

Adjustment on an asymptotic expansion of the distribution function with χ^2 -approximation

Hirokazu YANAGIHARA and Tetsuji TONDA

(Received June 11, 2001)

(Revised February 8, 2002)

ABSTRACT. This paper is concerned with an adjustment on an asymptotic expansion of the distribution function, whose limiting distribution is a chi-squared distribution, up to the order n^{-1} . The distribution function is a monotone function and has the upper and lower bounds with 0 and 1, but an asymptotic expansion does not satisfy these properties. We consider to add a term of n^{-2} order to the asymptotic expansion so that the resulting one satisfies such properties. Note that our adjustment does not give an influence on the order of the remainder term in the asymptotic expansion. Our method of preserving monotoneity is based on the idea in Kakizawa (1996).

1. Introduction

The limiting distribution is often used as an approximate distribution, when it is difficult to obtain the exact distribution function for its complexity. However, its accuracy tends to be bad as the sample size n tends to be small. It is well known that an asymptotic expansion will improve accuracy of approximation compared with the limiting distribution in the small sample case. Suppose that a nonnegative random variate T has an asymptotic expansion such that

$$\begin{aligned} \mathbf{P}(T \leq x) &= G_r(x) + \frac{1}{n} \sum_{j=0}^k b_j G_{r+2j}(x) + o(n^{-1}) \\ &= P_{ae}(x) + o(n^{-1}), \end{aligned} \tag{1.1}$$

where $G_r(x)$ is the distribution function of a central χ^2 distribution with r degrees of freedom, coefficients b_j 's satisfy the relation $\sum_{j=0}^k b_j = 0$ and k is a certain positive integer. The approximation $P_{ae}(x)$ with the supplementary term $\sum_{j=0}^k b_j G_{r+2j}(x)/n$ will give a better approximation than the limiting dis-

2000 *Mathematics Subject Classification.* primary 62E20, secondly 60E05.

Key words and phrases. Bound, Chi-squared distribution, Hypothesis testing, Monotoneity, Nonnormality, P-value.

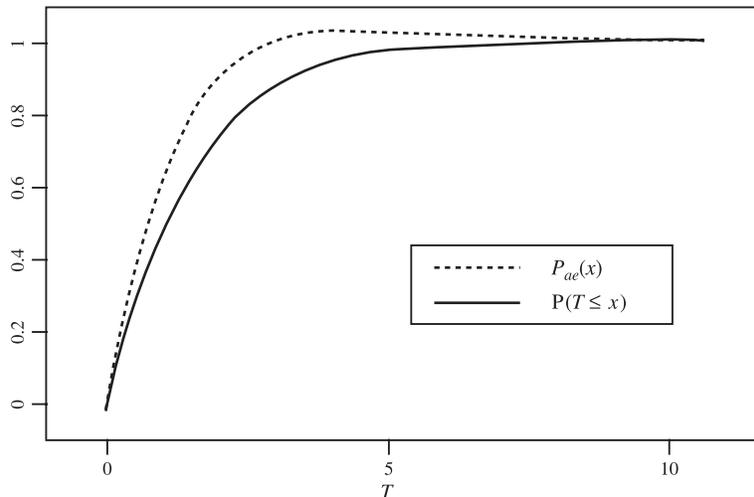


Figure 1.1. Example of an undesirable result on $P_{ae}(x)$

tribution $G_r(x)$. However, it doesn't always satisfy the basic properties of the distribution function: its monotonicity and upper & lower bound with 0 & 1. This fact leads to some problems when we use the approximation $P_{ae}(x)$ to get a p-value of certain statistic.

Figure 1.1 illustrates one of such problems. This example is a graph of $P_{ae}(x)$ in the case of the null distribution of the likelihood ratio statistic for testing equality of variances under nonnormality (see, Tonda and Wakaki (2001)) and its true distribution function which is obtained by Monte Carlo simulation. Note that $P_{ae}(x)$ exceeds 1, the upper bound of the distribution function. Such an undesirable phenomenon occurs under nonnormality more often than under normality. Our purpose is to avoid such a problem, considering to add an adjusting term of n^{-2} order to the asymptotic expansion such as (1.1) so that resulting one preserves the monotonicity and has the appropriate bounds. We call an expansion with such adjusting term an *adjusted asymptotic expansion*. Note that our adjustment does not give an influence on the order of the remainder term in the asymptotic expansion. In order to guarantee the monotonicity, we use an idea in Kakizawa (1996).

The present paper is organized in the following way. In §2 we derive the main result, i.e., an adjusted asymptotic expansion on χ^2 -approximation. In §3 we introduce some applications for our results in the cases of the null distributions of test statistics for testing the equality of variances and the sphericity of covariance. In §4 some simulation study on a former example in §3 is carried out.

2. Main result

In this section, we obtain an adjusted asymptotic expansion on χ^2 approximation. Suppose that a nonnegative random variate T has an asymptotic expansion as in (1.1). The adjusted asymptotic expansion $AP_{ae}(x)$ is defined by adding a supplementary term of order $O(n^{-2})$ as follows. Let

$$AP_{ae}(x) = d^{-1} \left\{ P_{ae}(x) + \frac{1}{n^2} a(x) \right\},$$

where d is defined by

$$d = \lim_{x \rightarrow \infty} \left\{ P_{ae}(x) + \frac{1}{n^2} a(x) \right\} = 1 + \frac{1}{n^2} \lim_{x \rightarrow \infty} a(x).$$

Here, $a(x)$ can be obtained by applying the idea in Kakizawa (1996) to our case:

$$\begin{aligned} a(x) &= \frac{1}{4} \int_0^x \{g_r(x)\}^{-1} \left\{ \frac{d}{dx} \left(\sum_{j=0}^k b_j G_{r+2j}(x) \right) \right\}^2 dx \\ &= \frac{1}{4} \int_0^x \{g_r(x)\}^{-1} \sum_{i=0}^k \sum_{j=0}^k b_i b_j g_{r+2i}(x) g_{r+2j}(x) dx \\ &= \frac{1}{4} \int_0^x p(x) dx, \end{aligned}$$

where $g_r(x)$ is the density function of a central χ^2 distribution with r degrees of freedom, which is defined by

$$g_r(x) = \frac{1}{2^{r/2} \Gamma(r/2)} e^{-x/2} x^{r/2-1}.$$

Then, $AP_{ae}(x)$ is to satisfy $dAP_{ae}(x)/dx \geq 0$ and $\lim_{x \rightarrow \infty} AP_{ae}(x) = 1$. Further, our adjustment does not give any influence on the order of the remainder term in the asymptotic expansion. Namely,

$$\begin{aligned} AP_{ae}(x) &= P_{ae}(x) + O(n^{-2}) \\ &= \mathbf{P}(T \leq x) + o(n^{-1}). \end{aligned}$$

Next we obtain explicit forms of $a(x)$ and d . Note that

$$g_{r+2i}(x) = \frac{x^i}{\prod_{\alpha=1}^i \{r + 2(\alpha - 1)\}} g_r(x) \quad (i \geq 1).$$

Therefore, we can obtain the following relation:

$$\frac{g_{r+2i}(x)}{g_r(x)} = \begin{cases} 1 & (i = 0), \\ \frac{x^i}{\prod_{\alpha=1}^i \{r + 2(\alpha - 1)\}} & (\text{otherwise}). \end{cases}$$

Using these results, we can derive

$$\begin{aligned} p(x) &= b_0 \sum_{j=0}^k b_j g_{r+2j}(x) + \sum_{i=1}^k \sum_{j=0}^k \frac{b_i b_j x^i g_{r+2j}(x)}{\prod_{\alpha=1}^i \{r + 2(\alpha - 1)\}} \\ &= \sum_{j=0}^k b_0 b_j g_{r+2j}(x) + \sum_{i=1}^k \sum_{j=0}^k b_i b_j \prod_{\alpha=1}^i \left\{ 1 + \frac{2j}{r + 2(\alpha - 1)} \right\} g_{r+2i+2j}(x). \end{aligned}$$

Therefore, the adjusted asymptotic expansion $AP_{ae}(x)$ can be obtained as in the following theorem.

THEOREM 2.1. *Suppose that the distribution function of a nonnegative variate T can be expanded as in (1.1), then an adjusted asymptotic expansion $AP_{ae}(x)$ can be given by*

$$\begin{aligned} AP_{ae}(x) &= d^{-1} \left\{ G_r(x) + \frac{1}{n} \sum_{j=0}^k b_j G_{r+2j}(x) \right. \\ &\quad \left. + \frac{1}{4n^2} \sum_{i=0}^k \sum_{j=0}^k b_i b_j c_{i,j} G_{r+2i+2j}(x) \right\}, \end{aligned} \quad (2.1)$$

where

$$d = 1 + \frac{1}{4n^2} \sum_{i=0}^k \sum_{j=0}^k b_i b_j c_{i,j},$$

and

$$c_{i,j} = \begin{cases} 1 & (i = 0), \\ \prod_{\alpha=1}^i \left(1 + \frac{2j}{r + 2\alpha - 2} \right) & (\text{otherwise}). \end{cases}$$

There is a different method of adjusting an asymptotic expansion, which is based on an inverse monotone function, i.e., an inverse improving transformation. Such improving transformations on χ^2 approximation were studied by Kakizawa (1996), Fujikoshi (1997), Fujisawa (1997) and Cordeiro et al. (1998), etc. The error in the χ^2 approximation is $o(n^{-1})$ by such a transformation, i.e.,

$$\mathbf{P}(f(T) \leq x) = G_r(x) + o(n^{-1}),$$

where $f(x)$ is a monotone increasing function. This gives an alternative adjustment defined by

$$P_{ii}(x) = G_r(f^{-1}(x)).$$

However, our method has some advantage in the comparison with the other methods. For example, our adjustment has a simple expression and always can be expressed as a closed form, and the distribution function of T can be given easily. On the other hand, $f^{-1}(x)$ cannot be expressed in a simple form because $f(x)$ in Fujikoshi (1997) and Fujisawa (1997) is based on the k -th polynomial, the one in Kakizawa (1996) is the $(2k - 1)$ -th polynomial, and the one in Cordeiro et al. (1998) is based on the distribution function of a normal distribution. From the monotoneity of such transformations p-values of T can be obtained as those of $f(T)$, but it is difficult to describe a distribution function of T because $f^{-1}(x)$ cannot be expressed in a simple form.

3. Some applications

In this section, we obtain the adjusted asymptotic expansions by applying THEOREM 2.1 to some test statistics. Particularly, we make sure that our method is valuable in both normal and nonnormal cases.

EXAMPLE 3.1. *The likelihood ratio statistic for testing equality of variances.*

First, we consider a nonnormal case. Under nonnormality, since the coefficients b_j 's in $P_{ae}(x)$ tend to be large, an undesirable phenomenon as in Figure 1.1 will occur more often by using $P_{ae}(x)$ as the distribution function. Therefore, it is necessary to adjust it as in our method in the nonnormal case.

Let y_{ij} be the j -th sample observation ($j = 1, \dots, n_i$) from the i -th population Π_i ($i = 1, \dots, q$) with mean μ_i and variance σ_i^2 . Under normality, the likelihood ratio statistic for testing the hypothesis $\sigma_1^2 = \dots = \sigma_q^2$ is defined by

$$T_1 = (n - q) \log \frac{S_e}{n - q} - \sum_{i=1}^q (n_i - 1) \log s_i^2,$$

where $n = \sum_{i=1}^q n_i$, $\bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$, $s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ and $S_e = \sum_{i=1}^q (n_i - 1) s_i^2$. Let $m_0 = \kappa_4 + 2$, where κ_4 is kurtosis of y_{ij} . From Tonda and Wakaki (2001), under some conditions on y_{ij} , the asymptotic expansion of the null distribution of $T = 2T_1/m_0$ under nonnormality can be given by

$$P(T \leq x) = G_{q-1}(x) + \frac{1}{n} \sum_{j=0}^3 b_j G_{q-1+2j}(x) + o(n^{-1}),$$

where

$$\begin{aligned} b_0 &= -a_1 + a_2 - a_5, & b_1 &= 3a_1 - 2a_2 - a_4 + a_5, \\ b_2 &= -3a_1 + a_2 - a_3 + a_4, & b_3 &= a_1 + a_3, \end{aligned}$$

and

$$\begin{aligned} a_1 &= \{4m_1^2 - 6qm_1^2 - 3q^2(-2m_0^2 + m_1)^2 + 5m_1^2\|\mathbf{r}^{-1}\|^2\}/24m_0^3, \\ a_2 &= m_2(1 - 2q + \|\mathbf{r}^{-1}\|^2)/8m_0^2, \\ a_3 &= -(m_0^2 - m_1)(-4 + 6q - 5\|\mathbf{r}^{-1}\|^2)/6m_0, \\ a_4 &= -\{(5 + 6q)m_0^2 - 6m_1 + (m_0^2 - 6m_1)\|\mathbf{r}^{-1}\|^2\}/12m_0, \\ a_5 &= \{-2m_0^2 + q(-2 + m_0 + 3m_0^2 - 2m_1) \\ &\quad + m_1 - (-2 + m_0)\|\mathbf{r}^{-1}\|^2\}/2m_0, \end{aligned}$$

and $\|\mathbf{r}^{-1}\|^2 = \sum_{i=1}^q n/n_i$. Then the coefficients $c_{i,j}$ in $AP_{ae}(x)$ can be given by

$$c_{i,j} = \begin{array}{c} i \backslash j \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ 0 & \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & \frac{q+1}{q-1} & \frac{q+3}{q-1} & \frac{q+5}{q-1} \\ 1 & \frac{q+3}{q-1} & \frac{(q+3)(q+5)}{(q-1)(q+1)} & \frac{(q+5)(q+7)}{(q-1)(q+1)} \\ 1 & \frac{q+5}{q-1} & \frac{(q+5)(q+7)}{(q-1)(q+1)} & \frac{(q+5)(q+7)(q+9)}{(q-1)(q+1)(q+3)} \end{array} \right) \\ 1 & \\ 2 & \\ 3 & \end{array} \end{array}.$$

Figure 3.1 shows the graph of $AP_{ae}(x)$, $P_{ae}(x)$, $G_{q-1}(x)$ and the true distribution function which is obtained by Monte Carlo simulation. From the figure, we can see that $AP_{ae}(x)$ solves some problems of $P_{ae}(x)$, that is to exceed 1 and not monotone, and gives a good approximation than the limiting distribution $G_{q-1}(x)$.

EXAMPLE 3.2. *Test criterion for the sphericity of a covariance matrix.*

Next, we consider a normal case. An undesirable phenomenon will happen even under normality.

Let \mathbf{y} be a $p \times 1$ random vector from a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ , and $\mathbf{y}_1, \dots, \mathbf{y}_n$ be n independent observation vectors of \mathbf{y} . We consider the test statistic (Nagao (1973)) for $H_0 : \Sigma = \sigma^2 I_p$ defined by

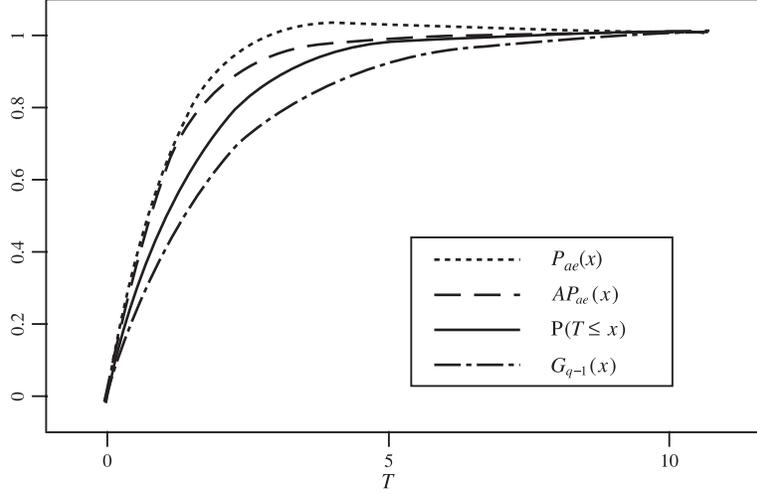


Figure 3.1. Illustration of $P_{ae}(x)$, $AP_{ae}(x)$, $G_{q-1}(x)$ and $P(T \leq x)$

$$T = \frac{p^2(n-1)}{2} \operatorname{tr} \left\{ \frac{1}{\operatorname{tr}(S)} S - \frac{1}{p} I_p \right\}^2,$$

where σ^2 is an unspecified constant, $S = \sum_{j=1}^n (y_j - \bar{y})(y_j - \bar{y})'$ and $\bar{y} = n^{-1} \sum_{j=1}^n y_j$. From Nagao (1973), an asymptotic expansion of the null distribution of T can be expanded as

$$P(T \leq x) = G_r(x) + \frac{1}{n} \sum_{j=0}^3 b_j G_{r+2j}(x) + o(n^{-1}),$$

where

$$b_0 = \frac{1}{24} (-2p^3 - 3p^2 + p + 436p^{-1}),$$

$$b_1 = \frac{1}{4} (p^3 + 2p^2 - p - 2 - 216p^{-1}),$$

$$b_2 = \frac{1}{8} (-2p^3 - 5p^2 + 7p + 12 + 420p^{-1}),$$

$$b_3 = \frac{1}{12} (p^3 + 3p^2 - 8p - 12 - 200p^{-1}),$$

and $r = p(p+1)/2 - 1$. Then the coefficients $c_{i,j}$ in $AP_{ae}(x)$ can be given by

Table 3.1. P-values based on several functions

Quantiles	P-values $\times 10^2$			
	$P(T \geq x)$	$1 - G_r(x)$	$1 - P_{ae}(x)$	$1 - AP_{ae}(x)$
13.43	10.0	14.4	10.4	10.8
15.26	5.00	8.40	5.12	5.51
19.56	1.00	2.08	0.63	0.90
25.67	0.10	0.23	-0.03	0.04

$$c_{i,j} = \begin{matrix} i \setminus j & 0 & 1 & 2 & 3 \\ 0 & \left(\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & \frac{r+2}{r} & \frac{r+4}{r} & \frac{r+6}{r} \\ 2 & 1 & \frac{r+4}{r} & \frac{(r+4)(r+6)}{r(r+2)} & \frac{(r+6)(r+8)}{r(r+2)} \\ 3 & 1 & \frac{r+6}{r} & \frac{(r+6)(r+8)}{r(r+2)} & \frac{(r+6)(r+8)(r+10)}{r(r+2)(r+4)} \end{array} \right) \\ 1 & & & & \\ 2 & & & & \\ 3 & & & & \end{matrix}.$$

Table 3.1 shows p-values which were obtained by using $P(T \geq x)$, $1 - G_r(x)$, $1 - P_{ae}(x)$ and $1 - AP_{ae}(x)$ based on some quantiles. We note that $P_{ae}(x)$ is to exceed 1 in the tail of distribution even if y is distributed as a normal distribution. $AP_{ae}(x)$ corrects such a fault in $P_{ae}(x)$.

4. Simulation study

In this section, we give a simulation study for the EXAMPLE 3.1 in §3. Our interest is to compare with several methods to obtain p-value. We considered the following five nonnormal models and one normal model with $q = 3$ and 5, and each sample size $n_j = 10$ ($1 \leq j \leq q$);

- (i) $X + YZ$, where X , Y and Z are independently distributed as $N(0, 1)$,
- (ii) symmetric uniform distribution $U(-5, 5)$,
- (iii) double exponential distribution $DE(0, 1)$,
- (iv) χ^2 distribution with 3 degrees of freedom,
- (v) χ^2 distribution with 8 degrees of freedom,
- (vi) normal distribution.

As for cumulants in each model, see Tonda and Wakaki (2001).

Tables 4.1 and 4.2 display p-values which were obtained by using $P(T \geq x)$, $1 - G_{q-1}(x)$, $1 - P_{ae}(x)$, $1 - AP_{ae}(x)$ and $1 - P_{it}(x)$ based on same quantiles. In this study, we use a monotone function $f(x)$ in Kakizawa (1996), which is defined by

Table 4.1. P-values in the case $q = 3$.

Models	Quantiles	P-value $\times 10^2$				
		$P(T \geq x)$	$1 - G_{q-1}(x)$	$1 - P_{ae}(x)$	$1 - AP_{ae}(x)$	$1 - P_{it}(x)$
Model (i) ($X + YZ$)	3.805	10.0	15.0	3.50	5.70	6.43
	4.929	5.00	8.54	1.04	2.74	3.25
	7.480	1.00	2.40	0.28	0.88	0.91
	10.93	0.10	0.45	0.33	0.44	0.30
Model (ii) ($U(-5, 5)$)	6.254	10.0	4.22	8.33	9.48	10.3
	8.501	5.00	1.34	3.42	4.30	5.43
	14.39	1.00	0.07	0.34	0.59	1.66
	23.50	0.10	0.00	0.01	0.02	0.79
Model (iii) ($DE(0, 1)$)	3.701	10.0	16.2	4.74	6.86	7.45
	4.738	5.00	9.72	1.18	3.11	3.72
	7.038	1.00	3.25	-0.60	0.55	0.89
	10.07	0.10	0.57	-0.24	0.05	0.12
Model (iv) (χ_3^2)	3.281	10.0	18.9	-1.59	4.34	5.31
	4.205	5.00	12.0	-2.78	2.20	2.87
	6.248	1.00	4.09	-1.16	1.05	0.93
	8.772	0.10	1.04	0.35	0.88	0.40
Model (v) (χ_8^2)	3.852	10.0	14.5	5.96	7.20	7.68
	4.972	5.00	8.11	2.20	3.26	3.71
	7.453	1.00	2.26	0.12	0.62	0.82
	10.97	0.10	0.41	0.01	0.11	0.14
Model (vi) ($N(0, 1)$)	4.831	10.0	8.58	9.51	9.54	9.56
	6.260	5.00	4.15	4.74	4.76	4.78
	9.606	1.00	0.75	0.91	0.92	0.93
	14.54	0.10	0.06	0.08	0.09	0.09

$$f(x) = x - \frac{2x}{n(q-1)}(d_1 + d_2x + d_3x^2) + \frac{x}{n^2(q-1)^2} \left\{ d_1^2 + 2d_1d_2x + \frac{2}{3}(2d_2^2 + 3d_1d_3)x^2 + 3d_2d_3x^3 + \frac{9}{5}d_3^2x^4 \right\},$$

where

$$d_1 = -b_0, \quad d_2 = \frac{(b_2 + b_3)}{(q+1)}, \quad d_3 = \frac{b_3}{(q+1)(q+3)}.$$

From the tables, we can see that $AP_{ae}(x)$ improves a fault of $P_{ae}(x)$, that is to exceed 1. Moreover, p-values based on $1 - AP_{ae}(x)$ shows the similar

Table 4.2. P-values in the case $q = 5$.

Models	Quantiles	P-value $\times 10^2$				
		$P(T \geq x)$	$1 - G_{q-1}(x)$	$1 - P_{ae}(x)$	$1 - AP_{ae}(x)$	$1 - P_{it}(x)$
Model (i) ($X + YZ$)	6.555	10.0	16.5	2.38	5.61	6.19
	8.024	5.00	9.24	1.10	3.31	3.40
	11.25	1.00	2.45	1.19	1.79	1.27
	16.17	0.10	0.35	0.76	1.06	0.67
Model (ii) ($U(-5, 5)$)	10.75	10.0	2.98	7.27	8.84	10.7
	13.56	5.00	0.89	2.82	3.88	5.94
	20.32	1.00	0.04	0.23	0.45	2.01
	30.44	0.10	0.00	0.00	0.01	1.10
Model (iii) ($DE(0, 1)$)	6.295	10.0	17.8	2.93	5.92	6.93
	7.613	5.00	10.6	0.26	2.73	3.56
	10.40	1.00	3.24	-0.37	0.68	0.95
	14.23	0.10	0.62	0.10	0.26	0.24
Model (iv) (χ_3^2)	5.683	10.0	22.5	-4.82	4.80	4.95
	6.919	5.00	14.2	-4.12	3.38	2.93
	9.560	1.00	4.70	0.08	2.76	1.30
	13.27	0.10	0.89	1.62	2.45	0.89
Model (v) (χ_8^2)	6.611	10.0	16.1	5.18	6.90	7.63
	8.036	5.00	9.24	2.08	3.41	3.96
	11.33	1.00	2.49	0.48	0.93	1.04
	15.26	0.10	0.41	0.27	0.32	0.26
Model (vi) ($N(0, 1)$)	8.109	10.0	8.87	10.0	10.0	10.1
	9.906	5.00	4.42	5.14	5.17	5.19
	13.85	1.00	0.82	1.01	1.03	1.04
	19.08	0.10	0.08	0.10	0.11	0.11

performances as $1 - P_{it}(x)$. We remark that, in the case of this test statistic, $AP_{ae}(x)$ tends to be conservative, that is, $P(T \geq x) \geq 1 - AP_{ae}(x)$.

Acknowledgement

The authors would like to thank Prof. Y. Fujikoshi and Dr. H. Wakaki, Hiroshima University, for several useful comments and suggestions.

References

- [1] G. M. Cordeiro, S. L. P. Ferrari, A. H. M. A. Cysneiros, A formula to improve score test statistic, *J. Statist. Comput. Simulation* **62** (1998), 123–136.

- [2] Y. Fujikoshi, A method for improving the large-sample chi-squared approximations to some multivariate analysis, *Amer. J. Math. Management Sci.* **17** (1997), 15–29.
- [3] H. Fujisawa, Improvement on chi-squared approximation by monotone transformation, *J. Multivariate Anal.* **60** (1997), 84–89.
- [4] Y. Kakizawa, Higher order monotone Bartlett type adjustment for some multivariate test statistics, *Biometrika* **83** (1996), 923–927.
- [5] H. Nagao, On some test criteria for covariance matrix, *Ann. Statist.* **1** (1973), 700–709.
- [6] T. Tonda and H. Wakaki, Asymptotic expansion of the null distribution of the likelihood ratio statistic for testing the equality of variances in a nonnormal one-way ANOVA model, TR No. 01–05, Statistical Research Group, Hiroshima University, 2001.

Hirokazu Yanagihara

*Department of Statistical Methodology
The Institute of Statistical Mathematics
Tokyo 106-8569, Japan
e-mail: hyanagi@ism.ac.jp*

Tetsuji Tonda

*Department of Mathematics
Graduate school of Science
Hiroshima University
Higashi-Hiroshima 739-8526, Japan
e-mail: ttetsuji@math.sci.hiroshima-u.ac.jp*