Ann. Inst. Statist. Math.
39 (1987), Part A, 299-306

NORMALIZING AND VARIANCE STABILIZING TRANSFORMATIONS OF MULTIVARIATE STATISTICS UNDER AN ELLIPTICAL POPULATION

TAKESI HAYAKAWA

(Received May 20, 1985; revised Feb. 6, 1986)

Summary

Normalizing and variance stabilizing transformations of a sample correlation, multiple correlation and canonical correlation coefficients are obtained under an elliptical population. It is shown that the Fisher's z-transformation is efficient for these statistics. A normalizing transformation is also studied for a latent root of a sample covariance matrix in an elliptical sample.

1. Introduction

Let T_n be a statistic whose distribution depends on the parameters n and $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_p)'$. Assume that there exist $\mu(\theta)$ and $\sigma(\theta)$ such that $\sqrt{n}\{T_n-\mu(\theta)\}/\sigma(\theta)$ has a limiting normal distribution with mean 0 and variance 1 as n tends to infinity, and that the rate of convergence to normality is

$$\mathrm{P}[\sqrt{n}\{T_n-\mu(\theta)\}/\sigma(\theta)\leqq x]=\Phi(x)+O(n^{-1/2}) ,$$

where $\Phi(x)$ is the standard normal distribution function. If there exists a strictly monotone function f such that

$$\mathrm{P}[\sqrt{n}\{f(T_n)-f(\mu(\theta))-c/n\}/\{\sigma(\theta)f'(\mu(\theta))\}\leqq x]=\Phi(x)+O(n^{-1}) ,$$

where c is an asymptotic bias of the transformed variate $f(T_n)$, then Konishi [8] called $f(T_n)$ as normalized transformation of T_n. It is known that a variance stabilizing transformation is obtained by solving the differential equation $\sigma(\theta)f'(\mu(\theta))=1$ for a continuous differentiable function f in a neighborhood of $T_n=\mu(\theta)$.

Konishi [7]-[9] studied these properties of the sample correlation, multiple correlation and canonical correlation coefficients for the normal population, and showed that the Fisher's z-transformation played the

Key words and phrases: Elliptical population, sample covariance matrix, correlation coefficient, multiple correlation coefficient, canonical correlation coefficient, latent root.

fundamental role when the tail part of distribution was considered. Konishi [10] also discussed a similar problem for intraclass correlation coefficient.

The limiting distribution of sample correlation coefficient under the elliptical population was studied by Devlin, Gnanadesikan and Kettenring [2] and Muirhead [12]. Muirhead and Waternaux [13] studied the limiting distributions of the multiple correlation and the canonical correlation coefficients. This paper shows that the Fisher's z-transformation has the similar properties for an elliptical population as well as the normal population.

2. Normalization and variance stabilization

Let $\boldsymbol{x}$ be a p-dimensional random vector with an elliptical density function of the form $c_p|V|^{-1/2}g((\boldsymbol{x}-\boldsymbol{\mu})'V^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$ where c_p is a positive constant, V is a positive definite symmetric matrix and g is a non-negative function. Then the characteristic function of $\boldsymbol{x}$ has the form $\exp(i\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'V\boldsymbol{t})$ for some function ψ. The exact form of $\psi(\boldsymbol{t}'V\boldsymbol{t})$ may be found in Hayakawa and Puri [6]. The expectation and the covariance matrix of $\boldsymbol{x}$ are $\boldsymbol{\mu}$ and $\Sigma=-2\psi'(0)V=\alpha V$, respectively. Without loss of generality it is assumed that $\boldsymbol{x}$ is centered at the origin.

2.1 *Correlation coefficient*

Let r be a correlation coefficient based on a sample of size n from the elliptical population.

Let $f(r)$ be a one-to-one and twice continuously differentiable function in the neighborhood of the population correlation coefficient ρ.

With the help of Theorem 2 of Bhattacharya and Ghosh [1], Fang and Krishnaiah [4] gave the asymptotic expansion of the probability density function of $f(r)$ for non-normal population. The asymptotic expansion of the distribution of the standardized quantity of $f(r)$ for an elliptical population is given by

$$\begin{aligned}(1)\qquad & \mathrm{P}[\sqrt{n}\{f(r)-f(\rho)-c/n\}/(1+\kappa)^{1/2}(1-\rho^2)f'(\rho)\leqq x]\\ &=\Phi(x)+\frac{(1+\kappa)^{1/2}}{\sqrt{n}}\left[-\frac{1}{2}\rho+\frac{c}{(1+\kappa)f'(\rho)(1-\rho^2)}\right.\\ &\qquad\left.+x^2\left\{\rho-\frac{1}{2}(1-\rho^2)\frac{f''(\rho)}{f'(\rho)}\right\}\right]\phi(x)+o(1/\sqrt{n})\end{aligned}$$

where $\phi(x)$ is the standard normal density function and

$$\kappa=[\psi''(0)-\{\psi'(0)\}^2]/\{\psi'(0)\}^2 .$$

It should be noted that κ only depends on the functional form of the characteristic function ψ, not depends on the population parameter. The 3κ is known as the kurtosis of the variable. Following a general procedure given by Konishi [8], we first search for a function which makes the coefficient of x^2 vanish, that is,

$$\rho-\frac{1}{2}(1-\rho^2)\frac{f''(\rho)}{f'(\rho)}=0 .$$

A particular solution of this differential equation is given by

$$f(\rho)=\frac{1}{2}\log\frac{1+\rho}{1-\rho} . \tag{2}$$

This is the Fisher's z-transformation ([3]) of ρ. By choosing $c=\rho(1+\kappa)/2$, we have

$$\mathrm{P}\left[\left(\frac{n}{1+\kappa}\right)^{1/2}\left\{\frac{1}{2}\log\frac{1+r}{1-r}-\frac{1}{2}\log\frac{1+\rho}{1-\rho}-\frac{1}{2}\frac{\rho}{n}(1+\kappa)\right\}\leq x\right]=\Phi(x)+O(1/n) . \tag{3}$$

This shows that the Fisher's z-transformation has a similar property for the elliptical population as well as the normal population.

The variance stabilizing transformation is obtained by solving

$$(1+\kappa)^{1/2}(1-\rho^2)f'(\rho)=1$$

and is given as

$$f(\rho)=\frac{1}{(1+\kappa)^{1/2}}\frac{1}{2}\log\frac{1+\rho}{1-\rho} . \tag{4}$$

This implies that the Fisher's z-transformation (4) of r is also a variance stabilizing transformation for the elliptical population.

2.2 *Canonical correlation coefficient*

Let $\boldsymbol{x}=(x_1,\cdots,x_p)'$ and $\boldsymbol{y}=(y_1,\cdots,y_q)'$, $p<q$ have a covariance structure

$$\Sigma=\begin{bmatrix} I_p & P & 0 \\ P & I_p & 0 \\ 0 & 0 & I_{q-p} \end{bmatrix}=\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{matrix} p \\ q \end{matrix}$$
$$\qquad\qquad\qquad\qquad\quad p \quad\; q$$

where $P=\mathrm{diag}(\rho_1, \rho_2,\cdots,\rho_p)$, $1>\rho_1>\rho_2>\cdots>\rho_p>0$ are the population canonical correlation coefficient. Let S be a sample variance-covariance matrix based on n observations from a $(p+q)$-variate elliptical distribution. Lawley [11] gave the asymptotic expansion of the sample canonical correlation coefficient r_i, $i=1, 2,\cdots, p$ as

$$r_i^2=a_{ii}+\sum_{j\neq i}\frac{a_{ij}a_{ji}}{\rho_i^2-\rho_j^2}+o_p(1/n)$$

where $A=(a_{ij})=S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$ and S_{ij}'s are partitioned matrices of S corresponding to Σ, respectively. Then the asymptotic expansion of the distribution function of a standardized quantity is given by

$$\text{(5)}\quad \mathrm{P}\left[\sqrt{n}\{f(r_i^2)-f(\rho_i^2)-c/n\}/2\rho_i(1-\rho_i^2)f'(\rho_i^2)(1+\kappa)^{1/2}\leqq x\right]$$
$$=\Phi(x)-\frac{(1+\kappa)^{1/2}}{2\rho_i\sqrt{n}}\left[p+q-2+\rho_i^2+2(1-\rho_i^2)\sum_{j\neq i}\frac{\rho_j^2}{\rho_i^2-\rho_j^2}\right.$$
$$-\frac{c}{(1-\rho_i^2)f'(\rho_i^2)(1+\kappa)}$$
$$\left.+\left\{1-3\rho_i^2+2\rho_i^2(1-\rho_i^2)\frac{f''(\rho_i^2)}{f'(\rho_i^2)}\right\}x^2\right]\phi(x)+o(1/\sqrt{n})\ .$$

The transformation which makes the coefficient of x^2 vanish is

$$f(\rho_i^2)=\frac{1}{2}\log\frac{1+\rho_i}{1-\rho_i}\ .$$

Thus by choosing

$$c=\frac{1}{2\rho_i}(1+\kappa)\left[p+q-2+\rho_i^2+2(1-\rho_i^2)\sum_{j\neq i}\frac{\rho_j^2}{\rho_i^2-\rho_j^2}\right]\ ,$$

we have

$$\text{(6)}\quad \mathrm{P}\left[\left(\frac{n}{1+\kappa}\right)^{1/2}\left\{\frac{1}{2}\log\frac{1+r_i}{1-r_i}-\frac{1}{2}\log\frac{1+\rho_i}{1-\rho_i}\right.\right.$$
$$\left.\left.-\frac{(1+\kappa)}{2n\rho_i}\left[p+q-2+\rho_i^2+2(1-\rho_i^2)\sum_{j\neq i}\frac{\rho_j^2}{\rho_i^2-\rho_j^2}\right]\right\}\leqq x\right]$$
$$=\Phi(x)+O(1/n)\ .$$

The variance stabilizing transformation is also obtained as

$$f(\rho_i^2)=\frac{1}{(1+\kappa)^{1/2}}\frac{1}{2}\log\frac{1+\rho_i}{1-\rho_i}\ .$$

Let R be the sample multiple correlation coefficient between x and $y=(y_1,\cdots,y_q)'$ based on n sample from a $(q+1)$ dimensional elliptical distribution with a population multiple correlation coefficient ρ_R. As the special case of the canonical correlation coefficient we have

$$\text{(7)}\quad \mathrm{P}\left[\left(\frac{n}{1+\kappa}\right)^{1/2}\left\{\frac{1}{2}\log\frac{1+R}{1-R}-\frac{1}{2}\log\frac{1+\rho_R}{1-\rho_R}-\frac{(1+\kappa)}{2\rho_R n}(q-1+\rho_R^2)\right\}\leqq x\right]$$
$$=\Phi(x)+O(1/n)\ .$$

2.3 *Latent root of the sample covariance matrix*

Let $l_1>l_2>\cdots>l_p$ be the latent roots of a sample covariance matrix S based on n observation from a p-variate elliptical population and let $\lambda_1>\lambda_2>\cdots>\lambda_p>0$ be the latent roots of the population covariance matrix Σ.

Fujikoshi [5] and Waternaux [15] studied the asymptotic expansion of the probability density function of latent root for non-normal population. By the similar argument discussed in subsection 2.1, the asymptotic expansion of the distribution of $f(l_i)$ is given as follows.

$$(8)\quad \begin{aligned} &\mathrm{P}[\sqrt{n}\{f(l_i)-f(\lambda_i)-c/n\}/\lambda_i f'(\lambda_i)(6b-1)^{1/2}\leqq x] \\ &=\Phi(x)-\frac{1}{\sqrt{(6b-1)n}}\left[2b\sum_{j\neq i}\frac{\lambda_j}{\lambda_i-\lambda_j}-\frac{1}{6(6b-1)}(-90a-18b+2)\right. \\ &\quad -\frac{c}{\lambda_i f'(\lambda_i)}+\left\{\frac{1}{6(6b-1)}(-90a-18b+2)\right. \\ &\quad \left.\left.+\frac{6b-1}{2}\lambda_i\frac{f''(\lambda_i)}{f'(\lambda_i)}\right\}x^2\right]\phi(x)+o(1/\sqrt{n}), \end{aligned}$$

where

$$a=\frac{4}{3}\frac{\psi'''(0)}{\alpha^3}, \qquad b=\frac{2\psi''(0)}{\alpha^2}, \qquad \alpha=-2\psi'(0).$$

The transformation which makes the coefficient of x^2 vanish is given by

$$(9)\qquad f(\lambda)=\begin{cases} \lambda^d/d, & d\neq 0, \\ \log\lambda, & d=0, \end{cases}$$

where

$$d=\frac{90a+108b^2-18b+1}{3(6b-1)^2}.$$

The correction term c is chosen as

$$c=\lambda^d\left[2b\sum_{j\neq i}\frac{\lambda_j}{\lambda_i-\lambda_j}+\frac{45a+9b-1}{3(6b-1)}\right].$$

The variance stabilization transformation is given by

$$(10)\qquad f(\lambda_i)=\frac{1}{(1+\kappa)^{1/2}}\log\lambda_i.$$

For the normal population we have $d=1/3$ and $c=\lambda_i^{1/3}[\sum_{j\neq i}\lambda_j/(\lambda_i-\lambda_j)-2/3]$, which agrees with the result by Konishi [8].

3. Numerical comparison

In this section we give some results of simulation. Let (X_1, X_2) be a bivariate standardized normal random vector with mean zero, variance one and correlation coefficient ρ, and S a chi-squared random variable with ν degrees of freedom which is independent with (X_1, X_2). Then a bivariate t random variable is defined as $(X_1(S/\nu)^{-1/2}, X_2(S/\nu)^{-1/2})$. We calculate a sample correlation coefficient r based on $n=100$ observations from the bivariate t distribution with $\nu=10$. 10^5 sample correlation coefficients are generated and 95 percentile points R of r are obtained for $\rho=0.1(0.2)0.9$. NR stands for $n^{1/2}(R-\rho)/\{(1-\rho^2)(1+\kappa)^{1/2}\}$ and FR stands for $\left(\frac{n}{1+\kappa}\right)^{1/2}\left\{\frac{1}{2}\log\frac{1+R}{1-R}-\frac{1}{2}\log\frac{1+\rho}{1-\rho}-\frac{1}{2}\frac{\rho}{n}(1+\kappa)\right\}$ where $\kappa=2/(\nu-4)$. Comparing the transformed values with the 95 percentile points 1.644 of the standardized normal distribution, FR shows remarkable agreement.

Table 1. The 95 percentile points of sample correlation coefficient under the bivariate t-population.

ρ	R	NR	FR
0.1	0.2838	1.607	1.652
0.3	0.4629	1.550	1.641
0.5	0.6304	1.505	1.640
0.7	0.7863	1.465	1.642
0.9	0.9313	1.426	1.644

Next we consider the case of a bivariate contaminated normal distribution

$$(1-\varepsilon)\phi(\boldsymbol{x}\,|\,\Delta)+\varepsilon\sigma^{-2}\phi(\boldsymbol{x}/\sigma\,|\,\Delta)\,,$$

where $\phi(\boldsymbol{x}\,|\,\Delta)$ is the standardized normal probability density function with zero means, one variances and covariance ρ. Srivastava and Awan [14] studied the exact probability density function of a sample correlation coefficient for a bivariate contaminated normal population with same covariance matrices. The 95 percentile points of 10^5 repetitions of r based on 100 samples are given in Table 2. It shows that the Fisher's z-transformation gives good enough approximation for small ε. However, the approximation becomes inaccurate as ε increases.

Table 2. The 95 percentile points of the sample correlation coefficient under the bivariate contaminated normal population with $\sigma=0.1$.

ρ	ε	R_0	NR	FR
0.1	0.1	0.2689	1.620	1.660
	0.5	0.3250	1.623	1.684
	0.9	0.5578	1.593	1.809
0.3	0.1	0.4517	1.583	1.668
	0.5	0.5001	1.569	1.692
	0.9	0.6890	1.472	1.808
0.5	0.1	0.6224	1.549	1.679
	0.5	0.6589	1.512	1.689
	0.9	0.8002	1.379	1.821
0.7	0.1	0.7807	1.502	1.673
	0.5	0.8037	1.451	1.676
	0.9	0.8896	1.280	1.802
0.9	0.1	0.9292	1.459	1.667
	0.5	0.9376	1.412	1.690
	0.9	0.9663	1.202	1.806

Acknowledgement

Thanks are due to the referee who gave useful comments for revising an original version and also to Mrs. Inomata for numerical calculation. This work was partly supported by C.S.I.R.O., Adelaide, Australia.

HITOTSUBASHI UNIVERSITY

REFERENCES

[1] Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion, *Ann. Statist.*, **6**, 434-451.

[2] Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1976). Some multivariate applications of elliptical distributions, *Essays in Probability and Statistics in Honor of Prof. J. Ogawa* (eds. S. Ikeda et al.), Shinko-Tusho, 365-393.

[3] Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample, *Metron*, **1**, 1-32.

[4] Fang, C. and Krishnaiah, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations, *J. Multivariate Anal.*, **12**, 39-63.

[5] Fujikoshi, Y. (1980). Asymptotic expansions for the distributions of the sample roots under nonnormality, *Biometrika*, **67**, 45-51.

[6] Hayakawa, T. and Puri, M. L. (1985). Asymptotic distributions of likelihood ratio criteria for testing latent roots and latent vectors of a covariance matrix under an elliptical population, *Biometrika*, **72**, 331-338.

[7] Konishi, S. (1978). An approximation to the distribution of the sample correlation

coefficient, *Biometrika*, **65**, 654-656.
[8] Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika*, **68**, 647-651.
[9] Konishi, S. (1984). Normalizing and variance stabilizing transformations of multivariate statistics (in Japanese), *Proc. Inst. Statist. Math.*, **32**, 159-171.
[10] Konishi, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations, *Ann. Inst. Statist. Math.*, **37**, 87-94.
[11] Lawley, D. N. (1959). Tests of significance in canonical analysis, *Biometrika*, **46**, 59-66.
[12] Muirhead, R. J. (1980). The effects of elliptical distributions on some standard procedures involving correlation coefficients: A Review, *Multivariate Statistical Analysis*. (ed. Gupta, R. P.), North Holland, 143-159.
[13] Muirhead, R. J. and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika*, **67**, 31-43.
[14] Srivastava, M. S. and Awan, H. M. (1984). On the robustness of the correlation coefficient in sampling from a mixture of two bivariate normals, *Commun. Statist.-Theo. Meth.*, **13**, 371-382.
[15] Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population, *Biometrika*, **63**, 639-645.