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

著者別名 | 赤平 昌文
---|---

権利 | 日本統計学会
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 $O(n^{-2})$, 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 $Z$-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 $(X_1, Y_1), \ldots, (X_n, Y_n)$ 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 $P(R \leq r_\alpha) = 1 - \alpha$, where

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

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. $Z = (1/2) \log((1 + R)/(1 - R))$. Indeed, since $Z$ is asymptotically normally distributed with mean $\xi = (1/2) \log((1 + \rho)/(1 - \rho))$ and variance $1/(n-3)$, the upper $100\alpha$ percentile $r_\alpha$ is asymptotically given by $\xi + (u_\alpha / \sqrt{n-3})$. 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
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 \( Z \)-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 \((X_i, Y_i), \cdots, (X_n, Y_n)\) 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 \(Y_i = aX_i + U_i\) (\(i = 1, \cdots, n\)) with \(a = \rho\sigma_2/\sigma_1\), 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 \(a\) and variance \(\sigma_2^2(1-\rho^2)/\sum_{i=1}^{n}(X_i - \bar{X})^2\). Let

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

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

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

where \(S_1 = \sqrt{S_1^2}\). Putting

\[
S_2^2 = \frac{1}{\sigma_2^2(1-\rho^2)} \left\{ \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - T^2 \sum_{i=1}^{n}(X_i - \bar{X})^2 \right\}
\]

\[
= \frac{1}{\sigma_2^2(1-\rho^2)} (1-R^2) \sum_{i=1}^{n}(Y_i - \bar{Y})^2,
\]

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

\[
R = T \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{\sigma_1 T S_1 \sqrt{1 - R^2}}{\sigma_2 \sqrt{1 - \rho^2} S_2},
\]

which yields

\[
\frac{R}{\sqrt{1 - R^2}} = \frac{\sigma_1 T S_1}{\sigma_2 \sqrt{1 - \rho^2} S_2} = \frac{Z}{S_2} + \frac{\rho}{\sqrt{1 - \rho^2}} \frac{S_1}{S_2},
\]

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

\[
P\{R \leq r\} = P\left\{ \frac{R}{\sqrt{1 - R^2}} \leq \frac{r}{\sqrt{1 - r^2}} \right\}
= P\left\{ \frac{Z}{S_2} + \frac{\rho}{\sqrt{1 - \rho^2}} \frac{S_1}{S_2} \leq \frac{r}{\sqrt{1 - r^2}} \right\}.
\]

Putting

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

we have

\[
(2.1) \quad P\{R \leq r\} = P\{Z + q S_1 - p S_2 \leq 0\}.
\]

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

\[
(2.2) \quad E(S_1) = \sqrt{n-1} b_{n-1}, \quad E(S_2) = \sqrt{n-2} b_{n-2},
\]

\[
Var(S_1) = (n-1)(1 - b_{n-1}^2), \quad Var(S_2) = (n-2)(1 - b_{n-2}^2),
\]

where

\[
b_\nu = \sqrt{\frac{2}{\nu}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}
\]

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

\[
(2.3) \quad E(Z + q S_1 - p S_2) = q \sqrt{n-1} b_{n-1} - p \sqrt{n-2} b_{n-2},
\]

\[
Var(Z + q S_1 - p S_2) = Var(Z) + q^2 Var(S_1) + p^2 Var(S_2)
= 1 + q^2 (n-1)(1 - b_{n-1}^2) + p^2 (n-2)(1 - b_{n-2}^2),
\]

it follows that \( Z + q S_1 - p S_2 \) is standardized as
\[ W := \frac{Z + q(S_1 - \sqrt{n-1} b_{n-1}) - p(S_2 - \sqrt{n-2} b_{n-2})}{(1 + q^2(n-1)(1-b_{n-1}^2) + p^2(n-2)(1-b_{n-2}^2))^{1/2}}, \]

which implies

\[ E(W) = 0, \quad \text{Var}(W) = 1. \]

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

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

\[ 1 - a = P(R \leq r_a) = P\left\{ \frac{Z + qS_1 - pS_2}{(1 + q^2(n-1)(1-b_{n-1}^2) + p^2(n-2)(1-b_{n-2}^2))^{1/2}} \leq r_a \right\}, \]

where \( p_a = \frac{r_a}{\sqrt{1-r_a^2}} \) and \( W_a \) denotes \( W \) with \( p_a \) 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

\[ E(S_1) = \sqrt{n-1} b_{n-1}, \quad E(S_2) = \sqrt{n-2} b_{n-2}, \quad E(S_1^2) = n-1, \quad E(S_2^2) = n-2, \]

\[ E(S_i^3) = (n-1)^{3/2} \left( 1 + \frac{1}{n-1} \right) b_{n-1}, \quad E(S_i^4) = (n-2)^{3/2} \left( 1 + \frac{1}{n-2} \right) b_{n-2}, \]

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,
\[ E[(S_1 - \sqrt{n-1} b_{n-1})^4] = 2(n-1)(1 - 2 b_{n-1})^2 + (n-1)^2(1 - b_{n-1}^2)(1 + 3 b_{n-1}^2), \]
\[ E[(S_2 - \sqrt{n-2} b_{n-2})^4] = 2(n-2)(1 - 2 b_{n-2})^2 + (n-2)^2(1 - b_{n-2}^2)(1 + 3 b_{n-2}^2), \]
it follows from (2.2) that
\[ (2.6) \quad E[(Z + q(S_1 - \sqrt{n-1} b_{n-1}) - p(S_2 - \sqrt{n-2} b_{n-2}))^4] = 3 + q^4[2(n-1)(1 - 2 b_{n-1})^2 + (n-1)^2(1 - b_{n-1}^2)(1 + 3 b_{n-1}^2)] + p^4[2(n-2)(1 - 2 b_{n-2})^2 + (n-2)^2(1 - b_{n-2}^2)(1 + 3 b_{n-2}^2)] + 6q^2(n-1)(1 - b_{n-1}^2) + 6p^2(n-2)(1 - b_{n-2}^2).
\]
From (2.3) and (2.6) we have
\[ \kappa_4(Z + qS_1 - pS_2) = E[(Z + q(S_1 - \sqrt{n-1} b_{n-1}) - p(S_2 - \sqrt{n-2} b_{n-2}))^4] - 3\{ \text{Var}(Z + q(S_1 - \sqrt{n-1} b_{n-1}) - p(S_2 - \sqrt{n-2} b_{n-2})) \}^2 = q^4[2(n-1)(1 - 2 b_{n-1})^2 + (n-1)^2(1 - b_{n-1}^2)(1 + 3 b_{n-1}^2)] - 3(n-1)^2(1 - b_{n-1}^2)^2 + p^4[2(n-2)(1 - 2 b_{n-2})^2 + (n-2)^2(1 - b_{n-2}^2)(1 + 3 b_{n-2}^2)] + 6q^2(n-1)(1 - b_{n-1}^2) + 6p^2(n-2)(1 - b_{n-2}^2) - 3(n-2)^2(1 - b_{n-2}^2)^2]. \]
Thus we complete the proof.

**Lemma 2.** For a sufficiently large \( n \)

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

**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
\[ (2.7) \quad b_\nu = \sqrt{\frac{2}{\nu}} \frac{\Gamma\left( \frac{\nu+1}{2} \right)}{\Gamma\left( \frac{\nu}{2} \right)} = 1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} - \frac{1}{128\nu^3} + O\left( \frac{1}{\nu^4} \right), \]
which yields
\[ (2.8) \quad 1 - b_\nu^2 = (1 - b_\nu)(1 + b_\nu) = (1 - b_\nu)(2 - (1 - b_\nu)) = \frac{1}{2\nu} - \frac{1}{8\nu^2} - \frac{1}{16\nu^3} + O\left( \frac{1}{\nu^4} \right). \]

From (2.3), (2.7) and (2.8) we have for a sufficiently large \( n \)
\[ \text{Var}(Z + qS_1 - pS_2) = 1 + q^2(n-1)(1-b_{n-1}^2) + p^2(n-2)(1-b_{n-2}^2) \]
\[ = 1 + (p^2 + q^2) \left( \frac{1}{2} - \frac{1}{8n} - \frac{1}{16n^2} + O\left( \frac{1}{n^3} \right) \right). \]

Since, by (2.7) and (2.8),
\[ \nu^{3/2} b_v \left\{ 2(b_v^2-1) + \frac{1}{\nu} \right\} = \frac{1}{4\sqrt{\nu}} \left( 1 + \frac{1}{4\nu} \right) + O\left( \frac{1}{\nu^{5/2}} \right), \]

it follows from Lemma 1 that
\[ \kappa_3(Z + qS_1 - pS_2) = \frac{q^2 - p^2}{4\sqrt{n}} \left( 1 + \frac{1}{4n} \right) + O\left( \frac{1}{n^{3/2}} \right). \]

From Lemma 1 and (2.8) we have
\[ \kappa_4(Z + qS_1 - pS_2) \]
\[ = q^4 \left[ 2(n-1) \left\{ -1 + \frac{1}{n-1} - \frac{1}{4(n-1)^2} - \frac{1}{8(n-1)^3} + O\left( \frac{1}{n^4} \right) \right\} \right. \]
\[ + (n-1)^2 \left\{ 4 \left( \frac{1}{2(n-1)} - \frac{1}{8(n-1)^2} - \frac{1}{16(n-1)^3} + O\left( \frac{1}{n^4} \right) \right) \right. \]
\[ - 6 \left( \frac{1}{4(n-1)^2} - \frac{1}{8(n-1)^3} + O\left( \frac{1}{n^4} \right) \right) \right] \]
\[ + p^4 \left[ 2(n-2) \left\{ -1 + \frac{1}{n-2} - \frac{1}{4(n-2)^2} - \frac{1}{8(n-2)^3} + O\left( \frac{1}{n^4} \right) \right\} \right. \]
\[ + (n-2)^2 \left\{ 4 \left( \frac{1}{2(n-2)} - \frac{1}{8(n-2)^2} - \frac{1}{16(n-2)^3} + O\left( \frac{1}{n^4} \right) \right) \right. \]
\[ - 6 \left( \frac{1}{4(n-2)^2} - \frac{1}{8(n-2)^3} + O\left( \frac{1}{n^4} \right) \right) \right] \]
\[ = O\left( \frac{1}{n^2} \right). \]

Thus we complete the proof.

**Lemma 3.** The third and fourth cumulants of \( W \) are given by
\[ \kappa_3(W) = \frac{q^3(n-1)^{3/2} b_{n-1} \left\{ 2(b_{n-1}^2-1) + \frac{1}{n-1} \right\} - p^3(n-2)^{3/2} b_{n-2} \left\{ 2(b_{n-2}^2-1) + \frac{1}{n-2} \right\} }{(1 + q^2(n-1)(1-b_{n-1}^2) + p^2(n-2)(1-b_{n-2}^2))^{3/2}} \]
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 \( p \) instead of \( \rho \) by \( W_a \).
THEOREM. The upper 100\(\alpha\) percentile \(r_a\) of the distribution of \(R\) can be derived from the formula

\[
(2.9) \quad \frac{-q\sqrt{n-1} b_{n-1} + p_a \sqrt{n-2} b_{n-2}}{(1 + q^2(n-1)(1 - b_{n-1}^2) + p_a^2(n-2)(1 - b_{n-2}^2))^{1/2}} = u_a + \frac{1}{6}(u_a^2 - 1)\kappa_3(W_a) + O\left(\frac{1}{n}\right),
\]

where \(p_a = r_a/\sqrt{1-r_a^2}\) and \(u_a\) is the upper 100\(\alpha\) percentile of the standard normal distribution and \(\kappa_3(W_a)\) is given in Lemma 3.

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

\[
\frac{-q\sqrt{n-1} b_{n-1} + p_a \sqrt{n-2} b_{n-2}}{(1 + q^2(n-1)(1 - b_{n-1}^2) + p_a^2(n-2)(1 - b_{n-2}^2))^{1/2}} = u_a + \frac{1}{6}\kappa_3(W_a)(u_a^2 - 1) + \frac{1}{24}\kappa_4(W_a)(u_a^2 - 3u_a) - \frac{1}{36}\kappa_3^2(W_a)(2u_a^3 - 5u_a) + O\left(\frac{1}{n}\right),
\]

where \(W_a\) denotes \(W\) with \(p_a\) 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 \(f(r_a) = u_a + \frac{1}{6}(u_a^2 - 1)\kappa_3(W_a)\)]
for a percentage point $t_a$ 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 $t_a$ 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.

\[
q \sqrt{n-1} b_{n-1} + p_a \sqrt{n-2} b_{n-2} + p_0 \sqrt{n} + \sigma(1) = u_a + o(1),
\]

which is called the first order approximation. Since, by (2.7), $\sqrt{\nu} b_0 = \sqrt{\nu - (1/2)}$, it follows that

\[
q \sqrt{n-1} b_{n-1} + p_a \sqrt{n-2} b_{n-2} = p_a \sqrt{n-5} - q \sqrt{n-3}.
\]

It also follows from (2.3) and Lemma 2 that

\[
1 + q^2(n-1)(1-b_{n-1}^2) + p_0(n-2)(1-b_{n-2}^2) = 1 + \frac{p_0^2 + q^2}{2} + O\left(\frac{1}{n}\right).
\]

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

\[
\frac{p_a \sqrt{2n-5} - q \sqrt{2n-3}}{\sqrt{p_0^2 + q^2 + 2}} = u_a,
\]

which yields the well-known normal approximation formula
\[
(2.13) \quad \rho_a = q \left( \frac{(2n-3)(2n-5)}{2n-5-u_a^2} + u_a \sqrt{\frac{(4n-8)q^2+2(2n-5)-u_a^2}{2n-5-u_a^2}} \right)
\]

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

Next we consider an approximation to the percentage point by Fisher’s Z-transformation. Let

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

Then it is known that \(Z\) 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 \(0 < \alpha < 1\)

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

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

\[
(2.14) \quad r_\alpha = \frac{e^{2z_\alpha} - 1}{e^{2z_\alpha} + 1}.
\]

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

\[
(2.15) \quad z_\alpha = \zeta + \frac{u_a}{\sqrt{n}} + \frac{\rho}{2n} + \frac{1}{12n\sqrt{n}} \left\{ u_a^2 + 3(3-\rho^2)u_a \right\} \\
+ \frac{1}{24n^2} \left\{ 4\rho^3u_a^2 + (15\rho - \rho^3) \right\} + \frac{1}{480n^3\sqrt{n}} \\
\times \left\{ u_a^2 + (80 + 30\rho^2 - 60\rho^4)u_a^2 + (375 - 21\rho^2 + 45\rho^4)u_a \right\}.
\]

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>
<thead>
<tr>
<th>(n)</th>
<th>(\alpha = .990)</th>
<th>(\alpha = .975)</th>
<th>(\alpha = .950)</th>
<th>(\alpha = .900)</th>
<th>(\alpha = .750)</th>
<th>(\alpha = .500)</th>
<th>(\alpha = .250)</th>
<th>(\alpha = .100)</th>
<th>(\alpha = .050)</th>
<th>(\alpha = .025)</th>
<th>(\alpha = .010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>.15</td>
<td>.05</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td>.02</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>.20</td>
<td>.10</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>.30</td>
<td>.05</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>
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 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
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 \( p=0.500 \)

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

REFERENCES


