Asymptotic representation of ratio statistics and its application

Yoshihiko MAESONO
Faculty of Economics
Kyushu University
Hakozaki, Fukuoka 812-8581, Japan

Abstract

Some statistics in common use take a form of a ratio of two statistics such as correlation coefficient, Pearson's coefficient of variation, cumulant estimators and so on. In this paper, obtaining an asymptotic representation of the ratio statistic until the third order term, we will discuss asymptotic mean squared errors of the ratio statistics and establish an Edgeworth expansion with remainder term $o(n^{-1})$.

AMS 1991 subject classifications. Primary 62E20; Secondary 60F05.

Key words and Phrases. asymptotic mean squared error, asymptotic U-statistics, Edgeworth expansion, correlation coefficient, H-decomposition.

1. Introduction

Let $X_1, \cdots, X_n$ be independently and identically distributed random vectors with distribution function F. Let $T_n = T_n(X_1, \cdots, X_n)$ and $S_n = S_n(X_1, \cdots, X_n)$ be statistics related to parameters t_n and s_n. 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. In this paper we will obtain an asymptotic representation of the ratio statistic with remainder term $n^{-1/2}R_{n,p}$ where

$$P\{|R_{n,p}| \geq n^{-1}(\log n)^{-1}\} = o(n^{-1}).$$

Using this asymptotic representation, an asymptotic mean squared error with remainder term $o(n^{-2})$ and an Edgeworth expansion with remainder term $o(n^{-1})$ are established. Applying the results to the sample correlation coefficient, we discuss asymptotic mean squared errors and Edgeworth expansions.

Let us assume that

$$\begin{aligned} T_n &= t_n + n^{-1}\delta_T + n^{-2}\sum_{i=1}^{n}\tau_0(X_i) + n^{-1}\sum_{i=1}^{n}\tau_1(X_i) + n^{-2}\sum_{C_{n,2}}\tau_2(X_i, X_j) \\ &\quad + n^{-3}\sum_{C_{n,3}}\tau_3(X_i, X_j, X_k) + n^{-1/2}R_{n,p} \qquad (1) \end{aligned}$$

and

$$\begin{aligned} S_n &= s_n + n^{-1}\delta_S + n^{-2}\sum_{i=1}^{n}\zeta_0(X_i) + n^{-1}\sum_{i=1}^{n}\zeta_1(X_i) + n^{-2}\sum_{C_{n,2}}\zeta_2(X_i, X_j) \\ &\quad + n^{-3}\sum_{C_{n,3}}\zeta_3(X_i, X_j, X_k) + n^{-1/2}R_{n,p} \qquad (2) \end{aligned}$$

where

$$E[\tau_0(X_1)] = E[\zeta_0(X_1)] = E[\tau_1(X_1)] = E[\zeta_1(X_1)] = 0, \qquad (3)$$

$$E[\tau_2(X_1, X_2)|X_1] = E[\zeta_2(X_1, X_2)|X_1] = 0 \ \ a.s., \qquad (4)$$

$$E[\tau_3(X_1, X_2, X_3)|X_1, X_2] = E[\zeta_3(X_1, X_2, X_3)|X_1, X_2] = 0 \ \ a.s. \qquad (5)$$

and δ_T and δ_S are constants. $\sum_{C_{n,k}}$ indicates that the summation is taken over all integers $i_1, \cdots, i_k$ satisfying $1 \leq i_1 < i_2 < \cdots < i_r \leq n$. Many statistics satisfy these assumptions, and Lai and Wang (1993) called them asymptotic U-statistics. Since we will consider the parameters depending on n, such as the variance and the central third moment of U-statistics etc., we assume that

$$t_n = t^{(0)} + n^{-1}t^{(1)} + O(n^{-2}) \tag{6}$$

and

$$s_n = s^{(0)} + n^{-1}s^{(1)} + O(n^{-2}) \quad (s^{(0)} \neq 0). \tag{7}$$

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. Knott and Frangos (1983) calculated an asymptotic variance $nVar(r_n)$ assuming the underlying distribution is bivariate normal. Here applying the asymptotic representation to the correlation coefficient, we will obtain general form of the asymptotic representation without assuming the normality. We also discuss the bias correction of the coefficient and its asymptotic mean squared error.

In Section 2, we will obtain the asymptotic representation of T_n/S_n with remainder term $n^{-1/2}R_{n,p}$ and we will discuss the asymptotic mean squared errors. Applying the Edgeworth expansion for the asymptotic U-statistics, which is obtained by Lai and Wang (1993), the Edgeworth expansion of T_n/S_n with remainder term $o(n^{-1})$ will be established. In Section 3, we will consider the application to the correlation coefficient and also discuss the correction of the bias.

2. Asymptotic representation, mean squared error and Edgeworth expansion

Using H-decomposition and the moment evaluation of them, we will obtain the asymptotic representation of T_n/S_n. Let us assume the following moment conditions

$$E[|\tau_0(X_1)|^3 + |\zeta_0(X_1)|^3] < \infty, \tag{8}$$

$$E[|\tau_1(X_1)|^{4+\varepsilon} + |\tau_2(X_1,X_2)|^{4+\varepsilon} + |\tau_3(X_1,X_2,X_3)|^{4+\varepsilon}] < \infty \tag{9}$$

and

$$E[|\zeta_1(X_1)|^{4+\varepsilon} + |\zeta_2(X_1,X_2)|^{4+\varepsilon} + |\zeta_3(X_1,X_2,X_3)|^{4+\varepsilon}] < \infty \tag{10}$$

for some $\varepsilon > 0$.

Let us define

$$
\begin{aligned}
\delta &= \frac{\delta_T}{s^{(0)}} - \frac{t^{(0)}\delta_S}{(s^{(0)})^2} - \frac{E[\tau_1(X_1)\zeta_1(X_1)]}{(s^{(0)})^2} + \frac{t^{(0)}E[\zeta_1^2(X_1)]}{(s^{(0)})^3}, \\
\eta_0(x) &= \frac{\tau_0(x)}{s^{(0)}} - \frac{t^{(0)}\zeta_0(x)}{(s^{(0)})^2} + \{\frac{E[\zeta_1^2(X_1)]}{(s^{(0)})^3} - \frac{\delta_S + s^{(1)}}{(s^{(0)})^2}\}\tau_1(x) \\
&\quad +\{\frac{2t^{(0)}s^{(1)} + 2t^{(0)}\delta_S + 2E[\tau_1(X_1)\zeta_1(X_1)]}{(s^{(0)})^3} \\
&\qquad -\frac{t^{(1)} + \delta_T}{(s^{(0)})^2} - \frac{3t^{(0)}E[\zeta_1^2(X_1)]}{(s^{(0)})^4}\}\zeta_1(x) \\
&\quad -\frac{E[\tau_1(X_2)\zeta_2(x, X_2) + \zeta_1(X_2)\tau_2(x, X_2)]}{(s^{(0)})^2} \\
&\quad -\frac{\tau_1(x)\zeta_1(x) - E[\tau_1(X_1)\zeta_1(X_1)]}{(s^{(0)})^2} \\
&\quad +\frac{t^{(0)}\{\zeta_1^2(x) - E[\zeta_1^2(X_1)] + 2E[\zeta_1(X_2)\zeta_2(x, X_2)]\}}{(s^{(0)})^3}, \\
\eta_1(x) &= \frac{\tau_1(x)}{s^{(0)}} - \frac{t^{(0)}\zeta_1(x)}{(s^{(0)})^2}, \\
\eta_2(x, y) &= \frac{\tau_2(x, y)}{s^{(0)}} - \frac{\tau_1(x)\zeta_1(y) + \tau_1(y)\zeta_1(x)}{(s^{(0)})^2} - \frac{t^{(0)}\zeta_2(x, y)}{(s^{(0)})^2} \\
&\quad +\frac{2t^{(0)}\zeta_1(x)\zeta_1(y)}{(s^{(0)})^3}, \\
\eta_3(x, y, z) &= \frac{\tau_3(x, y, z)}{s^{(0)}} - \frac{t^{(0)}\zeta_3(x, y, z)}{(s^{(0)})^2} \\
&\quad -\frac{\zeta_1(x)\tau_2(y, z) + \zeta_1(y)\tau_2(x, z) + \zeta_1(z)\tau_2(x, y)}{(s^{(0)})^2} \\
&\quad -\frac{\tau_1(x)\zeta_2(y, z) + \tau_1(y)\zeta_2(x, z) + \tau_1(z)\zeta_2(x, y)}{(s^{(0)})^2} \\
&\quad +\frac{2t^{(0)}\{\zeta_1(x)\zeta_2(y, z) + \zeta_1(y)\zeta_2(x, z) + \zeta_1(z)\zeta_2(x, y)\}}{(s^{(0)})^2}
\end{aligned}
$$

$$+\frac{2\{\tau_1(x)\zeta_1(y)\zeta_1(z)+\tau_1(y)\zeta_1(x)\zeta_1(z)+\tau_1(z)\zeta_1(x)\zeta_1(y)\}}{(s^{(0)})^3}$$
$$-\frac{6t^{(0)}\zeta_1(x)\zeta_1(y)\zeta_1(z)}{(s^{(0)})^4}$$

and

$$\begin{aligned} U_n &= \frac{t_n}{s_n}+n^{-1}\delta+n^{-2}\sum_{i=1}^{n}\eta_0(X_i)+n^{-1}\sum_{i=1}^{n}\eta_1(X_i) \\ &+ n^{-2}\sum_{C_{n,2}}\eta_2(X_i,X_j)+n^{-3}\sum_{C_{n,3}}\eta_3(X_i,X_j,X_k). \end{aligned}$$

Then we have the following representation.

[Theorem 1]. *Assume that the conditions* (1) $\sim$ (10) *are satisfied. Then we have*

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

Proof. See Appendix.

U_n is an approximation of the ratio statistic T_n/S_n until the third order term and we can study the asymptotic properties using U_n. At first we will consider the asymptotic mean squared error of U_n. It follows from the conditions (3), (4) and (5) that

$$E[\eta_0(X_1)]=E[\eta_0(X_1)]=0, \quad E[\eta_2(X_1,X_2)|X_1]=0 \ \ a.s.$$

and

$$E[\eta_3(X_1,X_2,X_3)|X_1,X_2]=0 \ \ a.s.$$

Thus we can obtain the asymptotic mean squared error $AMSE(T_n/S_n)$ as follows.

[Theorem 2]. *Under the same assumptions of Theorem 1, we have*

$$AMSE(\frac{T_n}{S_n})=E[U_n-\frac{t_n}{s_n}]^2=n^{-1}E[\eta_1^2(X_1)] \tag{11}$$
$$+n^{-2}\{\delta^2+2E[\eta_0(X_1)\eta_1(X_1)]+\frac{1}{2}E[\eta_2^2(X_1,X_2)]\}+O(n^{-3}).$$

Let us define

$$\begin{aligned}
&e_1 = E[\tau_1^2(X_1)], \quad e_2 = E[\zeta_1^2(X_1)], \quad e_3 = E[\tau_0(X_1)\tau_1(X_1)],\\
&e_4 = E[\tau_0(X_1)\zeta_1(X_1)], \quad e_5 = E[\tau_1(X_1)\zeta_0(X_1)], \quad e_6 = E[\zeta_0(X_1)\zeta_1(X_1)],\\
&e_7 = E[\tau_1(X_1)\zeta_1(X_1)], \quad e_8 = E[\zeta_1^3(X_1)], \quad e_9 = E[\tau_1(X_1)\zeta_1^2(X_1)],\\
&e_{10} = E[\tau_1^2(X_1)\zeta_1(X_1)], \quad e_{11} = E[\zeta_1(X_1)\zeta_1(X_2)\zeta_2(X_1, X_2)],\\
&e_{12} = E[\tau_1(X_1)\zeta_1(X_2)\zeta_2(X_1, X_2)], \quad e_{13} = E[\tau_1(X_1)\tau_1(X_2)\zeta_2(X_1, X_2)],\\
&e_{14} = E[\zeta_1(X_1)\zeta_1(X_2)\tau_2(X_1, X_2)], \quad e_{15} = E[\tau_1(X_1)\zeta_1(X_2)\tau_2(X_1, X_2)],\\
&e_{16} = E[\tau_2^2(X_1, X_2)], \quad e_{17} = E[\zeta_2^2(X_1, X_2)]
\end{aligned}$$

and

$$e_{18} = E[\tau_2(X_1, X_2)\zeta_2(X_1, X_2)].$$

Then from the direct computations, we can obtain an explicit form as follows.

$$E[\eta_1^2(X_1)] = \frac{1}{(s^{(0)})^2}e_1 - \frac{2t^{(0)}}{(s^{(0)})^3}e_7 + \frac{(t^{(0)})^2}{(s^{(0)})^4}e_2 \tag{12}$$

and

$$\begin{aligned}
&\delta^2 + 2E[\eta_0(X_1)\eta_1(X_1)] + \frac{1}{2}E[\eta_2^2(X_1, X_2)]\\
= \quad & \frac{\delta_T^2}{(s^{(0)})^2} + \frac{(t^{(0)})^2\delta_S^2}{(s^{(0)})^4} - \frac{2t^{(0)}\delta_T\delta_S}{(s^{(0)})^3} - \frac{2(s^{(1)} + \delta_S)}{(s^{(0)})^3}e_1\\
&+2\{\frac{t^{(0)}(t^{(1)} + 2\delta_T)}{(s^{(0)})^4} - \frac{(t^{(0)})^2(2s^{(1)} + 3\delta_S)}{(s^{(0)})^5}\}e_2 + \frac{1}{2(s^{(0)})^2}(4e_3 + e_{16})\\
&-\frac{t^{(0)}}{(s^{(0)})^3}(2e_4 + 2e_5 + e_{18}) + \frac{(t^{(0)})^2}{2(s^{(0)})^4}(4e_6 + e_{17})\\
&+2\{\frac{t^{(0)}(3s^{(1)} + 4\delta_S)}{(s^{(0)})^4} - \frac{t^{(1)} + 2\delta_T}{(s^{(0)})^3}\}e_7 - \frac{2(t^{(0)})^2}{(s^{(0)})^5}(e_8 + 3e_{11})\\
&+\frac{4t^{(0)}}{(s^{(0)})^4}(e_9 + 2e_{12} + e_{14}) - \frac{2}{(s^{(0)})^3}(e_{10} + e_{13} + 2e_{15})\\
&+\frac{3}{(s^{(0)})^4}(e_1e_2 + 2e_7^2) + \frac{9(t^{(0)})^2}{(s^{(0)})^6}e_2^2 - \frac{18t^{(0)}}{(s^{(0)})^5}e_2e_7.
\end{aligned} \tag{13}$$

Since the ratio of two statistics is an asymptotic U-statistic, using the Edgeworth expansion for U-statistics, we can obtain the Edgeworth expansion with remainder term $o(n^{-1})$. Let us assume the following conditions.

(C_1) $E\{|\eta_0(X_1)|^3 + |\eta_1(X_1)|^4 + |\eta_3(X_1, X_2, X_3)|^4\} < \infty$ and $E[\eta_1^2(X_1)] = u^2 > 0$

(C_2) $\limsup_{|t|\to\infty} |E[\exp\{it\eta_1(X_1)\}]| < 1$

$(C_3) E|\eta_2(X_1, X_2)|^s < \infty$ for some $s > 0$ and there exist K Borel functions q_ν:$\mathbf{R} \to \mathbf{R}$ such that $K(s-2) > 4s + (28s - 40)I_{\{E|\eta_3(X_1,X_2,X_3)|>0\}}$, $Eq_\nu^2(X_1) < \infty(\nu = 1, \cdots, K)$, and the covariance matrix of $(W_1, \cdots, W_K)$ is positive definite, where $W_\nu = (Lq_\nu)(X_1)$ and $(Lq_\nu)(y) = E[\eta_2(y, X_2)q_\nu(X_2)]$, and $I_{\{.\}}$ is an indicator function.

The condition C_3 is concerned with the number of nonzero eigen functions of $\eta_2(x, y)$. Alternatively Lai and Wang (1993) proved the validity of the Edgeworth expansion under the following condition $(\tilde{C}_3)$.

$(\tilde{C}_3)$ There exist constants c_ν and Borel functions w_ν:$\mathbf{R} \to \mathbf{R}$ such that $E[w_\nu(X_1)] = 0, E|w_\nu(X_1)|^s < \infty$ for some $s \geq 5$ and $\eta_2(X_1, X_2) = \sum_{\nu=1}^K c_\nu w_\nu(X_1)w_\nu(X_2)$ *a.s.*; moreover, for some $0 < \gamma < \min\{1, 2(1 - 11/(3s))\}$,

$$\limsup_{|t|\to\infty} \sup_{|c_1|+\cdots+|c_K|\leq|t|^{-\gamma}} |E[\exp(it\{g_1(X_1) + \sum_{\nu=1}^{K} c_\nu w_\nu(X_1)\})]| < 1.$$

Let us define

$$\begin{aligned}
\kappa_3 &= u^{-3}\{E[\eta_1^3(X_1)] + 3E[\eta_1(X_1)\eta_1(X_2)\eta_2(X_1, X_2)]\}, \\
\kappa_4 &= u^{-4}\{E[\eta_1^4(X_1)] - 3u^4 + 4E[\eta_1(X_1)\eta_1(X_2)\eta_1(X_3)\eta_3(X_1, X_2, X_3)] \\
&\quad +12E[\eta_1^2(X_1)\eta_1(X_2)\eta_2(X_1, X_2)] \\
&\quad +12E[\eta_1(X_1)\eta_1(X_2)\eta_2(X_1, X_3)\eta_2(X_2, X_3)]\}, \\
P_1(x) &= \frac{\kappa_3(x^2 - 1)}{6}
\end{aligned}$$

and

$$\begin{aligned}
P_2(x) &= u^{-2}\{E[\eta_0(X_1)\eta_1(X_1)] + \frac{E[\eta_2^2(X_1, X_2)]}{4}\}x + \frac{\kappa_4}{24}(x^3 - 3x) \\
&\quad +\frac{\kappa_3^2}{72}(x^5 - 10x^3 + 15x).
\end{aligned}$$

Then it follows from Lai and Wang (1993) that

$$P\{\frac{\sqrt{n}U_n}{u} - \frac{\delta}{\sqrt{n}u} \leq x\} = \Phi(x) - n^{-1/2}\phi(x)P_1(x) - n^{-1}\phi(x)P_2(x) + o(n^{-1}).$$

Thus expanding with respect to $\delta/\sqrt{n}u$, from the standard argument (see Petrov (1975) p.16), we have the Edgeworth expansion of $\sqrt{n}(T_n/S_n - t_n/s_n)/u$ as follows.

[Theorem 3]. *Assume that the conditions* (1)$\sim$ (10), C_1 *and* C_2 *hold. If either condition* C_3 *or* $\tilde{C}_3$ *is satisfied, we have*

$$\sup_x |P\{\frac{\sqrt{n}}{u}(\frac{T_n}{S_n} - \frac{t_n}{s_n}) \leq x\} - Q_n(x)| = o(n^{-1}).$$

where

$$Q_n(x) = \Phi(x) - n^{-1/2}\phi(x)\{P_1(x) + \frac{\delta}{u}\} - n^{-1}\phi(x)\{P_2(x) + \frac{\delta\kappa_3(x^3 - 3x)}{6u}\}.$$

[Remark]. For the Edgeworth expansion with remainder term $o(n^{-1/2})$, we do not need the condition C_3 nor $\tilde{C}_3$.

3. Correlation coefficient

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

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

Let us consider the estimation of the correlation coefficient ρ. 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 $\bar{Y} = \sum Y_i/n$ and $\bar{Z} = \sum Z_i/n$. Then $r_n = T_n/S_n$ is a sample correlation coefficient. Assuming that the underlying distribution is bivariate normal, Knott and Frangos (1983) calculated an asymptotic variance $nVar(r_n)$. Here applying the Theorem 1 and 2, we will obtain an asymptotic representation of r_n. Since we discuss the theoretical properties of the correlation coefficient, without loss of generality, we assume that $E[Y_1] = E[Z_1] = 0$. Let us define

$$\begin{aligned}
\tau_1(\mathbf{x}_1) &= y_1 z_1 - \rho\sigma_y\sigma_z, \qquad\qquad \tau_2(\mathbf{x}_1, \mathbf{x}_2) = -(y_1 z_2 + y_2 z_1), \\
\delta_S &= -\frac{E[(\sigma_z^2 Y_1^2 - \sigma_y^2 Z_1^2)^2]}{8\sigma_y^3\sigma_z^3}, \\
\zeta_0(\mathbf{x}_1) &= -\frac{(\sigma_z^4 y_1^4 + \sigma_y^4 z_1^4 - 2\sigma_y^2\sigma_z^2 y_1^2 z_1^2)}{8\sigma_y^3\sigma_z^3} \\
&+ \frac{E[3\sigma_z^4 Y_2^4 - \sigma_y^4 Z_2^4 - 2\sigma_y^2\sigma_z^2 Y_2^2 Z_2^2]}{16\sigma_y^5\sigma_z^3} y_1^2 \\
&+ \frac{E[3\sigma_y^4 Z_2^4 - \sigma_z^4 Y_2^4 - 2\sigma_y^2\sigma_z^2 Y_2^2 Z_2^2]}{16\sigma_y^3\sigma_z^5} z_1^2 \\
&+ \frac{E[\sigma_z^4 Y_2^3 - \sigma_y^2\sigma_z^2 Y_2 Z_2^2] y_1 + E[\sigma_y^4 Z_2^3 - \sigma_y^2\sigma_z^2 Y_2^2 Z_2] z_1}{2\sigma_y^3\sigma_z^3}, \\
\zeta_1(\mathbf{x}_1) &= \frac{\sigma_z^2 y_1^2 + \sigma_y^2 z_1^2 - 2\sigma_y^2\sigma_z^2}{2\sigma_y\sigma_z}, \\
\zeta_2(\mathbf{x}_1, \mathbf{x}_2) &= -\frac{(\sigma_z^2 y_1^2 - \sigma_y^2 z_1^2)(\sigma_z^2 y_2^2 - \sigma_y^2 z_2^2)}{4\sigma_y^3\sigma_z^3} - \frac{\sigma_z y_1 y_2}{\sigma_y} - \frac{\sigma_y z_1 z_2}{\sigma_z}
\end{aligned}$$

and

$$\begin{aligned}
&\zeta_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \{(\sigma_z^2 y_1^2 + \sigma_y^2 z_1^2 - 2\sigma_y^2\sigma_z^2)(\sigma_z^2 y_2^2 - \sigma_y^2 z_2^2)(\sigma_z^2 y_3^2 - \sigma_y^2 z_3^2) \\
&\quad + (\sigma_z^2 y_2^2 + \sigma_y^2 z_2^2 - 2\sigma_y^2\sigma_z^2)(\sigma_z^2 y_1^2 - \sigma_y^2 z_1^2)(\sigma_z^2 y_3^2 - \sigma_y^2 z_3^2) \\
&\quad + (\sigma_z^2 y_3^2 + \sigma_y^2 z_3^2 - 2\sigma_y^2\sigma_z^2)(\sigma_z^2 y_1^2 - \sigma_y^2 z_1^2)(\sigma_z^2 y_2^2 - \sigma_y^2 z_2^2)\}/\{8\sigma_y^5\sigma_z^5\} \\
&+ \{(\sigma_z^2 y_1^2 - \sigma_y^2 z_1^2)(\sigma_z^2 y_2 y_3 - \sigma_y^2 z_2 z_3) \\
&\quad + (\sigma_z^2 y_2^2 - \sigma_y^2 z_2^2)(\sigma_z^2 y_1 y_3 - \sigma_y^2 z_1 z_3) \\
&\quad + (\sigma_z^2 y_3^2 - \sigma_y^2 z_3^2)(\sigma_z^2 y_1 y_2 - \sigma_y^2 z_1 z_2)\}/\{2\sigma_y^3\sigma_z^3\}.
\end{aligned}$$

Then we have following representations.

[Lemma 1]. *If* $E[|Y_1|^{4+\varepsilon} + |Z_1|^{4+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$*, we have*

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

and

$$\begin{aligned} S_n &= \sigma_y\sigma_z + n^{-1}\delta_S + n^{-2}\sum_{i=1}^{n}\zeta_0(\mathbf{X}_i) + n^{-1}\sum_{i=1}^{n}\zeta_1(\mathbf{X}_i) \\ &+ n^{-2}\sum_{C_{n,2}}\zeta_2(\mathbf{X}_i,\mathbf{X}_j) + n^{-3}\sum_{C_{n,3}}\zeta_3(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_k) + n^{-1/2}R_{n,p}. \end{aligned}$$

Proof. See Appendix.

Thus from Theorem 1, we have the asymptotic representation of the correlation coefficient r_n and we can obtain the asymptotic mean squared error of r_n. Here we consider the case of the bivariate normal distribution

$$\mathbf{X}_i = \begin{pmatrix} Y_i \\ Z_i \end{pmatrix} \quad \sim \quad N(\boldsymbol{\theta}, \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_z \\ \rho\sigma_y\sigma_z & \sigma_z^2 \end{pmatrix}).$$

We can get

$$\zeta_0(\mathbf{x}_1) = -\frac{(\sigma_z^2 y_1^2 - \sigma_y^2 z_1^2)^2}{8\sigma_y^3\sigma_z^3} + \frac{(1-\rho^2)(\sigma_z^2 y_1^2 + \sigma_y^2 z_1^2)}{4\sigma_y^3\sigma_z^3}$$

and

$$\zeta_1(\mathbf{x}_1) = \frac{\sigma_z^2 y_1^2 + \sigma_y^2 z_1^2 - 2\sigma_y^2\sigma_z^2}{2\sigma_y\sigma_z}.$$

Then from direct computations, we have

$$\begin{aligned} &e_1 = e_2 = \sigma_y^2\sigma_z^2(1+\rho^2), \quad e_3 = e_4 = e_5 = 0, \quad e_6 = -\sigma_y^2\sigma_z^2\frac{(1-\rho^2)^2}{2}, \\ &e_7 = \sigma_y^2\sigma_z^2 2\rho, \quad e_8 = \sigma_y^3\sigma_z^3(2+6\rho^2), \quad e_9 = \sigma_y^3\sigma_z^3(6\rho+2\rho^3), \\ &e_{10} = \sigma_y^3\sigma_z^3(2+6\rho^2), \quad e_{11} = e_{12} = e_{13} = e_{14} = e_{15} = 0, \\ &e_{16} = \sigma_y^2\sigma_z^2(2+2\rho^2), \quad e_{17} = \sigma_y^2\sigma_z^2(3+\rho^4) \text{ and } e_{18} = \sigma_y^2\sigma_z^2 4\rho. \end{aligned}$$

Substituting these values, we have the asymptotic mean squared error

$$AMSE(r_n) = n^{-1}(1-\rho^2)^2 + n^{-2}(1-\rho^2)^2(1+\frac{23}{4}\rho^2).$$

Since the bias $\delta = \rho(\rho^2-1)/2$, we have the asymptotic variance of r_n

$$Var(r_n) \simeq AMSE(r_n) - n^{-2}\delta^2 = n^{-1}(1-\rho^2)^2 + n^{-2}(1-\rho^2)^2(1+\frac{11}{2}\rho^2).$$

This coincides with the result of Knott and Frangos (1983, p.502), who have obtained it backed by the computer program package.

Further using Theorem 1 and Theorem 2, we can discuss the bias correction and its asymptotic mean squared error. It is easy to see that

$$\begin{aligned}\delta &= \frac{\rho}{8}\{\frac{3\mu_{40}}{\sigma_y^4}+\frac{3\mu_{04}}{\sigma_z^4}+\frac{2\mu_{22}}{\sigma_y^2\sigma_z^2}\} \\ &- \frac{1}{2}\{\frac{\mu_{31}}{\sigma_y^3\sigma_z}+\frac{\mu_{13}}{\sigma_y\sigma_z^3}\}\end{aligned}$$

where $\mu_{40} = E[Y_1^4], \mu_{04} = E[Z_1^4], \mu_{22} = E[Y_1^2 Z_1^2], \mu_{31} = E[Y_1^3 Z_1]$ and $\mu_{13} = E[Y_1 Z_1^3]$. Using the moment method, we can obtain an estimator of $\mu_{k\ell}$ as follows

$$\hat{\mu}_{k\ell} = n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^k (Z_i - \bar{Z})^\ell.$$

Thus one estimator of the bias δ is given by

$$\begin{aligned}\hat{\delta} &= \frac{\hat{\rho}}{8}\{\frac{3\hat{\mu}_{40}}{\hat{\sigma}_y^4}+\frac{3\hat{\mu}_{04}}{\hat{\sigma}_z^4}+\frac{2\hat{\mu}_{22}}{\hat{\sigma}_y^2\hat{\sigma}_z^2}\} \\ &- \frac{1}{2}\{\frac{\hat{\mu}_{31}}{\hat{\sigma}_y{}^3\hat{\sigma}_z}+\frac{\hat{\mu}_{13}}{\hat{\sigma}_y\hat{\sigma}_z^3}\}\end{aligned}$$

where $\hat{\sigma}_y^2 = (n-1)^{-1}\sum(Y_i - \bar{Y})^2$ and $\hat{\sigma}_z^2 = (n-1)^{-1}\sum(Z_i - \bar{Z})^2$. Thus we can obtain the bias corrected estimator

$$r_n^* = r_n - \hat{\delta}/n.$$

Since each term of $\hat{\delta}$ is a ratio statistic again, using Theorem 1, we can obtain an asymptotic representation of $\hat{\delta}/n$ as

$$n^{-1}\hat{\delta} = n^{-1}\delta + n^{-2}\sum_{i=1}^{n}\varphi(\mathbf{X}_i) + n^{-1/2}R_{n,p}$$

where

$$\begin{aligned}\varphi(\mathbf{x}_1) &= \frac{\rho}{8}\{\frac{3(y_1^4-\mu_{40})}{\sigma_y^4}+\frac{3(z_1^4-\mu_{04})}{\sigma_z^4}+\frac{2(y_1^2z_1^2-\mu_{22})}{\sigma_y^2\sigma_z^2}\} \\ &+\{\frac{3\mu_{31}}{4\sigma_y^5\sigma_z}+\frac{\mu_{13}}{4\sigma_y^3\sigma_z^3}-\frac{6\mu_{40}}{\sigma_y^6}-\frac{2\mu_{22}}{\sigma_y^4\sigma_z^2}\}(y_1^2-\sigma_y^2)\end{aligned}$$

$$+\{\frac{\mu_{31}}{4\sigma_y^3\sigma_z^3}+\frac{3\mu_{13}}{4\sigma_y\sigma_z^5}-\frac{6\mu_{04}}{\sigma_z^6}-\frac{2\mu_{22}}{\sigma_y^2\sigma_z^4}\}(z_1^2-\sigma_z^2)$$
$$+\delta\eta_1(\mathbf{x}_1)-\frac{y_1^3z_1-\mu_{31}}{2\sigma_y^3\sigma_z}-\frac{y_1z_1^3-\mu_{13}}{2\sigma_y\sigma_z^3}.$$

¿From direct computation we have

$$2E[\varphi(\mathbf{X}_1)\eta_1(\mathbf{X}_1)]=(1-\rho^2)^2(3\rho^2-1).$$

Thus we have the asymptotic mean squared error of r_n^*

$$\begin{aligned}AMSE(r_n^*)&=AMSE(r_n)-n^{-2}\{\delta^2+2E[\varphi(\mathbf{X}_1)\eta_1(\mathbf{X}_1)]\}\\&=n^{-1}(1-\rho^2)^2+n^{-2}(1-\rho^2)^2(2+\frac{5}{2}\rho^2).\end{aligned}$$

Since $AMSE(r_n)-AMSE(r_n^*)=(1-\rho^2)^2(13\rho^2-4)/4$, if the underlying distribution is normal and $|\rho|\geq 2/\sqrt{13}(=0.555)$, r_n^* is superior than r_n from viewpoint of unbiasedness and mean squared error. It is possible to correct the bias by another methods like jackknife correction. But those corrected estimators will coincide with r_n^* until the remainder term $n^{-1/2}R_{n,p}$. Thus the asymptotic mean squared error takes the same one of r_n^*.

Further it is possible to improve the confidence interval of ρ using the Cornish-Fisher expansion based on Theorem 3.

Appendix

First we review the H-decomposition or $ANOVA$-decomposition which is a basic tool of the studies of the analysis of variance, the jackknife inference, etc. Let $\nu(x_1,\cdots,x_r)$ be a function which is symmetric in its arguments and $E[\nu(X_1,\cdots,X_r)]=0$. Let us define

$$\begin{aligned}\lambda_1(x_1)&=E[\nu(x_1,X_2,\cdots,X_r)],\\\lambda_2(x_1,x_2)&=E[\nu(x_1,x_2,\cdots,X_r)]-\lambda_1(x_1)-\lambda_1(x_2),\cdots,\end{aligned}$$

and

$$\lambda_r(x_1,x_2,\cdots,x_r)=\nu(x_1,x_2,\cdots,x_r)-\sum_{j=1}^{r-1}\sum_{C_{r,j}}\lambda_j(x_{i_1},x_{i_2},\cdots,x_{i_j}).$$

Then we can show that

$$E[\lambda_k(X_1, \cdots, X_k)|X_1, \cdots, X_{k-1}] = 0 \ \ a.s. \tag{14}$$

and

$$\sum_{C_{n,r}} \nu(X_{i_1}, \cdots, X_{i_r}) = \sum_{k=1}^{r} \binom{n-k}{r-k} \Lambda_k$$

where

$$\Lambda_k = \sum_{C_{n,k}} \lambda_k(X_{i_1}, \cdots, X_{i_k}).$$

Using the equation (14) and moment evaluations of martingales (von Bahr and Esséen (1965), and Dharmadhikari, Fabian and Jogdeo (1968)), we have the upper bounds of the absolute moments of Λ_k as follows.

[Lemma A1].
(1) *For* $1 \leq q \leq 2$, *if* $E|\nu(X_1, \cdots, X_r)|^q < \infty$, *there exists a positive constant* c, *which may depend on* ν *and* F *but not on* n, *such that*

$$E|\Lambda_k|^q = O(n^k). \tag{15}$$

(2) *For* $q \geq 2$, *if* $E|\nu(X_1, \cdots, X_r)|^q < \infty$, *we have*

$$E|\Lambda_k|^q = O(n^{(qk)/2}). \tag{16}$$

Hereafter in order to obtain evaluations of moments, we use the H-decomposition and the inequalities (15) and (16). It follows from Markov's inequality that if

$$E|R|^\beta = O(n^{-1-(3\beta)/2-\varepsilon}) \ \ for\ some \ \ \beta \geq 1 \ and \ \varepsilon > 0,$$

we have $R = n^{-1/2}R_{n,p}$. It is trivial that $cn^{-2} = n^{-1/2}R_{n,p}$ for constant c. Then from (15) and (16), we can obtain that if $E|\nu(X_1, \cdots, X_r)|^{2+\varepsilon} < \infty$ for $\varepsilon > 0$,

$$n^{-r-2} \sum_{C_{n,r}} \nu(X_{i_1}, \cdots, X_{i_r}) = n^{-1/2}R_{n,p}, \tag{17}$$

$$n^{-r-1} \sum_{C_{n,r}} \nu(X_{i_1}, \cdots, X_{i_r}) = n^{-2}\Lambda_1 + n^{-1/2}R_{n,p}, \tag{18}$$

$$n^{-r} \sum_{k=4}^{r} \binom{n-k}{r-k} \Lambda_k = n^{-1/2}R_{n,p} \tag{19}$$

and

$$n^{-r+1}\sum_{k=6}^{r}\binom{n-k}{r-k}\Lambda_k = n^{-1/2}R_{n,p}. \tag{20}$$

Since we need an approximation of the product of two statistics, we prepare the following lemma. Hereafter for the purpose of the simplicity, we use abbreviations $\tau_1(i), \tau_2(i,j), \cdots$ which represent $\tau_1(X_i), \tau_2(X_i, X_j), \cdots$.

[Lemma A2]. *Assume that the conditions* (1)$\sim$(10) *are satisfied. Then we have*

$$\begin{aligned} n^{-2}\sum_{i=1}^{n}\tau_1(i)\sum_{i=1}^{n}\zeta_1(i) &= n^{-1}E[\tau_1(X)\zeta_1(X)] + n^{-2}\sum_{i=1}^{n}\{\tau_1(i)\zeta_1(i) \\ &\quad -E[\tau_1(X)\zeta_1(X)]\} + n^{-2}\sum_{C_{n,2}}\{\tau_1(i)\zeta_1(j)+\tau_1(j)\zeta_1(i)\}, \end{aligned} \tag{21}$$

$$\begin{aligned} & n^{-3}\sum_{i=1}^{n}\tau_1(i)\sum_{C_{n,2}}\zeta_2(i,j) \\ = \; & n^{-2}\sum_{i=1}^{n}E[\tau_1(X)\zeta_2(X_i,X)|X_i] \\ & +n^{-3}\sum_{C_{n,3}}\{\tau_1(i)\zeta_2(j,k)+\tau_1(j)\zeta_2(i,k)+\tau_1(k)\zeta_2(i,j)\} \\ & +n^{-1/2}R_{n,p}, \end{aligned} \tag{22}$$

$$n^{-4}\sum_{C_{n,2}}\tau_2(i,j)\sum_{C_{n,2}}\zeta_2(i,j) = n^{-1/2}R_{n,p}, \tag{23}$$

$$n^{-4}\sum_{i=1}^{n}\tau_1(i)\sum_{C_{n,3}}\zeta_3(i,j,k) = n^{-1/2}R_{n,p}, \tag{24}$$

$$n^{-5}\sum_{C_{n,2}}\tau_2(i,j)\sum_{C_{n,3}}\zeta_3(i,j,k) = n^{-1/2}R_{n,p}, \tag{25}$$

$$n^{-6}\sum_{C_{n,3}}\tau_3(i,j,k)\sum_{C_{n,3}}\zeta_1(i,j,k) = n^{-1/2}R_{n,p} \tag{26}$$

and

$$\begin{aligned} n^{-3}\{\sum_{i=1}^{n}\zeta_1(i)\}^3 &= n^{-2}\sum_{i=1}^{n}3E[\zeta_1^2(X)]\zeta_1(i) \\ &+ n^{-3}\sum_{C_{n,3}}6\zeta_1(i)\zeta_1(j)\zeta_1(k) + n^{-1/2}R_{n,p} \end{aligned} \tag{27}$$

where X is an independent copy of X_i.
Proof. Here we will prove the equations (23) and (27). It follows from (16) that

$$\begin{aligned} & E|n^{-4} \sum_{C_{n,2}} \tau_2(i,j) \sum_{C_{n,2}} \zeta_2(i,j)|^{2+\varepsilon/2} \\ \leq \ & n^{-8-2\varepsilon} \{E| \sum_{C_{n,2}} \tau_2(i,j)|^{4+\varepsilon} E| \sum_{C_{n,2}} \zeta_2(i,j)|^{4+\varepsilon}\}^{1/2} \\ = \ & O(n^{-4-\varepsilon}) = O(n^{-1-3(2+\varepsilon/2)/2-\varepsilon/4}). \end{aligned}$$

Thus we get the equation (23). From direct computation, we have

$$\begin{aligned} & n^{-3}\{\sum_{i=1}^{n} \zeta_1(i)\}^3 = n^{-3} \sum_{i=1}^{n} \zeta_1^3(i) \\ & + n^{-3} \sum_{C_{n,2}} 3\{\zeta_1^2(i)\zeta_1(j) + \zeta_1^2(j)\zeta_1(i)\} + n^{-3} \sum_{C_{n,3}} 6\zeta_1(i)\zeta_1(j)\zeta(k). \end{aligned}$$

¿From the equation (15) we get

$$E|n^{-3} \sum_{i=1}^{n} \{\zeta_1^3(i) - E[\zeta_1^3(X)]\}|^{(4+\varepsilon)/3} = O(n^{-3-\varepsilon}) = O(n^{-1-(3/2)\dot{(}(4+\varepsilon)/3)-\varepsilon/2}).$$

Since $n^{-2}E[\zeta_1^3(X)] = n^{-1/2}R_{n,p}$, we have $n^{-3}\sum_{i=1}^{n}\zeta_1^3(i) = n^{-1/2}R_{n,p}$. Applying the H-decomposition to the second term, it follows from (16) that

$$n^{-3} \sum_{C_{n,2}} 3[\zeta_1^2(i)\zeta_1(j) + \zeta_1^2(j)\zeta_1(i)] = n^{-2} \sum_{i=1}^{n} 3E[\zeta_1^2(X)]\zeta_1(i) + n^{-1/2}R_{n,p}.$$

Thus we have the equation (27). We can similarly prove the others.

Using these Lemmas, we will obtain the asymptotic representations.

Proof of Theorem 1.

Using Taylor expansion of $(c+x)^{-1}$, we have

$$\begin{aligned} S_n^{-1} &= s_n^{-1} - s_n^{-2}(S_n - s_n) + s_n^{-3}(S_n - s_n)^2 \\ &- s_n^{-4}(S_n - s_n)^3 + (s_n + \vartheta)^{-5}(S_n - s_n)^4 \end{aligned}$$

where $0 \leq |\vartheta| \leq |S_n - s_n|$. It is easy to see that

$$-s_n^{-2}(S_n - s_n) = -n^{-1}\frac{\delta_S}{(s^{(0)})^2} - n^{-2}\sum_{i=1}^{n}\{\frac{\zeta_0(i)}{(s^{(0)})^2} - \frac{2s^{(1)}\zeta_1(i)}{(s^{(0)})^3}\}$$
$$-n^{-1}\sum_{i=1}^{n}\frac{\zeta_1(i)}{(s^{(0)})^2} - n^{-2}\sum_{C_{n,2}}\frac{\zeta_2(i,j)}{(s^{(0)})^2} - n^{-3}\sum_{C_{n,3}}\frac{\zeta_3(i,j,k)}{(s^{(0)})^2} + n^{-1/2}R_{n,p}.$$

It follows from Lemma A2 that under the moment conditions (8)∼(10),

$$s_n^{-3}(S_n - s_n)^2 = n^{-1}\frac{E[\zeta_1^2(X)]}{(s^{(0)})^3} + n^{-2}\sum_{C_{n,2}}\frac{2\zeta_1(i)\zeta_1(j)}{(s^{(0)})^3}$$
$$+n^{-2}\sum_{i=1}^{n}\frac{\zeta_1^2(i) - E[\zeta_1^2(X)] + 2\delta_S\zeta_1(i) + 2E[\zeta_1(X)\zeta_2(X_i,X)|X_i]}{(s^{(0)})^3}$$
$$+n^{-3}\sum_{C_{n,3}}\frac{2\{\zeta_1(i)\zeta_2(j,k) + \zeta_1(j)\zeta_2(i,k) + \zeta_1(k)\zeta_2(i,j)\}}{(s^{(0)})^3} + n^{-1/2}R_{n,p}.$$

Using this representation, it follows from Lemma A2 that

$$\begin{aligned} &-s_n^{-4}(S_n - s_n)^3 = -s_n^{-1}(S_n - s_n)s_n^{-3}(S_n - s_n)^2 \\ = \quad &-n^{-2}\sum_{i=1}^{n}\frac{3E[\zeta_1^2(X)]\zeta_1(i)}{(s^{(0)})^4} - n^{-3}\sum_{C_{n,3}}\frac{6\zeta_1(i)\zeta_1(j)\zeta_1(k)}{(s^{(0)})^4} + n^{-1/2}R_{n,p}. \end{aligned}$$

Again using this equation and Lemma A2, we have

$$(S_n - s_n)^4 = n^{-1/2}R_{n,p}.$$

Using Markov's inequality, we can show that

$$P\{|\vartheta| > \frac{|s^{(0)}|}{2}\} = o(n^{-1}) \quad or \quad P\{|s_n + \vartheta| \leq \frac{|s^{(0)}|}{2}\} = o(n^{-1}).$$

Since

$$\begin{aligned} &P\{|n^{1/2}(s_n + \vartheta)^{-5}(S_n - s_n)^4| \geq n^{-1}(\log n)^{-1}\} \\ \leq \quad &P\{|n^{1/2}\frac{32}{(s^{(0)})^5}(S_n - s_n)^4| \geq n^{-1}(\log n)^{-1}\} + P\{|s_n + \vartheta| \leq \frac{|s^{(0)}|}{2}\} \\ = \quad &o(n^{-1}), \end{aligned}$$

we have $(s_n+\vartheta)^{-5}(S_n-s_n)^4 = n^{-1/2}R_{n,p}$. Thus we obtain

$$\begin{aligned} S_n^{-1} &= s_n^{-1} + n^{-1}\delta_\nu + n^{-2}\sum_{i=1}^{n}\nu_0(i) + n^{-1}\sum_{i=1}^{n}\nu_1(i) + n^{-2}\sum_{C_{n,2}}\nu_2(i,j) \\ &+ n^{-3}\sum_{C_{n,3}}\nu_3(i,j,k) + n^{-1/2}R_{n,p} \end{aligned}$$

where

$$\begin{aligned} \delta_\nu &= \frac{E[\zeta_1^2(X)]}{(s^{(0)})^3} - \frac{\delta_S}{(s^{(0)})^2}, \\ \nu_0(x) &= -\frac{\zeta_0(x)}{(s^{(0)})^2} + \{\frac{2s^{(1)}+2\delta_S}{(s^{(0)})^3} - \frac{3E[\zeta_1^2(X)]}{(s^{0)})^4}\}\zeta_1(x) \\ &+ \frac{\zeta_1^2(x) - E[\zeta_1^2(X)] + 2E[\zeta_1(X)\zeta_2(x,X)]}{(s^{(0)})^3}, \\ \nu_1(x) &= -\frac{\zeta_1(x)}{(s^{(0)})^2}, \\ \nu_2(x,y) &= -\frac{\zeta_2(x,y)}{(s^{(0)})^2} + \frac{2\zeta_1(x)\zeta_1(y)}{(s^{(0)})^3} \end{aligned}$$

and

$$\begin{aligned} \nu_3(x,y,z) &= -\frac{\zeta_3(x,y,z)}{(s^{(0)})^2} - \frac{6\zeta_1(x)\zeta_1(y)\zeta_1(z)}{(s^{(0)})^4} \\ &+ \frac{2\{\zeta_1(x)\zeta_2(y,z) + \zeta_1(y)\zeta_2(x,z) + \zeta_1(z)\zeta_2(x,y)\}}{(s^{(0)})^3}. \end{aligned}$$

Similarly, using Lemma A2, we can show that

$$\begin{aligned} &T_n s_n^{-1} = \frac{t_n}{s_n} + n^{-1}\frac{\delta_T}{s^{(0)}} + n^{-2}\sum_{i=1}^{n}\{\frac{\tau_0(i)}{s^{(0)}} - \frac{s^{(1)}\tau_1(i)}{(s^{(0)})^2}\} + n^{-1}\sum_{i=1}^{n}\frac{\tau_1(i)}{s^{(0)}} \\ &+n^{-2}\sum_{C_{n,2}}\frac{\tau_2(i,j)}{s^{(0)}} + n^{-3}\sum_{C_{n,3}}\frac{\tau_3(i,j,k)}{s^{(0)}} + n^{-1/2}R_{n,p}, \\ &T_n n^{-1}\delta_\nu = n^{-1}t^{(0)}\delta_\nu + n^{-2}\sum_{i=1}^{n}\delta_\nu\tau_1(i) + n^{-1/2}R_{n,p}, \end{aligned}$$

$$
\begin{aligned}
&T_n n^{-2}\sum_{i=1}^{n}\nu_0(i) = n^{-2}\sum_{i=1}^{n} t^{(0)}\nu_0(i) + n^{-1/2}R_{n,p},\\
&T_n n^{-1}\sum_{i=1}^{n}\nu_1(i) = n^{-1}E[\tau_1(X)\nu_1(X)] + n^{-2}\sum_{i=1}^{n}\{(t^{(1)}+\delta_T)\nu_1(i)\\
&\qquad +E[\nu_1(X)\tau_2(X_i,X)|X_i] + \tau_1(i)\nu_1(i) - E[\tau_1(X)\nu_1(X)]\}\\
&+n^{-1}\sum_{i=1}^{n} t^{(0)}\nu_1(i) + n^{-2}\sum_{C_{n,2}}\{\tau_1(i)\nu_1(j)+\tau_1(j)\nu_1(i)\}\\
&+n^{-3}\sum_{C_{n,3}}\{\nu_1(i)\tau_2(j,k)+\nu_1(j)\tau_2(i,k)+\nu_1(k)\tau_2(i,j)\} + n^{-1/2}R_{n,p},\\
&T_n n^{-2}\sum_{C_{n,2}}\nu_2(i,j) = n^{-2}\sum_{i=1}^{n}E[\tau_1(X)\nu_2(X_i,X)|X_i] + n^{-2}\sum_{C_{n,2}} t^{(0)}\nu_2(i,j)\\
&+n^{-3}\sum_{C_{n,3}}\{\tau_1(i)\nu_2(j,k)+\tau_1(j)\nu_2(i,k)+\tau_1(k)\nu_2(i,j)\} + n^{-1/2}R_{n,p}
\end{aligned}
$$

and

$$
T_n n^{-3}\sum_{C_{n,3}}\nu_3(i,j,k) = n^{-3}\sum_{C_{n,3}} t^{(0)}\nu_3(i,j,k) + n^{-1/2}R_{n,p}.
$$

Combining the above equations, we have the desired result.

Proof of Lemma 1.

Applying H-decomposition, we have

$$
\begin{aligned}
(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 &= \sigma_y^2 + n^{-1}\sum_{i=1}^{n}\{Y_i^2-\sigma_y^2\} + n^{-2}\sum_{C_{n,2}}\{-2Y_iY_j\}\\
&\quad +n^{-1/2}R_{n,p},\\
(n-1)^{-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2 &= \sigma_z^2 + n^{-1}\sum_{i=1}^{n}\{Z_i^2-\sigma_z^2\} + n^{-2}\sum_{C_{n,2}}\{-2Z_iZ_j\}\\
&\quad +n^{-1/2}R_{n,p}
\end{aligned}
$$

and

$$
\begin{aligned}
(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})(Z_i-\bar{Z}) &= \rho\sigma_y\sigma_z + n^{-1}\sum_{i=1}^{n}\{Y_iZ_i-\rho\sigma_y\sigma_z\}\\
&+ \; n^{-2}\sum_{C_{n,2}}\{-Y_iZ_j-Y_jZ_i\} + n^{-1/2}R_{n,p}.
\end{aligned}
$$

Using Lemma A2, we can show that

$$\begin{aligned}
&(n-1)^{-2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2\sum_{i=1}^{n}(Z_i-\bar{Z})^2 \\
= \; &\sigma_y^2\sigma_z^2 + n^{-1}\delta_\psi + n^{-2}\sum_{i=1}^{n}\psi_0(\mathbf{X}_i) + n^{-1}\sum_{i=1}^{n}\psi_1(\mathbf{X}_i) + n^{-2}\sum_{C_{n,2}}\psi_2(\mathbf{X}_i,\mathbf{X}_j) \\
&+n^{-3}\sum_{C_{n,3}}\psi_3(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_k) + n^{-1/2}R_{n,p}
\end{aligned}$$

where

$$\begin{aligned}
&\delta_\psi = E[Y_1^2Z_1^2] - \sigma_y^2\sigma_z^2, \\
&\psi_0(\mathbf{x}_1) = (y_1^2-\sigma_y^2)(z_1^2-\sigma_z^2) - \delta - E[Y_2^2Z_2]z_1 - E[y_2Z_2^2]y_1, \\
&\psi_1(\mathbf{x}_1) = \sigma_z^2(y_1^2-\sigma_y^2) + \sigma_y^2(z_1^2-\sigma_z^2), \\
&\psi_2(\mathbf{x}_1,\mathbf{x}_2) = (y_1^2-\sigma_y^2)(z_2^2-\sigma_z^2) + (y_2^2-\sigma_y^2)(z_1^2-\sigma_z^2) - \sigma_z^2y_1y_2 - \sigma_y^2z_1z_2
\end{aligned}$$

and

$$\begin{aligned}
\psi_3(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3) \;=\; &-\{(y_1^2-\sigma_y^2)z_2z_3 + (y_2^2-\sigma_y^2)z_1z_3 + (y_3^2-\sigma_y^2)z_1z_2 \\
&+(z_1^2-\sigma_z^2)y_2y_3 + (z_2^2-\sigma_z^2)y_1y_3 + (z_3^2-\sigma_z^2)y_1y_2\}.
\end{aligned}$$

Using Taylor expansion of $\sqrt{x}$, we can show that

$$\begin{aligned}
&S_n = \sigma_y\sigma_z + n^{-1}\{\frac{\delta_\psi}{2\sigma_y\sigma_z} - \frac{E[\psi_1^2(\mathbf{X})]}{8\sigma_y^3\sigma_z^3}\} \\
+ \; &n^{-2}\sum_{i=1}^{n}\{\frac{\psi_0(i)}{2\sigma_y\sigma_z} - \frac{\psi_1^2(i) - E[\psi_1^2(\mathbf{X})] + 2\delta_\psi\psi_1(i) + 2E[\psi_1(\mathbf{X})\psi_2(i,\mathbf{X})|\mathbf{X}_i]}{8\sigma_y^3\sigma_z^3} \\
&+\frac{3E[\psi_1^2(\mathbf{X})]\psi_1(i)}{16\sigma_y^5\sigma_z^5}\} \\
+ \; &n^{-1}\sum_{i=1}^{n}\frac{\psi_1(i)}{2\sigma_y\sigma_z} + n^{-2}\sum_{C_{n,2}}\{\frac{\psi_2(i,j)}{2\sigma_y\sigma_z} - \frac{\psi_1(i)\psi_1(j)}{4\sigma_y^3\sigma_z^3}\} \\
+ \; &n^{-3}\sum_{C_{n,3}}\{\frac{\psi_3(i,j,k)}{2\sigma_y\sigma_z} - \frac{\psi_1(i)\psi_2(j,k) + \psi_1(j)\psi_2(i,k) + \psi_1(k)\psi_2(i,j)}{4\sigma_y^3\sigma_z^3} \\
&+\frac{3\psi_1(i)\psi_1(j)\psi_1(k)}{8\sigma_y^5\sigma_z^5}\} + n^{-1/2}R_{n,p}.
\end{aligned}$$

where $\mathbf{X}$ is an independent copy of $\mathbf{X}_i$. Thus we have the desired result.

References

[1] Dharmadhikari, S.W., Fabian, V. and Jogdeo, K.(1968). *Bounds on the moments of martingales,* Ann. Math. Statist., 39, 1719-1723.

[2] Knott, M. and Frangos, C.C. (1983). *Variance estimation for the jackknife using von Mises expansions*, Biometrika, 70, 501-504.

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

[4] Petrov, V.V. (1975). *Sums of Independent Random Variables*, Springer Berlin.

[5] von Bahr, B and Esséen C.G. (1965). *Inequalities for the rth absolute moment of a sum of random variables,* $1 \leq r \leq 2$, Ann. Math. Statist., 36, 299-303.