HIGHER ORDER EFFICIENCY OF LINEAR COMBINATIONS OF U-STATISTICS AS ESTIMATORS OF ESTIMABLE PARAMETERS

Toshifumi NOMACHI, Masao KONDO and Hajime YAMATO

Received Octorber 16, 2001; revised December 4, 2001

Abstract. As an estimator of an estimable parameter, we introduce a new statistic which is given by a linear combination of U-statistics. This statistic is identical with the V-statistic for the kernel of degree 1 or 2. For the kernel of degree larger than 2, the new statistic and the V-statistic have no difference in the mean of the second order efficiency. We shall compare these two statistics by the fourth order efficiency and give two examples.

Mathematics Subject Classification: Primary 62G05; Secondary 62G20.
Key Words and Phrases: Estimable parameter, higher order efficiency, linear combination of U-statistics, V-statistics.

1 Introduction Let $\theta(F)$ be a regular functional or an estimable parameter of a distribution F and $g(x_1, ..., x_k)$ be its kernel of degree k. In this paper we assume that the kernel g is symmetric and not degenerate. Let $X_1, ..., X_n$ be a random sample of size n from the distribution F. As an estimator of $\theta(F)$, Toda and Yamato (2001) introduces a linear combination Y_n of U-statistics as follows. Let $w(r_1, \dots, r_j; k)$ be a nonnegative and symmetric function of positive integers $r_1, \dots, r_j$ such that $j = 1, \dots, k$ and $r_1 + \cdots + r_j = k$, where k is the degree of the kernel g and fixed. We assume that at least one of $w(r_1, \dots, r_j; k)$'s is positive. For $j = 1, \dots, k$, let $g_{(j)}(x_1, ..., x_j)$ be the kernel given by

$$g_{(j)}(x_1, \dots, x_j) = \frac{1}{d(k,j)} \sum\nolimits^{+}_{r_1+\cdots+r_j=k} w(r_1, \dots, r_j; k) g(\underbrace{x_1, \dots, x_1}_{r_1}, \dots, \underbrace{x_j, \dots, x_j}_{r_j}),$$

where the summation $\sum^{+}_{r_1+\cdots+r_j=k}$ is taken over all positive integers $r_1, ..., r_j$ satisfying $r_1 + \cdots + r_j = k$ with j and k fixed and $d(k,j) = \sum^{+}_{r_1+\cdots+r_j=k} w(r_1, \dots, r_j; k)$ for $j = 1, 2, ..., k$. Let $U_n^{(j)}$ be the U-statistic associated with this kernel $g_{(j)}(x_1, \dots, x_j; k)$ for $j = 1, \dots, k$. The kernel $g_{(j)}(x_1, \dots, x_j; k)$ is symmetric because of the symmetry of $w(r_1, \dots, r_j; k)$. If $d(k,j)$ is equal to zero for some j, then the associated $w(r_1, \dots, r_j; k)$'s are equal to zero. In this case, we let the corresponding statistic $U_n^{(j)}$ be zero.
The statistics Y_n is given by

$$Y_n = \frac{1}{D(n,k)} \sum_{j=1}^{k} d(k,j) \binom{n}{j} U_n^{(j)}, \tag{1.1}$$

2000 *Mathematics Subject Classification.* Primary 62G05; Secondary 62G20.
Key words and phrases. Estimable parameter, higher order efficiency, linear combination of U-statistics, V-statistics.

where $D(n,k)=\sum_{j=1}^{k} d(k,j)\binom{n}{j}$. Since w's are nonnegative and at least one of them is positive, $D(n,k)$ is positive. Note that $U_n^{(k)}=U_n$ for $w(1,\dots,1;k)>0$, because of $g_{(k)}=g$.

For example, let w be the function given by $w(1,1,\dots,1;k)=1$ and $w(r_1,\dots,r_j;k)=0$ for positive integers $r_1,\dots,r_j$ such that $j=1,\dots,k-1$ and $r_1+\cdots+r_j=k$. Then $d(k,k)=1$, $d(k,j)=0$ $(j=1,\dots,k-1)$ and $D(n,k)=\binom{n}{k}$. Thus the corresponding statistic Y_n is equal to U-statistic U_n, which is given by

$$U_n=\binom{n}{k}^{-1}\sum_{1\le j_1<\cdots<j_k\le n} g(X_{j_1},\dots,X_{j_k}), \tag{1.2}$$

where $\sum_{1\le j_1<\cdots<j_k\le n}$ denotes the summation over all integers $j_1,\dots,j_k$ satisfying $1\le j_1<\cdots<j_k\le n$.

Let w be the function given by $w(r_1,\dots,r_j;k)=1$ for positive integers $r_1,\dots,r_j$ such that $j=1,\dots,k$ and $r_1+\cdots+r_j=k$. For the the equation $r_1+\cdots+r_j=k$ with j and k fixed, the number of its solutions is $\binom{k-1}{j-1}$. Hence we have $d(k,j)=\binom{k-1}{j-1}$ for $j=1,\dots,k$, and $D(n,k)=\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n}{j}=\binom{n+k-1}{k}$. Thus the kernel $g_{(j)}(x_1,\dots,x_j)$ is equal to

$$\binom{k-1}{j-1}^{-1}\sum\nolimits^{+}_{r_1+\cdots+r_j=k} g(\underbrace{x_1,\dots,x_1}_{r_1},\dots,\underbrace{x_j,\dots,x_j}_{r_j}).$$

In terms of the U-statistic $U_n^{(j)}$ associated with this kernel for $j=1,\dots,k$, the statistic Y_n given by (1.1) is written as

$$\binom{n+k-1}{k}^{-1}\sum_{j=1}^{k}\binom{k-1}{j-1}\binom{n}{j}U_n^{(j)},$$

which is equal to the LB-statistic B_n given by

$$B_n=\binom{n+k-1}{k}^{-1}\sum_{r_1+\cdots+r_n=k} g(\underbrace{X_1,\dots,X_1}_{r_1},\dots,\underbrace{X_n,\dots,X_n}_{r_n}), \tag{1.3}$$

where $\sum_{r_1+\cdots+r_n=k}$ denotes the summation over all non-negative integers $r_1,...,r_n$ satisfying $r_1+\cdots+r_n=k$. (See Yamato (1977) and Nomachi and Yamato(2001).)

Let w be the function given by $w(r_1,\dots,r_j;k)=k!/(r_1!\cdots r_j!)$ for positive integers $r_1,\dots,r_j$ such that $j=1,\dots,k$ and $r_1+\cdots+r_j=k$. The Stirling number of the second kind $\mathcal{S}(k,j)$ satisfies the relation $j!\mathcal{S}(k,j)=\sum^{+}_{r_1+\cdots+r_j=k}k!/(r_1!\cdots r_j!)$. Hence we have $d(k,j)=j!\mathcal{S}(k,j)$ for $j=1,\dots,k$. Thus we have $D(n,k)=\sum_{j=1}^{k}\mathcal{S}(k,j)(n)_j=n^k$ since $\mathcal{S}(k,j)$ satisfies the relation $t^k=\sum_{j=1}^{k}\mathcal{S}(k,j)(t)_j$, where $(t)_j=t(t-1)\cdots(t-j+1)$. (For Stirling number, see, for example, Charalambides and Singh (1988).) Therefore the kernel $g_{(j)}(x_1,\dots,x_j)$ is equal to

$$\frac{1}{j!\mathcal{S}(k,j)}\sum\nolimits^{+}_{r_1+\cdots+r_j=k}\frac{k!}{r_1!\cdots r_j!}g(\underbrace{x_1,\dots,x_1}_{r_1},\dots,\underbrace{x_j,\dots,x_j}_{r_j}).$$

In terms of the U-statistic $U_n^{(j)}$ associated with this kernel for $j = 1, \ldots, k$, the statistic Y_n given by (1.1) is written as

$$\frac{1}{n^k}\sum_{j=1}^{k}\mathcal{S}(k,j)(n)_j U_n^{(j)},$$

which is equal to the V-statistic V_n given by

$$V_n = \frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} g(X_{j_1},\ldots,X_{j_k}). \tag{1.4}$$

(See Yamato and Toda (2001), Lee(1990), p.183-184 and Koroljuk and Borovskich (1994), p.40). Nomachi and Yamato (2001) gives asymptotic comparisons of these statistics, that is, U-statistic, V-statistic and LB-statistic, by deficiency.

In Section 2, we introduce a new statistic S_n by choosing a weight function w different from the above. The weight function is based on the Stirling number of the first kind. In Section 3, by using H-decomposition of U-statistic we derive an asymptotic expansion of the statistic Y_n. Making use of it, we evaluate the mean square error of Y_n asymptotically. For the kernel of degrees 1 and 2, the two statistics V_n and S_n are same. For the kernel of degree lager than 2, these statistics have the same second order efficiency or deficiency. In Section 4, we compare the statistics V_n and S_n by the fourth order efficiency. We give two examples.

2 New statistic As stated in Section 1, the LB-statistic is the Y-statistic determined by the weight $w(r_1,\ldots,r_j;k) = 1$ ($j = 1,\ldots,k$ and $r_1 + \cdots + r_j = k$), which is the uniform weight over all the set $(r_1,\ldots,r_j)$'s satisfying $j = 1,\ldots,k$ and $r_1 + \cdots + r_j = k$. The V-statistic is the Y-statistic determined by the weight $w(r_1,\ldots,r_j;k) = k!/(r_1!\cdots r_j!)$ ($j = 1,\ldots,k$ and $r_1 + \cdots + r_j = k$), which gives more weight over the set $(r_1,\ldots,r_j)$ associated with large j than over the set associated with small j. Now we shall consider a Y-statistic associated with an intermediate weight w between the above two weight functions.

We consider a new statistic S_n, as an estimator of θ given by (1.1), by choosing

$$w(r_1,\ldots,r_j;k) = \frac{k!}{r_1\cdots r_j} \tag{2.1}$$

for positive integers $r_1,\ldots,r_j$ such that $j = 1,\ldots,k$ and $r_1 + \cdots + r_j = k$. The Stirling number of the first kind $s(k,j)$ has the expression $j! \mid s(k,j) \mid = \sum^{+}_{r_1+\cdots+r_j=k} k!/(r_1\cdots\ r_j)$ (see, for example, Charalambides and Singh (1988)). Hence we have

$$d(k,j) = j! \mid s(k,j) \mid, \quad j = 1,\ldots,k. \tag{2.2}$$

Thus we have

$$D(n,k) = \sum_{j=1}^{k} \mid s(k,j) \mid (n)_j. \tag{2.3}$$

Therefore the kernel $g_{(j)}(x_1,\dots,x_j)$ is given by

$$g_{(j)}(x_1,\dots,x_j)=\frac{1}{j!\mid s(k,j)\mid}\sum\nolimits^{+}_{r_1+\cdots+r_j=k}\frac{k!}{r_1\cdots r_j}g(\underbrace{x_1,\dots,x_1}_{r_1},\dots,\underbrace{x_j,\dots,x_j}_{r_j}).$$

Since positive integers $r_1,\dots,r_k$ satisfying $r_1+\cdots+r_k=k$ are $r_1=\cdots=r_k=1$, we have

$$g_{(k)}(x_1,\dots,x_k)=g(x_1,\dots,x_k).$$

Since positive integers $r_1,\dots,r_{k-1}$ satisfying $r_1+\cdots+r_{k-1}=k$ are $\{r_1,\dots,r_{k-1}\}=\{2,1,\dots,1\}$,

$$g_{(k-1)}(x_1,\dots,x_{k-1})=\frac{1}{k-1}\big[g(x_1,x_1,x_2,\dots,x_{k-1})+\cdots+g(x_1,\dots,x_{k-2},x_{k-1},x_{k-1})\big].$$

We note that these two kernels are same for any weight w such that $w(1,1,\dots,1;k)$ and $w(2,1,\dots,1;k)$ are positive.

In terms of the U-statistic $U_n^{(j)}$ associated with this kernel for $j=1,\dots,k$, the statistic S_n is given by

$$S_n=\frac{1}{D(n,k)}\sum_{j=1}^{k}\mid s(k,j)\mid (n)_j U_n^{(j)}, \tag{2.4}$$

where $D(n,k)$ is given by (2.3). For the degree $k=1,2$, the weight function w are same for the V-statistic and S-statistic and so these two statistics are identical. For the degree $k=3,4$, we have

$$D(n,3)=n(n^2+1),\quad D(n,4)=n(n^3+4n+1).$$

In general, we can not write S_n in an explicit form, differently from the U-statistic, the LB-statistic and the V-statistic given by (1.2), (1.3) and (1.4), respectively. But, in some case we can write S_n in an explicit form. For example, let us consider the kernel of degree $k=3$. The corresponding S-statistic S_n is denoted as follows, using the V-statistic V_n.

$$S_n=\frac{n^2}{n^2+1}V_n+\frac{1}{n(n^2+1)}\sum_{j=1}^{n}g(X_j,X_j,X_j).$$

Especially, for the third central moment of the distribution F, the new statistic S_n is given by

$$S_n=\frac{n}{n^2+1}\sum_{i=1}^{n}(X_i-\bar{X})^3,$$

where $\bar{X}$ is the sample mean of $X_1,\dots,X_n$ (see Section 4.2). For this parameter, the U-statistic U_n, the V-statistic V_n and the LB-statistic B_n are

$$U_n=\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(X_i-\bar{X})^3,\quad V_n=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^3,$$

$$B_n=\frac{n}{(n+1)(n+2)}\sum_{i=1}^{n}(X_i-\bar{X})^3,$$

respectively (see, Nomachi and Yamato (2001), p.96).

By the recurrence relation of absolute or signless Stirling number of the first kind which is $| s(n+1,k) | = | s(n,k-1) | + n | s(n,k) |$, we can get the recurrence relation of $D(n,k)$ as follows.

$$D(n,k+1) = nD(n-1,k) + kD(n,k), \quad n = k+1, k+2, \cdots, \ k = 1,2,\ldots \tag{2.5}$$

The values of Stirling number of the first kind are, for example, $s(k,k) = 1$, $s(k,k-1) = -k(k-1)/2$, $s(k,k-2) = k(k-1)(k-2)(3k-1)/24$, for $k = 1,2,\ldots$. By using these values, we can get the following asymptotic evaluation of $D(n,k)$,

$$D(n,k) = n^k \Big[1 + \frac{k(k-1)(k-2)}{6} n^{-2} + o(n^{-2}) \Big]. \tag{2.6}$$

3 H-decomposition and MSE In the followings we consider the kernel g of degree $k \geq 3$. For the kernel $g_{(j)}(x_1, \ldots, x_j)$ $(j = 1, \ldots, k)$ associated with the Y-statistic given by (1.1), we put

$$\theta_j = Eg_{(j)}(X_1, \ldots, X_j),$$

and

$$\psi_{(j),c}(x_1, \ldots, x_c) = E\big[g_{(j)}(X_1, \ldots, X_j) \mid X_1 = x_1, \ldots, X_c = x_c\big], \quad c = 1, ..., j.$$

We note that if $w(1, \ldots, 1; k) > 0$ then by the reason stated in Section 2 we have $g_{(k)}(x_1, \ldots, x_k) = g(x_1, \ldots, x_k)$ and so $\theta_k = \theta$. For $j = 1, \ldots, k$ and $c = 2, 3, ..., k$, we put

$$h^{(1)}_{(j)}(x_1) = \psi_{(j),1}(x_1) - \theta_j,$$

$$h^{(c)}_{(j)}(x_1, \ldots, x_c) = \psi_{(j),c}(x_1, \ldots, x_c) - \sum_{i=1}^{c-1} \sum_{1 \leq l_1 \cdots < l_i \leq c} h^{(i)}_{(j)}(x_{l_1}, \ldots, x_{l_i}) - \theta_j.$$

Let $H^{(c)}_{(j),n}$ be the U-statistic associated with the kernel $h^{(c)}_{(j)}$, that is,

$$H^{(c)}_{(j),n} = \binom{n}{c}^{-1} \sum_{1 \leq l_1 < \cdots < l_c \leq n} h^{(c)}_{(j)}(X_{l_1}, \ldots, X_{l_c}). \tag{3.1}$$

Then, the U-statistic $U^{(j)}_n$ associated with the kernel $g_{(j)}$ can be written as follows,

$$U^{(j)}_n = \theta_j + \sum_{c=1}^{j} \binom{j}{c} H^{(c)}_{(j),n}.$$

This form is known as H-decomposition in the context of U-statistics, because it is due to Hoeffding (see, for example, Lee (1990), p.26). In general, this kind of decomposition is well-known as ANOVA decomposition (see, for example, Efron (1982), p.22). Using this

H-decomposition to the right-hand side of (1.1), the statistic Y_n can be written as follows.

$$\begin{aligned} Y_n - \theta &= \frac{1}{D(n,k)} \sum_{j=1}^{k} d(k,j) \binom{n}{j} (U_n^{(j)} - \theta_j + \theta_j - \theta) \\ &= \frac{d(k,k)}{D(n,k)} \binom{n}{k} \left\{ \binom{k}{1} H_{(k),n}^{(1)} + \binom{k}{2} H_{(k),n}^{(2)} + \binom{k}{3} H_{(k),n}^{(3)} \right\} \\ &\quad + \frac{d(k,k-1)}{D(n,k)} \binom{n}{k-1} \left\{ \binom{k-1}{1} H_{(k-1),n}^{(1)} + \binom{k-1}{2} H_{(k-1),n}^{(2)} + \theta_{k-1} - \theta \right\} \\ &\quad + \frac{d(k,k-2)}{D(n,k)} \binom{n}{k-2} \left\{ \binom{k-2}{1} H_{(k-2),n}^{(1)} + \theta_{k-2} - \theta \right\} + R_{1,n}, \end{aligned}$$

where $R_{1,n}$ is the residual term.
We assume that $d(k,k) = w(1,\ldots,1;k) > 0$. Since $\sum_{j=1}^{k} d(k,j)\binom{n}{j}/D(n,k) = 1$, we can write the ratios $d(k,j)\binom{n}{j}/D(n,k)$ $(j = k, k-1, k-2)$ such that

$$\frac{d(k,k)}{D(n,k)} \binom{n}{k} = 1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2} + o(\frac{1}{n^2}), \tag{3.2}$$

$$\frac{d(k,k-1)}{D(n,k)} \binom{n}{k-1} = \frac{\beta_1}{n} + \frac{\beta_{22}}{n^2} + o(\frac{1}{n^2}), \tag{3.3}$$

$$\frac{d(k,k-2)}{D(n,k)} \binom{n}{k-2} = \frac{\beta_2}{n^2} + o(\frac{1}{n^2}), \tag{3.4}$$

where $\beta_1 (\geq 0)$, $\beta_2 (\geq 0)$, β_{21}, and β_{22} are constants and $\beta_2 + \beta_{21} + \beta_{22} = 0$. For the U-statistic U_n given by (1.2), we have $\beta_1 = \beta_2 = \beta_{21} = \beta_{22} = 0$. For the LB-statistic B_n given by (1.3),

$$\begin{aligned} &\beta_1 = k(k-1), \ \ \beta_{21} = \frac{1}{2} k^2 (k-1)^2, \\ &\beta_{22} = -k(k-1)^3, \ \ \beta_2 = \frac{1}{2} k(k-1)^2 (k-2). \end{aligned}$$

For the V-statistic V_n given by (1.4),

$$\begin{aligned} &\beta_1 = \frac{1}{2} k(k-1), \ \ \beta_{21} = \frac{1}{24} k(k-1)(k-2)(3k-1), \\ &\beta_{22} = -\frac{1}{4} k(k-1)^2 (k-2), \ \ \beta_2 = \frac{1}{24} k(k-1)(k-2)(3k-5). \end{aligned}$$

For the new statistic S_n given by (2.1),

$$\begin{aligned} &\beta_1 = \frac{1}{2} k(k-1), \ \ \beta_{21} = \frac{1}{24} k(k-1)(k-2)(3k-5), \\ &\beta_{22} = -\frac{1}{4} k(k-1)^2 (k-2), \ \ \beta_2 = \frac{1}{24} k(k-1)(k-2)(3k-1). \end{aligned}$$

For the V-statistic V_n and the new statistic S_n, the corresponding β_1 and β_{22} are same, and β_{21} and β_2 are different. Using the relations (3.2), (3.3) and (3.4), we can write $Y_n - \theta$ asymptotically as follows.

$$Y_n - \theta = (1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2}) \left\{ \binom{k}{1} H_{(k),n}^{(1)} + \binom{k}{2} H_{(k),n}^{(2)} + \binom{k}{3} H_{(k),n}^{(3)} \right\} \tag{3.5}$$

$$+\left(\frac{\beta_1}{n}+\frac{\beta_{22}}{n^2}\right)\left\{\binom{k-1}{1}H^{(1)}_{(k-1),n}+\binom{k-1}{2}H^{(2)}_{(k-1),n}+\theta_{k-1}-\theta\right\}$$
$$+\frac{\beta_2}{n^2}\left\{\binom{k-2}{1}H^{(1)}_{(k-2),n}+\theta_{k-2}-\theta\right\}+R^*_{2,n},$$

where

$$E[R^*_{2,n}]=O(\frac{1}{n^4}).$$

Therefore, for $k \geq 3$ we can evaluate the mean squared error of Y_n asymptotically which is given by the following. Its proof is given in Appendix.

Proposition 3.1 *We suppose that $E[g(X_{i_1},\dots,X_{i_k})]^2 < \infty$ for $1 \leq i_1 \leq \cdots \leq i_k \leq k$ and $w(1,\dots,1;k) > 0$. Then, for $k \geq 3$ the mean squared error of Y_n given by (1.1) is asymptotically given by*

$$E(Y_n-\theta)^2=\frac{1}{n}k^2\delta^2_{k,(1)}$$
$$+\frac{1}{n^2}\{-2\beta_1k^2\delta^2_{k,(1)}+\frac{k^2(k-1)^2}{2}\delta^2_{k,(2)}+\beta_1^2(\theta_{k-1}-\theta)^2+2\beta_1k(k-1)\zeta^{(1)}_{k-1}\}$$
$$+\frac{1}{n^3}\{(\beta_1^2+2\beta_{21})k^2\delta^2_{k,(1)}+k^2(k-1)^2(\frac{1}{2}-\beta_1)\delta^2_{k,(2)}+(k-1)^2\beta_1^2\delta^2_{k-1,(1)}$$
$$+2k(k-1)(\beta_{22}-\beta_1^2)\zeta^{(1)}_{k-1}+k(k-1)^2(k-2)\beta_1\zeta^{(2)}_{k-1}+2k(k-2)\beta_2\zeta^{(1)}_{k-2}$$
$$+2\beta_1\beta_{22}(\theta_{k-1}-\theta)^2+2\beta_1\beta_2(\theta_{k-1}-\theta)(\theta_{k-2}-\theta)+\frac{k^2(k-1)^2(k-2)^2}{6}\delta^2_{k,(3)}\}+O(\frac{1}{n^4})$$

where

$$\delta^2_{c,(i)}=Var[h^{(i)}_{(c)}(X_1,\dots,X_i)] \text{ for } c=k, i=1,2,3 \text{ and } c=k-1, i=1,$$

$$\zeta^{(i)}_{k-j}=Cov\left[h^{(i)}_{(k)}(X_1,\dots,X_i),h^{(i)}_{(k-j)}(X_1,\dots,X_i)\right] \text{ for } i=1,j=1,2 \text{ and } i=2,j=1.$$

4 Higher order efficiency

4.1 Fourth order efficiency For the V-statistic and the S-statistic, the associated β_1 and $g_{(k-1)}$ are identical, respectively. Therefore, by Proposition 3.1, their mean squared errors are same up to the order n^{-2}. Thus, the V-statistic and the S-statistic have no difference in the mean of the limiting risk deficiency. (For limiting risk deficiency, see Lehmann (1983) p.350 and Nomachi and Yamato (2001).) Or the V-statistic and the S-statistic have no difference in the mean of the second order efficiency. Therefore, we shall compare these two statistics by the fourth order efficiency (FOE). We shall define FOE by taking the second order efficiency into consideration (for the second order efficiency, see, for example, Mikulski (1982)). Since we consider the efficiency in the class of nonparametric distributions, we define FOE as follows.

Definition 4.1 *Let the mean squared errors of two statistics $Y_{i,n}$ $(i=1,2)$ be $MSE(Y_{i,n})$ $(i=1,2)$, respectively. If there exists the finite limit*

$$\lim_{n\to\infty} n^2\left\{1-\frac{MSE(Y_{1,n})}{MSE(Y_{2,n})}\right\},$$

then we say that this limiting value is the fourth order efficiency of $Y_{2,n}$ with respect to $Y_{1,n}$, which is denoted by $FOE(Y_{2,n},Y_{1,n})$.

Lemma 4.2 *If the two statistics $Y_{i,n}$, $(i=1,2)$ have the mean squared errors given by*

$$MSE(Y_{i,n}) = \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_{3,i}}{n^3} + o(\frac{1}{n^3}), \quad a_1 \neq 0,$$

then FOE of $Y_{2,n}$ with respect to $Y_{1,n}$ is given by

$$FOE(Y_{2,n}, Y_{1,n}) = \frac{a_{3,2} - a_{3,1}}{a_1}.$$

Let $Y_{1,n}$ and $Y_{2,n}$ be statistics given by (1.1) with different weight function w's which have the same value only for β_1 in the relations (3.2) and (3.3). About the values used in Proposition 3.1, the values associated with $Y_{1,n}$ are denoted by $\beta_{2,1}$, $\beta_{21,1}$, $\beta_{22,1}$, $\zeta^{(1)}_{k-2,1}$, and $\theta_{k-2,1}$. The values associated with $Y_{2,n}$ are denoted by $\beta_{2,2}$, $\beta_{21,2}$, $\beta_{22,2}$, $\zeta^{(1)}_{k-2,2}$, and $\theta_{k-2,2}$. For $Y_{1,n}$ and $Y_{2,n}$ having positive weights $w(1,\ldots,1;k)$'s and $w(2,1,\ldots,1;k)$'s, the associated $g_{(k)}$ is equal to g, and

$g_{(k-1)}(x_1,\ldots,x_{k-1}) = [g(x_1,x_1,x_2,\ldots,x_{k-1})+\cdots+g(x_1,\ldots,x_{k-2},x_{k-1},x_{k-1}]/(k-1)$.

Thus, the corresponding values of θ_k, θ_{k-1} and $\zeta^{(1)}_{k-1}$ are same, respectively.

Theorem 4.3 *We suppose that $E[g(X_{i_1},\ldots,X_{i_k})]^2 < \infty$ for $1 \leq i_1 \leq \cdots \leq i_k \leq k$. Let $Y_{1,n}$ and $Y_{2,n}$ be statistics given by (1.1) with different weight function w's, where β_1 have the same value in the relations (3.2) and (3.3), and the associated $w(1,\ldots,1;k)$'s and $w(2,1,\ldots,1;k)$'s are positive. Then, FOE of $Y_{2,n}$ with respect to $Y_{1,n}$ is given by*

$$FOE(Y_{2,n}, Y_{1,n}) = \frac{A}{k^2\delta^2_{k,(1)}},$$

where

$$A = 2\Big\{k^2(\beta_{21,2}-\beta_{21,1})\delta^2_{k,(1)} + (\beta_{22,2}-\beta_{22,1})\big[k(k-1)\zeta^{(1)}_{k-1} + \beta_1(\theta_{k-1}-\theta)^2\big]$$

$$+k(k-2)(\beta_{2,2}\zeta^{(1)}_{k-2,2} - \beta_{2,1}\zeta^{(1)}_{k-2,1}) + \beta_1(\theta_{k-1}-\theta)\big[\beta_{2,2}(\theta_{k-2,2}-\theta) - \beta_{2,1}(\theta_{k-2,1}-\theta)\big]\Big\}.$$

In case of $k=3$, we have $g_{(k-2)}(x_1) = g_{(1)}(x_1) = g(x_1,x_1,x_1)$ for any w such that $d(3,1) = w(3;3) > 0$. Thus we have $\theta_{1,1} = \theta_{1,2} = \theta_1$. We have also $\zeta^{(1)}_{1,1} = \zeta^{(1)}_{1,2}$, which we denote by $\zeta^{(1)}_1$. From Theorem 4.3 we have the following.

Corollary 4.4 *Let the degree of the kernel g be $k=3$. Let $Y_{1,n}$ and $Y_{2,n}$ be statistics given by (1.1) with different and positive weght function w's, where β_1 have the same value in the relations (3.2) and (3.3). Then,*

$$FOE(Y_{2,n}, Y_{1,n}) = \frac{2}{9\delta^2_{3,(1)}}\Big\{9(\beta_{21,2}-\beta_{21,1})\delta^2_{3,(1)} + (\beta_{22,2}-\beta_{22,1})\big[6\zeta^{(1)}_2 + \beta_1(\theta_2-\theta)^2\big]$$

$$+(\beta_{2,2}-\beta_{2,1})\big[3\zeta^{(1)}_1 + \beta_1(\theta_2-\theta)(\theta_1-\theta)\big]\Big\}.$$

4.2 FOE of S-statistic with respect to V-statistic As stated at the first paragraph of subsection 4.1, the V-statistic and the S-statistic have no difference in the mean of the second order efficiency. Since the V-statistic and the S-statistic have the same values for β_1 and β_{22}, by Theorem 4.3 we have the following.

Proposition 4.5 *For the kernel with the degree $k \geq 3$, we have*

$$FOE(S_n, V_n) = \frac{(k-1)(k-2)}{24\delta^2_{k,(1)}}\Big\{2(k-2)\big[(3k-1)\zeta^{(1)}_{k-2,2} - (3k-5)\zeta^{(1)}_{k-2,1}\big]$$
$$-8k\delta^2_{k,(1)} + (k-1)(\theta_{k-1} - \theta)\big[(3k-1)\theta_{k-2,2} - (3k-5)\theta_{k-2,1} - 4\theta\big]\Big\},$$

where $\zeta^{(1)}_{k-2,1}$ and $\theta_{k-2,1}$ are the values associated with V_n and $\zeta^{(1)}_{k-2,2}$ and $\theta_{k-2,2}$ are the values associated with S_n.

In the following, we consider the kernel g of degree $k = 3$. Then we have

$$g_{(3)}(x_1, x_2, x_3) = g(x_1, x_2, x_3),$$

$$g_{(2)}(x_1, x_2) = \frac{1}{2}\big[g(x_1, x_1, x_2) + g(x_1, x_2, x_2)\big], \;\; g_{(1)}(x_1) = g(x_1, x_1, x_1).$$

By the U-statistics $U^{(3)}_n$, $U^{(2)}_n$ and $U^{(1)}_n$, associated with the kernel $g_{(3)}$, $g_{(2)}$ and $g_{(1)}$, respectively, we have

$$V_n = \frac{1}{n^2}\big[U^{(1)}_n + 3(n-1)U^{(2)}_n + (n-1)(n-2)U^{(3)}_n\big],$$

and

$$S_n = \frac{1}{(n^2+1)}\big[2U^{(1)}_n + 3(n-1)U^{(2)}_n + (n-1)(n-2)U^{(3)}_n\big] = \frac{n^2}{n^2+1}V_n + \frac{1}{n^2+1}U^{(1)}_n,$$

where $U^{(1)}_n = \sum_{j=1}^n g(X_j, X_j, X_j)/n$. From Corollary 4.4, we have the following.

Proposition 4.6 *For the kernel with the degree $k = 3$,*

$$FOE(S_n, V_n) = \frac{2}{3\delta^2_{3,(1)}}\Big\{\zeta^{(1)}_1 - 3\delta^2_{3,(1)} + (\theta_2 - \theta)(\theta_1 - \theta)\Big\},$$

where $\delta^2_{3,(1)} = Var[h^{(1)}_{(3)}(X_1)]$, $\zeta^{(1)}_1 = Cov[h^{(1)}_{(3)}(X_1), h^{(1)}_{(1)}(X_1)]$, $h^{(1)}_{(1)}(x_1) = g_{(1)}(x_1) - \theta_1$, $h^{(1)}_{(3)}(x_1) = E[g(X_1, X_2, X_3) \mid X_1 = x_1] - \theta$, $\theta = Eg(X_1, X_2, X_3)$, $\theta_2 = Eg(X_1, X_1, X_2)$, and $\theta_1 = Eg(X_1, X_1, X_1)$.

Now we shall give two examples of Proposition 4.6.

Example 1 We consider the probability weighted moments, $\theta = \int x[F(x)]^2 dF(x)$. Since its kernel is $g(x_1, x_2, x_3) = Max\{x_1, x_2, x_3\}/3$, for the S-statistic S_n and the V-statistic V_n we have $g_{(1)}(x_1) = g(x_1, x_1, x_1) = x_1/3$, $g_{(2)}(x_1, x_2) = \{g(x_1, x_1, x_2) + g(x_1, x_2, x_2)\}/2 = Max\{x_1, x_2\}/3$ (Lee(1990); p.9). Let $U^{(1)}_n$, $U^{(2)}_n$, and $U^{(3)}_n$ be the U-statistics associated with the kernels $g_{(1)}(x_1) = x_1/3$, $g_{(2)}(x_1, x_2) = Max\{x_1, x_2\}/3$, and $g(x_1, x_2, x_3) = Max\{x_1, x_2, x_3\}/3$, respectively. The corresponding S-statistic S_n is

$$S_n = \frac{(n-1)(n-2)}{n^2+1}U^{(3)}_n + \frac{2}{n(n^2+1)}\sum_{i=1}^n (i-1)X_{(i)} + \frac{2}{3(n^2+1)}\bar{X}.$$

The corresponding V-statistics is

$$V_n = \frac{(n-1)(n-2)}{n^2}U^{(3)}_n + \frac{2}{n^3}\sum_{i=1}^n (i-1)X_{(i)} + \frac{1}{3n^2}\bar{X},$$

where $X_{(1)} < \cdots < X_{(n)}$ is the order statistics of the sample $\{X_1, \ldots, X_n\}$(Nomachi and Yamato (2001)).

For the sample of size n from the uniform distribution $U(-\tau, \tau)$, we have $\theta = \tau/6$, $\theta_2 = \tau/9$, $\theta_1 = 0$, $\delta^2_{3,(1)} = \tau^2/252$, $\zeta^{(1)}_1 = \tau^2/90$. Thus we get

$$FOE(S_n, V_n) = \frac{64}{45}.$$

In this example, V_n is prefer to S_n in the mean of fourth order efficiency.

Example 2 We consider the third central moment, $\theta = \int_{-\infty}^{\infty} (x-\mu)^3 dF(x)$ where μ is the mean of the distribution F. Its kernel is given by

$$g(x_1, x_2, x_3) = \frac{x_1^3 + x_2^3 + x_3^3}{3} - \frac{1}{2}\left\{x_1^2(x_2 + x_3) + x_2^2(x_3 + x_1) + x_3^2(x_1 + x_2)\right\} + 2x_1x_2x_3$$

(Koroljuk and Borovskich (1994); p.18–19). Since $g(x_1, x_2, x_2) = -g(x_1, x_1, x_2)$, $g(x_1, x_1, x_1) = 0$, $g_{(2)}(x_1, x_2) = 0$, and $g_{(1)}(x_1) = 0$ for this kernel, we have $U_{(2)} = 0$, $U_{(1)} = 0$ with probability one.
The corresponding V-statistics is

$$V_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3.$$

Since $g_{(1)}(x_1) = 0$, the corresponding S-statistic S_n is

$$S_n = \frac{n}{n^2+1} \sum_{i=1}^{n} (X_i - \bar{X})^3.$$

We consider a continuous distribution which is symmetric about mean and have the 6-th moment. Then we have $\theta = 0$, $\theta_2 = 0$, $\theta_1 = 0$ and $\zeta_1^{(1)} = 0$. Since $\delta^2_{3,(1)} = Var[h^{(1)}_{(3)}(X_1)]$ is finite, we get

$$FOE(S_n, V_n) = -2.$$

In this example, S_n is prefer to V_n in the mean of fourth order efficiency.

5 Appendix : Values of β's

The values of β's, given in Section 3, for LB-statistic, V-statistic and S-statistic can be derived as follows:
For LB-statistic,

$$\frac{d(k,j)}{D(n,k)} \binom{n}{j} = \frac{(n)_j}{[n]_k} \cdot \frac{k!}{j!} \binom{k-1}{j-1}, \quad j = 1, 2, \ldots, k,$$

where $[n]_k = n(n+1)\cdots(n+k-1) = \sum_{i=1}^{k} \mid s(k,i) \mid n^i$.
For V-statistic,

$$\frac{d(k,j)}{D(n,k)} \binom{n}{j} = \frac{\mathcal{S}(k,j)(n)_j}{n^k}, \quad j = 1, 2, \ldots, k.$$

For S-statistic,

$$\frac{d(k,j)}{D(n,k)} \binom{n}{j} = \frac{\mid s(k,j) \mid (n)_j}{D(n,k)}, \quad j = 1, 2, \ldots, k,$$

where $D(n,k) = n^k\{1 + [k(k-1)(k-2)/6]n^{-2} + o(n^{-2})\}$ by (2.6). By applying the values of the Stirling numbers to the above, we can get the values of β_1, β_{21}, β_{22}, β_2. The Stirling numbers of the first kind $s(k,k)$, $s(k,k-1)$, $s(k,k-2)$ $(k = 3, 4, \dots)$ are given immediately before (2.6). For the Stirling numbers of second kind,

$$\mathcal{S}(k,k) = 1,\ \mathcal{S}(k,k-1) = \frac{k(k-1)}{2},\ \mathcal{S}(k,k-2) = \frac{k(k-1)(k-2)(3k-5)}{24},\ k = 3, 4, \dots$$

Proof of Proposition 3.1 Any two components of H-decomposition are uncorrelated. By the representation of covariance of U-statistics (see, for example, Koroljuk and Borovskich (1994) and Lee(1990)) , we can get

$$Cov(H^{(i)}_{(k),n}, H^{(j)}_{(k-1),n}) = 0,\ \ i = 1, 2, 3,\ \ j = 1, 2\ \ (i \neq j),$$

$$Cov(H^{(i)}_{(k),n}, H^{(1)}_{(k-2),n}) = 0,\ \ i = 2, 3,$$

and

$$Cov(H^{(2)}_{(k-1),n}, H^{(1)}_{(k-2),n}) = 0,\ \ Cov(H^{(1)}_{(k-1),n}, H^{(1)}_{(k-2),n}) = O(n^{-1}).$$

We use these results to the right-hand side of the squared (3.5). Then we get

$$\begin{aligned}
&E(Y_n - \theta)^2 \\
&= (1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2})^2 \left\{ \binom{k}{1}^2 Var(H^{(1)}_{(k),n}) + \binom{k}{2}^2 Var(H^{(2)}_{(k),n}) + \binom{k}{3}^2 Var(H^{(3)}_{(k),n}) \right\} \\
&+ (\frac{\beta_1}{n} + \frac{\beta_{22}}{n^2})^2 \left\{ \binom{k-1}{1}^2 Var(H^{(1)}_{(k-1),n}) + \binom{k-1}{2}^2 Var(H^{(2)}_{(k-1),n}) + (\theta_{k-1} - \theta)^2 \right\} \\
&+ 2(1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2})(\frac{\beta_1}{n} + \frac{\beta_{22}}{n^2}) \left\{ \binom{k}{1}\binom{k-1}{1} Cov(H^{(1)}_{(k),n}, H^{(1)}_{(k-1),n}) \right. \\
&\qquad \left. + \binom{k}{2}\binom{k-1}{2} Cov(H^{(2)}_{(k),n}, H^{(2)}_{(k-1),n}) \right\} \\
&+ 2(1 - \frac{\beta_1}{n} + \frac{\beta_{21}}{n^2}) \frac{\beta_2}{n^2} \binom{k-2}{1}\binom{k}{1} Cov(H^{(1)}_{(k),n}, H^{(1)}_{(k-2),n}) + \\
&+ 2(\frac{\beta_1}{n} + \frac{\beta_{22}}{n^2}) \frac{\beta_2}{n^2} (\theta_{k-1} - \theta)(\theta_{k-2} - \theta) + O(\frac{1}{n^4}).
\end{aligned}$$

By using the representation of variances and covariances of U-statistics (see, for example, Lee(1990) and Koroljuk and Borovskich (1994)) to each terms of the right-hand side , we can get Proposition 3.1. □

Acknowledgments
The authors would like to express their hearty thanks to the referee for his careful reading of the manuscript and the useful comments.

References

[1] Charalambides, Ch. A. and Singh, J. (1988). A review of the Stirling numbers, their generalizations and statistical applications. *Commun. Statist.- Theory Meth.*, **17**, 2533–2595.

[2] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* SIAM Philadelphia, Pennsylvania.

[3] Koroljuk, V.S. and Borovskich, Yu.V. (1994). *Theory of U-statistics.* Kluwer Academic Publishers, Dordrecht.

[4] Lee, A.J. (1990). *U-statistics.* Marcel Dekker, New York.

[5] Lehmann, E.L. (1983). *Theory of point estimation.* John Wiley, New York.

[6] Mikulski, P.W. (1982). Efficiency, Second-order, in: Kotz, S. and Johnson, N.L., ed., *Encyclopedia of Statistical Science*, **2**. John Wiley and Sons, New York, 469–472.

[7] Nomachi, T. and Yamato, H. (2001). Asymptotic comparisons of U-statistics, V-statistics and limits of Bayes estimates by deficiencies. *J. Japan Statist. Soc.*, **31**, 85–98.

[8] Toda, K. and Yamato, H. (2001). Berry-Esseen bounds for some statistics including LB-statistic and V-statistic. *J. Japan Statist. Soc.*, **31**, No. 2, (To appear.)

[9] Yamato, H. (1977). Relations between limiting Bayes estimates and U-statistics for estimable parameters. *J. Japan Statist. Soc.*, **7**, 57–66.

First author:
Miyakonojo College of Technology, Miyakonojo, Miyazaki 885–8567, Japan.

Second and third authors:
Department of Mathematics and Computer Science, Kagoshima University, Kagoshima 890–0065, Japan