Edgeworth Expansion and Normalizing Transformation of Ratio Statistics and Their Application

Maesono, Yoshihiko
Faculty of Mathematics, Kyushu University

出版情報：Communications in Statistics : Theory and Methods. 39 (8/9), pp.1344-1358, 2010-01. Taylor & Francis
バージョン：

Edgeworth Expansion and Normalizing Transformation of Ratio Statistics and their Application

Yoshihiko MAESONO
Faculty of Mathematics, Kyushu University
e-mail address: maesono@math.kyushu-u.ac.jp

Abstract

Some statistics in common use take a form of a ratio of two statistics, such as sample correlation coefficient, Pearson's coefficient of variation, cumulant estimators and so on. In this paper, using an asymptotic representation of the ratio statistics, we will obtain an Edgeworth expansion and a normalizing transformation with remainder term $o(n^{-1/2})$. The Edgeworth expansion is based on a studentized ratio statistic, which is studentized by a consistent variance estimator. Applying these results to the sample correlation coefficient, we obtain the normalizing transformation and an asymptotic confidence interval of the correlation coefficient without assuming specific underlying distribution. This normalizing transformation is an extension of the Fisher's z-transformation.

Keywords: Correlation coefficient; Edgeworth expansion; Fisher's z-transformation; H-decomposition; Normalizing transformation.

Mathematics Subject Classification: Primary 62E20; Secondary 62G20.

1 Introduction

Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be independently and identically distributed random vectors with distribution function F. Let $T_n = T_n(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$, $S_n = S_n(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n) \in \boldsymbol{R}$ be real valued statistics related to parameters $\lambda, \theta \in \boldsymbol{R}$. Some statistics in common use take a form of a ratio of two statistics, T_n/S_n, such as sample correlation coefficients, cumulant estimators, Pearson's coefficient of variation, odds ratio, etc. Maesono (2005) obtained an asymptotic representation of T_n/S_n and discussed asymptotic mean squared errors. In this paper, using the asymptotic representation and a variance estimator, we will obtain an asymptotic representation of a studentized ratio statistic with remainder term $(\log n)R_n$ where

$$P\{|R_n| \geq n^{-1/2}(\log n)^{-3/2}\} = o(n^{-1/2}).$$

It is easy to see that if the approximation of the distribution of T_n has a finite differential function, we have

$$|P\{T_n \le x + (\log n)R_n\} - P\{T_n \le x\}| = o(n^{-1/2}).$$

Thus we can ignore $(\log n)R_n$ when we discuss the Edgeworth expansion until the order $n^{-1/2}$. Using this asymptotic representation, an Edgeworth expansion and a normalizing transformation with remainder term $o(n^{-1/2})$ are established. Applying the results to the sample correlation coefficient, we propose an extension of the Fisher's z-transformation.

Let us assume that

$$T_n = \lambda + n^{-1}\delta_T + n^{-1}\sum_{i=1}^{n} \tau_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}} \tau_2(\boldsymbol{X}_i, \boldsymbol{X}_j) + n^{-1/2}R_n \tag{1}$$

and

$$S_n = \theta + n^{-1}\delta_S + n^{-1}\sum_{i=1}^{n} \zeta_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}} \zeta_2(\boldsymbol{X}_i, \boldsymbol{X}_j) + n^{-1/2}R_n \tag{2}$$

where δ_T and δ_S are constants. Since we consider the ratio λ/θ, we assume

$$\theta \neq 0. \tag{3}$$

$\tau_1(\cdot), \zeta_1(\cdot), \tau_2(\cdot,\cdot)$ and $\zeta_2(\cdot,\cdot)$ are real-valued functions which satisfies

$$E[\tau_1(\boldsymbol{X}_1)] = E[\zeta_1(\boldsymbol{X}_1)] = 0, \tag{4}$$

$$E[\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|\boldsymbol{X}_1] = E[\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|\boldsymbol{X}_1] = 0 \ \ a.s. \tag{5}$$

$\sum_{C_{n,k}}$ indicates that the summation is taken over all integers $i_1, \cdots, i_k$ satisfying $1 \le i_1 < i_2 < \cdots < i_r \le n$. Many statistics satisfy these assumptions (1)∼(5), and Lai and Wang (1993) called them asymptotic U-statistics.

A typical example of the ratio statistic T_n/S_n is the correlation coefficient r_n which is constituted from a covariance estimator and variance estimators. Maesono (2005) obtained an asymptotic representation of the ratio statistic T_n/S_n and discussed the asymptotic mean squared errors. In this paper we will discuss variance estimation of the ratio statistic and obtain an asymptotic representation of a studentized ratio statistic. Using the asymptotic representation, we will establish an Edgeworth expansion and a normalizing transformation, which improves coverage probabilities of confidence intervals and is an extension of the Fisher's z-transformation.

In Section 2, we will discuss the estimation of the asymptotic variance of T_n/S_n, and obtain asymptotic representations of the variance estimator and the studentized ratio statistic. We will also establish the Edgeworth expansion of the studentized ratio statistic and the normalizing transformation with remainder term $o(n^{-1/2})$. In Section 3, we will discuss estimators of unknown parameters, which appear in the Edgeworth expansion and the normalizing transformation. In Section 4, we will study an application to the sample correlation coefficient, and discuss coverage probabilities of confidence intervals, based on the Fisher's z-transformation and normalizing transformation by simulation in Section 5.

2 Asymptotic representation and Edgeworth expansion

Using H-decomposition and the moment evaluation, Maesono (2005) has obtained the asymptotic representation of T_n/S_n. Let us assume the following moment conditions

$$E[|\tau_1(\boldsymbol{X}_1)|^{4+\varepsilon} + |\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|^4 + |\zeta_1(\boldsymbol{X}_1)|^{4+\varepsilon} + |\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|^4] < \infty \tag{6}$$

for some $\varepsilon > 0$.

Let us define

$$\begin{aligned}
\delta &= \frac{\delta_T}{\theta} - \frac{\lambda\delta_S + E[\tau_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_1)]}{\theta^2} + \frac{\lambda E[\zeta_1^2(\boldsymbol{X}_1)]}{\theta^3}, \\
\eta_1(x) &= \frac{\tau_1(x)}{\theta} - \frac{\lambda\zeta_1(x)}{\theta^2}, \\
\eta_2(x,y) &= \frac{\tau_2(x,y)}{\theta} - \frac{\tau_1(x)\zeta_1(y) + \tau_1(y)\zeta_1(x) + \lambda\zeta_2(x,y)}{\theta^2} \\
&\quad + \frac{2\lambda\zeta_1(x)\zeta_1(y)}{\theta^3}
\end{aligned}$$

and

$$U_n = \frac{\lambda}{\theta} + n^{-1}\delta + n^{-1}\sum_{i=1}^n \eta_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}} \eta_2(\boldsymbol{X}_i, \boldsymbol{X}_j). \tag{7}$$

When we obtain an asymptotic representation of the ratio statistic, we use a large deviation probability for a U-statistic (see Helmers (1991), and Malevich and Abdalimov (1979)). For instance, we can show that

$$\begin{aligned}
& P\Big\{\Big|n^{-1/2}\sum_{i=1}^n \tau_1(\boldsymbol{X}_i) + n^{-3/2}\sum_{C_{n,2}} \tau_2(\boldsymbol{X}_i, \boldsymbol{X}_j)\Big||R_n| \geq n^{-1/2}(\log n)^{-1}\Big\} \\
\leq\ & P\Big\{\Big|n^{-1/2}\sum_{i=1}^n \tau_1(\boldsymbol{X}_i) + n^{-3/2}\sum_{C_{n,2}} \tau_2(\boldsymbol{X}_i, \boldsymbol{X}_j)\Big| \geq (\log n)^{1/2}\Big\} \\
& + P\{|R_n| \geq n^{-1/2}(\log n)^{-3/2}\} = o(n^{-1/2}).
\end{aligned}$$

Modifying the result of Maesono (2005), under the conditions (1) $\sim$ (6), we have the following representation

$$\frac{T_n}{S_n} = U_n + n^{-1/2}(\log n)^{1/2}R_n.$$

U_n is an approximation of the ratio statistic T_n/S_n, and we can study the asymptotic properties of the ratio statistic. It follows from the conditions (4) and (5) that

$$E[\eta_1(\boldsymbol{X}_1)] = 0, \quad \text{and} \quad E[\eta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|\boldsymbol{X}_1] = 0 \;\; a.s.$$

The asymptotic variance of $\sqrt{n}(T_n/S_n - \lambda/\theta)$ is given by

$$\xi^2 = E[\eta_1^2(\boldsymbol{X}_1)] = \frac{m_1}{\theta^2} - \frac{2\lambda m_3}{\theta^3} + \frac{\lambda^2 m_2}{\theta^4}$$

where

$$m_1 = E[\tau_1^2(\boldsymbol{X}_1)], \quad m_2 = E[\zeta_1^2(\boldsymbol{X}_1)] \;\text{ and }\; m_3 = E[\tau_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_1)].$$

For studentization, there are several variance estimators of ξ^2. Maesono (1998) has obtained Edgeworth expansions of studentized statistics with residual term $o(n^{-1})$. He discusses the studentizations based on several variance estimators, and has shown that the difference of each expansions appears in n^{-1} order term. Here we discuss the Edgeworth expansion of the studentized ratio statistic until the order $n^{-1/2}$, and it is relatively easy to get an asymptotic representation of the jackknife variance estimator. So, we will discuss a jackknife type variance estimator in the next section. We assume the estimator $\hat{\xi}^2$ satisfies

$$\hat{\xi}^2 = \xi^2 + n^{-1}\sum_{i=1}^{n} b(\boldsymbol{X}_i) + R_n \tag{8}$$

where

$$b(x) = \eta_1^2(x) - \xi^2 + 2E[\eta_1(\boldsymbol{X}_2)\eta_2(x, \boldsymbol{X}_2)].$$

As shown in the following sections, some variance estimators satisfy the assumption (8). Let us consider the studentized ratio statistic

$$\frac{\sqrt{n}(T_n/S_n - \lambda/\theta)}{\hat{\xi}}.$$

Let us define

$$\begin{aligned} g_1(x) &= \frac{1}{\xi}\eta_1(x), \\ g_2(x,y) &= \frac{1}{\xi}\eta_2(x,y) - \frac{1}{2\xi^3}[\eta_1(x)b(y) + \eta_1(y)b(x)] \end{aligned}$$

and $\nu = E[\eta_1(\boldsymbol{X}_1)b(\boldsymbol{X}_1)]$. Then, similarly as Maesono (1999), we have an asymptotic representation of the studentized ratio statistic as follows.

[Theorem 1]. *Assume the conditions* (1)$\sim$(6) *and* (8) *are satisfied, and* $E[\eta_1^2(\boldsymbol{X}_1)] = \xi^2 > 0$.

(i) *We have*

$$
\begin{aligned}
\frac{\sqrt{n}(T_n/S_n-\lambda/\theta)}{\hat{\xi}} &= n^{-1/2}\Big(\frac{\delta}{\xi}-\frac{\nu}{2\xi^3}\Big)+n^{-1/2}\sum_{i=1}^{n} g_1(\boldsymbol{X}_i) \\
&\quad + n^{-3/2}\sum_{C_{n,2}} g_2(\boldsymbol{X}_i,\boldsymbol{X}_j)+(\log n)R_n.
\end{aligned}
$$

(ii) *If* $\limsup_{|t|\to\infty}|E[\exp\{it\eta_1(\boldsymbol{X}_1)\}]|<1$, *we have*

$$
\sup_x\left|P\left\{\frac{\sqrt{n}(T_n/S_n-\lambda/\theta)}{\hat{\xi}}\le x\right\}-Q_n(x)\right|=o(n^{-1/2})
$$

where

$$
\begin{aligned}
Q_n(x) &= \Phi(x)+n^{-1/2}\phi(x)\Big[\frac{1}{6\xi^3}\Big\{(2x^2+1)E[\eta_1^3(\boldsymbol{X}_1)] \\
&\quad +3(x^2+1)E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1,\boldsymbol{X}_2)]\Big\}-\frac{\delta}{\xi}\Big].
\end{aligned}
$$

Proof. Similarly as the proof of Lemma 4 of Maesono (1999), using the Taylor expansion, we get

$$
\xi\hat{\xi}^{-1}=1-n^{-1}\sum_{i=1}^{n}\frac{b(\boldsymbol{X}_i)}{2\xi^2}+(\log n)^{1/2}R_n.
$$

Under the conditions (1)∼(6) and (8), it is easy to show that

$$
n^{-1/2}\frac{\delta}{\hat{\xi}}=n^{-1/2}\frac{\delta}{\xi}\xi\hat{\xi}^{-1}=n^{-1/2}\frac{\delta}{\xi}+(\log n)^{1/2}R_n.
$$

Furthermore, similarly as the proof of Theorem 1 of Maesono (1999), we can obtain the asymptotic representation (i) of the studentized ratio statistic.

Since the ratio of two statistics is the asymptotic U-statistic, using the Edgeworth expansion for U-statistics, we can obtain the Edgeworth expansion of the studentized ratio statistic with remainder term $o(n^{-1/2})$ in (ii).

Based on the asymptotic representation (i) of Theorem 1, we will obtain a normalizing transformation of the ratio statistic. Hall (1992) and Fujioka and Maesono (2000) discussed the normalizing transformation. Applying their results to the studentized ratio statistic, we will obtain the normalizing transformation with remainder term $o(n^{-1/2})$. Let us define

$$
p=-\frac{1}{6}E[g_1^3(\boldsymbol{X}_1)]-\frac{1}{2}E[g_1(\boldsymbol{X}_1)g_1(\boldsymbol{X}_2)g_2(\boldsymbol{X}_1,\boldsymbol{X}_2)]
$$

and

$$
q=\frac{1}{6}E[g_1^3(\boldsymbol{X}_1)]+\frac{1}{2}E[g_1(\boldsymbol{X}_1)g_1(\boldsymbol{X}_2)g_2(\boldsymbol{X}_1,\boldsymbol{X}_2)]-\Big(\frac{\delta}{\xi}-\frac{\nu}{2\xi^3}\Big).
$$

Then we have the normalizing transformation

$$\pi(s) = s + \frac{\hat{p}}{\sqrt{n}}s^2 + \frac{\hat{q}}{\sqrt{n}} + \frac{\hat{p}^2}{3n}s^3 \tag{9}$$

where $\hat{p}$ and $\hat{q}$ are consistent estimator of p and q. From Fujioka and Maesono (2000), we have the following theorem.

[Theorem 2]. *Assume the same conditions of* (ii) *in Theorem 1. If*

$$\frac{\hat{p}}{\sqrt{n}} = \frac{p}{\sqrt{n}} + (\log n)^{1/2}R_n \quad \textit{and} \quad \frac{\hat{q}}{\sqrt{n}} = \frac{q}{\sqrt{n}} + (\log n)^{1/2}R_n,$$

we have

$$\sup_x \left| P\left\{ \pi\left(\frac{\sqrt{n}(T_n/S_n - \lambda/\theta)}{\hat{\xi}} \right) \leq x \right\} - \Phi(x) \right| = o(n^{-1/2}).$$

From direct computation, we get

$$p = \frac{1}{3\xi^3}E[\eta_1^3(\boldsymbol{X}_1)] + \frac{1}{2\xi^3}E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)] \tag{10}$$

and

$$q = \frac{1}{6\xi^3}E[\eta_1^3(\boldsymbol{X}_1)] + \frac{1}{2\xi^3}E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)] - \frac{\delta}{\xi}. \tag{11}$$

Further, let us define

$$\begin{aligned}
&m_4 = E[\tau_1^3(\boldsymbol{X}_1)], \quad m_5 = E[\zeta_1^3(\boldsymbol{X}_1)], \quad m_6 = E[\tau_1^2(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_1)],\\
&m_7 = E[\tau_1(\boldsymbol{X}_1)\zeta_1^2(\boldsymbol{X}_1)], \quad m_8 = E[\tau_1(\boldsymbol{X}_1)\tau_1(\boldsymbol{X}_2)\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)],\\
&m_9 = E[\zeta_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_2)\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)], \quad m_{10} = E[\tau_1(\boldsymbol{X}_1)\tau_1(\boldsymbol{X}_2)\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)],\\
&m_{11} = E[\zeta_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_2)\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)], \quad m_{12} = E[\tau_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_2)\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)]
\end{aligned}$$

and

$$m_{13} = E[\tau_1(\boldsymbol{X}_1)\zeta_1(\boldsymbol{X}_2)\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)].$$

Then, from direct computation, we have

$$\begin{aligned}
\delta &= \frac{\delta_T}{\theta} - \frac{\lambda\delta_S + m_3}{\theta^2} + \frac{\lambda m_2}{\theta^3},\\
E[\eta_1^3(\boldsymbol{X}_1)] &= \frac{m_4}{\theta^3} - \frac{3\lambda m_6}{\theta^4} + \frac{3\lambda^2 m_7}{\theta^5} - \frac{\lambda^3 m_5}{\theta^6}
\end{aligned}$$

and

$$\begin{aligned}
& E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)]\\
= \; & \frac{m_8}{\theta^3} - \frac{1}{\theta^4}(2\lambda m_{12} + 2m_1 m_3 + \lambda m_{10})\\
& + \frac{\lambda}{\theta^5}(\lambda m_{11} + 2m_1 m_2 + 6m_3^2 + 2\lambda m_{13}) - \frac{\lambda^2}{\theta^6}(9m_2 m_3 + \lambda m_9) + \frac{4\lambda^3 m_2^2}{\theta^7}.
\end{aligned}$$

In the next section, we will discuss the jackknife type estimators of the variance ξ^2 and these unknown parameters.

3 Jackknife type estimators

Let $T_n^{(i)}$ and $S_n^{(i)}$ denote corresponding statistics computed from a sample of $n-1$ points with $\boldsymbol{X}_i$ left out. Let us define

$$\hat{t}_1(i) = T_n^{(i)} - T_n \quad \text{and} \quad \hat{s}_1(i) = S_n^{(i)} - S_n.$$

Then jackknife estimators of m_1, m_2 and m_3 are given by

$$\hat{m}_1 = (n-1)\sum_{i=1}^{n} \hat{t}_1^2(i), \qquad \hat{m}_2 = (n-1)\sum_{i=1}^{n} \hat{s}_1^2(i)$$

and

$$\hat{m}_3 = (n-1)\sum_{i=1}^{n} \hat{t}_1(i)\hat{s}_1(i).$$

Using these estimators, we have the jackknife type variance estimator

$$\hat{\xi}^2 = \frac{\hat{m}_1}{S_n^2} - \frac{2T_n\hat{m}_3}{S_n^3} + \frac{T_n^2\hat{m}_2}{S_n^4}.$$

In order to satisfy the condition (8), we have to assume the stronger conditions as follows. Let us assume

$$\begin{aligned} T_n &= \lambda + n^{-1}\delta_T + n^{-2}\sum_{i=1}^{n}\tau_0(\boldsymbol{X}_i) + n^{-1}\sum_{i=1}^{n}\tau_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}}\tau_2(\boldsymbol{X}_i,\boldsymbol{X}_j) \\ &\quad + n^{-3}\sum_{C_{n,3}}\tau_3(\boldsymbol{X}_i,\boldsymbol{X}_j,\boldsymbol{X}_k) + n^{-1/2}\tilde{R}_n \end{aligned} \tag{12}$$

and

$$\begin{aligned} S_n &= \theta + n^{-1}\delta_S + n^{-2}\sum_{i=1}^{n}\zeta_0(\boldsymbol{X}_i) + n^{-1}\sum_{i=1}^{n}\zeta_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}}\zeta_2(\boldsymbol{X}_i,\boldsymbol{X}_j) \\ &\quad + n^{-3}\sum_{C_{n,3}}\zeta_3(\boldsymbol{X}_i,\boldsymbol{X}_j,\boldsymbol{X}_k) + n^{-1/2}\tilde{R}_n \end{aligned} \tag{13}$$

where

$$E[\tau_0(\boldsymbol{X}_1)] = E[\zeta_0(\boldsymbol{X}_1)] = E[\tau_1(\boldsymbol{X}_1)] = E[\zeta_1(\boldsymbol{X}_1)] = 0, \tag{14}$$

$$E[\tau_2(\boldsymbol{X}_1,\boldsymbol{X}_2)|\boldsymbol{X}_1] = E[\zeta_2(\boldsymbol{X}_1,\boldsymbol{X}_2)|\boldsymbol{X}_1] = 0 \;\; a.s., \tag{15}$$

$$E[\tau_3(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)|\boldsymbol{X}_1,\boldsymbol{X}_2] = E[\zeta_3(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)|\boldsymbol{X}_1,\boldsymbol{X}_2] = 0 \;\; a.s. \tag{16}$$

Further we assume that for some $d > 0$

$$E|\tilde{R}_n|^{2(1+d)} = O(n^{-3(1+d)}), \tag{17}$$

$$E[|\tau_0(\boldsymbol{X}_1)|^{2+\varepsilon} + |\zeta_0(\boldsymbol{X}_1)|^{2+\varepsilon}] < \infty, \tag{18}$$

$$E[|\tau_1(\boldsymbol{X}_1)|^{4+\varepsilon} + |\tau_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|^4 + |\tau_3(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)|^{8/3}] < \infty \tag{19}$$

and

$$E[|\zeta_1(\boldsymbol{X}_1)|^{4+\varepsilon} + |\zeta_2(\boldsymbol{X}_1, \boldsymbol{X}_2)|^4 + |\zeta_3(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)|^{8/3}] < \infty \tag{20}$$

for some $\varepsilon > 0$.

Similarly as Maesono (1999), we have the following lemma.

[Lemma 1]. *Assume the conditions* (12)$\sim$(20)*. Then the jackknife type variance estimator* $\hat{\xi}^2$ *satisfies the equation* (8)*.*

Proof. Let $\boldsymbol{X}$ denote an independent random vector of $\{\boldsymbol{X}_i\}$, and have same distribution of $\boldsymbol{X}_i$. Then, it follows from Maesono (1999) that

$$\begin{aligned}
\hat{m}_1 &= m_1 + n^{-1}\sum_{i=1}^{n}\{\tau_1^2(\boldsymbol{X}_i) - m_1 + 2E[\tau_1(X)\tau_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i]\} + R_n, \\
\hat{m}_2 &= m_2 + n^{-1}\sum_{i=1}^{n}\{\zeta_1^2(\boldsymbol{X}_i) - m_2 + 2E[\zeta_1(X)\zeta_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i]\} + R_n
\end{aligned}$$

and

$$\begin{aligned}
\hat{m}_3 &= m_3 + n^{-1}\sum_{i=1}^{n}\{\tau_1(\boldsymbol{X}_i)\zeta(\boldsymbol{X}_i) - m_3 + E[\tau_1(X)\zeta_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i] \\
&\qquad + E[\zeta_1(X)\tau_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i]\} + R_n.
\end{aligned}$$

From direct computation, we can show that

$$\begin{aligned}
S_n^2 &= \theta^2 + n^{-1}\sum_{i=1}^{n} 2\theta\zeta_1(\boldsymbol{X}_i) + R_n, \\
S_n^3 &= \theta^3 + n^{-1}\sum_{i=1}^{n} 3\theta^2\zeta_1(\boldsymbol{X}_i) + R_n, \\
S_n^4 &= \theta^4 + n^{-1}\sum_{i=1}^{n} 4\theta^3\zeta_1(\boldsymbol{X}_i) + R_n, \\
T_n^2 &= \lambda^2 + n^{-1}\sum_{i=1}^{n} 2\lambda\tau_1(\boldsymbol{X}_i) + R_n, \\
2T_n\hat{m}_3 &= 2\lambda m_3 + n^{-1}\sum_{i=1}^{n}\Big(2m_3\tau_1(\boldsymbol{X}_i) + 2\lambda\Big\{\tau_1(\boldsymbol{X}_i)\zeta(\boldsymbol{X}_i) - m_3 \\
&\qquad + E[\tau_1(X)\zeta_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i] + E[\zeta_1(X)\tau_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i]\Big\}\Big) + R_n
\end{aligned}$$

and

$$\begin{aligned}
T_n^2\hat{m}_2 &= \lambda^2 m_2 + n^{-1}\sum_{i=1}^{n}\Big(2\lambda m_2\tau_1(\boldsymbol{X}_i) + \lambda^2\Big\{\zeta^2(\boldsymbol{X}_i) - m_2 \\
&\qquad + 2E[\zeta_1(X)\zeta_2(\boldsymbol{X}_i, \boldsymbol{X})|\boldsymbol{X}_i]\Big\}\Big) + R_n.
\end{aligned}$$

Note that $\hat{\xi}^2$ is a linear combination of the ratio statistics of the above statistics. Applying the asymptotic representation of the ratio statistic in (7), we have

$$\begin{aligned}
\frac{\hat{m}_1}{S_n^2} &= \frac{m_1}{\theta^2} + n^{-1}\sum_{i=1}^{n}\Big\{\frac{\tau_1^2(\boldsymbol{X}_i) - m_1}{\theta^2} + \frac{2E[\tau_1(X)\tau_2(\boldsymbol{X}_i,\boldsymbol{X})|\boldsymbol{X}_i]}{\theta^2} \\
&\quad -\frac{2m_1}{\theta^3}\zeta_1(\boldsymbol{X}_i)\Big\} + (\log n)^{1/2}R_n, \\
\frac{2T_n\hat{m}_3}{S_n^3} &= \frac{2\lambda m_3}{\theta^3} + n^{-1}\sum_{i=1}^{n}\Big(2\lambda\Big\{\frac{\tau_1(\boldsymbol{X}_i)\zeta_1(\boldsymbol{X}_i) - m_3}{\theta^3} \\
&\quad +\frac{E[\tau_1(X)\zeta_2(\boldsymbol{X}_i,\boldsymbol{X})|\boldsymbol{X}_i]}{\theta^3} + \frac{E[\zeta_1(X)\tau_2(\boldsymbol{X}_i,\boldsymbol{X})|\boldsymbol{X}_i]}{\theta^3}\Big\} \\
&\quad +\frac{2m_3}{\theta^3}\tau_1(\boldsymbol{X}_i) - \frac{6\lambda m_3}{\theta^4}\zeta_1(\boldsymbol{X}_i)\Big) + (\log n)^{1/2}R_n
\end{aligned}$$

and

$$\begin{aligned}
\frac{T_n^2\hat{m}_2}{S_n^4} &= \frac{\lambda^2 m_2}{\theta^4} + n^{-1}\sum_{i=1}^{n}\Big(\frac{\lambda^2\{\zeta_1^2(\boldsymbol{X}_i) - m_2\}}{\theta^4} + \frac{2\lambda^2E[\zeta_1(X)\zeta_2(\boldsymbol{X}_i,\boldsymbol{X})|\boldsymbol{X}_i]}{\theta^4} \\
&\quad +\frac{2\lambda m_2}{\theta^4} - \frac{4\lambda^2 m_2}{\theta^5}\zeta_1(\boldsymbol{X}_i)\Big) + (\log n)^{1/2}R_n.
\end{aligned}$$

Combining the above evaluation, we have the desired result.

Assuming the existence of the moments of the symmetric statistic, Maesono (1999) has obtained the asymptotic representation of the jackknife variance estimator. It is too restrictive to assume the existence of the moments of the ratio statistic, and then we assume the conditions (12)$\sim$(20).

Similarly as $\hat{t}_1(i)$ and $\hat{s}_1(i)$, let us define

$$\hat{t}_2(i,j) = -[nT_n - (n-1)(T_n^{(i)} + T_n^{(j)}) + (n-2)T_n^{(i,j)}]$$

and

$$\hat{s}_2(i,j) = -[nS_n - (n-1)(S_n^{(i)} + S_n^{(j)}) + (n-2)S_n^{(i,j)}]$$

where $T_n^{(i,j)}$ and $S_n^{(i,j)}$ are computed from a sample of $n-2$ points with $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ lef out. Then jackknife estimators of unknown parameter $m_4 \sim m_{13}$ are given by

$$\hat{m}_4 = \frac{(n-1)^3}{n}\sum_{i=1}^{n}\hat{t}_1^3(i), \quad \hat{m}_5 = \frac{(n-1)^3}{n}\sum_{i=1}^{n}\hat{s}_1^3(i),$$
$$\hat{m}_6 = \frac{(n-1)^3}{n}\sum_{i=1}^{n}\hat{t}_1^2(i)\hat{s}_1(i), \quad \hat{m}_7 = \frac{(n-1)^3}{n}\sum_{i=1}^{n}\hat{t}_1(i)\hat{s}_1^2(i),$$

$$\hat{m}_8 = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{t}_1(i)\hat{t}_1(j)\hat{t}_2(i,j),$$

$$\hat{m}_9 = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{s}_1(i)\hat{s}_1(j)\hat{s}_2(i,j),$$

$$\hat{m}_{10} = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{t}_1(i)\hat{t}_1(j)\hat{s}_2(i,j),$$

$$\hat{m}_{11} = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{s}_1(i)\hat{s}_1(j)\hat{t}_2(i,j),$$

$$\hat{m}_{12} = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{t}_1(i)\hat{s}_1(j)\hat{t}_2(i,j)$$

and

$$\hat{m}_{13} = \frac{(n-1)^2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{t}_1(i)\hat{s}_1(j)\hat{s}_2(i,j).$$

Jackknife estimators of the biases δ_T and δ_S are also given by

$$\hat{\delta}_T = n(n-1)(\overline{T}_n - T_n) \quad \text{and} \quad \hat{\delta}_S = n(n-1)(\overline{S}_n - S_n)$$

where

$$\overline{T}_n = n^{-1} \sum_{i=1}^{n} T_n^{(i)} \quad \text{and} \quad \overline{S}_n = n^{-1} \sum_{i=1}^{n} S_n^{(i)}.$$

Substituting these values to p and q, we can obtain the normalizing transformation $\pi(s)$. Alternatively, if we know more precise structures of ξ^2, p and q, it is possible to make another estimators $\hat{\xi}^2$, $\hat{p}$ and $\hat{q}$. In the next section, we will discuss another estimators in the case of the sample correlation coefficient.

4 Correlation coefficient

Let $\{\boldsymbol{X}_i\}_{i \geq 1}$ be two dimensional random vectors, and putting $\boldsymbol{X}_i^t = (Y_i, Z_i)$, we denote

$$Var(\boldsymbol{X}_1) = Var\Big[\begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix}\Big] = \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_z \\ \rho\sigma_y\sigma_z & \sigma_z^2 \end{pmatrix}.$$

Let us consider the sample correlation coefficient. Define

$$T_n = (n-1)^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y})(Z_i - \bar{Z})$$

and

$$S_n = \{(n-1)^{-2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \sum_{i=1}^{n} (Z_i - \bar{Z})^2\}^{1/2}$$

where $\bar{Y} = n^{-1}\sum Y_i$ and $\bar{Z} = n^{-1}\sum Z_i$. Then $r_n = T_n/S_n$ is the sample correlation coefficient. Fujisawa (2000) discussed the normalizing transformation of the coefficient r_n. He obtained the transformation when the underlying distributions are a bivariate normal and an elliptical distribution. His transformation deeply depends on the underlying distribution. Here applying the Theorem 2 and 3, we will obtain an Edgeworth expansion and a normalizing transformation without assuming the specific underlying distribution. Let us define $\tilde{y}_i = y_i - E(Y_1),\ \tilde{z}_i = z_i - E(Z_1)$,

$$\begin{aligned} \tau_1(\boldsymbol{X}_1) &= \tilde{y}_1\tilde{z}_1 - \rho\sigma_y\sigma_z, \qquad\qquad \tau_2(\boldsymbol{X}_1,\boldsymbol{X}_2) = -(\tilde{y}_1\tilde{z}_2 + \tilde{y}_2\tilde{z}_1), \\ \delta_S &= -\frac{E[(\sigma_z^2\tilde{Y}_1^2 - \sigma_y^2\tilde{Z}_1^2)^2]}{8\sigma_y^3\sigma_z^3}, \\ \zeta_1(\boldsymbol{X}_1) &= \frac{\sigma_z^2\tilde{y}_1^2 + \sigma_y^2\tilde{z}_1^2 - 2\sigma_y^2\sigma_z^2}{2\sigma_y\sigma_z} \end{aligned}$$

and

$$\zeta_2(\boldsymbol{X}_1,\boldsymbol{X}_2) = -\frac{(\sigma_z^2\tilde{y}_1^2 - \sigma_y^2\tilde{z}_1^2)(\sigma_z^2\tilde{y}_2^2 - \sigma_y^2\tilde{z}_2^2)}{4\sigma_y^3\sigma_z^3} - \frac{\sigma_z\tilde{y}_1\tilde{y}_2}{\sigma_y} - \frac{\sigma_y\tilde{z}_1\tilde{z}_2}{\sigma_z}.$$

Maesono (2005) obtained the following representation.
[Lemma 2]. *If $E[|Y_1|^{4+\varepsilon} + |Z_1|^{4+\varepsilon}] < \infty$ for some $\varepsilon > 0$, we have*

$$T_n = \rho\sigma_y\sigma_z + n^{-1}\sum_{i=1}^n \tau_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}} \tau_2(\boldsymbol{X}_i,\boldsymbol{X}_j) + n^{-1/2}R_n$$

and

$$S_n = \sigma_y\sigma_z + n^{-1}\delta_S + n^{-1}\sum_{i=1}^n \zeta_1(\boldsymbol{X}_i) + n^{-2}\sum_{C_{n,2}} \zeta_2(\boldsymbol{X}_i,\boldsymbol{X}_j) + n^{-1/2}R_n.$$

Let us define

$$\mu_{k\ell} = \frac{E[\{Y_1 - E(Y_1)\}^k\{Z_1 - E(Z_1)\}^\ell]}{\sigma_y^k\sigma_z^\ell}, \quad k = 0,1,\cdots,6; \ell = 0,1,\cdots,6.$$

Then from direct computation, we have

$$\delta = -\frac{\mu_{31}}{2} - \frac{\mu_{13}}{2} + \Big(\frac{3\mu_{40}}{8} + \frac{3\mu_{04}}{8} + \frac{\mu_{22}}{4}\Big)\rho. \tag{21}$$

Similarly, we can get

$$\begin{aligned} E[\eta_1^3(\boldsymbol{X}_1)] &= \mu_{33} - \Big(\frac{3\rho}{2} + \frac{3\rho^3}{8}\Big)(\mu_{42} + \mu_{24}) \\ &\quad + \frac{3\rho^2}{2}\mu_{33} + \frac{3\rho^2}{4}(\mu_{51} + \mu_{15}) - \frac{\rho^3}{8}(\mu_{60} + \mu_{06}) \end{aligned} \tag{22}$$

and

$$
\begin{aligned}
& E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1,\boldsymbol{X}_2)] \\
= \; & -\Big(2+\frac{5\rho^2}{2}\Big)\mu_{21}\mu_{12}+\Big(2\rho+\frac{\rho^3}{4}\Big)(\mu_{21}^2+\mu_{12}^2) \\
& +\Big(\rho+\frac{\rho^3}{2}\Big)(\mu_{30}\mu_{12}+\mu_{21}\mu_{03})-\frac{3\rho^2}{2}(\mu_{30}\mu_{21}+\mu_{12}\mu_{03})-\frac{\rho^2}{2}\mu_{30}\mu_{03} \\
& +\frac{\rho^3}{4}(\mu_{30}^2+\mu_{03}^2)+\Big(\rho+\frac{\rho^3}{2}\Big)\mu_{22}^2+\frac{3\rho}{2}\mu_{31}\mu_{13}+\frac{\rho^3}{8}\mu_{40}\mu_{04} \\
& -\Big(1+\frac{3\rho^2}{2}\Big)(\mu_{22}\mu_{31}+\mu_{22}\mu_{13})+\Big(\frac{\rho}{2}+\frac{\rho^3}{2}\Big)(\mu_{40}\mu_{22}+\mu_{22}\mu_{04}) \\
& +\frac{5\rho}{4}(\mu_{31}^2+\mu_{13}^2)-\rho^2(\mu_{40}\mu_{31}+\mu_{13}\mu_{04}) \\
& -\frac{\rho^2}{2}(\mu_{40}\mu_{13}+\mu_{31}\mu_{04})+\frac{3\rho^3}{16}(\mu_{40}^2+\mu_{04}^2). \qquad (23)
\end{aligned}
$$

It is easy to make estimators of the above unknown parameters. An estimator of the correlation ρ is sample correlation coefficient, and estimators of $\mu_{k,\ell},(k=0,1,\cdots,6;\ell=0,1,\cdots,6)$ are given by

$$
\hat{\mu}_{k,\ell}=\frac{n^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^k(Z_i-\bar{Z})^\ell}{\hat{\sigma}_y^k\hat{\sigma}_z^\ell},\quad k=0,1,\cdots,6;\ell=0,1,\cdots,6 \qquad (24)
$$

where

$$
\hat{\sigma}_y^2=(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 \quad \text{and} \quad \hat{\sigma}_z^2=(n-1)^{-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2.
$$

5 Simulation

Here we will compare confidence intervals based on the Fisher's z-transformation and the normalizing transformation. Substituting the estimators in (24) to δ, $E[\eta_1^3(\boldsymbol{X}_1)]$ and $E[\eta_1(\boldsymbol{X}_1)\eta_1(\boldsymbol{X}_2)\eta_2(\boldsymbol{X}_1,\boldsymbol{X}_2)]$ in (21), (22) and (23), we can get the estimators $\hat{p}$ and $\hat{q}$ in (10) and (11). Then we can construct normalizing transformation $\pi(s)$ in (9).

Using the inversion of Hall (1992), we can get the confidence interval based on the normalizing transformation i.e.

$$
1-\alpha\approx P\Big\{\pi^{-1}(-z_{\alpha/2})\le\rho\le\pi^{-1}(z_{\alpha/2})\Big\} \qquad (25)
$$

where

$$
\pi^{-1}(t)=\frac{\sqrt{n}}{\hat{p}}\Big\{1+\frac{3\hat{p}}{\sqrt{n}}\Big(t-\frac{\hat{q}}{\sqrt{n}}\Big)\Big\}^{1/3}-\frac{\sqrt{n}}{\hat{p}}.
$$

The confidence intervals based on simple normal approximation and the Fisher's z-transformation are given by

$$
1-\alpha\approx P\Big\{r_n-n^{-1/2}\hat{\xi}z_{\alpha/2}\le\rho\le r_n+n^{-1/2}\hat{\xi}z_{\alpha/2}\Big\} \qquad (26)
$$

and

$$1-\alpha \approx P\Big\{\frac{e^{2z_1}-1}{e^{2z_1}+1} \le \rho \le \frac{e^{2z_2}-1}{e^{2z_2}+1}\Big\} \tag{27}$$

where

$$z_1 = \frac{1}{2}\log\frac{1+r_n}{1-r_n} - \frac{z_{\alpha/2}}{\sqrt{n-3}} \quad \text{and} \quad z_2 = \frac{1}{2}\log\frac{1+r_n}{1-r_n} + \frac{z_{\alpha/2}}{\sqrt{n-3}}.$$

Here we consider the random vector $\boldsymbol{X}_i^t = (Y_i, Z_i)$ with

$$Y_i = V_i + K_i \quad \text{and} \quad Z_i = W_i + K_i$$

where $\{V_i\}, \{W_i\}$ and $\{K_i\}$ are all independently and identically distributed random variables, i.e. the underlying distributions of $\{V_i\}, \{W_i\}$ and $\{K_i\}$ are same. Thus the correlation coefficient $\rho = 0.5$. Here we consider the cases that the underlying distributions of V_i are normal, χ^2 with 2-dgrees of freedom and log-normal. Table 1 shows the estimated coverage probability based on 1,000,000 times repetition when underlying distribution is normal. "Simple" denotes the simple normal approximation in (26). "N-T" and "Fisher" mean the confidence intervals (25) and (27).

Table 1. Normal distribution

	$1-\alpha$	$n=20$	$n=50$	$n=100$
Simple	0.90	0.846	0.879	0.889
Fisher		0.901	0.900	0.900
N-T		0.849	0.877	0.888
Simple	0.95	0.897	0.929	0.939
Fisher		0.950	0.950	0.949
N-T		0.915	0.934	0.941
Simple	0.99	0.951	0.976	0.983
Fisher		0.989	0.990	0.990
N-T		0.973	0.984	0.987

Table 2 and 3 are simulation results when the underlying distributions are χ^2 with 2-degrees of freedom and log-normal, respectively.

Table 2. χ^2 with 2-degrees of freedom

	$1-\alpha$	$n=20$	$n=50$	$n=100$
Simple	0.90	0.735	0.753	0.756
Fisher		0.796	0.777	0.768
N-T		0.818	0.844	0.859
Simple	0.95	0.802	0.826	0.832
Fisher		0.870	0.854	0.845
N-T		0.886	0.912	0.921
Simple	0.99	0.881	0.912	0.921
Fisher		0.952	0.943	0.938
N-T		0.953	0.974	0.979

Table 3. Log-normal distribution

	$1-\alpha$	$n=20$	$n=50$	$n=100$
Simple	0.90	0.605	0.557	0.513
Fisher		0.654	0.570	0.517
N-T		0.739	0.795	0.806
Simple	0.95	0.682	0.639	0.593
Fisher		0.743	0.655	0.598
N-T		0.808	0.872	0.888
Simple	0.99	0.785	0.762	0.722
Fisher		0.866	0.789	0.731
N-T		0.885	0.945	0.962

In the case of the normal distribution, the Fisher's z-transformation is superior to the other methods, but the normalizing transformation (N-T) method works well. The asymptotic variance of the Fisher's z-transformation is $\xi^2/[n(1-\rho^2)^2]$. So if the underlying distribution is normal, $\xi^2 = (1-\rho^2)^2$ and then the asymptotic variance does not depend on unknown parameter. If the underlying distribution is not normal, the asymptotic variance of the Fisher's z-transformation depends on unknown parameters and the approximation (27) does not work. For any underlying distributions, the normalizing transformation works well, and so we recommend the normalizing transformation method for constructing confidence intervals of the correlation coefficient.

Acknowledgements

The author would like to thank the referees for their valuable comments which improve the manuscript significantly.

References

[1] Fujioka, Y and Maesono, Y. (2000). *Higher order normalizing transformations for asymptotic U-statistics,* Jour. Statist. Plan. & Inf., 83, 47-74.

[2] Fujisawa, H. (2000). *Variance stabilizing transformation and studentization for estimator of correlation coefficient,* Statist. Probab. Lett., 47, 213-217.

[3] Hall, P. (1992). *On the removal of skewness by transformation,* Jour. Royal. Statist. Soc. B, 54, 221-228.

[4] Helmers, R. (1991). *On the Edgeworth expansion and the bootstrap approximation for a studentized U-statistic,* Ann. Statist., 19, 470-484.

[5] Lai, T.L. and Wang, J.Q. (1993). *Edgeworth expansion for symmetric statistics with applications to bootstrap methods*, Statistica Sinica, 3, 517-542.

[6] Maesono, Y. (1998). *Asymptotic comparisons of several variance estimators and their effects for studentizations*, Ann. Inst. Stat. Math., 50, 451-470.

[7] Maesono, Y. (1999). *On the normal approximations of a studentized symmetric statistic*, Mathematica Japonica, 50, 433-449.

[8] Maesono, Y. (2005). *Asymptotic representations of ratio statistics and their mean squared errors*, Jour. of the Japan Statist. Soci., 35, 73-97.

[9] Malevich, T.L. and Abdalimov B. (1979). *Large deviation probabilities for U-statistics,* Theory of Probab. Appl., 24, 215-219.