AISM (2007) 59: 531–556
DOI 10.1007/s10463-006-0055-7

Solomon W. Harrar · Arjun K. Gupta

Asymptotic expansion for the null distribution of the F-statistic in one-way ANOVA under non-normality

Received: 9 February 2004 /
Published online: 17 June 2006

Abstract In this paper we derive the asymptotic expansion of the null distribution of the F-statistic in one-way ANOVA under non-normality. The asymptotic framework is when the number of treatments is moderate but sample size per treatment (replication size) is small. This kind of asymptotics will be relevant, for example, to agricultural screening trials where large number of cultivars are compared with few replications per cultivar. There is also a huge potential for the application of this kind of asymptotics in microarray experiments. Based on the asymptotic expansion we will devise a transformation that speeds up the convergence to the limiting distribution. The results indicate that the approximation based on limiting distribution are unsatisfactory unless number of treatments is very large. Our numerical investigations reveal that our asymptotic expansion performs better than other methods in the literature when there is skewness in the data or even when the data comes from a symmetric distribution with heavy tails.

Keywords Analysis of variance, Edgeworth expansion, Cumulants, Characteristic function, Asymptotic expansion, Non-normality

1 Introduction

In the univariate one-way ANOVA model the response on the jth replication of the ith treatment group can be described as,

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}; \quad i = 1, \cdots, k \quad \text{and } j = 1, \cdots, n_i,$$

S.W. Harrar
Department of Mathematics and Statistics, South Dakota State University,
Brookings, SD 57007, USA

A.K. Gupta (✉)
Department of Mathematics and Statistics, Bowling Green State University,
Bowling Green, OH 43403, USA
E-mail: gupta@bgnet.bgsu.edu

where μ is the general mean effect, α_i is the ith treatment effect and ε_{ij} is iid with mean 0 and variance σ^2.

Let $S_h = 1/(k-1)\sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$ and $S_e = 1/(n-k)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(n_i - 1)S_i$, where $\bar{y}_{i.} = 1/n_i\sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{..} = 1/n\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$, $n = \sum_{i=1}^{k} n_i$ and $S_i = 1/(n_i-1)\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$, be the between and within group mean squares.

It is well known that when $\varepsilon_{ij} \sim N(0,\sigma^2)$ the statistic $F = S_h/S_e$ provides UMPU and UMPI(with respect to linear transformation) test for the hypothesis $H_0 : \alpha_i = 0; i = 1, \ldots, k$ and it has $\mathcal{F}(k-1, n-k)$ distribution. Here $\mathcal{F}(k-1, n-k)$ means F distribution with $k-1$ and $n-k$ degrees of freedom. On the other hand it is known that F converges in distribution to χ^2_{k-1} when $n \to \infty$ even when the errors do not have normal distribution. In a recent paper, Fujikoshi, Ohmae and Yanagihara (1999) took this limiting distribution result further by giving asymptotic expansion up to the order $o(n^{-1})$ in terms of central χ^2 distribution. The extension to univariate linear model was done by Yanagihara (2003). In a same context, Yanagihara (2000) derived the asymptotic expansion of James' (1954) test statistic suited for ANOVA under heterogeneous variances. All these results are based the assumption that Huber's condition (Huber, 1973) holds. In the one way ANOVA model, Huber's conditions reduces to $n/n_i = O(1)$. This condition means that all the replication sizes n_i are large. For the same testing problem, Fisher and Hall (1990) provide a bootstrap method for obtaining the critical values which they claim to lead to a test size accurate to the order $O(n^{-2})$. However, it is not clear whether Huber's condition is needed or not. Their simulation results do not provide any evidence of adequate performance when Huber's condition is not met.

On the other hand, it is of interest to know the distribution of F when k is large and n_i's (replications) are small and the distribution of ε_{ij} is not known. In agricultural screening trials it is quite common to meet a situation where large number of cultivars are compared with few replications per cultivar. For instance, a perennial ryegrass cultivar screening trial, conducted by the Nova Scotia Department of Agriculture, Canada (http://www.gov.ns.ca/nsaf), involved the screening of sixteen forage type cultivars for winter hardiness and disease resistance in a completely randomized design with four blocks. Data from such experiments are characterized by fewer replications and the presence of extreme observations, as a result the assumption of normality is unlikely to hold. For more examples see Akritas and Arnold(2000) and Brownie and Boos(1994). Needless to say, the condition $n/n_i = O(1)$, commonly known as Huber condition, fails in the asymptotic framework of this paper.

In this large number of treatments setup, the limiting null distribution of F in balanced one-way and two-way (CRBD) designs was shown to be normal by Boos and Brownie (1995). They have also noted that the null distribution of the F-statistic is invariant to the distribution of the data. This result was generalized to one-way and two-way fixed, random and mixed effects models by Akritas and Arnold (2000). In addition, Akritas and Arnold's (2000) results show that the distribution invariance of F-statistic does not hold in the unbalanced and non-null cases. Bathke (2002) derived similar results in the null case for m-way balanced ANOVA model for any fixed integer $m \geq 1$. All these results were extended to MANOVA models by Gupta, Harrar and Fujikoshi (2005, 2006).

For the balanced univariate one–way fixed effects model, the limiting distribution results cited in the previous paragraph assert that the distribution of $\sqrt{k}(F-1)$ converges to $N(0, 2n/(n-1))$ as $k \to \infty$. When the data come from a normal distribution this result has the implication that F distribution can be approximated by normal distribution when both the numerator and denominator degrees of freedom are large. Table 1, however, clearly shows that the discrepancy between the quantiles obtained from F and its normal approximation can be considerably large. The approximated 5% quantiles are all outside of 5% of the actual values. The situation is much worse for 1% quantiles. Numerical results also show that (see Sect. 6) k has to be very large for the limiting distribution to give reasonable approximation under non-normality. In other words, the rate of convergence to the limiting distribution is rather slow. In the more practical situation where k is moderate the actual test size from the limiting distribution will differ significantly from the desired size. Hence in such situations some finiteness adjustment (correction) must be made to the limiting distribution to get a reasonable approximation for the distribution of F-statistic.

It is the aim of this paper to provide such correction by accounting for the effects of skewness and kurtosis present in the data. This is achieved by including terms of order up to $1/k$ in the balanced case and up to $1/\sqrt{k}$ in the unbalanced case, in the asymptotic expansion of the distribution of F. In Sect. 2 we derive Edgeworth expansions for some statistics needed in the subsequent sections. We take up the problem of deriving the asymptotic expansion of the joint distribution of $\sqrt{k}(S_h - 1)$ and $\sqrt{k}(S_e - 1)$ in Sect. 3. The main result of the paper, namely the asymptotic expansion of $\sqrt{k}(F - 1)$ will be given in Sect. 4 for the balanced case. Based on the asymptotic expansion, we will devise a transformation on F which improves the rate of convergence to the limiting normal distribution. Such a transformation is known as normalization transformation. The more general case, the unbalanced case, will be treated in Sect. 5. We will assess the gain in improvement from the asymptotic expansion through a simulation study which will be the subject of Sect. 6. Numerical example will also be given in that section to illustrate the application of our results. We make some concluding remarks in Sect. 7.

2 Edgeworth expansions

In this section we derive the Edgeworth expansion for the joint distribution of some useful statistics. For the sake of convenience we assume $\sigma^2 = 1$. This does not

Table 1 Comparison of quantiles of F and its normal approximations (f_1 and f_2 are numerator and denominator degrees of freedoms, respectively.)

f_1	f_2	5%		1%	
$k-1$	$k(n-1)$	$F_\alpha - Z_\alpha$	$\frac{F_\alpha - Z_\alpha}{F_\alpha}$ (%)	$F_\alpha - Z\alpha$	$\frac{F_\alpha - Z\alpha}{F_\alpha}$ (%)
9	40	0.30	14	0.72	25
14	60	0.19	10	0.44	19
19	80	0.14	8	0.32	15
24	100	0.11	7	0.25	12
29	120	0.09	6	0.20	11
34	140	0.07	5	0.17	9

cause loss of generality because F-statistic is invariant to scale transformation. Instructively, we consider the case $n_1 = \cdots = n_k = n$ first. We take up the more general case in Sect. 5.

Let us define,

$$Z = \sqrt{k}\bar{y}_{..}, \quad V = \sqrt{k}(S_h - 1) \quad \text{and } W = \sqrt{k}(S_e - 1). \tag{1}$$

To find the asymptotic expansion of the joint distribution of (Z, V, W), we would like, first, to define a new random variable $\tilde{V} = \sqrt{k}\left((1/k)\sum_{i=1}^{k} n\bar{y}_{i.}^2 - 1\right)$ such that $(Z, \tilde{V}, W)$ can be expressed as a localized sum of identically and independently distributed random vectors. It is an easy matter to check that,

$$V = \tilde{V} + \frac{1}{\sqrt{k}}(1 - nZ^2) + \frac{1}{k}\tilde{V} + o_p\left(\frac{1}{k}\right). \tag{2}$$

In what follows we give Edgeworth expansion for the distribution of $(Z, \tilde{V}, W)$. Let us put $\mathbf{x}_i = \left(\bar{y}_{i.}, n\bar{y}_{i.}^2, S_i\right)$. It is obvious that $E(\mathbf{x}_i) = (0, 1, 1)'$ and $\mathbf{x}_i$'s are identically and independently distributed. Moreover, it is noted that the joint characteristic function of $(Z, \tilde{V}, W) = 1/\sqrt{k}\sum_{i=1}^{k}(\mathbf{x}_i - E(\mathbf{x}_i))$ admits asymptotic expansion in powers of $k^{-1/2}$ (Bhattacharya and Rao, 1976; Hall, 1992) and the expansion is given below.

Lemma 2.1 *Under the assumption,*

$$A1 : E\left(\varepsilon_{11}^8\right) < \infty$$

the characteristic function of $(Z, \tilde{V}, W)$ *can be expanded as*

$$C_{(Z,\tilde{V},W)}(t_1, t_2, t_3) = \exp\left\{\frac{i^2}{2}\sum_{a,b}^{3} K_{ab}t_at_b\right\}\left[1 + \frac{i^3}{6\sqrt{k}}\sum_{a,b,c}^{3} K_{abc}t_at_bt_c \right. \tag{3}$$

$$+\frac{i^4}{24k}\sum_{a,b,c,d}^{3} K_{abcd}t_at_bt_ct_d \tag{4}$$

$$\left. +\frac{i^6}{72k}\sum_{a,b,c,d,e,f}^{3} K_{abc}K_{def}t_at_bt_ct_dt_et_f + o\left(\frac{1}{k}\right)\right], \tag{5}$$

where K_{ab}, K_{abc} *and* K_{abcd} *are, respectively, the second, third and fourth order cumulants of* $\mathbf{x}_i$ *defined by,*

$$\begin{aligned} K_{ab} &= \mu_{ab}, \quad K_{abc} = \mu_{abc}, \quad K_{abcd} = \mu_{abcd} - \mu_{ab}\mu_{cd} - \mu_{ac}\mu_{bd} - \mu_{ad}\mu_{bc} \\ \mu_a &= E[x_{ia}], \quad \mu_{ab} = E\left[(x_{ia} - \mu_a)(x_{ib} - \mu_b)\right] \\ \mu_{abc} &= E\left[(x_{ia} - \mu_a)(x_{ib} - \mu_b)(x_{ic} - \mu_c)\right] \textit{and} \\ \mu_{abcd} &= E[(x_{ia} - \mu_a)(x_{ib} - \mu_b)(x_{ic} - \mu_c)(x_{id} - \mu_d)], \end{aligned}$$

where x_{ia} *is the ath entry of* $\mathbf{x}_i$.

As it happens our approach depends only on some of the K_{ab}, K_{abc} and K_{abcd}. In the following lemma we compute those we need.

Lemma 2.2 *Let K_{ab}, K_{abc} and K_{abcd} be the second, third and fourth order mixed cumulants of $\mathbf{x}_i$. Then*

$$
\begin{aligned}
K_{11} &= \frac{1}{n}, \quad K_{22} = \frac{1}{n}\kappa_4 + 2, \quad K_{33} = \frac{1}{n}\kappa_4 + \frac{2}{n-1}, \\
K_{12} &= K_{13} = \frac{1}{n}\kappa_3, \quad K_{23} = \frac{1}{n}\kappa_4, \quad K_{111} = \frac{1}{n}\kappa_3, K_{113} = \frac{1}{n^2}\kappa_4 \\
K_{112} &= \frac{1}{n^2}\kappa_4 + \frac{2}{n}, \quad K_{122} = \frac{1}{n^2}\kappa_5 + \frac{8}{n}\kappa_3, \quad K_{133} = \frac{1}{n^2}\kappa_5 + \frac{4}{n(n-1)}\kappa_3 \\
K_{123} &= \frac{1}{n^2}\kappa_5 + \frac{2}{n}\kappa_3, \quad K_{222} = \frac{1}{n^2}\kappa_6 + \frac{12}{n}\kappa_4 + \frac{10}{n}\kappa_3^2 + 8, \\
K_{223} &= \frac{1}{n^2}\kappa_6 + \frac{4}{n}\kappa_4 + \frac{4}{n}\kappa_3^2, \quad K_{233} = \frac{1}{n^2}\kappa_6 + \frac{4}{n(n-1)}\kappa_4 + \frac{2(n+1)}{n(n-1)}\kappa_3^2 \\
K_{333} &= \frac{1}{n^2}\kappa_6 + \frac{12}{n(n-1)}\kappa_4 + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2} \\
K_{2222} &= \frac{1}{n^3}\kappa_8 + \frac{24}{n^2}\kappa_6 + \frac{56}{n^2}\kappa_5\kappa_3 + \frac{32}{n^2}\kappa_4^2 + \frac{144}{n}\kappa_4 + \frac{240}{n}\kappa_3^2 + 48 \\
K_{3333} &= \frac{1}{n^3}\kappa_8 + \frac{24}{n^2(n-1)}\kappa_6 + \frac{32(n-2)}{n^2(n-1)^2}\kappa_5\kappa_3 + \frac{8(4n^2-9n+6)}{n^2(n-1)^3}\kappa_4^2 \\
&\quad + \frac{144}{n(n-1)^2}\kappa_4 + \frac{96(n-2)}{n(n-1)^3}\kappa_3^2 + \frac{48}{(n-1)^3} \\
K_{2333} &= \frac{1}{n^3}\kappa_8 + \frac{12}{n^2(n-1)}\kappa_6 + \frac{6n^2+20n-34}{n^2(n-1)^2}\kappa_5\kappa_3 + \frac{20n-28}{n^2(n-1)^2}\kappa_4^2 \\
&\quad + \frac{24}{n(n-1)^2}\kappa_4 + \frac{24(n+1)}{n(n-1)^2}\kappa_3^2 \\
K_{2233} &= \frac{1}{n^3}\kappa_8 + \frac{4}{n(n-1)}\kappa_6 + \frac{12n+4}{n^2(n-1)}\kappa_5\kappa_3 + \frac{4n+8}{n^2(n-1)}\kappa_4^2 \\
&\quad + \frac{8n+24}{n(n-1)}\kappa_3^2 - \frac{(n+1)}{(n-1)} \quad \textit{and} \\
K_{2223} &= \frac{1}{n^3}\kappa_8 + \frac{12}{n^2}\kappa_6 + \frac{26}{n^2}\kappa_5\kappa_3 + \frac{12}{n^2}\kappa_4^2 + \frac{24}{n}\kappa_4 + \frac{48}{n}\kappa_3^2,
\end{aligned}
$$

where κ_r is the rth cumulant of y_{11} ($\kappa_1 = 0$ and $\kappa_2 = 1$).

Proof Let $k_1 = \bar{y}_{1.}$ and $k_2 = S_1^2$. These are known as the first two Fisher's k-statistics. Write K_{abcd} in terms of the raw moments nk_1^2 and k_2. Next express $k_1^{2a}k_2^b$, in terms of polykays $k_{rst\ldots}$ (David, Kendall, and Barton, 1966, Table 2.3, pp. 196–200). Using the key property of polykays, $E(k_{rst\ldots}) = \kappa_r\kappa_s\kappa_t\ldots$, the desired results follow after quite a bit of algebra. □

The results of the Lemma have also been checked by using the formulae of Tan and Cheng (1981) for computing mixed cumulants of linear and quadratic forms.

To find the asymptotic expansion for the distribution of $\sqrt{k}(F-1)$ our approach proceeds as follows. First we expand the test statistic as a function of $(Z, \tilde{V}, W)$. This is followed by an expansion of the characteristic function and formal inversion of the expanded characteristic function. In this process, the following Lemma plays a key role.

Lemma 2.3 *Under the assumption* A1, *we have the following expansion results.*

1. *The characteristic function of* $\tilde{V} - W$ *can be expanded as,*

$$C_{\tilde{V}-W}(t) = \exp\left\{\frac{i^2}{2}\tau t^2\right\}\left[1 + \frac{(it)^3}{6\sqrt{k}}b_0 + \frac{(it)^4}{24k}b_1 + \frac{(it)^6}{72k}b_0^2 + o\left(\frac{1}{k}\right)\right] \tag{6}$$

where

$$\begin{aligned} b_0 &= \frac{4n}{(n-1)^2}\kappa_3^2 + \left(8 - \frac{8}{(n-1)^2}\right) \\ b_1 &= \frac{48}{n}\left[1 + \frac{1}{(n-1)^2}\right]\kappa_4^2 + \frac{96}{n(n-1)}\left[n - \frac{1}{n-1}\right]\kappa_3^2 \\ &\quad + \left[48 + \frac{48}{(n-1)^3} - \frac{6(n+1)}{n-1}\right] \\ \tau &= \frac{2n}{n-1} \end{aligned}$$

2. *The joint characteristic function of* Z *and* $\tilde{V} - W$ *can be expanded as,*

$$\begin{aligned} C_{Z,\tilde{V}-W}(t_1, t) &= \exp\left\{\frac{i^2}{2}\left(\frac{1}{n}t_1^2 + \tau t^2\right)\right\} \\ &\quad \times\left[1 + \frac{i^3}{6\sqrt{k}}\left[\frac{\kappa_3}{n^2}t_1^3 + \frac{6}{n}t_1^2 t + \frac{12}{n-1}\kappa_3 t_1 t^2 + b_0 t^3\right] + o\left(\frac{1}{\sqrt{k}}\right)\right]. \end{aligned} \tag{7}$$

3. *The joint characteristic function of* $\tilde{V}$ *and* $\tilde{V} - W$ *can be expanded as,*

$$C_{\tilde{V},\tilde{V}-W}(t_1, t) = \exp\left\{\frac{i^2}{2}\left[\left(\frac{1}{n}\kappa_4 + 2\right)t_1^2 + 4t_1 t + \tau t^2\right]\right\} + o(1). \tag{8}$$

4. *The joint characteristic function of* V *and* W *can be expanded as,*

$$\begin{aligned} C_{V,W}(t_1, t_2) &= \exp\left\{\frac{i^2}{2}\left[\left(\frac{1}{n}\kappa_4 + 2\right)t_1^2 + \frac{2}{n}\kappa_4 t_1 t_2 + \left(\frac{1}{n}\kappa_4 + \frac{2}{n-1}\right)t_2^2\right]\right\} \\ &\quad \times\left[1 + \frac{i^3}{\sqrt{k}}\left[l_1 t_1^3 + l_2 t_1^2 t_2 + l_3 t_1 t_2^2 + l_4 t_2^3\right] + o\left(\frac{1}{\sqrt{k}}\right)\right], \end{aligned} \tag{9}$$

where

$$l_1 = \frac{1}{6}\left(\frac{1}{n^2}\kappa_6 + \frac{12}{n}\kappa_4 + \frac{4}{n}\kappa_3^2 + 8\right)$$
$$l_2 = \frac{1}{2}\left(\frac{1}{n^2}\kappa_6 + \frac{4}{n}\kappa_4\right)$$
$$l_3 = \frac{1}{2}\left(\frac{1}{n^2}\kappa_6 + \frac{4}{n(n-1)}\kappa_4 + \frac{4}{n(n-1)}\kappa_3^2\right)$$
$$l_4 = \frac{1}{6}\left(\frac{1}{n^2}\kappa_6 + \frac{12}{n(n-1)}\kappa_4 + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}\right)$$

5. *The joint characteristic function of* W *and* $V - W$ *can be expanded as,*

$$C_{W,V-W}(t_1,t) = \exp\left\{\frac{i^2}{2}\left[\tau t^2 - \frac{4}{n-1}t_1 t + \left(\frac{1}{n}\kappa_4 + \frac{2}{n-1}\right)t_1^2\right]\right\} \times\left[1 + \frac{i^3}{\sqrt{k}}\left(m_1 t^3 + m_2 t_1 t^2 + m_3 t_1^2 t + m_4 t_1^3\right) + o\left(\frac{1}{\sqrt{k}}\right)\right] \tag{10}$$

where

$$m_1 = \frac{4n}{6(n-1)^2}\kappa_3^2 + \frac{8}{6}\left(1 - \frac{1}{(n-1)^2}\right)$$
$$m_2 = \frac{2}{n-1}\kappa_4 - \frac{2}{(n-1)^2}\kappa_3^2 + \frac{4}{(n-1)^2}$$
$$m_3 = -\frac{4}{n^2(n-1)}\kappa_4 + \frac{2}{n^2(n-1)^2}\kappa_3^2 - \frac{4}{n(n-1)^2}$$
$$m_4 = \frac{1}{6n^2}\kappa_6 + \frac{2}{n(n-1)}\kappa_4 + \frac{4(n-2)}{6n(n-1)^2}\kappa_3^2 + \frac{8}{6(n-1)^2}.$$

Proof It is clear that $C_{\tilde{V}-W}(t) = C_{Z,\tilde{V},W}(0,t,-t)$. Then (6) follows by applying Lemmas 2.1 and 2.2, and some algebra. Along similar lines, (7) and (8) follow by noting that $C_{Z,\tilde{V}-W}(t_1,t) = C_{Z,\tilde{V},W}(t_1,t,-t)$ and $C_{\tilde{V},\tilde{V}-W}(t_1,t) = C_{Z,\tilde{V},W}(0,t_1+t,-t)$. In order to establish (9), observe that

$$C_{Z,V,W}(t,t_1,t_2) = \left[1 + \frac{i}{\sqrt{k}}t_1\left(1 + n\frac{\partial^2}{\partial t^2}\right) + o\left(\frac{1}{\sqrt{k}}\right)\right]C_{Z,\tilde{V},W}(t,t_1,t_2). \tag{11}$$

To see this using (2), first write,

$$C_{(Z,V,W)}(t,t_1,t_2) = E\left(\mathrm{e}^{itZ+it_1\tilde{V}+it_2W}\mathrm{e}^{it_1\frac{1}{\sqrt{k}}(1-nZ^2)+o_p(\frac{1}{\sqrt{k}})}\right).$$

Then expanding the second exponential in power series and taking expectation yields (11). Hence, (9) follows by setting $t = 0$. Finally, (10) follows from the fact that $C_{W,V-W}(t_1,t) = C_{V,W}(t,t_1-t)$. □

3 Asymptotic expansion of the joint distribution of S_h and S_e

In this section we derive asymptotic expansion for the joint distribution of S_h and S_e. Knowledge of this distribution may be useful, for example, in assessing the performance of estimators of σ^2 which are functions of both S_h and S_e.

It is apparent from (9) that the limiting distribution function of (V, W) is bivariate normal with mean (0, 0) and covariance Γ given by,

$$\Gamma = \begin{pmatrix} K_{22} & K_{23} \\ K_{23} & K_{33} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\kappa_4 + 2 & \frac{1}{n}\kappa_4 \\ \frac{1}{n}\kappa_4 & \frac{1}{n}\kappa_4 + \frac{2}{n-1} \end{pmatrix}. \tag{12}$$

By inverting (9) term by term, one gets the expansion of joint density of (V, W) up to the order $k^{-1/2}$. However, we need the following condition (see Hall, 1992, p. 78; Bhattacharya and Rao, 1976, p. 199),

$$A2 : \text{There exists an } r \geq 1 \text{ such that} \int_{\mathbb{R}^3} |C_{\bar{y}_{1.}, n\bar{y}_{1.}^2, S_1}(t_1, t_2, t_3)|^r \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 < \infty$$

for uniform validity of the density expansion.

We find the inversion formula,

$$\begin{aligned}(2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} & \mathrm{e}^{-it_2 v - it_3 w} (-it_2)^{\alpha_2} (-it_3)^{\alpha_3} \mathrm{e}^{\frac{i^2}{2} \sum_{a,b=2}^{3} K_{ab} t_a t_b} \mathrm{d}t_2 \mathrm{d}t_3 \\ &= \left(\frac{\partial}{\partial v}\right)^{\alpha_2} \left(\frac{\partial}{\partial w}\right)^{\alpha_3} \phi_{(0,\Gamma)}(v, w), \end{aligned} \tag{13}$$

handy to invert (9), where $\phi_{(0,\Gamma)}$ is the density of a bivariate normal distribution with mean 0 and covariance Γ, and α_2 and α_3 are nonnegative integers.

Finally we get the expansion of the joint density of (V, W) as summarized in the following Theorem.

Theorem 3.1 *Under the assumptions A1 and A2, the joint probability density function of (V,W) can be expanded as,*

$$f(v, w) = \phi_2(v, w; 0, \Gamma) \left[1 + \frac{n}{\sqrt{k}} g_{(1)}(v, w) + o\left(\frac{1}{\sqrt{k}}\right)\right] \tag{14}$$

where

$$
\begin{aligned}
g_{(1)}(v,w) &= \frac{1}{6}\left[\frac{1}{n^3}\kappa_6 + \frac{12}{n^2}\kappa_4 + \frac{4}{n^2}\kappa_3^2 + \frac{8}{n}\right] g_1(v,w) \\
&\quad + \frac{1}{2}\left[\frac{1}{n^3}\kappa_6 + \frac{4}{n^2}\kappa_4\right] g_2(v,w) \\
&\quad + \frac{1}{2}\left[\frac{1}{n^3}\kappa_6 + \frac{4}{n(n-1)}\kappa_4 + \frac{4}{n(n-1)}\kappa_3^2\right] g_3(v,w) \\
&\quad + \frac{1}{6}\left[\frac{1}{n^3}\kappa_6 + \frac{12}{n^2(n-1)}\kappa_4 + \frac{4(n-2)}{n^2(n-1)^2}\kappa_3^2 + \frac{8}{n(n-1)^2}\right] g_4(v,w) \\
g_1(v,w) &= \xi^{-6}\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right]^3 \\
&\quad - 3\xi^{-4}\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right]\left(\frac{1}{n}\kappa_4 + \frac{2}{n-1}\right) \\
g_2(v,w) &= \xi^{-6}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right]^2 \\
&\quad - \xi^{-4}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]\left(\frac{1}{n}\kappa_4 + \frac{2}{n-1}\right) \\
&\quad + 2\xi^{-4}\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right]\left(\frac{1}{n}\kappa_4\right) \\
g_3(v,w) &= \xi^{-6}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]^2\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right] \\
&\quad - \xi^{-4}\left[\frac{1}{n}(\kappa_4+2)t_0 + \frac{2}{n-1}u\right]\left(\frac{1}{n}\kappa_4 + 2\right) \\
&\quad + 2\xi^{-4}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]\left(\frac{1}{n}\kappa_4\right) \\
g_4(v,w) &= \xi^{-6}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]^3 \\
&\quad - 3\xi^{-4}\left[2u - \frac{1}{n}(\kappa_4+2)t_0\right]\left(\frac{1}{n}\kappa_4 + 2\right) \\
\xi^2 &= \frac{2}{n-1}(\kappa_4+2),\ t_0 = v - w \ \textit{ and } \ u = \frac{n-1}{n}w + \frac{1}{n}v.
\end{aligned}
$$

Proof The Theorem follows from (9) and (13), and some algebra. □

The notations t_0 and u are introduced for the important reason that they are transformations which lead to asymptotic independence. This fact will be exploited in Sect. 4.

It is clear from Theorem 3.1 that S_h and S_e are not asymptotically uncorrelated unless $\kappa_4 = 0$. It can also be shown that the correlation between S_h and S_e does not depend on κ_3. It is apparent from Fig. 1 that the correlations reach their

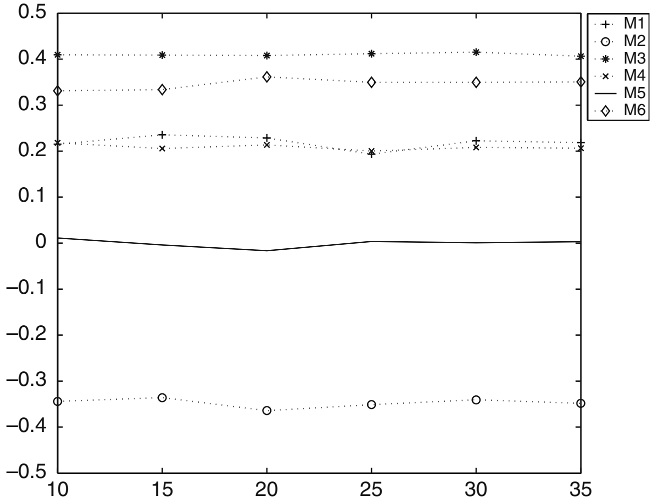

Fig. 1 Empirical correlations between S_h and S_e for models $M1$–$M6$ (See Sect. 6 for description of the models $M1$–$M6$ in the figure.)

limiting values pretty quickly and, hence, the influence of k seems to be not considerable. The figure also reveals that the strength of correlation increases with the the absolute magnitude of κ_4.

4 Asymptotic expansion of the distribution of the test statistic

As mentioned earlier, we first expand the test statistic $F = S_h/S_e$ in terms of V and W leaving the error to the order $k^{-3/2}$. Based on this expansion, we derive an expansion for the characteristic function of $\sqrt{k}(F-1)$. Then by inverting the expanded characteristic function term by term our final result follows.

4.1 Expansion of characteristic function of $\sqrt{k}(F-1)$

The statistic $\sqrt{k}(F-1)$ can be expanded as,

$$\sqrt{k}(F-1) = (V-W) - \frac{1}{\sqrt{k}}W(V-W) + \frac{1}{k}W^2(V-W) + o_p\left(\frac{1}{k}\right). \quad (15)$$

It is not difficult to see from (15) that the characteristic function of $\sqrt{k}(F-1)$ can be expanded as,

$$C_{\sqrt{k}(F-1)}(t) = C_0(t) + \frac{1}{\sqrt{k}}C_1(t) + \frac{1}{k}C_2(t) + o\left(\frac{1}{k}\right),$$

where

$$\begin{aligned}
C_0(t) &= E[\mathrm{e}^{it(V-W)}]\\
C_1(t) &= E[(-it)W(V-W)\mathrm{e}^{it(V-W)}]\\
C_2(t) &= E\left[\left\{(it)W^2(V-W)+\frac{1}{2}(it)^2W^2(V-W)^2\right\}\mathrm{e}^{it(V-W)}\right].
\end{aligned}$$

4.1.1 Evaluation of $C_0(t)$

Using (2), the statistic $V-W$ can be expressed as,

$$V-W=(\tilde{V}-W)+\frac{1}{\sqrt{k}}(1-nZ^2)+\frac{1}{k}\tilde{V}+o_p\left(\frac{1}{k}\right).$$

As a result, the characteristic function of $V-W$ can be expanded as,

$$C_{V-W}(t)=B_0(t)+\frac{1}{\sqrt{k}}B_1(t)+\frac{1}{k}B_2(t)+o\left(\frac{1}{k}\right), \tag{16}$$

where

$$B_0(t)=E\left[\mathrm{e}^{it(\tilde{V}-W)}\right] \tag{17}$$

$$B_1(t)=E\left[(it)(1-nZ^2)\mathrm{e}^{it(\tilde{V}-W)}\right] \tag{18}$$

$$B_2(t)=E\left[\left\{(it)\tilde{V}+\frac{1}{2}(it)^2(1-nZ^2)^2\right\}\mathrm{e}^{it(\tilde{V}-W)}\right]. \tag{19}$$

Directly from (6) we get,

$$B_0(t)=\exp\left\{\frac{i^2}{2}\tau t^2\right\}\left[1+\frac{(it)^3}{6\sqrt{k}}b_0+\frac{(it)^4}{24k}b_1+\frac{(it)^6}{72k}b_0^2+o\left(\frac{1}{k}\right)\right]. \tag{20}$$

Since,

$$E\left[Z^2e^{it(\tilde{V}-W)}\right]=\frac{1}{i^2}\frac{\partial^2}{\partial t_1^2}C_{Z,\tilde{V}-W}(t_1,t)\Big|_{t_1=0}$$

using (7) one obtains,

$$=\frac{1}{n}\mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[1+\frac{i^3}{6\sqrt{k}}\left(b_0t^3-12t\right)+o\left(\frac{1}{\sqrt{k}}\right)\right].$$

Then it follows from (21) that,

$$B_1(t)=\mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[-2\frac{(it)^2}{\sqrt{k}}+o\left(\frac{1}{\sqrt{k}}\right)\right]. \tag{21}$$

Also using (8),

$$E[\tilde{V}\mathrm{e}^{it(\tilde{V}-W)}] = \frac{1}{i}\frac{\partial}{\partial t_1}C_{\tilde{V},\tilde{V}-W}(t_1,t)\Big|_{t_1=0}$$
$$= (2it)\mathrm{e}^{\frac{i^2}{2}\tau t^2} + o(1).$$

Along similar lines using (6) and (7),

$$E\left[(1-nZ^2)^2\mathrm{e}^{it(\tilde{V}-W)}\right] = 2\mathrm{e}^{\frac{i^2}{2}\tau t^2} + o(1).$$

Thus,

$$B_2(t) = 3(it)^2\mathrm{e}^{\frac{i^2}{2}\tau t^2} + o(1). \tag{22}$$

Finally, combining (20), (21) and (22) according to (16) we get,

$$C_0(t) = \mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[1 + \frac{1}{\sqrt{k}}b_0\frac{(it)^3}{6} + \frac{1}{k}\left[\frac{b_0^2}{72}(it)^6 + \frac{b_1}{24}(it)^4 + (it)^2\right] + o\left(\frac{1}{k}\right)\right]. \tag{23}$$

4.1.2 ***Evaluation of*** $C_1(t)$

Here also using (9) observe that,

$$E[W(V-W)\mathrm{e}^{it(V-W)}] = \frac{1}{i^2}\frac{\partial^2}{\partial t_1 \partial t}C_{W,V-W}(t_1,t)\Big|_{t_1=0}.$$

Hence,

$$C_1(t) = \mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[\frac{\tau}{n}(it) + \frac{\tau^2}{n}(it)^3 + \frac{1}{\sqrt{k}}\left[a_0(it)^6 - a_1(it)^4 - a_2(it)^2\right] + o\left(\frac{1}{\sqrt{k}}\right)\right], \tag{24}$$

where

$$a_0 = \tau^2\left[\frac{4}{6(n-1)^2}\kappa_3^2 + \frac{8}{6n}\left(1 - \frac{1}{(n-1)^2}\right)\right]$$
$$a_1 = \tau\left[\frac{2}{n-1}\kappa_4 - \frac{16}{6(n-1)^2}\kappa_3^2 - \frac{32n-88}{6(n-1)^2}\right]$$
$$a_2 = \frac{4}{(n-1)}\kappa_4 - \frac{4}{(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}.$$

4.1.3 Evaluation of $C_2(t)$

It is not hard to see that,

$$E[W^2(V-W)\mathrm{e}^{it(V-W)}] = E_{T_0,U}\left[\left(U-\frac{1}{n}T_0\right)^2 T_0\mathrm{e}^{itT_0}\right] + o(1),$$

where $U = ((n-1)/n)W + (1/n)V$ and $T_0 = V - W$ are independently distributed as $N\left(0, \frac{1}{n}(\kappa_4+2)\right)$ and $N(0, \tau)$, respectively. This was noted in section 3. Hence, after some simplification, we get,

$$E\left[(it)W^2(V-W)\mathrm{e}^{it(V-W)}\right] = \mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[\frac{1}{n^2}\tau^3(it)^4 \right.$$
$$\left. + \left[\frac{1}{n}(\kappa_4+2) + \frac{1}{n^2}\right](it)^2\right] + o(1).$$

Similarly,

$$E\left[\frac{(it)^2}{2}W^2(V-W)^2\mathrm{e}^{it(V-W)}\right] = \mathrm{e}^{\frac{i^2}{2}\tau t^2}\left[\frac{\tau^4}{2n^2}(it)^6 + \left[\frac{1}{2n}(\kappa_4+2)\tau^2 + \frac{3\tau^3}{n^2}\right](it)^4 \right.$$
$$\left. + \left[\frac{1}{2n}(\kappa_4+2)\tau + \frac{3\tau^2}{2n^2}\right](it)^2\right] + o(1).$$

Therefore,

$$C_2(t) = e^{\frac{i^2}{2}\tau t^2}\left[c_0(it)^6 + c_1(it)^4 + c_2(it)^2\right] + o(1), \tag{25}$$

where

$$c_0 = \frac{\tau^4}{2n^2},\ c_1 = \frac{\tau^2}{2n}\kappa_4 + \frac{n\tau^2+4\tau^3}{n^2} \text{ and } c_2 = \frac{3\tau}{2n}\kappa_4 + \frac{9\tau^2}{2n^2} + \frac{3\tau}{n}.$$

4.2 Main results

From (23), (35) and (25), and substituting $\frac{t}{\sqrt{\tau}}$ for t we get the following Lemma.

Lemma 4.1 *Under the assumptions A1 and A2, the characteristic function of $\sqrt{\frac{k}{\tau}}(F-1)$ can be expanded as,*

$$C_{\sqrt{\frac{k}{\tau}}(F-1)}(t) = \exp\left\{\frac{i^2}{2}t^2\right\}\left[1 - \frac{1}{\sqrt{k}}\sum_{j=1}^{2} d_j^{(0)}(-it)^{2j-1} + \frac{1}{k}\sum_{j=1}^{3} d_j^{(1)}(-it)^{2j}\right]$$
$$+ o\left(\frac{1}{k}\right), \tag{26}$$

where

$$
\begin{aligned}
d_1^{(0)} &= \frac{\tau^{1/2}}{n}\\
d_2^{(0)} &= \frac{1}{6}\left(\frac{4n}{(n-1)^2}\kappa_3^2 + 8 - \frac{8}{(n-1)^2} + \frac{6\tau^2}{n}\right)\tau^{-3/2}\\
d_1^{(1)} &= -\frac{1}{2n}\kappa_4 + \frac{2}{n(n-1)}\kappa_3^2 + \frac{n^2+4n+5}{2n(n-1)}\\
d_2^{(1)} &= \frac{(n^2-2n+2)}{2n^3}\kappa_4^2 + \frac{1}{2n}\kappa_4 + \frac{3n^3-2n^2+3}{3n^3(n-1)}\kappa_3^2 + \frac{21n^3+107n^2+59n-3}{48n^2(n-1)}\\
d_3^{(1)} &= \frac{(12\tau^2+n)}{18\tau^3(n-1)^2}\kappa_3^2 + \frac{\tau}{18n^2} + \frac{(n-2)}{18\tau^2(n-1)} + \frac{4(n-2)}{3\tau(n-1)^2}.
\end{aligned}
$$

Inverting (26) formally, we get our main result which is summarized below in Theorem 4.1. For the uniform validity of the inversion we need a condition known as Cramer's condition (e.g. Hall, 1992, p. 45).

Theorem 4.1 *Under the assumption $A1$ and Cramer's condition,*

$$
A3: \limsup_{\|(t_1,t_2,t_3)\|\to\infty} \left| E\left[e^{it_1\bar{y}_{1.}+it_2\bar{y}_{1.}^2+it_3S_1}\right]\right| < 1
$$

the distribution function of $\sqrt{k}(F-1)$ can be expanded as,

$$
P\left(\sqrt{\frac{k}{\tau}}(F-1)\le x\right) = \Phi(x) - \frac{1}{\sqrt{k}}\sum_{j=1}^{2} d_j^{(0)}\Phi^{(2j-1)}(x)
+\frac{1}{k}\sum_{j=1}^{3} d_j^{(1)}\Phi^{(2j)}(x) + o\left(\frac{1}{k}\right), \tag{27}
$$

where $d_j^{(0)}$ and $d_j^{(1)}$ are as defined in Lemma 4.1 and $\Phi^{(j)}(x)$ is the jth derivative of the CDF $\Phi(x)$ of $N(0,1)$.

Remark 4.1 The result given in Theorem 4.1 depends only on cumulants of y_{11} up to fourth order. Hence, it is conjectured that $A1$ in this Theorem can be weakened as,

$$
A1^*: E(Y_{11}^4) < \infty.
$$

It may be noted that $\Phi^{(j)}(x) = (-1)^{j+1}H_{j-1}(x)\phi(x)$ where $H_j(x)$ is the jth Hermite polynomial and $\phi(x)$ is the density of $N(0,1)$. The first seven Hermite polynomials are, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2-1$, $H_3(x) = x^3-3x$, $H_4(x) = x^4-6x^2+3$, $H_5(x) = x^5-10x^3+15x$ and $H_6(x) = x^6-15x^4+45x^2-15$. Consequently, (27) can, alternatively, be expressed as,

$$
P\left(\sqrt{\frac{k}{\tau}}(F-1)\le x\right) = \Phi(x) + \frac{1}{\sqrt{k}}p_1(x)\phi(x) + \frac{1}{k}p_2(x)\phi(x) + o\left(\frac{1}{k}\right), \tag{28}
$$

where

$$p_1(x) = -\left[d_1^{(0)} H_0(x) + d_2^{(0)} H_2(x)\right] \text{ and}$$
$$p_2(x) = -\left[d_1^{(1)} H_1(x) + d_2^{(1)} H_3(x) + d_3^{(1)} H_5(x)\right],$$

Suppose f is a function such that,

$$P\left(\sqrt{\frac{k}{\tau}}(F-1) \leq f(z)\right) = P(Z \leq z)$$

where $Z \sim N(0, 1)$. Then, the Cornish-Fisher expansion can be obtained.

Corollary 4.1 *Under the assumptions of Theorem 4.1, $f(z)$ can be expanded as,*

$$f(z) = f_E(z) + o\left(\frac{1}{k}\right),$$

where

$$f_E(z) = z - \frac{1}{\sqrt{k}} p_1(z) + \frac{1}{k}\left[p_1(z)p_1'(z) - \frac{1}{2} z p_1(z)^2 - p_2(z)\right].$$

Based on the asymptotic expansion (28), it can be shown (Xu and Gupta, 2005) that the following transformation speeds up convergence to the limiting distribution.

Corollary 4.2 *Under the assumptions of Theorem 4.1, we have,*

$$F^* + \frac{1}{\sqrt{k}} p_1(F^*) + \frac{1}{k}\left[p_2(F^*) + \frac{F^*}{2} p_1(F^*)^2\right] \stackrel{d}{=} Z,$$

where $F^* = \sqrt{k/\tau}(F-1)$, $Z \sim N(0, 1)$ and $U \stackrel{d}{=} V$ means $|P(U \leq x) - P(V \leq x)| = o(1/k)$.

5 Unbalanced model

In the unbalanced version of one-way ANOVA model the sample sizes in the k treatment groups are not necessarily equal. We need the following assumption to develop our results.

$$B1 : \bar{n} = \frac{1}{k}\sum_{i=1}^{k} n_i = O(1), \quad \underline{\underline{n}} = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{n_i^2} = O(1) \quad \text{and } \underline{n} = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{n_i} = O(1).$$

As in the balanced case define,

$$Z = \sqrt{k}\bar{y}_{..}, \quad V = \sqrt{k}(S_h - 1) \quad \text{and } W = \sqrt{k}(S_e - 1).$$

Here also it can be checked that,

$$V = \tilde{V} + \frac{1}{\sqrt{k}}\left(1 - \bar{n}Z^2\right) + o_p\left(\frac{1}{\sqrt{k}}\right)$$

where $\tilde{V} = \sqrt{k}\left((1/k)\sum_{i=1}^{k} n_i \bar{y}_{i.}^2 - 1\right)$. Note also that we can write $W = \sqrt{k}((1/k)\sum_{i=1}^{k} \tilde{S}_i^2 - 1)$ where $\tilde{S}_i^2 = 1/(\bar{n}-1)\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$.

Let $\mathbf{x}_i = \left((n_i/\bar{n})\bar{y}_{i.}, n_i\bar{y}_i.^2 - 1, \tilde{S}_i - 1\right)$. Then, $(Z, \tilde{V}, W) = \sqrt{k}\bar{\mathbf{x}}_k$. Note that $\mathbf{x}_i$'s are not identically distibuted and hence we need more assumptions for the asymptotic expansions (see Bhattacharya and Rao, 1976). Further let $\rho_k = k^{-1}\sum_{i=1}^{k} E\|\mathbf{x}_i\|^5$ and $g_m(\mathbf{t}) = \prod_{j=m+1}^{m+p} |E(i\mathbf{t}'\mathbf{x}_j)|$where $m = 0, 1, 2, 3, \cdots$ and $\mathbf{t} = (t_1, t_2, t_3)$. We need the following assumptions for later use.

$$B2 : E\|\mathbf{x}_i\|^5 < \infty, \quad B3 : \sup_{k\geq 1} \rho_k < \infty$$

$$B4 : \text{There exists some integer } p \text{such that } \sup_{m>0} \int g_m(\mathbf{t})d\mathbf{t} < \infty \text{ and}$$

$$\sup\{g_m(\mathbf{t}) : \|\mathbf{t}\| > b, m \geq 0\} < 1 \text{ for all } b > 0.$$

The Edgeworth expansion of the joint characteristic function of $(Z, \tilde{V}, W)$ is given in the following result.

Lemma 5.1 *Under the assumptions $B1$ and $B2$, the joint characteristic function of $(Z, \tilde{V}, W)$ can be expanded as,*

$$C_{Z,\tilde{V},W}(t_1, t_2, t_3) = \exp\left\{\frac{i^2}{2}\sum_{a,b}^{3} \bar{K}_{ab}t_a t_b\right\}\left[1 + \frac{i^3}{6\sqrt{k}}\sum_{a,b,c}^{3} \bar{K}_{abc}t_a t_b t_c + o\left(\frac{1}{\sqrt{k}}\right)\right],$$

where

$$\bar{K}_{11} = \frac{1}{\bar{n}}, \quad \bar{K}_{12} = \bar{K}_{13} = \frac{1}{\bar{n}}\kappa_3, \quad \bar{K}_{22} = \underline{n}\kappa_4 + 2, \quad \bar{K}_{23} = \frac{1 - \underline{n}}{\bar{n} - 1}\kappa_4,$$

$$\bar{K}_{33} = \frac{\bar{n} - 2 + \underline{\underline{n}}}{(\bar{n}-1)^2}\kappa_4 + \frac{2}{\bar{n}-1}, \quad \bar{K}_{222} = \underline{\underline{n}}\kappa_6 + 12\underline{n}\kappa_4 + 10\underline{n}\kappa_3^2 + 8$$

$$\bar{K}_{333} = \frac{\bar{n} - 3 + 3\underline{n} - \underline{\underline{n}}}{(\bar{n}-1)^3}\kappa_6 + \frac{12(\bar{n} - 2 + \underline{n})}{(\bar{n}-1)^3}\kappa_4 + \frac{4(\bar{n} - 3 + 2\underline{n})}{(\bar{n}-1)^3}\kappa_3^2 + \frac{8}{(\bar{n}-1)^2},$$

$$\bar{K}_{223} = \frac{\underline{n} - \underline{\underline{n}}}{\bar{n}-1}\kappa_6 + \frac{4(1 - \underline{n})}{\bar{n}-1}\kappa_4 + \frac{4(1 - \underline{n})}{\bar{n}-1}\kappa_3^2$$

$$\bar{K}_{233} = \frac{1 - 2\underline{n} + \underline{\underline{n}}}{(\bar{n}-1)^2}\kappa_6 + \frac{4(1 - \underline{n})}{(\bar{n}-1)^2}\kappa_4 + \frac{2(\bar{n} - \underline{n})}{(\bar{n}-1)^2}\kappa_3^2$$

As for the balanced case, by using the above Lemma we can establish the following Lemma.

Lemma 5.2 *Under the assumptions $B1$ and $B2$,*

1. *The joint characteristic function of V and W can be expanded as,*

$$C_{V,W}(t_2, t_3) = \exp\left\{\frac{i^2}{2}\sum_{a,b=2}^{3} \bar{K}_{ab} t_a t_b\right\}\Big[1+ \tag{29}$$

$$\frac{i^3}{\sqrt{k}}\left[m_1 t_2^3 + m_2 t_2^2 t_3 + m_3 t_2 t_3^2 + m_4 t_3^3\right] + o\left(\frac{1}{\sqrt{k}}\right)\Big], \tag{30}$$

where

$$m_1 = \frac{\underline{\underline{n}}}{6}\kappa_6 + 2\underline{n}\kappa_4 + \frac{(10\bar{n}\underline{n} - 6)}{6\bar{n}}\kappa_3^2 + \frac{4}{3},$$
$$m_2 = \frac{(\underline{n} - \underline{\underline{n}})}{2(\bar{n} - 1)}\kappa_6 + \frac{2(1 - \underline{n})}{(\bar{n} - 1)}\kappa_4 + \frac{2(1 - \bar{n}\underline{n})}{\bar{n}(\bar{n} - 1)}\kappa_3^2,$$
$$m_3 = \frac{(1 - 2\underline{n} + \underline{\underline{n}})}{2(\bar{n} - 1)^2}\kappa_6 + \frac{2(1 - \underline{n})}{(\bar{n} - 1)^2}\kappa_4 + \frac{(2\bar{n} - \bar{n}\underline{n} - 1)}{\bar{n}(\bar{n} - 1)}\kappa_3^2,$$
$$m_4 = \frac{\bar{n} - 3 + 3\underline{n} - \underline{\underline{n}}}{6(\bar{n} - 1)^3}\kappa_6 + \frac{2(\bar{n} - 2 + \underline{n})}{(\bar{n} - 1)^3}\kappa_4 + \frac{2(\bar{n} - 3 + 2\underline{n})}{3(\bar{n} - 1)^3}\kappa_3^2 + \frac{4}{3(\bar{n} - 1)^2}.$$

2. *The characteristic function of $V - W$ can be expanded as,*

$$C_{V-W}(t) = exp\left\{\frac{i^2}{2}\tau^* t^2\right\}\left[1 + \frac{1}{\sqrt{k}}(m_1 - m_2 + m_3 - m_4)(it)^3 + o\left(\frac{1}{\sqrt{k}}\right)\right] \tag{31}$$

where $\tau^* = \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\kappa_4 + \frac{2\bar{n}}{\bar{n}-1}$.

3. *The joint characteristic function of $T_0 = V - W$ and $U = \frac{1}{\bar{n}}V + \frac{\bar{n}-1}{\bar{n}}W$ can be expanded as,*

$$C_{T_0,U}(t, t_1) = exp\left\{\frac{i^2}{2}\left[\tau^* t^2 + \frac{1}{\bar{n}}(\kappa_4 + 2)t_1^2\right]\right\} + o(1). \tag{32}$$

It can clearly be seen from (29) that the asymptotic distribution of S_h and S_e is normal in the unbalanced case also. It may be noted that the $1/\sqrt{k}$ term depends on the sixth order cumulant κ_6. This indicates that the joint distribution of S_h and S_e is very sensitive to the distribution of the data unless k is very large.

The characteristic function of $\sqrt{k}(F - 1)$ can be expanded as,

$$C_{\sqrt{k}(F-1)}(t) = C_0(t) + \frac{1}{\sqrt{k}}C_1(t) + o\left(\frac{1}{\sqrt{k}}\right), \tag{33}$$

where

$$C_0(t) = E\left[\mathrm{e}^{it(V-W)}\right]$$
$$C_1(t) = E\left[-itW(V - W)\mathrm{e}^{it(V-W)}\right].$$

We obtain, directly from (31), that

$$C_0(t) = exp\left\{\frac{i^2}{2}\tau^* t^2\right\}\left[1 - \frac{1}{\sqrt{k}}m_0(-it)^3 + o\left(\frac{1}{\sqrt{k}}\right)\right], \tag{34}$$

where $m_0 = m_1 - m_2 + m_3 - m_4$.

Also by using (32), one has,

$$\begin{aligned} C_1(t) &= \frac{it}{n}E_{T_0}\left[T_0^2 \mathrm{e}^{itT_0}\right] + o(1) \\ &= -\left[\frac{\tau^{*2}}{n}(-it)^3 + \frac{\tau^*}{n}(-it)\right]\mathrm{e}^{\frac{i^2}{2}\tau^* t^2} + o(1). \end{aligned} \tag{35}$$

Then, combining (34) and (35) according to (33) and replacing t with $t/\sqrt{\tau}^*$, we get the the expanded characteristic function of $\sqrt{\frac{k}{\tau^*}}(F-1)$ as summarized in the following Lemma.

Lemma 5.3 *Under the assumptions $B1$ and $B2$, the characteristic function of $\sqrt{\frac{k}{\tau^*}}(F-1)$ can be expanded as,*

$$C_{\sqrt{\frac{k}{\tau^*}}(F-1)}(t) = \exp\left\{\frac{i^2}{2}t^2\right\}\left[1 - \frac{1}{\sqrt{k}}\sum_{j=1}^{2} d_j^{(0)}(-it)^{2j-1}\right] + o\left(\frac{1}{\sqrt{k}}\right), \tag{36}$$

where

$$\begin{aligned} d_1^{(0)} &= \frac{\tau^{*\frac{1}{2}}}{\bar{n}} \\ d_2^{(0)} &= \tau^{*-3/2}\left[\left(\frac{\tau^{*2}}{\bar{n}} + \frac{4}{3} - \frac{4}{3(\bar{n}-1)^2}\right)\right. \\ &\quad + \left(\frac{5}{3}\underline{n} - \frac{2(1-\underline{n})}{\bar{n}-1} + \frac{(\bar{n}-\underline{n})}{(\bar{n}-1)^2} - \frac{2(\bar{n}-3+2\underline{n})}{3(\bar{n}-1)^3}\right)k_3^2 \\ &\quad + \left(2\underline{n} - \frac{2(1-\underline{n})}{(\bar{n}-1)} + \frac{2(1-\underline{n})}{(\bar{n}-1)^2} - \frac{2(\bar{n}-2+\underline{n})}{(\bar{n}-1)^3}\right)k_4 \\ &\quad \left. + \left(\frac{1}{6}\underline{\underline{n}} - \frac{(\underline{n}-\underline{\underline{n}})}{2(\bar{n}-1)} + \frac{(1-2\underline{n}+\underline{\underline{n}})}{2(\bar{n}-1)^2} - \frac{(\bar{n}-3+3\underline{n}-\underline{\underline{n}})}{6(\bar{n}-1)^3}\right)k_6\right]. \end{aligned}$$

Inverting (36) we get the following result.

Theorem 5.1 *Under the assumptions $B1$-$B4$, the CDF of $\sqrt{k/\tau^*}(F-1)$ can be expanded as,*

$$P\left(\sqrt{\frac{k}{\tau^*}}(F-1) \le x\right) = \Phi(x) + \frac{1}{\sqrt{k}}p_1(x)\phi(x) + o\left(\frac{1}{\sqrt{k}}\right),$$

where

$$p_1(x) = -\left[d_1^{(0)}H_0(x) + d_2^{(0)}H_2(x)\right].$$

It may be noted that by using similar approach terms of order $1/k$ can be included in the asymptotic expansion. However, the calculation is tedious and the results get too messy. Moreover, the final results will depend on cumulants up to eighth order (κ_8).

Suppose f is a function such that,

$$P\left(\sqrt{\frac{k}{\tau^*}}(F-1) \le f(z)\right) = P(Z \le z),$$

where $Z \sim N(0,1)$. Similar to the balanced case, the asymptotic expansion for percentiles may be obtained.

Corollary 5.1 *Under the assumptions of Theorem 5.1, $f(z)$ can be expanded as,*

$$f(z) = f_E(z) + o\left(\frac{1}{\sqrt{k}}\right),$$

where

$$f_E(z) = z - \frac{1}{\sqrt{k}} p_1(z).$$

Along the similar lines as in the balanced case the following transformation can improve the approximation by the limiting distribution.

Corollary 5.2 *Under the assumptions of Theorem 5.1, we have,*

$$F^* + \frac{1}{\sqrt{k}} p_1(F^*) \stackrel{d}{=} Z,$$

where $F^* = \sqrt{\frac{k}{\tau^*}}(F-1)$ and $U \stackrel{d}{=} V$ means $|P(U \le x) - P(V \le x)| = o\left(\frac{1}{\sqrt{k}}\right)$.

In practice the population values of κ_3, κ_4 and κ_6 may not be known. One can use consistent estimates of these quantities given by,

$$\hat{\kappa}_3 = \frac{1}{N\hat{\sigma}^3} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^3$$

$$\hat{\kappa}_4 = \frac{(N+1)}{N(N-1)\hat{\sigma}^4} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^4 - 3 \quad \text{and}$$

$$\hat{\kappa}_6 = \frac{1}{N\hat{\sigma}^6} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^6 - 10\hat{\kappa}_3^2 - 15\hat{\kappa}_4 - 15,$$

where $N = \sum_{i=1}^{k} n_i$ and $\hat{\sigma}^2 = 1/N \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$.

It is known (see, for example, Mardia, 1970) that $\hat{\kappa}_3$ is an unbiased estimator of κ_3 under normality. The estimator $\hat{\kappa}_4$ has been indicated by Browne (1984) to be an unbiased estimator of κ_4 under normality. This estimator also arises as a special case of the family of estimators devised by Yanagihara (2005). We have not seen any similar result for $\hat{\kappa}_6$. In fact, $\hat{\kappa}_6$ is obtained by using the plug-in approach as in Mardia (1970).

6 Numerical results

In this section we compare our asymptotic expansion results with the limiting distribution and other results in the literature. We will also apply our results to two real data sets.

6.1 Simulation design

In the simulation experiment, we compare the upper 5% percentiles and 5% achieved test sizes when the null distribution is approximated by the following methods.

SIM: Numerically using simulation.
LMT: The limiting distribution, i.e. the first term in the asymptotic expansion of this paper.
HG: The asymptotic expansion of this paper.
EHG: The asymptotic expansion of this paper when the cumulants of the population are estimated from the sample.
FOY: The asymptotic expansion of Fujikoshi, Ohmae and Yanagihara (1999).
FH: The bootstrap method of Fisher and Hall (1990).
F: The $\mathcal{F}(k-1, \sum_{i=1}^{k} n_i - k)$ distribution.

To that effect, we generate our data from the following six populations which have also been used in Fujikoshi, Ohmae and Yanagihara (1999) and Yanagihara(2003).

M1: $X + YZ$ where X,Y,Z are independent $N(0, 1)$; $\kappa_3 = 0$; $\kappa_4 = 1.5$; $\kappa_5 = 0$; $\kappa_6 = 15$; $\kappa_8 = 315$.
M2: Symmetric uniform $U[-5, 5]$; $\kappa_3 = 0$; $\kappa_4 = -1.2$; $\kappa_5 = 0$; $\kappa_6 = 6.86$; $\kappa_8 = -86.4$.
M3: χ^2 with 3 degrees of freedom; $\kappa_3 = 1.63$; $\kappa_4 = 4$; $\kappa_5 = 13.06$; $\kappa_6 = 53.33$; $\kappa_8 = 1,493.3$.
M4: χ^2 with 8 degrees of freedom; $\kappa_3 = 1$; $\kappa_4 = 1.5$; $\kappa_5 = 3$; $\kappa_6 = 7.5$; $\kappa_8 = 78.75$.
M5: Normal $N(0, 1)$; $\kappa_3 = 0$; $\kappa_4 = 0$; $\kappa_5 = 0$; $\kappa_6 = 0$; $\kappa_8 = 0$
M6: Double Exponential $DE(0, 1)$; $\kappa_3 = 0$; $\kappa_4 = 3$; $\kappa_5 = 0$; $\kappa_6 = 30$; $\kappa_8 = 630$.

The models $M1$ and $M6$ are symmetric heavy tailed where as $M2$ is symmetric light tailed distribution. Models $M3$ and $M4$ are skewed where as $M4$ has longer tail.

We conduct our simulation study in the balanced as well as in the unbalanced cases. In the balanced case we take$n_1 = n_2 = \cdots = n_k = 5$ and we consider five values of k; i.e. 10, 15, 20, 25, 30. In the unbalanced case, we consider six values of k; i.e 15, 20, 25, 30, 40, 45 and we choose the replication sizes depending on the value of k as described in the Table 2.

6.2 Simulation results and discussion

In Table 3 we display the 5% percentiles and 5% achieved test sizes for the balanced case. It is clear from this table that our asymptotic expansion results (HG

Table 2 Replication sizes in the simulation study for the unbalanced case

k	Replication sizes
15	$n_1 = \cdots = n_{10} = 2, n_{11} = \cdots = n_{15} = 6$
20	$n_1 = \cdots = n_{15} = 2, n_{16} = \cdots = n_{20} = 6$
25	$n_1 = \cdots = n_{15} = 2, n_{16} = \cdots = n_{25} = 6$
30	$n_1 = \cdots = n_{15} = 2, n_{16} = \cdots = n_{30} = 6$
35	$n_1 = \cdots = n_{20} = 2, n_{21} = \cdots = n_{35} = 6$
40	$n_1 = \cdots = n_{25} = 2, n_{26} = \cdots = n_{40} = 6$
45	$n_1 = \cdots = n_{30} = 2, n_{31} = \cdots = n_{45} = 6$

Table 3 Upper 5% percentiles and actual 5% test sizes in the balanced case

Model	k	Upper 5% percentile					Actual 5% sizes ($\times 100$)					
		SIM	HG	EHG	FOY	$\mathcal{F}$	LMT	HG	EHG	FOY	FH	$\mathcal{F}$
M1	10	2.197	2.185	2.213	2.102	2.248	9.1	5.1	5.0	5.5	4.9	4.7
	15	2.062	2.071	2.090	1.997	2.107	8.5	4.9	4.9	5.4	4.8	4.7
	20	2.024	2.006	2.023	1.939	2.031	7.9	5.1	5.1	5.6	5.1	4.9
	25	1.944	1.963	1.961	1.901	1.982	7.4	4.9	4.9	5.3	4.9	4.7
	30	1.971	1.932	1.937	1.874	1.947	7.9	5.3	5.3	5.7	5.4	5.2
M2	10	2.342	2.219	2.219	2.157	2.248	9.6	5.5	5.5	6.0	5.6	5.5
	15	2.107	2.094	2.094	2.033	2.107	8.6	5.1	5.1	5.4	5.0	5.0
	20	2.049	2.023	2.023	1.965	2.031	8.3	5.2	5.2	5.5	5.2	5.1
	25	2.047	1.977	1.977	1.922	1.982	8.5	5.5	5.5	6.0	5.7	5.5
	30	1.961	1.944	1.944	1.891	1.947	7.8	5.1	5.1	5.6	5.4	5.1
M3	10	2.222	2.191	2.228	2.032	2.248	8.9	5.2	5.1	6.0	4.9	4.9
	15	2.102	2.086	2.105	1.960	2.107	8.3	5.1	5.1	5.8	5.2	5.0
	20	1.961	2.025	2.038	1.918	2.031	7.4	4.6	4.6	5.3	4.8	4.6
	25	1.982	1.983	1.990	1.890	1.982	7.5	5.0	5.0	5.5	5.2	5.0
	30	1.943	1.953	1.957	1.868	1.947	7.6	5.0	5.0	5.5	5.2	5.0
M4	10	2.230	2.215	2.217	2.095	2.248	9.0	5.1	5.0	5.7	4.9	4.9
	15	2.030	2.095	1.985	1.996	2.107	8.2	4.9	4.8	5.2	4.7	4.5
	20	2.023	2.027	2.025	1.941	2.031	8.1	5.0	5.0	5.5	5.1	4.9
	25	1.967	1.982	1.980	1.904	1.982	7.8	4.9	4.9	5.4	4.8	4.9
	30	1.905	1.950	1.945	1.877	1.947	7.4	4.7	4.7	5.3	5.0	4.7
M5	10	2.259	2.210	2.216	2.133	2.248	9.4	5.3	5.3	5.7	5.1	5.1
	15	2.067	2.087	2.092	2.017	2.107	8.2	4.9	4.9	5.3	4.9	4.8
	20	2.056	2.019	2.019	1.954	2.031	8.0	5.2	5.2	5.6	5.3	5.1
	25	1.938	1.973	1.975	1.912	1.982	7.5	4.7	4.7	5.2	5.0	4.7
	30	1.932	1.941	1.938	1.883	1.947	7.5	4.9	4.9	5.4	5.2	4.9
M6	10	2.108	2.145	2.081	2.072	2.248	8.5	4.8	4.7	5.2	4.4	4.4
	15	2.040	2.044	2.077	1.977	2.107	8.3	5.0	4.9	5.3	4.7	4.5
	20	1.999	1.986	1.988	1.924	2.031	7.9	5.1	5.0	5.5	4.9	4.8
	25	1.935	1.947	1.971	1.889	1.982	7.7	4.9	4.9	5.4	4.9	4.7
	30	1.911	1.919	1.936	1.864	1.947	7.2	4.9	4.9	5.3	4.9	4.8

Note that $z_{0.05} = 1.645$. Number of simulations is 20,000 (standard error 0.154). The percentiles displayed for EHG are the averages over the 20,000 simulations.

and EHG) perform quite well for all the models and the value of k as small as 10 or 15.

In particular, for the model M6 (Double Exponential) which is heavy tailed our expansion reaches the desired size quickly compared to the other methods. The approximation based on the bootstrap of Fisher and Hall (1990) and the $\mathcal{F}$ distribution tend to be conservative for smaller values of k. The conservative nature of the $\mathcal{F}$ approximation for heavy tailed populations was also noted by Donaldson (1968).

For the lighter tailed distributions M1 and M2, all the methods, except the one based on the limiting distribution, do almost equally well. Indeed, our asymptotic expansion hits close to the desired size a little more frequently than the others. In the case of M5 where the $\mathcal{F}$ distribution is appropriate, we see that our expansion performs comparably well.

In general, we do not recommend using just the limiting distribution at least when the total sample size is less than 150. The approximation based on the expansion of Fujikoshi, Ohmae and Yanagihara (1999) turns out to be liberal in most cases. This should not be surprising because this expansion is not designed to work in the large k and small n_i's situation. In fact, we learn from this simulation that Huber's condition (Huber, 1973) is important for the validity of the expansion results of Fujikoshi, Ohmae and Yanagihara (1999).

An interesting observation is that the upper percentiles and actual test sizes of our asymptotic expansion with true population cumulant values (HG) and with the estimated cumulants (EHG) do strikingly agree. This assures that in the most practical situation where κ_3 and κ_4 are not known, the practitioner can use the sample estimates to apply the results of this paper.

Figure 2 contains charts to illustrates the effect of the transformation given in Corollary 4.2. We see that the convergence to the limiting distribution has been greatly speeded up by using the transformation.

In the unbalanced case (Table 4), the numerical results are in strong support of our asymptotic expansion for all values of k considered. The approximation based on the limiting distribution (LMT) tends to be too liberal for all the models and hence needs to be avoided unless k is very large (at least bigger than 45). In the case when there is a skewness in the population as in M3 and M4 or when the population has a symmetric heavy tailed distribution as M6, the approximation based on our asymptotic expansion is the best compared to all the other approximations. It is also clear that the results of our asymptotic expansions based on the estimated cumulants strikingly agree with those based on the true cumulants.

In general, we do not recommend to use the approximations based on the asymptotic expansion of Fujikoshi, Ohmae and Yanagihara (1999), the bootstrap method of Fisher and Hall (1990) and the $\mathcal{F}$ approximation when the normality assumption is suspected. It may be noted that the behavior of these approximations is unclear as k gets larger. In a more extensive simulation study, which is not reported here, we noticed that the particular values chosen for the n_i's do not matter as long as the values are small.

6.3 Real-data examples

Our first example deals with the comparison of eighteen clones of potato for their resistance to bacterial wilt caused by Pseudomonas Solanacearum. Data on yield were collected for all the clones from four locations with known prevalence of the bacteria population. That is, in this study $n = 4$ and $k = 18$. Inspection of the data revealed the presence of extreme observations which justify the violation of the normality assumption. Moreover, $\hat{\kappa}_3 = 0.4077$ and $\hat{\kappa}_4 = -1.2045$ indicate

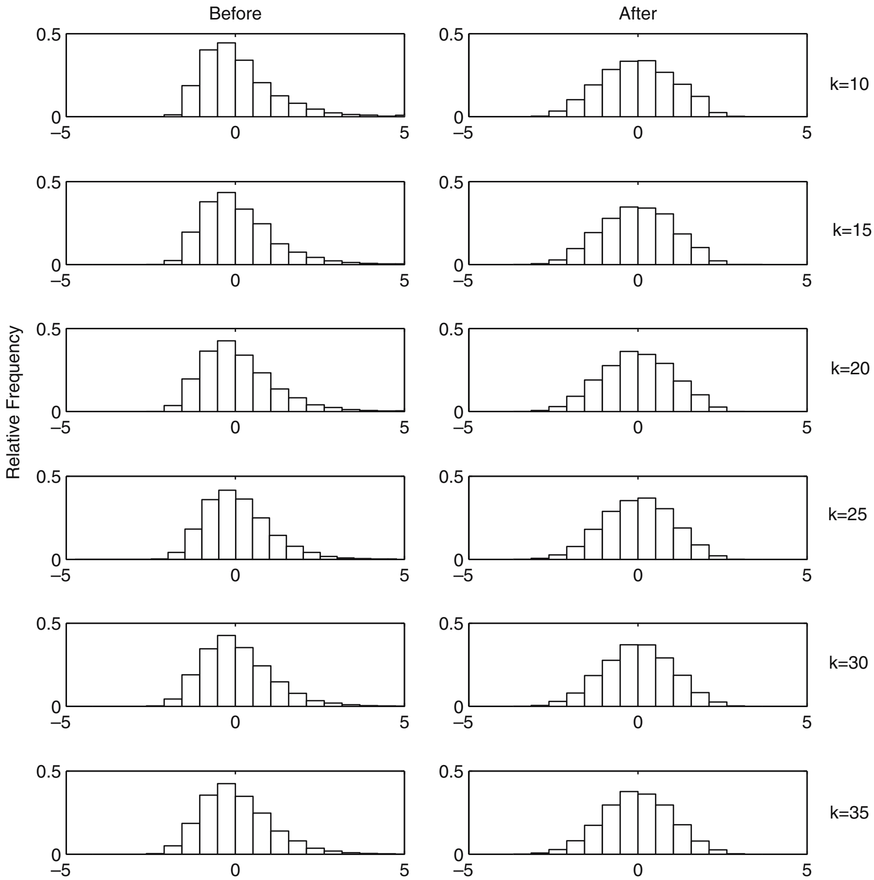

Fig. 2 Sampling distribution of $\sqrt{\frac{k}{\tau}}(F-1)$ before and after transformation in Corollary 4.2 for the model $M3$

non–normality . The calculated value of the test statistics is $F^* = -1.7203$. The upper 5% percentile based on our asymptotic expansion is 2.1116, hence we fail to reject the null hypothesis and conclude that the 18 clones are similar in their resistance.

The second example is based on a publicly available data from a spike-in microarray experiment. Sixteen probe sets were selected (from about 12,000 probe sets) and spiked-in with different concentrations of synthesized RNA on three Affymetrix chips. In this example we would like to test if the expression values of the probe sets are the same after spiked-in with the same concentration of RNA. We calculated RMA expression values (Irizarry et al. 2003) of the sixteen probes in each of the three Affymetrix chips used in the experiment. Notice that $n = 3$ and $k = 16$. The readers are referred to Cope, Irizarry, Jaffee, Wu and Speed (2004) for more details about this data set. The calculated value of the test statistic is $F^* = 180.9294$ and the estimates of the third and fourth cumulants are $\hat{\kappa}_3 = 0.4524$ and $\hat{\kappa}_4 = -0.4438$. The upper 5% percentile based on our asymptotic

Table 4 Upper 5% percentiles and actual 5% test sizes in the unbalanced case

Model	k	Upper 5% percentile					Actual 5% sizes ($\times 100$)					
		SIM	HG	EHG	FOY	$\mathcal{F}$	LMT	HG	EHG	FOY	FH	$\mathcal{F}$
M1	15	2.274	2.130	2.127	1.984	2.158	9.5	5.9	5.9	6.7	6.3	5.7
	20	2.209	2.092	2.093	1.912	2.106	9.2	5.6	5.6	6.8	6.3	5.5
	25	2.005	2.005	2.001	1.883	1.972	8.0	5.0	5.0	5.9	6.0	5.2
	30	2.007	1.954	1.950	1.859	1.913	7.8	5.4	5.4	6.1	6.3	5.7
	40	1.964	1.934	1.932	1.809	1.879	7.7	5.2	5.2	6.2	6.3	5.6
	45	1.952	1.925	1.924	1.789	1.867	7.7	5.2	5.2	6.3	6.5	5.6
M2	15	2.291	2.074	2.077	2.088	2.350	9.7	6.3	6.3	6.2	5.3	4.7
	20	2.140	2.045	2.048	2.011	2.290	9.0	5.6	5.6	5.9	5.0	4.3
	25	2.012	1.962	1.963	1.954	2.142	8.1	5.4	5.4	5.4	5.0	4.2
	30	1.968	1.918	1.919	1.914	2.060	7.9	5.4	5.4	5.4	5.0	4.4
	40	1.940	1.899	1.900	1.872	2.043	7.6	5.3	5.3	5.5	5.1	4.3
	45	1.981	1.892	1.893	1.855	2.033	7.9	5.3	5.2	5.9	5.5	4.6
M3	15	2.281	2.281	2.219	1.902	20.17	9.1	5.0	5.3	7.1	6.8	6.4
	20	2.229	2.223	2.185	1.851	1.970	8.9	5.0	5.2	7.3	7.1	6.4
	25	2.011	2.120	2.082	1.852	1.846	7.8	4.5	4.6	6.2	6.8	6.2
	30	2.014	2.055	2.022	1.844	1.801	7.8	4.7	5.0	6.2	7.1	6.5
	40	2.015	2.025	2.003	1.796	1.757	7.7	4.9	5.0	6.6	7.4	6.8
	45	2.042	2.012	1.994	1.777	1.745	8.1	5.1	5.2	6.9	7.8	7.1
M4	15	2.269	2.177	2.153	1.980	2.158	9.5	5.5	5.7	6.8	6.4	5.7
	20	2.182	2.136	2.121	1.915	2.106	9.3	5.3	5.4	6.7	6.3	5.5
	25	2.053	2.038	2.025	1.890	1.972	8.1	5.1	5.2	6.0	6.2	5.5
	30	2.040	1.982	1.972	1.868	1.913	8.2	5.4	5.5	6.2	6.6	5.9
	40	1.983	1.961	1.954	1.820	1.879	7.8	5.1	5.2	6.2	6.5	5.7
	45	2.010	1.952	1.946	1.801	1.867	8.1	5.1	5.1	6.6	6.8	6.0
M5	15	2.233	2.098	2.101	2.038	2.259	9.4	5.7	5.7	6.1	5.6	4.9
	20	2.180	2.068	2.070	1.963	2.202	9.4	5.7	5.7	6.5	5.8	4.9
	25	2.056	1.979	1.980	1.919	2.061	8.4	5.6	5.5	5.9	5.8	5.0
	30	1.962	1.931	1.932	1.887	1.991	7.8	5.2	5.2	5.5	5.5	4.8
	40	1.950	1.914	1.914	1.841	1.965	7.5	5.3	5.3	5.9	5.7	4.9
	45	1.953	1.907	1.907	1.823	1.954	7.9	5.3	5.3	6.1	5.8	5.0
M6	15	2.189	2.157	2.152	1.939	2.070	9.2	5.2	5.2	6.7	6.6	5.8
	20	2.163	2.114	2.116	1.868	2.021	9.1	5.3	5.3	7.2	6.7	6.0
	25	2.082	2.026	2.021	1.852	1.894	8.3	5.4	5.4	6.6	7.0	6.3
	30	1.974	1.973	1.967	1.836	1.843	7.9	5.0	5.1	6.1	6.8	6.1
	40	1.960	1.951	1.948	1.783	1.803	7.7	5.1	5.1	6.4	6.9	6.2
	45	1.971	1.941	1.939	1.762	1.791	7.6	5.3	5.3	6.7	7.1	6.4

Note that $z_{0.05} = 1.645$. Number of simulations is 20,000 (standard error 0.154). The percentiles displayed for EHG are the averages over the 20,000 simulations

expansion is 2.2685. Hence, even though the probe sets were spiked-in with the same concentration of RNA, they have significantly different expression levels.

7 Concluding remarks

Asymptotic expansion for the distribution of F-statistic in one-way ANOVA was derived to the order $1/k$ in the balanced case and to the order $1/\sqrt{k}$ in the unbalanced case under general conditions. Even if our results assume existence of eighth moment, the results involve only cumulants up to forth order in the balanced case. It may be possible that similar results could be derived under minimal moment requirement as in Hall (1987). We have also derived asymptotic expansion for the joint distribution of the treatment mean square and errors mean square up to order $1/\sqrt{k}$. The numerical results do clearly show the excellent performance of the approximation from our asymptotic expansion, in particular, when the parent population is heavy tailed.

In the unbalanced case, however, the numerical results are in favor of our asymptotic expansion in most of the populations sampled. Hence, one needs to be cautious in using previous results when the normality assumption is suspected unless the replication sizes are large. If possible one is advised to maintain equal replication sizes for all the treatments to reduce the effect of non–normality on previous approximation results.

Acknowledgements The authors would like to thank the two anonymous referees for critically reading the paper and providing valuable suggestions which have led to a significant improvement over the original manuscript.

References

1. Akritas, M. A., Arnold, S. F. (2000). Asymptotics for analysis of variance when the number of levels is large. *Journal of the American Statistical association*, *95*, 212–226
2. Bathke, A. (2002). ANOVA for large number of treatments, *Mathematical Methods of Statistics*, *11*, 118–132
3. Bhattacharya, R. N., Rao, R. R. (1976). Normal approximation and asymptotic expansions. New York: Wiley
4. Boos, D. D., Brownie, C. (1995). ANOVA and rank tests when number of treatments is large. *Statistics and Probability Letters*, *23*, 183–191
5. Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37*, 62–83
6. Brownie, C., Boos, D. D. (1994). Type I error robustness of ANOVA and ANOVA on ranks when the number of treatments is large. *Bimetrics*, *50*, 542–549
7. Cope, L. M., Irizarry, R. A., Jaffee, H. A., Wu, Z., Speed, T. P. (2004). A benchmark for affymetrix genechip expression measures. *Bioinformatics*, 20, 323–331
8. David, F. N., Kendall, M. G., Barton, D. E. (1966). Symmetric function and allied tables. London: Cambridge University Press
9. Donaldson, T.S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the neumerator and denominator of the F-ratio. *Journal of the American Statistical Association*, *63*, 660–676
10. Fisher, N. I., Hall, P. (1990). On bootstrap hypothesis testing. *Australian Journal of Statistics*, *32*, 177–190
11. Fujikoshi, Y., Ohmae, M., Yanagihara, H. (1999). Asymptotic approximations of the null distribution of the one-way ANOVA test statistic under nonormality. *Journal of Japan Statistical Society*, *29*, 147–161
12. Gupta, A. K., Harrar,S., Fujikoshi, Y. (2005). MANOVA for large hypothesis degrees of freedom under non-normality. *Test* (in press)
13. Gupta, A. K., Harrar, S., Fujikoshi, Y. (2006). Asymptotics for testing hypothesis in some variance components model under nonnormality. *Journal of Multivariate Analysis*, *97*, 148–178
14. Hall, P. (1987). Edgeworth expansion for Students's statistics under minimal moment conditions. *Annals of Probability*, *15*, 920–931
15. Hall, P. (1992). The bootstrap and edgeworth expansions. Berlin Heidelberg New York: Springer
16. Huber, P. J. (1973), Robust regression: asymptotics, cojectures and montecarlo. *Annals of Statistics*, *1*, 799–821
17. Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, *4*, 249–264

18. James, G. S. (1954). Test for Linear hypothesis in univariate and multivariate analysis when ratios of the population variances are unknown. *Biometrika*, *41*, 19–43
19. Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530
20. Tan, W. Y., Cheng, S. S. (1981). Cumulants of bilinear forms and quadratic forms and their application. *Communications in Statistics–Theory and Methods*, *A10*, 283–298
21. Yanagihara, H. (2000). Asymptotic expansion of the null distribution of the one-way ANOVA test statistic for heteroscedastic case under nonnormality. *Communications in Statistics–Theory and Methods*, *29*, 463–476
22. Yanagihara, H. (2003). Asymptotic expansion of the null distribution of test statistic for linear hypothesis in non-normal linear model. *Journal of Multivariate Analysis*, *84*, 222–246
23. Yanagihara, H. (2005). A family of estimators for multivariate kurtosis in a nonnormal linear regression model. *Journal of Multivariate Analysis* (in press)
24. Xu, J., Gupta, A. K. (2005). Confidence intervals for the mean value of response function in generalized linear models. *Statistica Sinica*, *15*, 1081–1096