HIGHER ORDER APPROXIMATION OF THE PROBABILITY DISTRIBUTION OF THE RATIO ESTIMATOR FOR A REGRESSION MODEL

Shinichi Kawai*

In the present paper, we consider the problem of estimating a ratio $\rho = E(Y)/E(X)$ in a regression model $Y = \alpha + \beta X + U$. We obtain the higher order approximation of the probability distribution of the usual ratio estimator based on the sample means. In the gamma, lognormal and exponential cases, the approximation is numerically compared with the normal one and the empirical distribution. We also consider the higher order approximation of the percentage point and the construction of the confidence interval by using the approximation.

Key words and phrases: Cornish-Fisher expansion, Edgeworth expansion, higher order approximation, ratio estimator, regression model.

1. Introduction

When we want to estimate the ratio $\rho$ of the means of two random variables $X$ and $Y$, i.e. $\rho = E(Y)/E(X)$, it often happens that $X$ and $Y$ are correlated. In this case, we can assume a linear regression model $Y = \alpha + \beta X + U$, and we can utilize this information in the estimation of $\rho$.

Many authors have studied the estimation of the ratio $\rho$ using the linear regression model. For example, Durbin (1959), Rao (1965), Rao and Webster (1966), Gray and Schucany (1972), Rao (1988), Akahira and Kawai (1990), Kawai and Akahira (1994) have discussed the estimation of $\rho$ from the viewpoint of the jackknife method proposed by Quenouille (1956), which is based on dividing the sample at random into groups. Their papers addressed the problem of the optimum choice of the number of blocks and presented the comparison of the jackknife estimator with other estimators for $\rho$.

In the present work, we consider the probability distribution and the percentage point of the ratio estimator. Since it is difficult to obtain them exactly, some approximations are needed. One approach is to get the empirical distribution function by running the Monte Carlo simulation in computers. The simulation becomes more accurate when the number of Monte Carlo trials increases, but the cost gets higher. Another approach is to get the approximate formula by the asymptotic expansions of the type of Edgeworth and Cornish-Fisher. The derivation of the approximate formula is complicated but once it is obtained, we only calculate it by substituting the appropriate values. It takes less time and fewer computational complexity than the Monte Carlo simulation.

*National Research Institute for Earth Science and Disaster Prevention, 3-1 Tenno-dai, Tsukuba, Ibaraki 305-0006, Japan.

In this paper, we consider the higher order approximation of the probability distribution of the ratio estimator of $\rho$ in the regression model. Indeed, using the Edgeworth expansion, we obtain the approximation for the distribution of the usual ratio estimator based on the sample means up to the order $o(1/n)$, when the size of a sample is $n$. In the gamma, lognormal and exponential distribution cases, we numerically compare the higher order approximation with the normal one and the empirical distribution function. We also consider the higher order approximation of the percentage point by using the Cornish-Fisher expansion. As an application of the higher order approximation, we consider the construction of the confidence interval.

2. Higher order approximation of the distribution of the ratio estimator

Suppose that $(X_1,Y_1), \ldots, (X_n,Y_n)$ are a random sample of size $n$. We consider the problem of estimating the ratio $\rho = E(Y_i) / E(X_i)$. Let

$$Y_i = \alpha + \beta X_i + U_i \ (i = 1, \ldots, n),$$

where $X_1, \ldots, X_n, U_1, \ldots, U_n$ are mutually independent and $X_1, \ldots, X_n$ and $U_1, \ldots, U_n$ are i.i.d. samples respectively. Here, $k_0 := E(X_i)$, $k_1 := V(X_i) > 0$, $k_2 := E(X_i^3)$, $k_3 := E(X_i^4)$, $E(U_i) = 0$, $\delta := V(U_i)$, $\eta := E(U_i^3)$, and $\gamma := E(U_i^4)$, where $\delta = O(1)$. Suppose that the value of $k_0$ is either positive or negative. Let $\bar{X} = \sum_{i=1}^n X_i / n$, $\bar{Y} = \sum_{i=1}^n Y_i / n$, and $\bar{U} = \sum_{i=1}^n U_i / n$. One of the common estimators for the ratio $\rho$ is a ratio estimator

$$R := \frac{\bar{Y}}{\bar{X}} = \beta + \frac{\alpha + \bar{U}}{\bar{X}}.$$

Assume that $k_0 > 0$. Then, the cumulative distribution function (c.d.f.) of $\sqrt{n} (R - \rho)$ is given by

$$F_R(r) := P\{ \sqrt{n}(R - \rho) \leq r \}$$

$$= P\left\{ \sqrt{n} \left( \frac{\bar{Y}}{\bar{X}} - \rho \right) \leq r, \sqrt{n}\bar{X} \leq 0 \right\}$$

$$+ P\left\{ \sqrt{n} \left( \frac{\bar{Y}}{\bar{X}} - \rho \right) \leq r, \sqrt{n}\bar{X} > 0 \right\}. \noindent

If $\limsup_{|t| \to \infty} |E(e^{itX_1})| < 1$ and $E(|X_1|^4) = k_3 < \infty$, the Edgeworth expansion

$$P\left\{ \sqrt{\frac{n}{k_1}} (\bar{X} - k_0) \leq x \right\}$$

$$= \Phi(x) + \frac{1}{\sqrt{n}} \phi(x)Q_1(x) + \frac{1}{n} \phi(x)Q_2(x) + o \left( \frac{1}{n} \right)$$

$$+ o \left( \sqrt{\frac{n}{k_1}} \right)$$

is valid.
approximation of distribution of ratio estimator 67

is valid uniformly in \( x \) (see, e.g., Feller (1966)), where

\[
Q_1(x) = -\frac{\kappa_X^{(3)}}{6k_1^{3/2}}(x^2 - 1) \quad \text{and} \quad Q_2(x) = -\frac{\kappa_X^{(4)}}{24k_1^2}(x^3 - 3x) - \frac{\kappa_X^{(3)^2}}{72k_1^3}(x^5 - 10x^3 + 15x).
\]

Here, \( \kappa_X^{(3)} \) and \( \kappa_X^{(4)} \) are the third and fourth cumulant of \( X_1 \), respectively. When \( x = -k_0 \sqrt{n/k_1} \), the sum of the first three terms of the right hand side of (2.2) is \( o(1/n) \). Then

\[
P\left\{ \sqrt{n} \left( \frac{\bar{Y}}{\bar{X}} - \rho \right) \leq r, \, \sqrt{n} \bar{X} \leq 0 \right\} \leq P\left\{ \sqrt{n} \bar{X} \leq 0 \right\} = P\left\{ \sqrt{\frac{n}{k_1}} (\bar{X} - k_0) \leq -k_0 \sqrt{\frac{n}{k_1}} \right\} = o\left( \frac{1}{n} \right).
\]

It follows from (2.1) that

\[
F_R(r) = P\left\{ \sqrt{n} \left( \frac{\bar{Y}}{\bar{X}} - \rho \right) \leq r, \, \sqrt{n} \bar{X} > 0 \right\} + o\left( \frac{1}{n} \right)
= P\left\{ \bar{U} - \left( \frac{\alpha}{k_0} + \frac{r}{\sqrt{n}} \right) \bar{X} \leq -\alpha, \sqrt{n} \bar{X} > 0 \right\} + o\left( \frac{1}{n} \right).
\]

Putting \( c = (\alpha / k_0) + (r / \sqrt{n}) \), we have

\[
0 \leq P\{ \bar{U} - c\bar{X} \leq -\alpha \} - P\{ \bar{U} - c\bar{X} \leq -\alpha, \sqrt{n} \bar{X} > 0 \}
= P\{ \bar{U} - c\bar{X} \leq -\alpha, \sqrt{n} \bar{X} \leq 0 \}
\leq P\{ \sqrt{n} \bar{X} \leq 0 \} = o\left( \frac{1}{n} \right).
\]

From (2.3) we obtain

\[
F_R(r) = P\left\{ \bar{U} - \left( \frac{\alpha}{k_0} + \frac{r}{\sqrt{n}} \right) \bar{X} \leq -\alpha \right\} + o\left( \frac{1}{n} \right).
\]

Putting

\[
W := \bar{U} - \left( \frac{\alpha}{k_0} + \frac{r}{\sqrt{n}} \right) \bar{X},
\]

we have

\[
\mu_W := E(W) = -\alpha - \frac{1}{\sqrt{n}} (k_0 r),
\]

\[
\sigma_W^2 := V(W) = \frac{1}{n} A + \frac{1}{n \sqrt{n}} B + \frac{1}{n^2} C,
\]
where

\[ A := \delta + \frac{k_1}{k_0} \alpha^2, \quad B := 2 \frac{k_1}{k_0} r \alpha, \quad C := k_1 r^2. \]

Let

\[ Z := \frac{W - \mu_W}{\sigma_W}. \]

Then \( E(Z) = 0, \; V(Z) = 1. \) Let the third and fourth cumulant of \( Z \) be denoted by \( \kappa_3 \) and \( \kappa_4 \), respectively. If \( \limsup_{|t_1| \to \infty, |t_2| \to \infty} |E \left[ e^{i(t_1 X_1 + t_2 U_1)} \right]| < 1, \) \( E(|X_1|^4) = k_3 < \infty \) and \( E(|U_1|^4) = \gamma < \infty \), the Edgeworth expansion

(2.6) \( P\{Z \leq z\} = \Phi(z) - \phi(z) \times \left\{ \frac{\kappa_3}{6} (z^2 - 1) + \frac{\kappa_4}{24} (z^3 - 3z) + \frac{\kappa_3^2}{72} (z^5 - 10z^3 + 15z) \right\} + o \left( \frac{1}{n} \right) \)

is valid uniformly in \( z \) (see, e.g., Bhattacharya and Ghosh (1978) and Hall (1992)), where

\[ z = \frac{-\alpha - \mu_W}{\sigma_W}, \]

\[ \kappa_3 = E \left[ (Z - E(Z))^3 \right] = E(Z^3), \]

\[ \kappa_4 = E \left[ (Z - E(Z))^4 \right] - 3 \{ V(Z) \}^2 = E(Z^4) - 3, \]

\( \Phi(z) \) is the distribution function and \( \phi(z) \) is the density function of the standard normal distribution, respectively. We obtain from (2.4) and (2.6),

(2.7) \( F_R(r) = \Phi(z) - \phi(z) \times \left\{ \frac{\kappa_3}{6} (z^2 - 1) + \frac{\kappa_4}{24} (z^3 - 3z) + \frac{\kappa_3^2}{72} (z^5 - 10z^3 + 15z) \right\} + o \left( \frac{1}{n} \right) \).

Then we have

(2.8) \( \kappa_3 = E \left[ \left( \frac{W - \mu_W}{\sigma_W} \right)^3 \right] = \sigma_W^{-3} \kappa_W^{(3)}, \)

and

(2.9) \( \kappa_4 = E \left[ \left( \frac{W - \mu_W}{\sigma_W} \right)^4 \right] - 3 = \sigma_W^{-4} \left\{ E[(W - \mu_W)^4] - 3(\sigma_W^2)^2 \right\} \]

\[ = \sigma_W^{-4} \kappa_W^{(4)}, \]
where \( \kappa_W^{(3)} \) and \( \kappa_W^{(4)} \) are the third and fourth cumulant of \( W \) respectively. By the Taylor expansion, we obtain from (2.5),

\[
\sigma_W^{-3} = \frac{n \sqrt{n}}{A^{3/2}} \left\{ 1 + \frac{1}{\sqrt{n}} \left( -3 \frac{k_1}{A k_0} \right) r + o \left( \frac{1}{\sqrt{n}} \right) \right\},
\]

(2.10)

\[
\sigma_W^{-4} = \frac{n^2}{A^2} \left\{ 1 + o(1) \right\}.
\]

(2.11)

We also have

\[
\kappa^{(i)}_W = \frac{1}{n^{i-1}} \left\{ \kappa^{(i)}_U + (-1)^i \left( \frac{\alpha}{k_0} + \frac{r}{\sqrt{n}} \right)^i \kappa^{(i)}_X \right\} \quad (i = 3, 4),
\]

(2.12)

where \( \kappa^{(i)}_U \) and \( \kappa^{(i)}_X \) are the \( i \)th cumulant of \( U_1 \) and \( X_1 \) respectively \( (i = 3, 4) \) which are calculated as follows.

\[
\kappa^{(3)}_X = k_2 - 3k_1k_0 - k_0^3,
\]

\[
\kappa^{(3)}_U = \eta,
\]

\[
\kappa^{(4)}_X = k_3 - 4k_2k_0 + 6k_1k_0^2 + 3k_0^4 - 3k_1^2,
\]

\[
\kappa^{(4)}_U = \gamma - 3\delta^2.
\]

Then \( \kappa_W^{(3)} \) and \( \kappa_W^{(4)} \) are expanded as

\[
\kappa^{(3)}_W = \frac{1}{n^2} \left[ \left\{ \kappa^{(3)}_U - \left( \frac{\alpha}{k_0} \right)^3 \kappa^{(3)}_X \right\}
\right.
\]

\[
+ \frac{1}{\sqrt{n}} (-3) \left( \frac{\alpha}{k_0} \right)^2 \kappa^{(3)}_X r + o \left( \frac{1}{\sqrt{n}} \right) \right],
\]

(2.13)

\[
\kappa^{(4)}_W = \frac{1}{n^3} \left[ \left\{ \kappa^{(4)}_U + \left( \frac{\alpha}{k_0} \right)^4 \kappa^{(4)}_X \right\} + o(1) \right].
\]

(2.14)

Substituting (2.10), (2.11), (2.13) and (2.14) into (2.8) and (2.9), we have

\[
\kappa_3 = \frac{1}{A^{3/2}} \left\{ \frac{1}{\sqrt{n}} D + \frac{1}{n} E r + o \left( \frac{1}{n} \right) \right\},
\]

(2.15)

\[
\kappa_4 = \frac{1}{n} F + o \left( \frac{1}{n} \right),
\]

(2.16)

where

\[
D := \kappa^{(3)}_U - \left( \frac{\alpha}{k_0} \right)^3 \kappa^{(3)}_X,
\]

\[
E := -\frac{3 k_1}{A k_0} \kappa^{(3)}_U - \left( \frac{\alpha}{k_0} \right)^3 \kappa^{(3)}_X - 3 \left( \frac{\alpha}{k_0} \right)^2 \kappa^{(3)}_X,
\]

\[
F := \kappa^{(4)}_U + \left( \frac{\alpha}{k_0} \right)^4 \kappa^{(4)}_X.
\]
We also obtain from (2.15),

\[(2.17) \quad \kappa_3^2 = \frac{1}{n} \frac{D^2}{A^3} + o \left( \frac{1}{n} \right) . \]

Hence, the higher order approximation of \( F_R(r) \) is given by (2.7) with (2.15), (2.16) and (2.17) in which the values of \( A, D, E, \) and \( F \) are obtained in the above.

We can also consider the higher order approximation of the percentage point of the ratio estimator by using the Cornish-Fisher expansion. We use the same notations defined above except for the functions of \( r \) which are denoted explicitly as \( z(r), \mu_W(r), \sigma_W(r), B(r), C(r), \) and \( \kappa_3(r) \).

Let \( r_p \) be the upper 100\( p \) percentile of the distribution of \( \sqrt{n} (R - \rho) \), i.e.

\[ F_R(r_p) = P\{ \sqrt{n} (R - \rho) \leq r_p \} = 1 - p . \]

Then \( F_R(r_p) = P\{ Z \leq z(r_p) \} + o \left( n^{-1} \right) \), where

\[(2.18) \quad z(r_p) = \frac{-\alpha - \mu_W(r_p)}{\sigma_W(r_p)} = \frac{k_0 r_p}{\sqrt{A + \frac{1}{\sqrt{n}} B(r_p) + \frac{1}{n} C(r_p)}} . \]

Hence the upper 100\( p \) percentile of the distribution of \( \sqrt{n} (R - \rho) \) asymptotically follows from that of the distribution of \( Z \). Using the Cornish-Fisher expansion, we obtain

\[(2.19) \quad z(r_p) = u_p + \frac{\kappa_3(r_p)}{6} (u_p^2 - 1) + \frac{\kappa_4}{24} (u_p^3 - 3u_p) + \frac{\kappa_5^2}{36} (-2u_p^3 + 5u_p) + o \left( \frac{1}{n} \right) , \]

where \( u_p \) is the upper 100\( p \) percentile of the standard normal distribution, i.e. \( \Phi(u_p) = 1 - p \). The higher order approximation of \( r_p \) is obtained by solving the equation (2.19) with (2.15), (2.16), (2.17), and (2.18). The Newton’s method, an iterative method, is applied to equation (2.19) to find the value of \( r_p \) numerically.

If \( k_0 < 0 \), then, in a similar way to the case \( k_0 > 0 \), we have

\[ F_R(r) = P \left\{ \bar{U} - \left( \frac{\alpha}{k_0} + \frac{r}{\sqrt{n}} \right) \bar{X} \geq -\alpha \right\} + o \left( \frac{1}{n} \right) \]

\[ = P\{ Z \geq z \} + o \left( \frac{1}{n} \right) . \]

**Remark.** By expanding the left hand side of (2.19), some other approximations are obtained. If the left hand side of (2.19) is expanded up to the order of \( n^{-1} \), the cubic equation about \( r_p \) is obtained. If both sides of (2.19) are multiplied by \( \sqrt{n} \sigma_W \) and expanded up to the order of \( n^{-1} \), the quadratic equation about \( r_p \) is obtained. These equations can be solved explicitly.
By utilizing the higher order approximation of the percentage point, we consider the construction of the confidence interval. Since

\[ P \left\{ r_{1-(p/2)} \leq \sqrt{n}(R - \rho) \leq r_{p/2} \right\} = 1 - p, \]

the two-sided 100(1 − p)% confidence interval for \( \rho \), which is denoted by \( I \), is

\[ I = \left[ R - n^{-1/2} r_{p/2}, R - n^{-1/2} r_{1-(p/2)} \right]. \]

If some parameters are unknown, they are estimated from the sample \((X_1, Y_1), \ldots, (X_n, Y_n)\). \( r_{p/2} \) and \( r_{1-(p/2)} \) are also estimated by solving the equation (2.19) in which unknown parameters are replaced by the estimated values. In this case, the confidence interval, which is denoted by \( \hat{I} \), is

\[ \hat{I} = \left[ R - n^{-1/2} \hat{r}_{p/2}, R - n^{-1/2} \hat{r}_{1-(p/2)} \right], \tag{2.20} \]

where \( \hat{r}_{p/2} \) and \( \hat{r}_{1-(p/2)} \) are the estimated values of \( r_{p/2} \) and \( r_{1-(p/2)} \) respectively.

3. Examples

Some examples are shown to see the accuracy of the approximation. Since the true c.d.f. of \( \sqrt{n}(R - \rho) \), \( F_R(r) \), is not so easy to calculate, for suitable values of parameters, we obtain the empirical distribution function (e.d.f.) by simulation. The e.d.f. of \( \sqrt{n}(R - \rho) \), \( \hat{F}_R(r) \), is defined by

\[ \hat{F}_R(r) := \frac{\# \left\{ \sqrt{n}(R - \rho) \leq r \right\}}{b} \]

where \( b \) is the number of simulation times and \( \# \left\{ \sqrt{n}(R - \rho) \leq r \right\} \) is the number of values of \( \sqrt{n}(R - \rho) \) which are less than or equal to \( r \). We regard this as the true c.d.f. and compare it with the approximations of \( F_R(r) \). We consider two kinds of approximations of \( F_R(r) \), that is, one is the Edgeworth expansion and another one is the normal distribution. These three values are shown in the figures of the examples below.

In the following examples, suppose that \( U_1, \ldots, U_n \) are identically and independently distributed random variables according to the normal distribution with mean 0 and variance \( \sigma^2 \). Then it follows that \( E(U_i) = 0 \ (i = 1, \ldots, n) \), \( \delta = \sigma^2, \eta = 0 \), and \( \gamma = 3\sigma^4 \). Let \( \alpha = 2, \beta = 1, \sigma = 1 \), and the repeated number \( b \) of simulation be 10,000.

Example 3.1 (Gamma case). Suppose that \( X_1, \ldots, X_n \) are identically and independently distributed random variables according to the gamma distribution with the density

\[ \frac{1}{\Gamma(h)} e^{-x} x^{h-1} \]

for \( x > 0 \) and 0 otherwise, where \( h > 0 \). Then we have \( k_0 = k_1 = h, k_2 = (h + 2)(h + 1)h \), and \( k_3 = h^4 + 6h^3 + 11h^2 + 6h \).
Figure 1. Comparison between \( \hat{F}_R(r) \) and the approximations of \( F_R(r) \) for the gamma case when \( h = 1, 2 \) with \( n = 10, \alpha = 2, \beta = 1, \sigma^2 = 1 \), where \( b = 10000 \).

Figure 2. Comparison between \( \hat{F}_R(r) \) and the approximations of \( F_R(r) \) for the gamma case when \( n = 10, 20 \) with \( h = 0.5 \), \( \alpha = 2, \beta = 1, \sigma^2 = 1 \), where \( b = 10000 \).

Figure 1 shows the comparison between the empirical distribution function \( \hat{F}_R(r) \) and the approximations of \( F_R(r) \) in case of \( h = 1, 2 \) with \( n = 10 \). Figure 2 also shows the comparison in case of \( n = 10, 20 \) with \( h = 0.5 \).

Higher order approximation of the percentage point is applied to this gamma case. Since the true value of the percentage point of \( F_R(r) \) is hard to calculate, we regard the upper 100\( p \) percentile point of \( \hat{F}_R(r) \), which is the value of \( b \times (1 - p) + 1 \)th order statistics: \( r_p = r_{b(1-p)+1} \), as the true value of \( F_R(r) \).

The upper 5 percentile point of \( \hat{F}_R(r) \), \( r_{(0.05)} \), is compared with the two approximations of \( r_{0.05} \) of \( F_R(r) \). One is the Cornish-Fisher expansion (2.19) and another one is the normal distribution, which satisfies the equation \( z(r_p) = u_p \).

Table 1 shows the approximations when \( h = 1(0.5)5 \) with \( n = 10 \).

Example 3.2 (Lognormal case). Suppose that \( X_1, \ldots, X_n \) are identically and independently distributed random variables according to the lognormal dis-
Table 1. Comparison between the upper 5 percentile point of $\hat{F}_R(r)$, $r_{(0.05)}$, and the approximations of $r_{0.05}$ for the gamma case when $h = 1(0.5)5$ with $n = 10$, $\alpha = 2$, $\beta = 1$, $\sigma^2 = 1$, where $b = 10000$.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$r_{(0.05)}$ of $F_{R}(r)$</th>
<th>Higher order Approximation of $r_{0.05}$</th>
<th>Error from $r_{(0.05)}$</th>
<th>Normal Approximation of $r_{0.05}$</th>
<th>Error from $r_{(0.05)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.668303</td>
<td>5.60528</td>
<td>-0.063023</td>
<td>7.24965</td>
<td>1.581347</td>
</tr>
<tr>
<td>1.5</td>
<td>3.008991</td>
<td>2.93707</td>
<td>-0.071921</td>
<td>3.42589</td>
<td>0.416899</td>
</tr>
<tr>
<td>2</td>
<td>1.920881</td>
<td>1.89113</td>
<td>-0.029751</td>
<td>2.10446</td>
<td>0.183579</td>
</tr>
<tr>
<td>2.5</td>
<td>1.349946</td>
<td>1.35846</td>
<td>0.008514</td>
<td>1.47162</td>
<td>0.121674</td>
</tr>
<tr>
<td>3</td>
<td>1.030939</td>
<td>1.04394</td>
<td>0.013001</td>
<td>1.11154</td>
<td>0.080601</td>
</tr>
<tr>
<td>3.5</td>
<td>0.8527138</td>
<td>0.839629</td>
<td>-0.0130848</td>
<td>0.883396</td>
<td>0.0306822</td>
</tr>
<tr>
<td>4</td>
<td>0.69100905</td>
<td>0.697767</td>
<td>0.0066765</td>
<td>0.727797</td>
<td>0.0367065</td>
</tr>
<tr>
<td>4.5</td>
<td>0.6071027</td>
<td>0.594297</td>
<td>-0.0128057</td>
<td>0.615828</td>
<td>0.0087253</td>
</tr>
<tr>
<td>5</td>
<td>0.5126955</td>
<td>0.51592</td>
<td>0.0032245</td>
<td>0.531902</td>
<td>0.0192065</td>
</tr>
</tbody>
</table>

distribution with the density

$$
\frac{1}{\sqrt{2\pi }\sigma _{LN}}e^{-1/2((\log x-\zeta )^2/\sigma _{LN}^2)}
$$

for $x > 0$ and 0 otherwise, where $\sigma _{LN} > 0$. The $r$th moment of $X$ about zero is $E(X^r) = e^{r\zeta +r^2\sigma _{LN}^2/2}$. We now consider the case of $\zeta = 0$. Then we have $k_0 = e^{(1/2)\sigma _{LN}^2}$, $k_1 = e^{2\sigma _{LN}^2 - e^{\sigma _{LN}^2}}$, $k_2 = e^{(9/2)\sigma _{LN}^2}$, $k_3 = e^{8\sigma _{LN}^2}$.

Figure 3 shows the comparison between the empirical distribution function $\hat{F}_R(r)$ and the approximations of $F_R(r)$ in case of $\sigma _{LN} = 0.5, 0.8, 1$ with $n = 20$. Figure 4 also shows the comparison in case of $n = 10, 30, 50$ with $\sigma _{LN} = 1$.

**Example 3.3 (Exponential case).** Suppose that $X_1, \ldots, X_n$ are identically and independently distributed random variables according to the exponential distribution with the location parameter $\theta _E$ and the scale parameter $\sigma _E$. The density function is

$$
\frac{1}{\sigma _E}e^{-(x-\theta _E)/\sigma _E}
$$

for $x > \theta _E$ and 0 otherwise, where $\sigma _E > 0$. Then we have $k_0 = \theta _E + \sigma _E$, $k_1 = \sigma _E^2$, $k_2 = \theta _E^3 + 3\theta _E^2\sigma _E + 6\theta _E\sigma _E^2 + 6\sigma _E^3$, $k_3 = \theta _E^4 + 4\theta _E^3\sigma _E + 12\theta _E^2\sigma _E^2 + 24\theta _E\sigma _E^3 + 24\sigma _E^4$.

As an example of $k_0 < 0$, we consider the case of $\theta _E = -1$ and $\sigma _E = 0.5$, which results in $k_0 = -0.5$. Figure 5 shows the comparison between the empirical distribution function $\hat{F}_R(r)$ and the approximations of $F_R(r)$ in case of $n = 10, 20$. In this case, the Edgeworth approximation is better than the normal one for the empirical distribution.

We also consider the construction of the confidence interval for $\rho$. If all parameters are unknown, they are estimated from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. Suppose that the estimators of $\alpha, \beta, \sigma^2, \theta_E$ and $\sigma_E$ are $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2, \hat{\theta}_E$ and $\hat{\sigma}_E$ respectively. $\hat{\alpha}$ and $\hat{\beta}$ are the least square estimators of the regression model
Figure 3. Comparison between $\hat{F}_R(r)$ and the approximations of $F_R(r)$ for the lognormal case when $\sigma_{LN} = 0.5, 0.8, 1$ with $n = 20, \alpha = 2, \beta = 1, \sigma^2 = 1$, where $b = 10000$.

Figure 4. Comparison between $\hat{F}_R(r)$ and the approximations of $F_R(r)$ for the lognormal case when $n = 10, 30, 50$ with $\sigma_{LN} = 1, \alpha = 2, \beta = 1, \sigma^2 = 1$, where $b = 10000$.

Figure 5. Comparison between $\hat{F}_R(r)$ and the approximations of $F_R(r)$ for the exponential case when $n = 10, 20$ with $\theta_E = -1, \sigma_E = 0.5, \alpha = 2, \beta = 1, \sigma^2 = 1$, where $b = 10000$. 
Table 2. Coverage errors of the two-sided 50(10)90,95 and 99 % confidence intervals \( \hat{I} \) in (2.20) when \( n = 10, 30, 50\) and 100 for the exponential case with \( \alpha = 2, \beta = 1, \sigma^2 = 1, \theta_E = 0 \) and \( \sigma_E = 1 \), where \( b = 10000 \).

<table>
<thead>
<tr>
<th>( 100(1 - p)% )</th>
<th>( n = 10 )</th>
<th>( n = 30 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher order</td>
<td>normal</td>
<td>higher order</td>
<td>normal</td>
<td>higher order</td>
</tr>
<tr>
<td>99%</td>
<td>-7.36%</td>
<td>-8.38%</td>
<td>-3.56%</td>
<td>-4.19%</td>
</tr>
<tr>
<td>95%</td>
<td>-7.46%</td>
<td>-8.00%</td>
<td>-3.18%</td>
<td>-3.62%</td>
</tr>
<tr>
<td>90%</td>
<td>-5.92%</td>
<td>-5.91%</td>
<td>-1.74%</td>
<td>-1.84%</td>
</tr>
<tr>
<td>80%</td>
<td>-2.41%</td>
<td>-1.43%</td>
<td>-0.59%</td>
<td>-0.20%</td>
</tr>
<tr>
<td>70%</td>
<td>-1.22%</td>
<td>-0.21%</td>
<td>-0.01%</td>
<td>0.24%</td>
</tr>
<tr>
<td>60%</td>
<td>-0.75%</td>
<td>-0.80%</td>
<td>0.20%</td>
<td>0.18%</td>
</tr>
<tr>
<td>50%</td>
<td>-0.85%</td>
<td>-1.68%</td>
<td>0.38%</td>
<td>-0.02%</td>
</tr>
</tbody>
</table>

The two-sided 100(1 – \( p \))% confidence interval for \( \rho \) is given by (2.20).

To see the accuracy of the confidence interval, the coverage error is calculated by the Monte Carlo simulation. When the true value of \( \rho \) is denoted by \( \rho_0 \), the coverage error is

\[
\#\{ \rho_0 \in \hat{I} \} / b - (1 - p),
\]

where \( b \) is the number of simulation times and \( \#\{ \rho_0 \in \hat{I} \} \) is the number of \( \hat{I} \) which includes \( \rho_0 \).

Table 2 shows the coverage errors of the two-sided 50(10)90,95 and 99 % confidence intervals in case of \( n = 10, 30, 50\) and 100 with \( \theta_E = 0 \) and \( \sigma_E = 1 \) for \( b = 10000 \) Monte Carlo trials. All parameters are assumed to be unknown and estimated from the data. Each percentage point is calculated by both higher order and normal approximation. From Table 2, the Cornish-Fisher approximation improves the normal one when 1 – \( p \) = 0.95 and above.

Acknowledgements

The author wishes to thank Professor Masafumi Akahira of the University of Tsukuba for his advice and guidance and the referees for their helpful comments.

References


