1)$	$\exp(\mu_1 + \frac{1}{2}\sigma_1^2)$
$E(X_2)$	$\mu_2 + 1$
$Var(X_1)$	$\exp(2\mu_1 + \sigma_1^2)(\exp(\sigma_1^2) - 1)$
$Var(X_2)$	σ_2^2
$Cov(X_1, X_2)$	$\rho\sigma_1\sigma_2 \exp(\mu_1 + \frac{1}{2}\sigma_1^2)$
	$\vec{\lambda} = (\frac{1}{2}, \frac{1}{2})$
$E(X_1)$	$(\frac{1}{2}\mu_1 + 1)^2 + \frac{1}{4}\sigma_1^2$
$E(X_2)$	$(\frac{1}{2}\mu_2 + 1)^2 + \frac{1}{4}\sigma_2^2$
$Var(X_1)$	$\sigma_1^2(\frac{1}{8}\sigma_1^2 + \frac{1}{4}\mu_1^2 + \mu_1 + 1)$
$Var(X_2)$	$\sigma_2^2(\frac{1}{8}\sigma_2^2 + \frac{1}{4}\mu_2^2 + \mu_2 + 1)$
$Cov(X_1, X_2)$	$\frac{1}{8}(\rho\sigma_1\sigma_2)^2 + \rho\sigma_1\sigma_2\frac{1}{2}(\frac{1}{2}\mu_1\mu_2 + \mu_1\mu_2 + 2)$
	$\vec{\lambda} = (\frac{1}{2}, 1)$
$E(X_1)$	$(\frac{1}{2}\mu_1 + 1)^2 + \frac{1}{4}\sigma_1^2$
$E(X_2)$	$\mu_2 + 1$
$Var(X_1)$	$\sigma_1^2(\frac{1}{8}\sigma_1^2 + \frac{1}{4}\mu_1^2 + \mu_1 + 1)$
$Var(X_2)$	σ_2^2
$Cov(X_1, X_2)$	$\rho\sigma_1\sigma_2(\frac{1}{2}\mu_1 + 1)$
	$\vec{\lambda} = (1, 1)$
$E(X_1)$	$\mu_1 + 1$
$E(X_2)$	$\mu_2 + 1$
$Var(X_1)$	σ_1^2
$Var(X_2)$	σ_2^2
$Cov(X_1, X_2)$	$\rho\sigma_1\sigma_2$

Table 2
Constant b_1, b_2 and $ARE(r_y, r_x)$ values for $\mu_1 = \mu_2 = 1$ and $\sigma_1 = \sigma_2 = 3$

$\vec{\lambda}$	b_1	b_2	ARE
(0, 0)	$b_{1i} = \exp(\mu_1 + \sigma_1^2/2)\sigma_1^i/i!$	$b_{2i} = \exp(\mu_2 + \sigma_2^2/2)\sigma_2^i/i!$	1800.4
(0, 1/2)	$b_{1i} = \exp(\mu_1 + \sigma_1^2/2)\sigma_1^i/i!$	$b_{21} = \sigma_2(\mu_2/2 + 1)$ $b_{22} = \sigma_2^2/4$	1350.3
(0, 1)	$b_{1i} = \exp(\mu_1 + \sigma_1^2/2)\sigma_1^i/i!$	$b_{21} = \sigma_2$	900.2
(1/2, 1/2)	$b_{11} = \sigma_1(\mu_1/2 + 1)$ $b_{12} = \sigma_1^2/4$	$b_{21} = \sigma_2(\mu_2/2 + 1)$ $b_{22} = \sigma_2^2/4$	22.7
(1/2, 1)	$b_{11} = \sigma_1(\mu_1/2 + 1)$ $b_{12} = \sigma_1^2/4$	$b_{21} = \sigma_2$	1.5
(1, 1)	$b_{11} = \sigma_1$	$b_{21} = \sigma_2$	1.0

$\rho_{ijkh} + \frac{1}{4}\rho_{ij}\rho_{kh}(\rho_{iikk} + \rho_{jjkk} + \rho_{iihh} + \rho_{jjhh}) - \frac{1}{2}\rho_{ij}(\rho_{iikh} + \rho_{jjkh}) - \frac{1}{2}\rho_{kh}(\rho_{ijkk} + \rho_{ijhh})$. This result was first obtained by Pearson and Filton [17] for the multivariate normal distribution. A good treatment of this subject appears in Steiger and Hakstian [19].

Let $W_j = (X_j - E(X_j))/\sqrt{Var(X_j)}$. It follows that $\sqrt{n}(r_x - \rho_x)$ has an asymptotic normal distribution with mean zero and $n\,Var(r_x) = E(W_1^2 W_2^2) + \rho_x^2(E(W_1^4)/4 + 2E(W_1^2 W_2^2) + E(W_2^4)) - \rho_x(E(W_1^3 W_2) + E(W_1 W_2^3))$. Note that when $\rho_x = 0$, X_1 and X_2 are independent and $Var(r_x)$ simplifies to $1/n$. Further, it is known that under joint normality, $\sqrt{n}(r_y - \rho_y)$ has an asymptotic normal distribution with mean zero and the variance $(1 - \rho_y^2)^2$ [2]. The Pitman's asymptotic efficiency of r_y to r_x is defined as $ARE(r_y, r_x) = \lim_{n\to\infty}\left\{\frac{[\partial E(r_y)/\partial\rho_y]^2_{|\rho_y=0}}{Var(r_y)_{|\rho=0}} \times \frac{Var(r_x)_{|\rho_y=0}}{[\partial E(r_x)/\partial\rho_y]^2_{|\rho_y=0}}\right\}$. The following lemma shows that $ARE(r_y, r_x)$ is at least 1.

Lemma. *Suppose X_1 and X_2 have a bivariate p.d.f. $f(\vec{x} \mid \vec{\mu}, \Sigma, \vec{\lambda})$. For testing $H_0 : \rho_x = 0$ against $H_1 : \rho_x = \tau/\sqrt{n}$ for some $\tau > 0$, $ARE(r_y, r_x) \geqslant 1$ with equality holding at $\vec{\lambda} =$ (1, 1).*

Proof. Using the chain rule, we have $\partial E(r_x)/\partial\rho_y = \partial E(r_x)/\partial\rho_x \times \partial\rho_x/\partial\rho_y$. Because $\partial E(r_y)/\partial\rho_{y|\rho_y=0} = 1$ and $Var(r_y)\ |_{\rho_y=0} = 1/n$, we have $\partial E(r_x)/\partial\rho_y = \partial\rho_x/\partial\rho_y$. Note $\rho_x = h(\rho_y) = 0$ if and only if $\rho_y = 0$. Thus, $H_0 : \rho_x = 0$ is equivalent to $H_0 : h(\rho_y) = 0$, which is true whenever $\rho_y = 0$. Also, $\partial h(\rho_y)/\partial\rho_y\ |_{\rho_y=0} = b_{11}b_{21}/(v_1 v_2)$ where $v_j = \sqrt{\sum_{i=1}^{\infty} b_{ji}^2 i!}$ for $j = 1, 2$ and $Var(r_x) = Var(r_y) = 1$ under the null hypothesis. The Pitman's ARE of the test statistic based on r_y to the one based on r_x is

$$ARE(r_y, r_x) = \frac{(\sum_{i=1}^{\infty} b_{1i}^2 i!)(\sum_{j=1}^{\infty} b_{2j}^2 j!)}{b_{11}^2 b_{21}^2}.$$

Thus, $ARE(r_y, r_x) \geqslant 1$ with equality holding at $\vec{\lambda} = (1, 1)$, in which case $b_{j1} = \sigma_j$, and $b_{jk} = 0$ for $j = 1, 2$ and $k > 1$. $\square$

When (X_1, X_2) have a bivariate lognormal distribution, i.e., $\lambda_1 = \lambda_2 = 0$, we have $\rho_x = (\exp(\rho_y\sigma_1\sigma_2) - 1)/\sqrt{e^{\sigma_1} - 1}\sqrt{e^{\sigma_2} - 1}$ and $ARE(r_y, r_x) = (e^{\sigma_1} - 1)(e^{\sigma_2} - 1)/\sigma_1^2\sigma_2^2 > 1$.

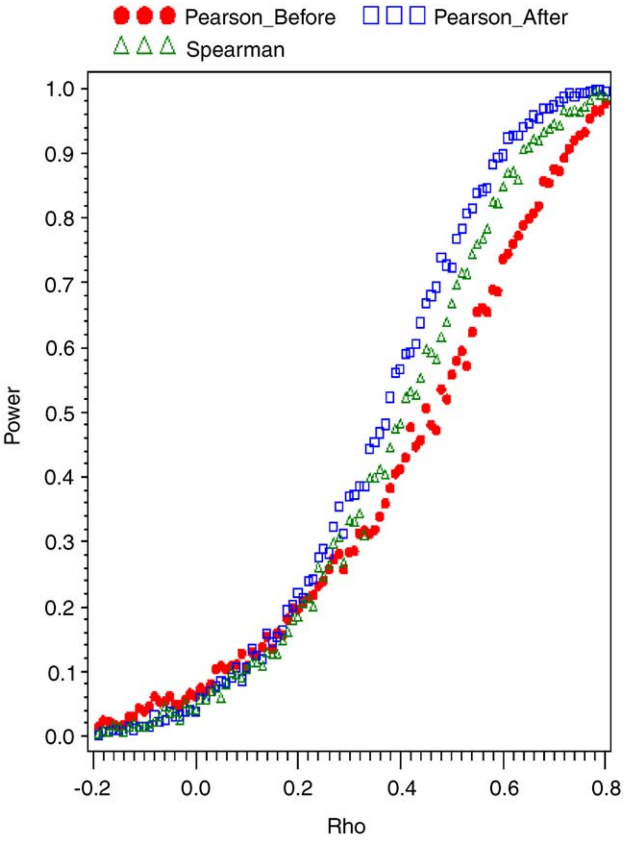

Fig. 1. Empirical power functions for $n = 20$.

4. Application

The lemma in the previous section shows that it is more efficient to conduct test independence on transformed observations. How does the procedure perform when the sample sizes are finite? More importantly how does it perform when the transformation parameter $\vec{\lambda}$ is estimated from the data? Figs. 1–3 show the power of test of independence using Pearson correlation coefficient before and after a Box–Cox transformation. We have also included the power of the Spearman rank correlation. To obtain the figures, we generated (x_{1i}, x_{2i}), $i = 1, .., n$ from a bivariate lognormal distribution with mean vector zero, unit variances and correlation ρ. The power of test of independence based on r_x and r_y were computed over 1000 simulations with $n = 20, 30, 40$ and $\rho = -0.2, \ldots, 0.8$. To obtain the scale to which we need to transform, we maximized the likelihood function and obtained an estimate for λ in interval $(0, 1)$. The figures and the asymptotic ARE show that we can increase the power of the normal theory test of independence by a Box–Cox transformation to normality. It is interesting to compare the performance of Spearman correlation. Figs. 1–3 show that the empirical

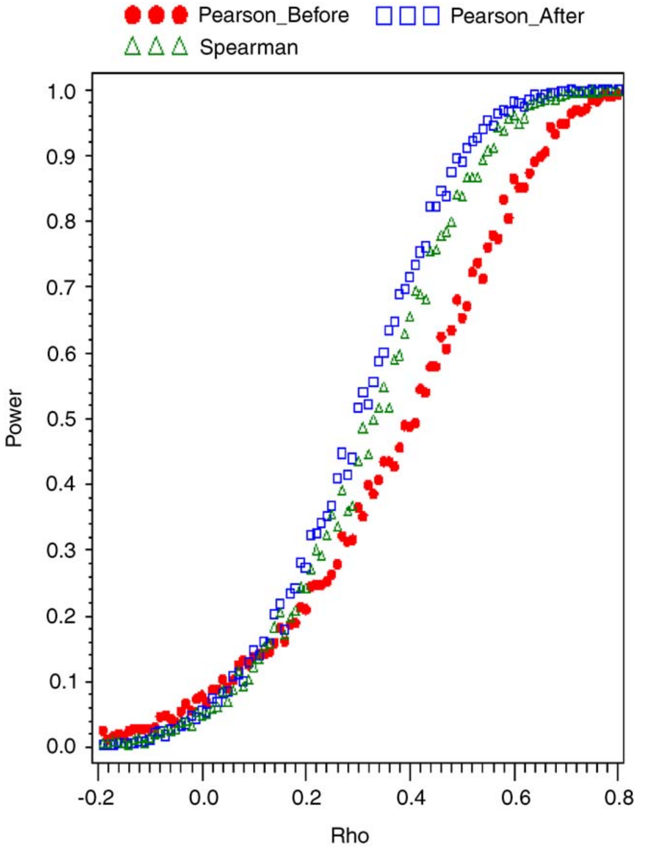

Fig. 2. Empirical power functions for $n = 30$.

power curves of the Spearman correlation are enclosed by those of Pearson correlation before and after a transformation.

Andrews [3] shows that the maximum likelihood estimate of the Box–Cox transformation parameter is sensitive to outliers. It can, however, be argued that all efficient methods depend critically on the extreme observations. In this case, extreme observations contain pertinent information for selecting the best power transformation. As noted by a referee, in our testing problem robustness to outliers may be a greater issue than efficiency. A clear lesson from the figures is that one should perform the test of independence on the original scale using Spearman rank correlation if testing independence is all that is intended and one does not wish to perform further analysis based on normal theory. In such cases, Spearman correlation is recommenced as Pearson correlation is notorious for the effects of extreme observations. Kowalski [11] demonstrates the effects of non-normality of the bivariate parent distribution on the distribution of the Perason correlation. These effects are why one may consider a transformation to normality. For example, Kowalski [10] and Kowalski and Tarter [12] assume that normal correlation analysis is robust with respect to the kinds of non-normality

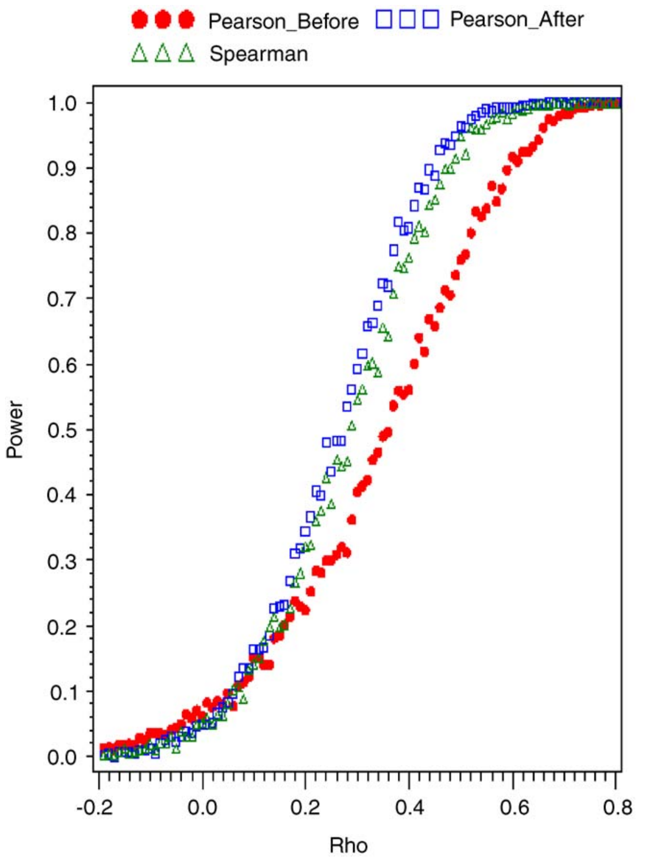

Fig. 3. Empirical power functions for $n = 40$.

possible when the marginals are constrained to be normal. Kowalski [10] notes that for a wide range of bivariate distributions, transformed correlation agrees more closely with the normal theory distribution of the sample correlation coefficient for a wide range of values of the correlation.

In many cases of interest, testing for independence is only a prelude to further analysis after a transformation to normality [8]. Box–Cox transformation is used when further investigation such as regression analysis on the transformed data is needed following the rejection of the null hypothesis of independence based on r_y. In order to see whether or not efficiency comparisons in the presence of outliers are affected we performed the following experiment with samples from a bivariate log-normal distribution with at point mass at (a) $(10^{-6}, 10^{-6})$ (Fig. 4) and (b) $(20, 20)$ (Fig. 5). With probability 0.90, we generated a random sample of size 30 from a bivariate lognormal distribution with mean vector zero, unit variances and correlation ρ and with probability 0.10 we generated from the point mass. Effects of the contamination on the power comparison can be seen in Figs. 4 and 5. One can study the behavior of the sample correlation coefficient before and

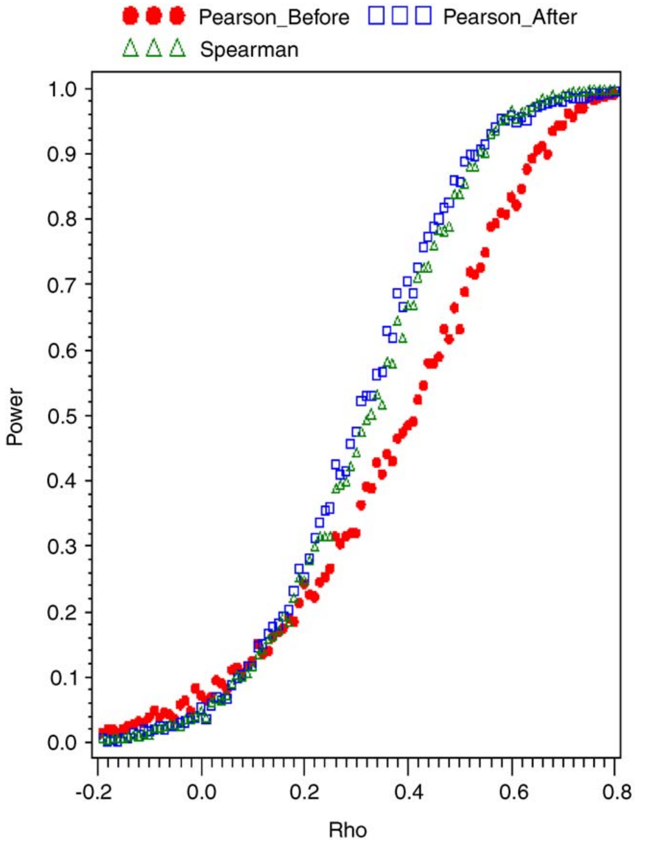

Fig. 4. Bivariate log-normal distribution with a point mass at $(10^{-6}, 10^{-6})$.

after the transformation through its influence function. We plan to investigate this further in another paper.

In many applications, we want to obtain correlated random variables with marginals that are positive and skewed. Bivariate lognormal distribution with $\vec{\lambda} = (0, 0)$ is frequently used to model such data. One can specify other values for $\vec{\lambda}$ and use the form of $h(\rho_y)$ to obtain the value of ρ_y required to induce the correlation ρ_x on observations $X_j = (\lambda_j Y_j + 1)^{1/\lambda_j}$. We generate a random sample of size n from (Y_1, Y_2), a bivariate normal distribution with mean $\vec{\mu}$ and covariance Σ. For specified $\vec{\lambda}$, we form X_j. The generated values (X_1, X_2) have means, variances and covariances determined from b's. The values of b's for some specified $\vec{\lambda}$ can be found in Tables 1 and 2. However, in general, numerical calculation may be required to obtain these.

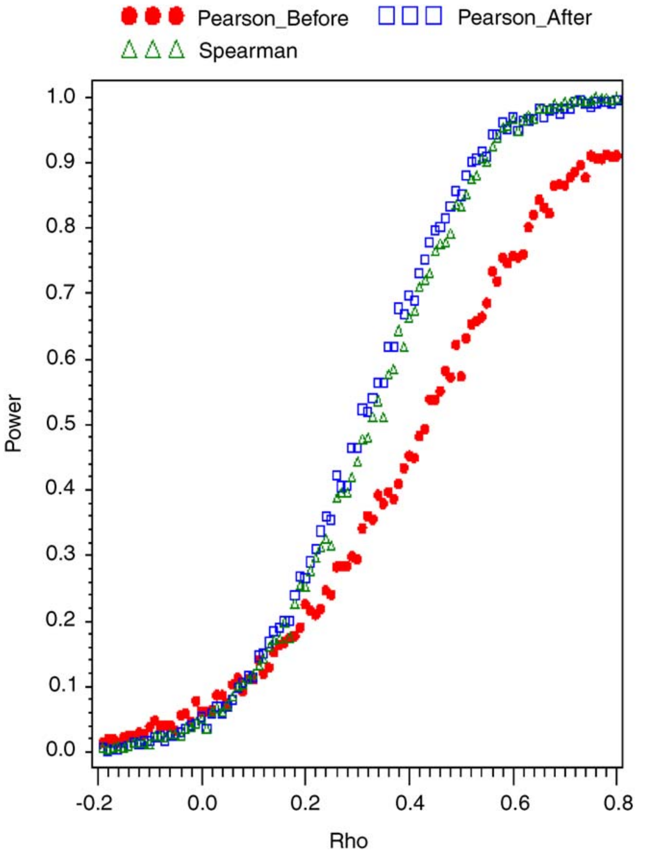

Fig. 5. Bivariate log-normal distribution with a point mass at (20, 20).

Acknowledgements

We thank the editor and two anonymous referees whose helpful comments and suggestions improved the paper.

References

[1] S.D. Al-Saadi, D.F. Scrimshaw, D.H. Young, Testing for independence of exponential variables, J. Statist. Comput. Simulation 1979, 9 (1982) 217–233.
[2] T.W. Anderson, An Introduction to Multivariate Analysis, second ed., Wiley, New York, 1984.
[3] D.F. Andrews, A note on the selection of a data transformation, Biometrika 58 (1971) 249–254.
[4] G.E. Box, D.R. Cox, An analysis of transformed data, J. Roy. Statist. Soc. B 39 (1964) 211–252.
[5] R.A. Fisher, Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika 10 (1915) 507–521.
[6] R. Gnanadesikan, Method for Statistical Data Analysis of Multivariate Observations, Wiley, New York, 1977.
[7] F. Hernandez, R.A. Johnson, The large sample behavior of transformations to normality, J. Amer. Statist. Assoc. 75 (372) (1980) 855–861.

[8] D.V. Hinkely, G. Runger, The analysis of transformed data, J. Amer. Statist. Assoc. 79 (1984) 303–320.
[9] A.F. Karr, Probability, Springer, New York, 1993.
[10] C.J. Kowalski, On the effects of non-normality on the distribution of the sample product-moment correlation coefficient, Appl. Statist. 21 (1972) 1–12.
[11] C.J. Kowalski, The performance of some rough tests for bivariate normality before and after coordinate transformation to normality, Technometrics 12 (1970) 517–544.
[12] C.J. Kowalski, M.E. Tarter, Co-ordinate transformations to normality and the power of normal tests for independence, Biometrika 56 (1) (1969) 139–148.
[13] S. Kotz, N. Balakrishnan, N.L. Johnson, Continuous Multivariate Distributions, vol. 1, Models and Applications, Second ed., John Wiley, 2000.
[14] H.O. Lancaster, The structure of bivariate distributions, Ann. Math. Statist. 29 (1958) 719–736.
[15] P.G. Moore, Transformations to normality using fractional powers of the variable, Amer. Statist. Assoc. J. (1957) 237–246.
[16] P.A.P. Moran, Testing for correlation between non-negative variates, Biometrika 54 (3,4) (1967) 385–394.
[17] K. Pearson, L.N.G. Filton, Mathematical contributions to the theory of evolution, Philos. Trans. Roy. Soc. London, Ser. A 191 (1898) 229–311.
[18] E.J.G. Pitman, Significance tests which may be applied to samples from any populations. II. The correlation coefficient test, J. Roy. Statist. Soc. Ser. B 4 (1937) 225–232.
[19] J.H. Steiger, A.R. Hakstian, The asymptotic distribution of elements of a correlation matrix: theory and application, British J. Math. Statist. Psychol. 35 (1982) 208–215.