J. Japan Statist. Soc.
Vol. 28 No. 1 1998 45-57

# A NEW HIGHER ORDER APPROXIMATION TO A PERCENTAGE POINT OF THE DISTRIBUTION OF THE SAMPLE CORRELATION COEFFICIENT

Masafumi Akahira* and Norio Torigoe**

A new higher order approximation formula for a percentage point of the distribution of the sample correlation coefficient is given up to the order $O(n^{-1})$, using the Cornish-Fisher expansion for the statistic based on a linear combination of a normal random variable and chi-random variables. The numerical comparison of the formula with others shows that it dominates the others and gives almost precise values in various cases even for the size $n=10$ of sample.

*Key Words and Phrases*: percentage point, (sample) correlation coefficient, Fisher's $\boldsymbol{Z}$-transformation, Cornish-Fisher expansion.

## 1. Introduction

Percentage points of the distribution of the sample correlation coefficient play an important role in the inference of the correlation coefficient $\rho$ of a bivariate normal distribution. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent and identically distributed random vectors according to the bivariate normal distribution. The density of the sample correlation coefficient $R$ can be obtained, but it is quite difficult to get a percentage point analytically since the density has a complicated form. Hence, it is useful to consider approximation formulae for a percentage point of the distribution of $R$ (see Johnson *et al.* [3], Chapter 32 and Shibata [7]). One of the well-known ways to obtain percentage points of the distribution of $R$ is a normal approximation (see, e.g. Ruben [6]). Indeed, for $0<\alpha<1$, we have $P\{R \leq r_\alpha\} \fallingdotseq 1-\alpha$, where

$$u_\alpha \fallingdotseq \frac{p_\alpha\sqrt{2n-5}-q\sqrt{2n-3}}{\sqrt{p_\alpha^2+q^2+2}}$$

with $p_\alpha := r_\alpha/\sqrt{1-r_\alpha^2}$ and $q := \rho/\sqrt{1-\rho^2}$, which yields the upper $100\alpha$ percentile $r_\alpha$ of the distribution of $R$ with the upper $100\alpha$ percentile $u_\alpha$ of the standard normal distribution. Another way is to use Fisher's transformation, *i.e.* $\boldsymbol{Z}=(1/2)\log((1+R)/(1-R))$. Indeed, since $\boldsymbol{Z}$ is asymptotically normally distributed with mean $\zeta=(1/2)\log((1+\rho)/(1-\rho))$ and variance $1/(n-3)$, the upper $100\alpha$ percentile $r_\alpha$ is asymptotically given by $\zeta+(u_\alpha/\sqrt{n-3})$. The higher order asymptotic expansion for the distribution of the sample correlation coefficient is derived by Niki and Konishi [4] from the Fisher transformation. It is extremely accurate and very complex as is stated in Johnson *et al.* [3].

In this paper, similar to Akahira [1], we derive a new approximation

Received February, 1997. Revised October, 1997. Accepted December, 1997.
* Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan.
** Department of Mathematical Sciences, School of Science, Tokai University, Kanagawa 259-1292, Japan.

formula of the percentage point up to the order $O(n^{-1})$ using the Cornish-Fisher expansion for the statistic based on a linear combination of a normal random variable and chi-random variables. In numerical calculations, the higher order approximation formula dominates the above normal approximation, the approximation by Fisher's $\boldsymbol{Z}$-transformation etc., and gives nearly precise values in various cases of $\alpha$ and $\rho$ even for $n=10$.

## 2. A new higher order approximation formula of a percentage point of the distribution of the sample correlation coefficient

In this section, first we derive a new higher order approximation formula of a percentage point of the distribution of the sample correlation coefficient using the Cornish-Fisher expansion.

Suppose that $(X_1, Y_1), \cdots, (X_n, Y_n)$ are independent and identically distributed random vectors according to a bivariate normal distribution with mean vector $(\mu_1, \mu_2)$ and variances $\sigma_1^2$ and $\sigma_2^2$ and correlation coefficient $\rho$. Without loss of generality we assume that $\mu_1=\mu_2=0$. Then, it is known that the distribution of the sample correlation coefficient

$$R:=\frac{\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2\sum_{i=1}^n(Y_i-\bar{Y})^2}}$$

depends on only $\rho$, where $\bar{X}=(1/n)\sum_{i=1}^n X_i$ and $\bar{Y}=(1/n)\sum_{i=1}^n Y_i$. Letting $Y_i=\alpha X_i+U_i(i=1, \cdots, n)$ with $\alpha=\rho\sigma_2/\sigma_1$, we see that, for each $i=1, \cdots, n$, $X_i$ and $U_i$ are independently and normally distributed with mean 0 and variances $\sigma_1^2$ and $\sigma_2^2(1-\rho^2)$, respectively. Putting

$$T:=\frac{\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^n(X_i-\bar{X})^2},$$

we see that the conditional distribution of $T$, given $X_1, \cdots, X_n$, is normal with mean $\alpha$ and variance $\sigma_2^2(1-\rho^2)/\sum_{i=1}^n(X_i-\bar{X})^2$. Let

$$Z:=\frac{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2}}{\sigma_2\sqrt{1-\rho^2}}(T-\alpha)=\frac{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2}}{\sigma_1\sqrt{1-\rho^2}}\left(\frac{\sigma_1}{\sigma_2}T-\rho\right).$$

It is seen that $Z$ is normally distributed with mean 0 and variance 1. Putting $S_1^2:=\sum_{i=1}^n(X_i-\bar{X})^2/\sigma_1^2$, we see that $S_1^2$ is distributed according to a chi-square distribution with $n-1$ degrees of freedom. Then we have

$$T=\frac{\sigma_2}{\sigma_1}\left(\sqrt{1-\rho^2}\frac{Z}{S_1}+\rho\right),$$

where $S_1:=\sqrt{S_1^2}$. Putting

$$\begin{aligned}S_2^2:&=\frac{1}{\sigma_2^2(1-\rho^2)}\left\{\sum_{i=1}^n(Y_i-\bar{Y})^2-T^2\sum_{i=1}^n(X_i-\bar{X})^2\right\}\\&=\frac{1}{\sigma_2^2(1-\rho^2)}(1-R^2)\sum_{i=1}^n(Y_i-\bar{Y})^2,\end{aligned}$$

we see that $S_2^2$ is independent of $X_1, \cdots, X_n$ and $Z$ and is distributed according to a chi-square distribution with $n-2$ degrees of freedom. We also have

$$R=T\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}=\frac{\sigma_1 TS_1\sqrt{1-R^2}}{\sigma_2\sqrt{1-\rho^2}\,S_2},$$

which yields

$$\frac{R}{\sqrt{1-R^2}}=\frac{\sigma_1 TS_1}{\sigma_2\sqrt{1-\rho^2}\,S_2}=\frac{Z}{S_2}+\frac{\rho}{\sqrt{1-\rho^2}}\frac{S_1}{S_2},$$

where $S_2:=\sqrt{S_2^2}$. Hence we obtain

$$\begin{aligned}P\{R\leq r\}&=P\left\{\frac{R}{\sqrt{1-R^2}}\leq\frac{r}{\sqrt{1-r^2}}\right\}\\&=P\left\{\frac{Z}{S_2}+\frac{\rho}{\sqrt{1-\rho^2}}\frac{S_1}{S_2}\leq\frac{r}{\sqrt{1-r^2}}\right\}.\end{aligned}$$

Putting

$$p:=\frac{r}{\sqrt{1-r^2}},\quad q:=\frac{\rho}{\sqrt{1-\rho^2}},$$

we have

$$P\{R\leq r\}=P\{Z+qS_1-pS_2\leq 0\}. \tag{2.1}$$

The above derivation is stated in Takeuchi [8] to obtain the exact density function of $R$ (see also Johnson *et al.* [3]). It is also noted that $Z$, $S_1^2$ and $S_2^2$ are independent.

In order to get a higher order approximation formula of a percentage point of the distribution of $R$ in a similar way to Akahira [1], we first have

$$\begin{aligned}&E(S_1)=\sqrt{n-1}\,b_{n-1},\quad E(S_2)=\sqrt{n-2}\,b_{n-2},\\&Var(S_1)=(n-1)(1-b_{n-1}^2),\quad Var(S_2)=(n-2)(1-b_{n-2}^2),\end{aligned} \tag{2.2}$$

where

$$b_\nu=\sqrt{\frac{2}{\nu}}\,\Gamma\left(\frac{\nu+1}{2}\right)\Big/\Gamma\left(\frac{\nu}{2}\right)$$

with the Gamma function

$$\Gamma(\nu)=\int_0^\infty x^{\nu-1}e^{-x}dx\quad\text{for }\nu>0,$$

and $Var(S)$ denotes the variance of $S$. Since

$$\begin{aligned}&E(Z+qS_1-pS_2)=q\sqrt{n-1}\,b_{n-1}-p\sqrt{n-2}\,b_{n-2},\\&Var(Z+qS_1-pS_2)=Var(Z)+q^2Var(S_1)+p^2Var(S_2)\\&\qquad=1+q^2(n-1)(1-b_{n-1}^2)+p^2(n-2)(1-b_{n-2}^2),\end{aligned} \tag{2.3}$$

it follows that $Z+qS_1-pS_2$ is standardized as

$$(2.4) \qquad W := \frac{Z+q(S_1-\sqrt{n-1}\,b_{n-1})-p(S_2-\sqrt{n-2}\,b_{n-2})}{\{1+q^2(n-1)(1-b_{n-1}^2)+p^2(n-2)(1-b_{n-2}^2)\}^{1/2}},$$

which implies

$$E(W)=0, \quad Var(W)=1.$$

Note that the statistic $W$ is based on a linear combination of a normal random variable and chi-random variables.

For any $\alpha$ with $0<\alpha<1$, there exists a $r_\alpha$ such that $P\{R\leq r_\alpha\}=1-\alpha$. The $r_\alpha$ is called the upper $100\alpha$ percentile of the distribution of the sample correlation coefficient $R$. Then we have from (2.1) and (2.4)

$$(2.5) \quad 1-\alpha=P\{R\leq r_\alpha\}=P\{Z+qS_1-p_\alpha S_2\leq 0\}$$
$$=P\left\{W_\alpha\leq\frac{-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\}^{1/2}}\right\},$$

where $p_\alpha=r_\alpha/\sqrt{1-r_\alpha^2}$ and $W_\alpha$ denotes $W$ with $p_\alpha$ instead of $p$. In a similar way to Akahira [1], we obtain an approximation formula of the percentage point up to the order $O(n^{-1})$, using the Cornish-Fisher expansion for the statistic $W$ (see also Akahira, *et al.* [2]). In order to do so we need the third and fourth cumulants of $W$.

LEMMA 1. *The third and fourth cumulants of* $Z+qS_1-pS_2$ *are given by*

$$\kappa_3(Z+qS_1-pS_2)$$
$$=q^3(n-1)^{3/2}b_{n-1}\left\{2(b_{n-1}^2-1)+\frac{1}{n-1}\right\}-p^3(n-2)^{3/2}b_{n-2}\left\{2(b_{n-2}^2-1)+\frac{1}{n-2}\right\}$$

*and*

$$\kappa_4(Z+qS_1-pS_2)$$
$$=2q^4[(n-1)\{-1+2(1-b_{n-1}^2)\}+(n-1)^2\{2(1-b_{n-1}^2)-3(1-b_{n-1}^2)^2\}]$$
$$+2p^4[(n-2)\{-1+2(1-b_{n-2}^2)\}+(n-2)^2\{2(1-b_{n-2}^2)-3(1-b_{n-2}^2)^2\}],$$

*respectively, for* $n\geq 3$.

PROOF. Since

$$E(S_1)=\sqrt{n-1}\,b_{n-1}, \quad E(S_2)=\sqrt{n-2}\,b_{n-2}, \quad E(S_1^2)=n-1, \quad E(S_2^2)=n-2,$$
$$E(S_1^3)=(n-1)^{3/2}\left(1+\frac{1}{n-1}\right)b_{n-1}, \quad E(S_2^3)=(n-2)^{3/2}\left(1+\frac{1}{n-2}\right)b_{n-2},$$

it follows that

$$\kappa_3(Z+qS_1-pS_2)$$
$$=E[\{Z+q(S_1-\sqrt{n-1}\,b_{n-1})-p(S_2-\sqrt{n-2}\,b_{n-2})\}^3]$$
$$=q^3\kappa_3(S_1)-p^3\kappa_3(S_2)$$
$$=q^3(n-1)^{3/2}b_{n-1}\left\{2(b_{n-1}^2-1)+\frac{1}{n-1}\right\}$$
$$-p^3(n-2)^{3/2}b_{n-2}\left\{2(b_{n-2}^2-1)+\frac{1}{n-2}\right\}.$$

Since, by Akahira [1], page 599,

$$E[(S_1-\sqrt{n-1}\,b_{n-1})^4]=2(n-1)(1-2b_{n-1}^2)+(n-1)^2(1-b_{n-1}^2)(1+3b_{n-1}^2),$$
$$E[(S_2-\sqrt{n-2}\,b_{n-2})^4]=2(n-2)(1-2b_{n-2}^2)+(n-2)^2(1-b_{n-2}^2)(1+3b_{n-2}^2),$$

it follows from (2.2) that

$$\begin{aligned}(2.6)\quad & E[\{Z+q(S_1-\sqrt{n-1}\,b_{n-1})-p(S_2-\sqrt{n-2}\,b_{n-2})\}^4]\\ &=3+q^4\{2(n-1)(1-2b_{n-1}^2)+(n-1)^2(1-b_{n-1}^2)(1+3b_{n-1}^2)\}\\ &\quad+p^4\{2(n-2)(1-2b_{n-2}^2)+(n-2)^2(1-b_{n-2}^2)(1+3b_{n-2}^2)\}\\ &\quad+6\{q^2(n-1)(1-b_{n-1}^2)+p^2q^2(n-1)(n-2)(1-b_{n-1}^2)(1-b_{n-2}^2)\\ &\quad+p^2(n-2)(1-b_{n-2}^2)\}.\end{aligned}$$

From (2.3) and (2.6) we have

$$\begin{aligned}&\kappa_4(Z+qS_1-pS_2)\\ &=E[\{Z+q(S_1-\sqrt{n-1}\,b_{n-1})-p(S_2-\sqrt{n-2}\,b_{n-2})\}^4]\\ &\quad-3\{Var(Z+q(S_1-\sqrt{n-1}\,b_{n-1})-p(S_2-\sqrt{n-2}\,b_{n-2}))\}^2\\ &=q^4[2(n-1)(1-2b_{n-1}^2)+(n-1)^2(1-b_{n-1}^2)\{4-3(1-b_{n-1}^2)\}\\ &\quad-3(n-1)^2(1-b_{n-1}^2)^2]+p^4[2(n-2)(1-2b_{n-2}^2)\\ &\quad+(n-2)^2(1-b_{n-2}^2)\{4-3(1-b_{n-2}^2)\}-3(n-2)^2(1-b_{n-2}^2)^2].\end{aligned}$$

Thus we complete the proof.

LEMMA 2. *For a sufficiently large $n$*

$$Var(Z+qS_1-pS_2)=1+(p^2+q^2)\left\{\frac{1}{2}-\frac{1}{8n}-\frac{1}{16n^2}+O\left(\frac{1}{n^3}\right)\right\},$$
$$\kappa_3(Z+qS_1-pS_2)=\frac{q^3-p^3}{4\sqrt{n}}\left(1+\frac{1}{4n}\right)+O\left(\frac{1}{n^{5/2}}\right),$$
$$\kappa_4(Z+qS_1-pS_2)=O\left(\frac{1}{n^2}\right).$$

PROOF. Since, by the Stirling formula,

$$\Gamma(\nu)=\sqrt{2\pi}\,\nu^{\nu-(1/2)}e^{-\nu}\left(1+\frac{1}{12\nu}+\frac{1}{288\nu^2}-\frac{139}{51840\nu^3}+O\left(\frac{1}{\nu^4}\right)\right),$$

we obtain

$$(2.7)\qquad b_\nu=\sqrt{\frac{2}{\nu}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}=1-\frac{1}{4\nu}+\frac{1}{32\nu^2}+\frac{1}{128\nu^3}+O\left(\frac{1}{\nu^4}\right),$$

which yields

$$\begin{aligned}(2.8)\qquad 1-b_\nu^2&=(1-b_\nu)(1+b_\nu)=(1-b_\nu)\{2-(1-b_\nu)\}\\ &=\frac{1}{2\nu}-\frac{1}{8\nu^2}-\frac{1}{16\nu^3}+O\left(\frac{1}{\nu^4}\right).\end{aligned}$$

From (2.3), (2.7) and (2.8) we have for a sufficiently large $n$

$$
\begin{aligned}
Var(Z+qS_1-pS_2)&=1+q^2(n-1)(1-b_{n-1}^2)+p^2(n-2)(1-b_{n-2}^2)\\
&=1+(p^2+q^2)\left\{\frac{1}{2}-\frac{1}{8n}-\frac{1}{16n^2}+O\left(\frac{1}{n^3}\right)\right\}.
\end{aligned}
$$

Since, by (2.7) and (2.8),

$$
\nu^{3/2}b_\nu\left\{2(b_\nu^2-1)+\frac{1}{\nu}\right\}=\frac{1}{4\sqrt{\nu}}\left(1+\frac{1}{4\nu}\right)+O\left(\frac{1}{\nu^{5/2}}\right),
$$

it follows from Lemma 1 that

$$
\kappa_3(Z+qS_1-pS_2)=\frac{q^3-p^3}{4\sqrt{n}}\left(1+\frac{1}{4n}\right)+O\left(\frac{1}{n^{5/2}}\right).
$$

From Lemma 1 and (2.8) we have

$$
\begin{aligned}
&\kappa_4(Z+qS_1-pS_2)\\
&=q^4\Bigg[2(n-1)\left\{-1+\frac{1}{n-1}-\frac{1}{4(n-1)^2}-\frac{1}{8(n-1)^3}+O\left(\frac{1}{n^4}\right)\right\}\\
&\quad+(n-1)^2\Bigg\{4\left(\frac{1}{2(n-1)}-\frac{1}{8(n-1)^2}-\frac{1}{16(n-1)^3}+O\left(\frac{1}{n^4}\right)\right)\\
&\quad-6\left(\frac{1}{4(n-1)^2}-\frac{1}{8(n-1)^3}+O\left(\frac{1}{n^4}\right)\right)\Bigg\}\Bigg]\\
&\quad+p^4\Bigg[2(n-2)\left\{-1+\frac{1}{n-2}-\frac{1}{4(n-2)^2}-\frac{1}{8(n-2)^3}+O\left(\frac{1}{n^4}\right)\right\}\\
&\quad+(n-2)^2\Bigg\{4\left(\frac{1}{2(n-2)}-\frac{1}{8(n-2)^2}-\frac{1}{16(n-2)^3}+O\left(\frac{1}{n^4}\right)\right)\\
&\quad-6\left(\frac{1}{4(n-2)^2}-\frac{1}{8(n-2)^3}+O\left(\frac{1}{n^4}\right)\right)\Bigg\}\Bigg]\\
&=O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Thus we complete the proof.

LEMMA 3. *The third and fourth cumulants of $W$ are given by*

$$
\begin{aligned}
&\kappa_3(W)\\
&=\frac{q^3(n-1)^{3/2}b_{n-1}\left\{2(b_{n-1}^2-1)+\frac{1}{n-1}\right\}-p^3(n-2)^{3/2}b_{n-2}\left\{2(b_{n-2}^2-1)+\frac{1}{n-2}\right\}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p^2(n-2)(1-b_{n-2}^2)\}^{3/2}}
\end{aligned}
$$

*for $n\geq 3$ and*

$$
\kappa_4(W)=O\left(\frac{1}{n^2}\right)
$$

*for a sufficiently large $n$.*

The proof is straightforward from (2.4) and Lemmas 1 and 2. Using the Cornish-Fisher expansion, we can obtain a higher order approximation formula of a percentage point of the distribution of the sample correlation coefficient $R$. We denote $W$ with $p_\alpha$ instead of $p$ by $W_\alpha$.

THEOREM. *The upper* $100\alpha$ *percentile* $r_\alpha$ *of the distribution of* $R$ *can be derived from the formula*

$$\frac{-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\}^{1/2}} = u_\alpha+\frac{1}{6}(u_\alpha^2-1)\kappa_3(W_\alpha)+O\left(\frac{1}{n}\right), \tag{2.9}$$

*where* $p_\alpha=r_\alpha/\sqrt{1-r_\alpha^2}$ *and* $u_\alpha$ *is the upper* $100\alpha$ *percentile of the standard normal distribution and* $\kappa_3(W_\alpha)$ *is given in Lemma 3.*

PROOF. From (2.5) and Lemmas 2, 3 we have by the Cornish-Fisher expansion

$$\begin{aligned}
&\frac{-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\}^{1/2}}\\
&\quad=u_\alpha+\frac{1}{6}\kappa_3(W_\alpha)(u_\alpha^2-1)+\frac{1}{24}\kappa_4(W_\alpha)(u_\alpha^3-3u_\alpha)\\
&\qquad-\frac{1}{36}\kappa_3^2(W_\alpha)(2u_\alpha^3-5u_\alpha)+o\left(\frac{1}{n^2}\right)\\
&\quad=u_\alpha+\frac{1}{6}(u_\alpha^2-1)\kappa_3(W_\alpha)+O\left(\frac{1}{n}\right),
\end{aligned}$$

where $W_\alpha$ denotes $W$ with $p_\alpha$ instead of $p$ and $\kappa_3(W)$ is given in Lemma 3. Thus we complete the proof.

The existence and uniqueness of the solution of the equation (2.9) may be guaranteed in a similar way to the discussion [2] on the approximation formula

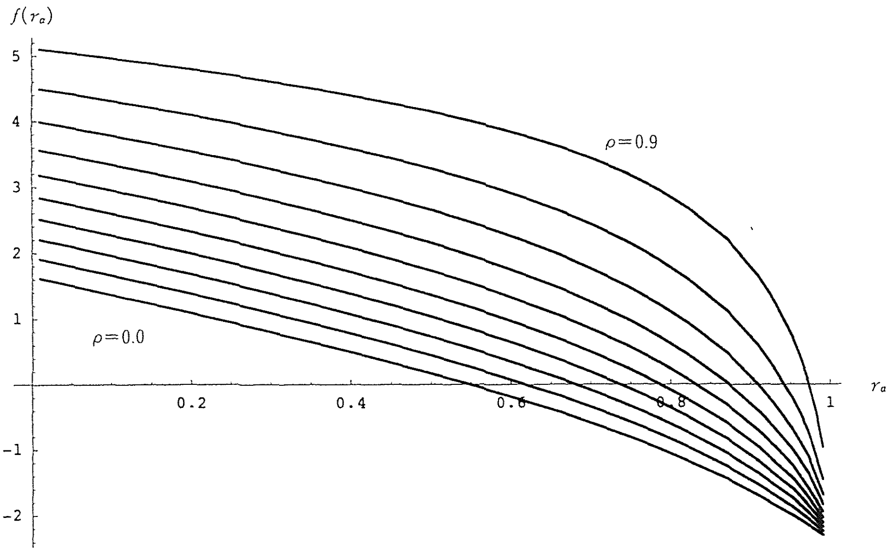


Figure 1. Graph of $f(r_\alpha):=u_\alpha+\dfrac{1}{6}(u_\alpha^2-1)\kappa_3(W_\alpha)-\dfrac{-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\}^{1/2}}$
($n=10$, $u_\alpha=1.64485$ ($\alpha=0.05$), $\rho=0.0$ (0.1) 0.9)

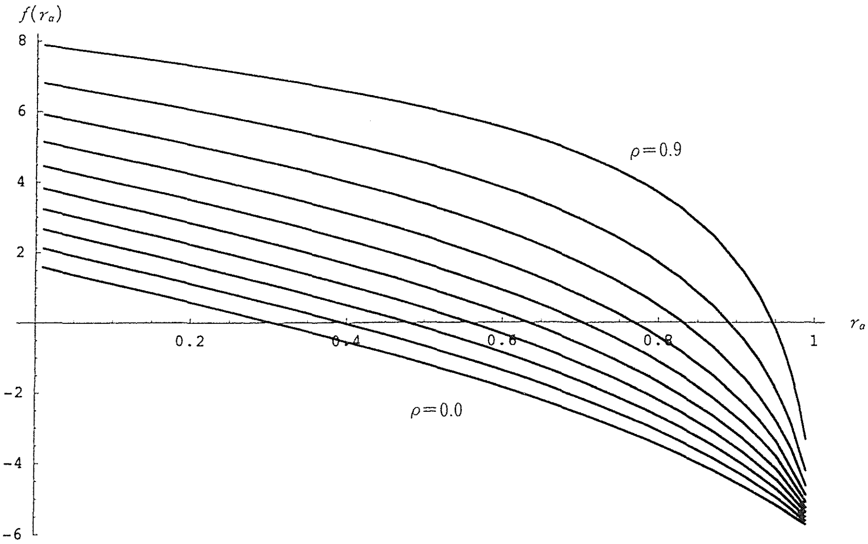


Figure 2. Graph of $f(r_\alpha)$ ($n=30$, $u_\alpha=1.64485$ ($\alpha=0.05$), $\rho=0.0\ (0.1)\ 0.9$)

for a percentage point $t_\alpha$ of the non-central $t$-distribution, because (2.9) is of a linear combination of the standard normal random variable $Z$ and two chi-random variables and the formula for $t_\alpha$ is of a linear combination of $Z$ and a chi-random variable. In this paper we numerically examine the existence and uniqueness of the solution of (2.9) for $\alpha=0.05$ and $n=10, 30$ (see Figures 1 and 2). A similar tendency to the above is also seen in the case when $\alpha=0.01, 0.10$ and $n=10, 20, 30$.

If we ignore the second term of the right-hand side of (2.9), *i.e.*

$$\frac{-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}}{\{1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\}^{1/2}}=u_\alpha+o(1), \tag{2.10}$$

which is called the first order approximation. Since, by (2.7), $\sqrt{\nu}\,b_\nu \fallingdotseq \sqrt{\nu-(1/2)}$, it follows that

$$-q\sqrt{n-1}\,b_{n-1}+p_\alpha\sqrt{n-2}\,b_{n-2}\fallingdotseq p_\alpha\sqrt{n-\frac{5}{2}}-q\sqrt{n-\frac{3}{2}}. \tag{2.11}$$

It also follows from (2.3) and Lemma 2 that

$$\begin{aligned}1+q^2(n-1)(1-b_{n-1}^2)+p_\alpha^2(n-2)(1-b_{n-2}^2)\\ =1+\frac{p_\alpha^2+q^2}{2}+O\left(\frac{1}{n}\right).\end{aligned} \tag{2.12}$$

From (2.10), (2.11) and (2.12) we have

$$\frac{p_\alpha\sqrt{2n-5}-q\sqrt{2n-3}}{\sqrt{p_\alpha^2+q^2+2}}\fallingdotseq u_\alpha,$$

which yields the well-known normal approximation formula

(2.13) $$p_\alpha \doteqdot \frac{q\sqrt{(2n-3)(2n-5)}+u_\alpha\sqrt{(4n-8)q^2+2(2n-5)-u_\alpha^2q^2-2u_\alpha^2}}{2n-5-u_\alpha^2}$$

(see, e.g. Yamauti *et al.* [10]).

Next we consider an approximation to the percentage point by Fisher's $\boldsymbol{Z}$-transformation. Let

$$\boldsymbol{Z}=\frac{1}{2}\log\left(\frac{1+R}{1-R}\right).$$

Then it is known that $\boldsymbol{Z}$ is asymptotically normally distributed with mean $(1/2)\log((1+\rho)/(1-\rho))$ and variance $1/(n-3)$ for a sufficiently large $n$. Hence we have for $0<\alpha<1$

$$1-\alpha=P\{R\leq r_\alpha\}=P\{\boldsymbol{Z}\leq z_\alpha\}\doteqdot\Phi(\sqrt{n-3}(z_\alpha-\zeta)),$$

where

$$z_\alpha=\frac{1}{2}\log\frac{1+r_\alpha}{1-r_\alpha},\quad \zeta=\frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

Then the upper $100\alpha$ percentile $r_\alpha$ of the distribution of $R$ is given by

(2.14) $$r_\alpha\doteqdot\frac{e^{2z_\alpha}-1}{e^{2z_\alpha}+1}.$$

with $z_\alpha=\zeta+(u_\alpha/\sqrt{n-3})$. The Cornish-Fisher expansion for the upper $100\alpha$ percentile of the distribution of $\boldsymbol{Z}$ is given by Winterbottom [9] as follows.

(2.15) $$\begin{aligned} z_\alpha\doteqdot\ &\zeta+\frac{u_\alpha}{\sqrt{n}}+\frac{\rho}{2n}+\frac{1}{12n\sqrt{n}}\{u_\alpha^3+3(3-\rho^2)u_\alpha\} \\ &+\frac{1}{24n^2}\{4\rho^3u_\alpha^2+(15\rho-\rho^3)\}+\frac{1}{480n^2\sqrt{n}} \\ &\times\{u_\alpha^5+(80+30\rho^2-60\rho^4)u_\alpha^3+(375-21\rho^2+45\rho^4)u_\alpha\}. \end{aligned}$$

From (2.14) and (2.15) we can get the upper $100\alpha$ percentile $r_\alpha$ of the distribution of $R$. See Johnson *et al.* [3] for other approximations to the distribution of the sample correlation coefficient.

## 3. Evaluation of the new approximation formula in comparison with others by numerical calculation

In this section we numerically compare the higher order approximation

Table 1. The maximum absolute errors ($\times 10^{-4}$) of the new higher order approximation formula (2.9) for $\rho=.000\ (.100)\ .900,\ .950$

| $n$ \ $\alpha$ | .990 | .975 | .950 | .900 | .750 | .250 | .100 | .050 | .025 | .010 |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 15 | 5 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 6 |
| 20 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 30 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

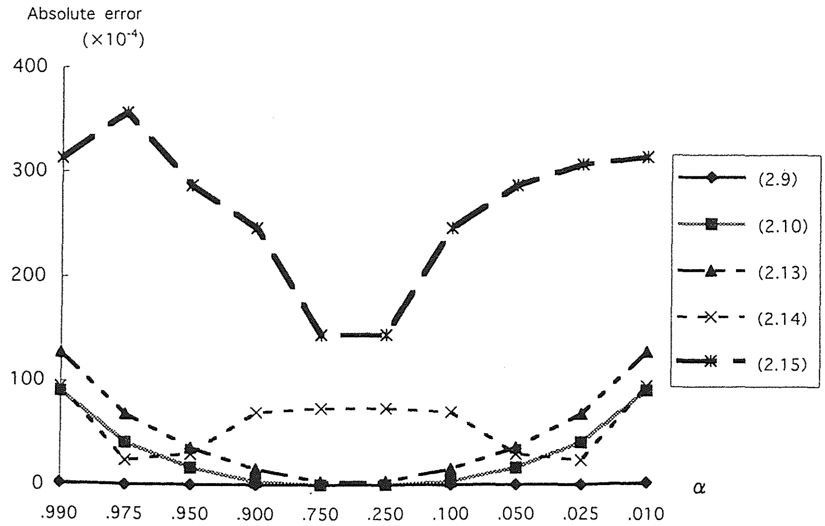


Figure 3. Comparison of the new higher order approximation formula (2.9) with others for $n=10$ and $\rho=0.000$

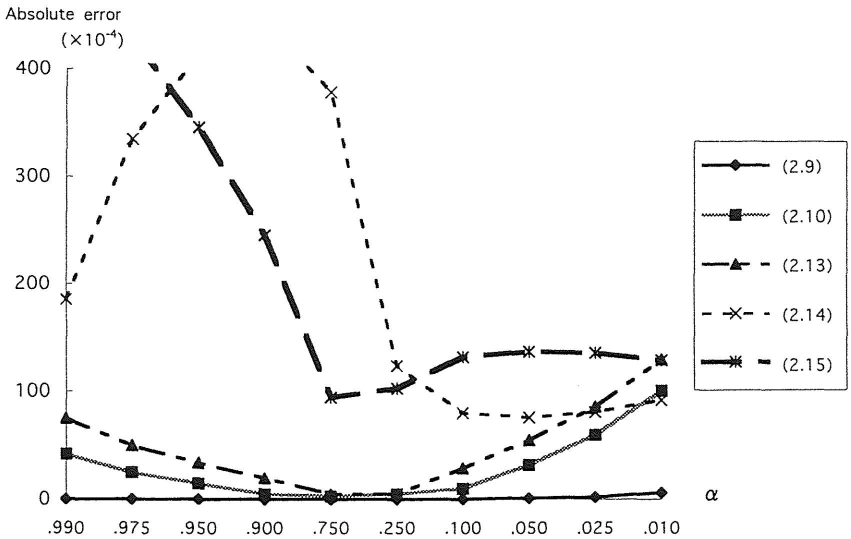


Figure 4. Comparison of the new higher order approximation formula (2.9) with others for $n=10$ and $\rho=0.500$

formula (2.9) with the first order approximation (2.10), the normal approximation (2.13), the approximation (2.14) by Fisher's $\boldsymbol{Z}$-transformation and (2.15) by Winterbottom [9] when $\alpha$=0.990, 0.975, 0.950, 0.900, 0.750, 0.500, 0.250, 0.100, 0.050, 0.025, 0.010, $\rho$=0.000, 0.100, 0.200, 0.300, 0.400, 0.500, 0.600, 0.700, 0.800, 0.900, 0.950, $n$=10, 20, 30. The errors of the approximation formula (2.9) are given in Table 1, where the true values of percentage points of the distribution of the sample correlation coefficient are referred from Odeh [5]. The values of (2.9) and (2.10) are calculated by Newton's method in *Mathematica* for Macintosh. As is seen in Table 1, the approximation formula (2.9) gives

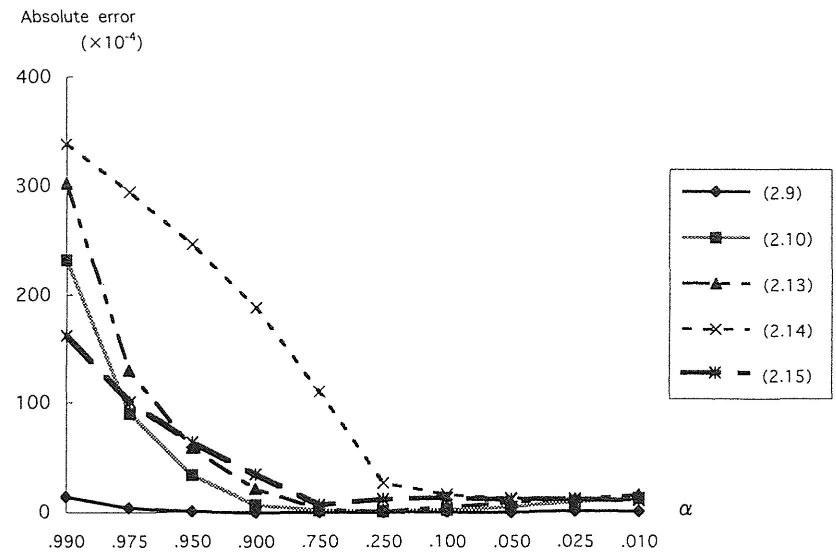


Figure 5. Comparison of the new higher order approximation formula (2.9) with others for $n=10$ and $\rho=0.950$

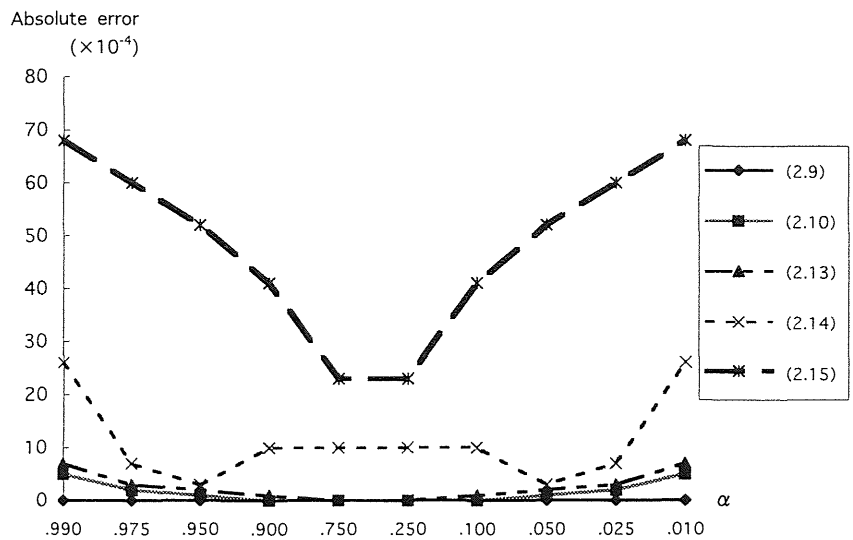


Figure 6. Comparison of the new higher order approximation formula (2.9) with others for $n=30$ and $\rho=0.000$

almost precise values for various cases of $\alpha$ and $\rho$ for not only $n=20$, 30, 50, 100, but also even $n=10$. The approximation formula (2.9) also dominates the others (see Figures 3 to 8). Hence, the formula (2.9) can be recommended as a useful one in the derivation of percentage points of the distribution of the sample correlation coefficient.


## Acknowledgements

The authors thank the referee for suggesting the forms of a table and figures.

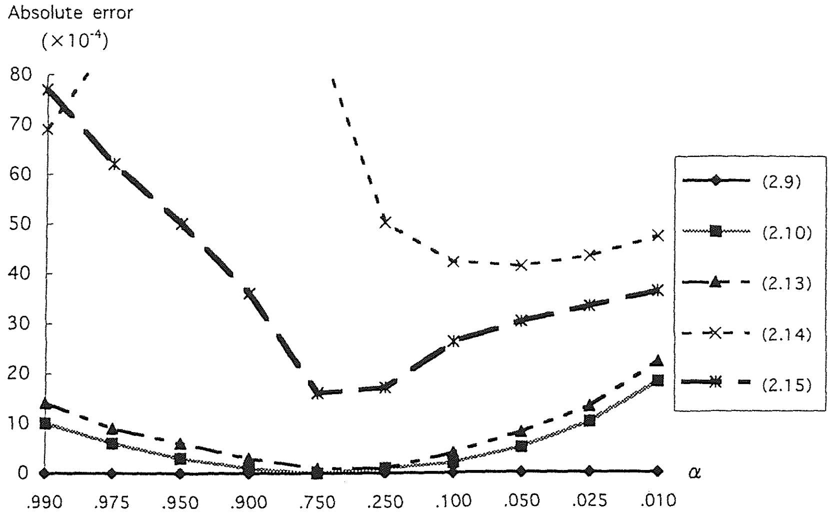


Figure 7. Comparison of the new higher order approximation formula (2.9) with others for $n=30$ and $\rho=0.500$

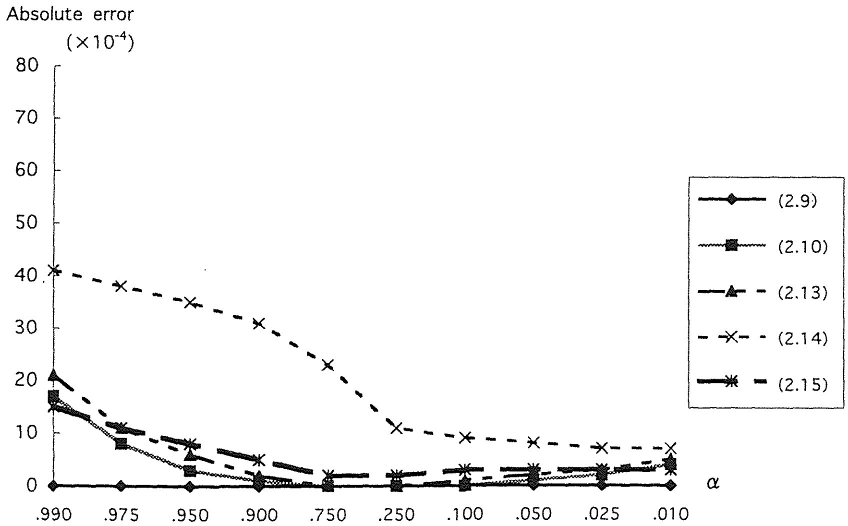


Figure 8. Comparison of the new higher order approximation formula (2.9) with others for $n=30$ and $\rho=0.950$

## REFERENCES


[1] Akahira, M. (1995). A higher order approximation to a percentage point of the noncentral $t$-distribution, *Commun. Statist.-Simula.*, **24** (3), 595-605.

[2] Akahira, M., Sato, M. and Torigoe, N. (1995). On the new approximation to noncentral $t$-distributions, *J. Japan Statist. Soc.*, **25** (1), 1-18.

[3] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2. Second edition,* Wiley, New York.

[4] Niki, N. and Konishi, S. (1984). Higher order asymptotic expansions for the distribution of the sample correlation coefficient, *Commun. Statist.-Simula.*, **13** (2), 169-182.

[5] Odeh, R. E. (1982). Critical values of the sample product-moment correlation coefficient

in the bivariate normal distribution, *Commun. Statist.-Simula.*, **11** , 1-26.
[ 6 ] Ruben, H. (1966). Some new results on the distribution of the sample correlation coefficient, *J. Roy. Statist. Soc., Ser. B*, 28, 513-525.
[ 7 ] Shibata, Y. (1981). *Normal Distributions*, (In Japanese), Tokyo Univ. Press, Tokyo.
[ 8 ] Takeuchi, K. (1975). *Probability Distributions and Statistical Analysis*, (In Japanese), Japanese Standard Association, Tokyo.
[ 9 ] Winterbottom, A. (1980). Estimation for the bivariate normal correlation coefficient using asymptotic expansions, *Commun. Statist.-Simula. Computa.*, **B9** (6), 599-609.
[10] Yamauti, Z. *et al.* (1977). *Concise Statistical Tables*, (In Japanese), Japanese Standard Association, Tokyo.