Analytic Bias Reduction for k–Sample Functionals [1]

by

Christopher S. Withers
Applied Mathematics Group
Industrial Research Limited
Lower Hutt, NEW ZEALAND

Saralees Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, UK

Abstract: We give analytic methods for nonparametric bias reduction that remove the need for computationally intensive methods like the bootstrap and the jackknife.

We call an estimate *pth order* if its bias has magnitude n_0^{-p} as $n_0 \to \infty$, where n_0 is the sample size (or the minimum sample size if the estimate is a function of more than one sample). Most estimates are only first order and require $O(N)$ calculations, where N is the total sample size. The usual bootstrap and jackknife estimates are second order but they are computationally intensive, requiring $O(N^2)$ calculations for one sample. By contrast Jaeckel's infinitesimal jackknife is an analytic second order one sample estimate requiring only $O(N)$ calculations. When pth order bootstrap and jackknife estimates are available, they require $O(N^p)$ calculations, and so become even more computationally intensive if one chooses $p > 2$.

For general p we provide analytic pth order nonparametric estimates that require only $O(N)$ calculations. Our estimates are given in terms of the von Mises derivatives of the functional being estimated, evaluated at the empirical distribution.

For products of moments an unbiased estimate exists: our form for this "polykay" is much simpler than the usual form in terms of power sums.

1   Introduction and Summary

Let $T(F)$ be any *smooth functional* of one or more unknown distributions F based on random samples from them. Bias reduction of estimates of $T(F)$, say $T(\widehat{F})$, has been a subject of considerable interest. Traditionally bias reduction has been based on well known resampling methods like bootstrapping and jackknifing in nonparametric settings, see Efron (1982). However, these methods may not be effective in complex situations when the sampling distribution of the statistic changes too abruptly with the parameter, or when this distribution is very skewed and has heavy tails. Also the robustness properties of F may not be preserved for $T(F)$ for all $T(\cdot)$.

Recently, various analytical methods have been developed for bias reduction in parametric settings. Withers (1987) developed methods for bias reduction based on Taylor series expansions.

---

[1] *AMS 1991 Subject Classification:* Primary 62G03; Secondary 62G20, 62G30.
Key Words and Phrases: Bias reduction; k–samples; Nonparametric; U–statistics; Unbiased estimate; Von Mises derivatives.

Sen (1988) established asymptotic normality of $\sqrt{n}\{T(\widehat{F}) - T(F)\}$ as $n \to \infty$ under suitable regularity conditions. Cabrera and Fernholz (1999, 2004) defined a *target estimator*: for a given T and a parametric family of distributions it is defined by setting the expected value of the statistic equal to the observed value. Cabrera and Fernholz (1999, 2004) established under suitable regularity conditions that the target estimator has smaller bias and mean squared error than the original estimator. See also Fernholz (2001).

This paper provides the first analytical methods for nonparametric bias reduction. We give three analytic methods for obtaining unbiased estimates (*UEs*) of any smooth functional $T(F)$. These UEs are in general infinite series which in practice need to be truncated. Let us define a *pth order estimate* as one with bias $O(n_0^{-p})$ as $n_0 \to \infty$, where n_0 is the minimum sample size. Our truncated pth order estimates require only $O(N)$ computations, where N is the total sample size. By contrast computer intensive methods, like the pth order bootstrap and jackknife estimates require $O((n_1 \cdots n_k)^p)$ calculations. Put another way, for fixed p, the computational efficiency of our analytic pth order estimate relative to the pth order bootstrap or jackknife estimate is $O(n_0^{p-1})$. So, our truncated estimates remove the need for these computationally intensive methods of nonparametric bias reduction. The downside is that the details must be worked out for each nonparametric functional of interest. This involves calculating the von Mises or functional derivatives of the functional up to order $2p-2$. When von Mises (1947) introduced these derivatives, he did not define them uniquely, nor did he give a method to obtain higher derivatives. This was rectified in Withers (1983): the second derivative is *not* the derivative of the first derivative, but requires a 'correction' term. von Mises did give a method for calculating the first derivative, also known as *the influence function* and this is well known and widely used. von Mises' expansion for say $T(\widehat{F})$ about $T(F)$ was extended to functionals of more than one distribution in Withers (1988). This introduced for the first time the partial von Mises derivatives and showed how to calculate them.

Suppose we observe k independent random samples of sizes $n = (n_1, \ldots, n_k)$ from k unknown distributions $F = (F_1, \ldots, F_k)$ on $R^{s_1}, \ldots, R^{s_k}$, where R is the real line. Let $\widehat{F} = (\widehat{F}_1, \ldots, \widehat{F}_k)$ be their k empirical (or sample) distributions. We shall give three pth order estimates of any smooth functional $T(F)$ in terms of the derivatives of $T(F)$ up to order $2p - 2$ evaluated at $\widehat{F}$.

As noted we derive these pth order estimates from three forms of UE for $T(F)$. These are all infinite series *unless* $T(F)$ is a polynomial in F, (for example, a polynomial in the moments of $F_1, \ldots, F_k$). Truncation of these series yields three forms of estimates of $T(F)$ of bias $O(n_0^{-p})$, where $n_0 = \min_{i=1}^k n_i$ and $p \geq 1$ is any specified integer.

We call these three forms of estimates the S, T and V estimates. For $p = 2$, all three forms of estimates have $k + 1$ terms. But for $p > 2$ the S estimate is the best choice, requiring fewer terms than the T estimate or the V estimate: see Section 5. The T estimate is its power series equivalent. The V estimate is an intermediate form for arriving at the S estimate.

If $T(F)$ is a product of moments or cumulants, then an unbiased estimate of it exists, and is given by our S estimate with the appropriate choice of p. Special cases include the UEs of the cumulants of Fisher (1929), the UEs of the central moments of James (1958), and the polykays of Wishart (1952) given in terms of the power sums via tables of the symmetric polynomials: see Stuart and Ord (1987, Section 12.22). Our S estimate gives these polykays in terms of the sample central moments and so is much more compact and avoids the need for these tables.

For $p = 2$ and $k = 1$ the relation of our S estimate to the *infinitesimal jackknife* of Jaeckel (1972) is given in Appendix A. Jaeckel gave formulas for second order estimates in terms of the

derivatives with respect to the weights of $T(\widehat{F})$, where $\widehat{F}$ is now the *weighted* empirical distribution. His formulas are equivalent to our second order one sample S and T estimates. Our formulas are given in terms of the second derivatives of $T(F)$.

For the case of *one sample* problems (functionals of only one distribution function), the S and T estimates given here were obtained by a much more laborious approach in Withers (1994a) starting with an expansion for $ET(\widehat{F})$ based on a generalised delta method; results were given up to $p = 4$. (This also contains more examples.) Here these results are extended to $p = 6$ for the general k-sample problem. The present method uses U-statistics and so bypasses the need in Withers (1994a) to differentiate functionals of derivatives.

At this point we give four simple one sample examples dealt with in Section 6:

- For univariate data, a second order estimate of the standard deviation $T(F) = \sigma$ is $T(\widehat{F})\{1 + m^{-1}s_1(\widehat{F})\}$, where $m = n$ or $n - 1$, $s_1(F) = (\beta_4 + 3)/8$ and β_4 is the kurtosis, that is the standardised fourth central moment.

- For univariate data, a second order estimate of $T(F) = \mu/\sigma$ is $T(\widehat{F}) + n^{-1}S_1(\widehat{F})$, where $S_1(F) = -\beta_3/2 - T(F)(3\beta_4 + 1)/8$ and β_3 is the skewness, that is the standardised third central moment.

- For bivariate data a second order estimate of the ratio of marginal means $T(F) = \mu_1/\mu_2$, is

$$(\overline{X_1}/\overline{X_2})\{1 + n^{-1}[\overline{X_1 X_2}/(\overline{X_1}\ \overline{X_2}) - \overline{X_2^2}/\overline{X_2}^2]\}.$$

  (The usual ratio estimate $T(\widehat{F}) = \overline{X_1}/\overline{X_2}$ has bias $O(n^{-1})$.)

- For multivariate data, we give a second order estimate of $T(F) = (\alpha'\mu)^q$, where μ is the mean vector and α is any given vector of the same dimension, is

$$(\alpha'\widehat{\mu})^q\{1 - \binom{q}{2}(\alpha'\widehat{\mu})^{-2}\alpha'\widehat{V}\alpha/(n - 1)\},$$

  where $\widehat{\mu}$ is the sample mean, and $\widehat{V}$ is the sample covariance.

Applications to skewness and kurtosis have already been given in Withers (1994b). A k-sample univariate example is given by interpreting μ_i in the last example as the mean of the ith distribution sampled, $\widehat{\mu}_i$ as the mean of the ith sample, and replacing $\alpha'\widehat{V}\alpha/(n - 1)$ by $\sum_{i=1}^{k} \alpha_i^2 \widehat{v}_i/(n_i - 1)$, where $\widehat{v}_i$ is the ith sample variance. All these examples allow for a possible initial transformation of the data $X \to h(X)$ say.

The three analytic methods for obtaining UEs are as follows. The simplest UE for $T(F)$ has the S *estimate*

$$\sum_{i_1,\ldots,i_k=0}^{\infty} S_{i_1\cdots i_k}(\widehat{F})/\{(n_1 - 1)_{i_1} \cdots (n_k - 1)_{i_k}\}, \tag{1.1}$$

where $(m)_i = m(m - 1) \cdots (m - i + 1) = m!/(m - i)!$. Clearly this can be transformed if desired to the T *estimate*

$$\sum_{i_1,\ldots,i_k=0}^{\infty} T_{i_1\cdots i_k}(\widehat{F})n_1^{-i_1} \cdots n_k^{-i_k}. \tag{1.2}$$

3

The coefficients $S_{i_1 \cdots i_k}(F)$ and $T_{i_1 \cdots i_k}(F)$ are functions of the partial von Mises derivatives of $T(F)$ of order up to $(2i_1, \ldots, 2i_k)$. The third form of UE for $T(F)$ has the V *estimate*

$$\sum_{r_1,\ldots,r_k=0}^{\infty} \widehat{V}_{r_1 \cdots r_k} / (r_1! \cdots r_k!), \tag{1.3}$$

where $V_{r_1 \cdots r_k}$ is determined by the partial derivatives of $T(F)$ of order $(r_1, \ldots, r_k)$. If $T(F)$ is a polynomial in F (such as a polynomial in the moments and cumulants of F), then the S and V forms of the UE reduce to finite sums.

In Section 2 we derive the V estimate (1.3) and its multivariate analogue using U-statistics and tables of the symmetric polynomials. Section 3 derives from it the S estimate (1.1). Section 4 derives the T estimate (1.2). The number of terms required for these estimates and the bootstrap estimate are compared in Section 5. Finally, Section 6 illustrates the three estimates using various examples, including the four listed above. We show in particular that our estimates consistently outperform those due to Sen (1988) and Cabrera and Fernholz (1999, 2004). Computer programs in MAPLE for the implementation of the V, S and T estimates for any p and k are given in Appendix B.

We shall often use bold to denote an integer vector, for example, $\boldsymbol{n}$ for $(n_1, n_2, \ldots, n_k)$, and $\boldsymbol{1}$ for $(1, \ldots, 1)$. Similarly, we write $\boldsymbol{r!} = r_1! \cdots r_k!$ and $(\boldsymbol{m})_{\boldsymbol{i}} = (m_1)_{i_1} \cdots (m_k)_{i_k}$. With this notation, (1.1)–(1.3) become

$$\sum_{\boldsymbol{i}=0}^{\infty} S_{\boldsymbol{i}}(\widehat{F})/(\boldsymbol{n}-\boldsymbol{1})_{\boldsymbol{i}}, \quad \sum_{\boldsymbol{i}=0}^{\infty} T_{\boldsymbol{i}}(\widehat{F})/\boldsymbol{n}^{-\boldsymbol{i}}, \quad \sum_{\boldsymbol{r}=0}^{\infty} \widehat{V}_{\boldsymbol{r}}/\boldsymbol{r!}.$$

We show that the truncated forms

$$S_{\boldsymbol{np}}(\widehat{F}) = \sum_{\boldsymbol{i}=0}^{\boldsymbol{p}-\boldsymbol{1}} S_{\boldsymbol{i}}(\widehat{F})/(\boldsymbol{n}-\boldsymbol{1})_{\boldsymbol{i}}, \quad T_{\boldsymbol{np}}(\widehat{F}) = \sum_{\boldsymbol{i}=0}^{\boldsymbol{p}-\boldsymbol{1}} T_{\boldsymbol{i}}(\widehat{F})/\boldsymbol{n}^{-\boldsymbol{i}}, \quad V_{\boldsymbol{np}}(\widehat{F}) = \sum_{\boldsymbol{r}=0}^{2\boldsymbol{p}-\boldsymbol{2}} \widehat{V}_{\boldsymbol{r}}/\boldsymbol{r!},$$

all have bias $O(n_1^{-p_1} + \cdots + n_k^{-p_k})$ as $\boldsymbol{n} \to \infty$.

2 The V Form of Unbiased Estimate

2.1 One Sample

Let us first consider the case of *one* distribution F on R^s and one sample $X_1, \ldots, X_n$. For G any distribution on R^s the von Mises-Taylor expansion of $T(F)$ about $T(G)$ is

$$T(F) = \sum_{r=0}^{\infty} V_r(F, G)/r!, \tag{2.1}$$

where $V_0(F, G) = T(G)$ and $V_r(F, G) = \int \cdots \int T_G(x_1 \cdots x_r) dF(x_1) \cdots dF(x_r)$ for $r \geq 1$, and $T_G(x_1 \cdots x_r)$ is the rth order (von Mises) derivative of $T(G)$, uniquely defined by (2.1) subject to the constraints that $T_G(x_1 \cdots x_r)$ is not altered by permuting $x_1, \ldots, x_r$, and $\int T_G(x_1 \cdots x_r) dG(x_1) = 0$.

The first derivative or *influence function* of $T(G)$ is just $T_G(x) = \lim_{\epsilon \to 0}(T(G_\epsilon) - T(G))/\epsilon$, where $G_\epsilon(y) = (1-\epsilon)G(y) + \epsilon I(y \leq x)$ and $I(A)$ is 1 or 0 for A true or false. A simple method of obtaining

$T_G(x_1 \cdots x_r)$ from $T_G(x_1 \cdots x_{r-1})$ was given in Withers (1983). For example, $S(G) = T_G(x)$ has derivative $S_G(y) = T_G(x, y) - T_G(y)$ so this gives the second derivative of $T(G)$ in terms of its first derivative. Similarly, $S(G) = T_G(x, y)$ has derivative $S_G(z) = T_G(x, y, z) - T_G(z, y) - T_G(x, z)$ so this gives the third derivative. The analogous formula holds for the derivative of the general rth order derivative, with r 'correction' terms subtracted.

For fixed G an UE of $V_r(F, G)$ is the U-statistic

$$V_r^n(\widehat{F}, G) = \sum_r T_G(X_{i_1} \cdots X_{i_r})/(n)_r, \tag{2.2}$$

where $\sum_r$ sums over all $(n)_r$ permutations of distinct $i_1, \ldots, i_r$ in $1, \ldots, n$. So,

$$
\begin{aligned}
nV_1^n(\widehat{F}, G) &= \sum_{i=1}^n T_G(X_i) = n \int T_G(x) d\widehat{F}(x), \\
(n)_2 V_2^n(\widehat{F}, G) &= \sum_{i \neq j}^n T_G(X_i, X_j) = \left(\sum_{i,j=1}^n - \sum_{i=j=1}^n \right) T_G(X_i, X_j) \\
&= n^2 \int \int T_G(x_1, x_2) d\widehat{F}(x_1) d\widehat{F}(x_2) - n \int T_G(x_1^2) d\widehat{F}(x_1),
\end{aligned}
$$

where $T_G(x_1^2) = T_G(x_1, x_1)$, and so on.

Note that $V_r^n(\widehat{F}, G)$ can be written down using the tables of the symmetric polynomials in Stuart and Ord (1987, Appendix Table 10, page 554) for $r \leq 6$ and David and Kendall (1949) for $r \leq 12$. The last column of these tables expresses the symmetric polynomials $[1^r] = \sum_r x_{i_1} \cdots x_{i_r}$, where $\sum_r$ sums over distinct $i_1, \ldots, i_r$ in $1, \ldots, n$ say, in terms of the power sum functions $(j) = \sum_{i=1}^n x_i^j$ for $1 \leq j \leq r$. For example, $[1^4] = -6(4) + 8(31) + 3(2^2) - 6(21^2) + (1^4)$, where for $\pi = (\pi_1, \ldots, \pi_m)$, $(\pi) = (\pi_1) \cdots (\pi_m)$. In general

$$[1^r] = \sum_\pi^r (\pi) c(\pi), \tag{2.3}$$

where $\sum_\pi^r$ sums over all partitions π of r, and the numerical coefficients $c(\pi)$ are provided in the last column of the tables. The MAPLE procedure symmpoly(...) in Appendix B expresses a given number of symmetric polynomials in terms of power sum functions.

However, the relation (2.3) remains true if $[1^r]$ is redefined as $\sum_r t(x_{i_1}, \ldots, x_{i_r})$ and $(\pi) = (\pi_1 \cdots \pi_m)$ is redefined as $\sum_{i_1=1}^n \cdots \sum_{i_m=1}^n t(x_{i_1}^{\pi_1}, \ldots, x_{i_m}^{\pi_m})$, where $x_1^{\pi_1}$ denotes π_1 arguments equal to x_1, not a power, and t is an arbitrary symmetric polynomial of r variables. In particular taking $x_i \equiv X_i$ and $t \equiv T_G$, $[1^r] = (n)_r V_r^n(\widehat{F}, G)$, and $(\pi) = T_{\widehat{F}G}[\pi] n^{m(\pi)}$, where $m = m(\pi)$ is the number of elements in π, and

$$T_{\widehat{F}G}[\pi] = \int \cdots \int T_G(x_1^{\pi_1} \cdots x_m^{\pi_m}) d\widehat{F}(x_1) \cdots d\widehat{F}(x_m). \tag{2.4}$$

So, (2.3) can be written

$$V_r^n(\widehat{F}, G) = (n)_r^{-1} \sum_\pi^r c(\pi) T_{\widehat{F}G}[\pi] n^{m(\pi)},$$

and an UE of $T(F)$ is $\sum_{r=0}^{\infty} V_r^n(\widehat{F}, G)/r!$. Since G is arbitrary we may now take $G = \widehat{F}$ and set $\widehat{V}_r = V_r^n(\widehat{F}, \widehat{F})$, $\widehat{T}[\pi] = T_{\widehat{F}\widehat{F}}[\pi]$, $T[\pi] = T_{FF}[\pi]$. For example, for second order estimates we shall need $\widehat{T}[2] = \int T_{\widehat{F}}(x, x) d\widehat{F}(x)$.

Our expression above for LHS (2.2) at $G = \widehat{F}$ now yields the following V form of UE for $T(F)$:

$$\sum_{r=0}^{\infty} \widehat{V}_r/r!, \tag{2.5}$$

where

$$\widehat{V}_r = (n)_r^{-1} \sum_{\pi}^{\sim r} c(\pi) \widehat{T}[\pi] n^{m(\pi)}, \tag{2.6}$$

and $\sum_{\pi}^{\sim r}$ is $\sum_{\pi}^{r}$ excluding partitions containing 1 because of the constraint $\int T_{\widehat{F}}(x_1 \cdots x_r) d\widehat{F}(x_1) = 0$. The first few V_r are

$$
\begin{aligned}
\widehat{V}_0 &= T(\widehat{F}), \ \widehat{V}_1 = 0, \ \widehat{V}_2 = -(n-1)^{-1}\widehat{T}[2], \\
\widehat{V}_3 &= 2(n-1)_2^{-1}\widehat{T}[3], \\
\widehat{V}_4 &= (n-1)_3^{-1}\{-6\widehat{T}[4] + 3\widehat{T}[2^2]n\}, \\
\widehat{V}_5 &= 2(n-1)_4^{-1}\{24\widehat{T}[5] - 20\widehat{T}[32]n\}, \\
\widehat{V}_6 &= (n-1)_5^{-1}\{-120\widehat{T}[6] + 90\widehat{T}[42]n + 40\widehat{T}[3^2]n - 15\widehat{T}[2^3]n^2\}.
\end{aligned}
$$

Using the $O_p(.)$ notation of Mann and Wald (1943), since $\widehat{V}_{2r-1}$ and $\widehat{V}_{2r}$ are $O_p(n^{-r})$, the UE (2.5) is $V_{n:2p-2}(\widehat{F}) + O_p(n^{-p})$, where

$$V_{n:j}(\widehat{F}) = \sum_{r=0}^{j} \widehat{V}_r/r!.$$

Note that $V_{n:2p-2}(\widehat{F})$ estimates $T(F)$ with bias $O(n^{-p})$: $V_{n:2}(\widehat{F}) = T(\widehat{F}) - (n-1)^{-1}\widehat{T}[2]/2$ has bias $O(n^{-2})$, $V_{n:4}(\widehat{F}) = V_{n:2}(\widehat{F}) + (n-1)_2^{-1}\widehat{T}[3]/3 + (n-1)_3^{-1}(-6\widehat{T}[4] + 3\widehat{T}[2^2]n)/24$ has bias $O(n^{-3})$, and so on. The MAPLE procedures Vestsum(...) and Vest(...) in Appendix B calculate (2.6) for any r and hence $V_{n:2p-2}$ for any p, so estimates of bias of any order can be obtained.

If $T(F)$ is a polynomial in F of degree p, for example, $\mu_p(F)$, $\kappa_p(F)$ or $\mu(F)^p$ then $T_G(x_1 \cdots x_r) = 0$ for $r > p$ so that $V_{np}(\widehat{F})$ is an UE.

2.2 More than One Sample

Now consider the case of $k \geq 1$ distributions, with k samples $\{X_{1j}, \ldots, X_{n_j j}\}, 1 \leq j \leq k$. The von Mises–Taylor expansion of $T(F)$ about $T(G)$ for $G = (G_1, \ldots, G_k)$ distributions on $R^{s_1}, \ldots, R^{s_k}$ is

$$T(F) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} V_{r_1 \cdots r_k}(F, G)/(r_1! \cdots r_k!) = \sum_{\boldsymbol{r}=\boldsymbol{0}}^{\infty} V_{\boldsymbol{r}}(F, G)/\boldsymbol{r}!, \tag{2.7}$$

where

$$
\begin{aligned}
V_{\boldsymbol{r}}(F, G) &= \int \cdots \int T_G \left(\begin{array}{ccccccc} 1\,1 & \cdots & 1 & \cdots & k\,k & \cdots k \\ x_1 x_2 & \cdots & x_{r_1} & \cdots & z_1 z_2 & \cdots z_{r_k} \end{array} \right) \\
&\quad dF_1(x_1) \cdots dF_1(x_{r_1}) \cdots dF_1(z_1) \cdots dF_k(z_{r_k})
\end{aligned}
$$

6

and $T_G\begin{pmatrix} a_1 \cdots a_r \\ x_1 \cdots x_r \end{pmatrix}$ is the partial (von Mises) derivative, defined for $a_1, \ldots, a_r$ in $1, \ldots, k$ and x_i in $R^{s_{a_i}}$. These were introduced in Withers (1988). They are determined uniquely by (2.7) and the two constraints

$$T_G\begin{pmatrix} a_1 & \cdots & a_r \\ x_1 & \cdots & x_r \end{pmatrix}$$

is not altered by permuting columns, and

$$\int T_G\begin{pmatrix} a_1 & \cdots & a_r \\ x_1 & \cdots & x_r \end{pmatrix} dF_{G_{a_1}}(x_1) = 0.$$

In practice they are determined by adapting the rules given above for the one distribution derivatives, namely: $T_G\begin{pmatrix} a \\ x \end{pmatrix}$ is just $S_{G_a}(x)$ for $S(G_a) = T(G)$, $U(G_b) = T_G\begin{pmatrix} a \\ x \end{pmatrix}$ has derivative

$$U_{G_b}(y) = T_G\begin{pmatrix} a & b \\ x & y \end{pmatrix} - T_G\begin{pmatrix} a \\ x \end{pmatrix}$$

and similarly for the derivative of the general derivative.

The term $V_{\mathbf{0}}(F, G)$ is interpreted as $T(G)$. For more details and examples see Withers (1988, 1994a). For a given G, an UE of $V_{\mathbf{r}}(F, G)$ is

$$V_{\mathbf{r}}^{\mathbf{n}}(\widehat{F}, G) = \sum_{r_1 \cdots r_k} T_G\begin{pmatrix} \overset{1\cdots1}{X_{i_1 1} \cdots X_{i_{r_1} 1}} & \cdots & \overset{k \cdots k}{X_{j_1 k} \cdots X_{j_{r_k} k}} \end{pmatrix} / (\mathbf{n})_{\mathbf{r}},$$

where $\sum_{r_1 \cdots r_k}$ sums over all $(n_1)_{r_1}$ permutations of distinct $i_1, \ldots, i_{r_1}$ in $\{1, \ldots, n_1\}, \ldots$ and all $(n_k)_{r_k}$ permutations of distinct $j_1, \ldots, j_{r_k}$ in $\{1, \ldots, n_k\}$. So,

$$\sum_{\mathbf{r}=0}^{\infty} V_{\mathbf{r}}^{n}(\widehat{F}, G) / \mathbf{r}!$$

is an UE of $T(F)$. Taking $G = \widehat{F}$,

$$\sum_{\mathbf{r}=0}^{\infty} \widehat{V}_{\mathbf{r}} / \mathbf{r}!$$

is a UE of $T(F)$, where $\widehat{V}_{\mathbf{r}} = V_{\mathbf{r}}^{\mathbf{n}}(\widehat{F}, \widehat{F})$. For $i_1, \ldots, i_k$ in $\{0, 1\}, \widehat{V}_{2\mathbf{r}-\mathbf{i}}$ is $O_p(n_1^{-r_1} \cdots n_k^{-r_k})$. So, $V_{\mathbf{n}:2\mathbf{p}-\mathbf{2}}(\widehat{F})$ estimates $T(F)$ with bias $O(n_1^{-p_1} + \cdots + n_k^{-p_k}) = O(n_0^{-p_0})$, where

$$V_{\mathbf{n}:\mathbf{j}}(\widehat{F}) = \sum_{\mathbf{r}=0}^{\mathbf{j}} \widehat{V}_{\mathbf{r}} / \mathbf{r}!, \quad n_0 = \min_{1 \le i \le k} n_i, p_0 = \min_{1 \le i \le k} p_i.$$

However for $p_0 > 1$, an estimate of $T(F)$ of bias $O(n_0^{-p_0})$ with fewer terms than $V_{\mathbf{n}:2p_0-\mathbf{2}}$ is

$$V_{\mathbf{n}p_0}(\widehat{F}) = \sum_{r_0=0}^{2p_0-2} \widehat{V}_{r_0} / r_0!, \tag{2.8}$$

where

$$\widehat{V}_{r_0} = r_0! \sum_{r_1 + \cdots + r_k = r_0} \widehat{V}_{\mathbf{r}} / \mathbf{r}!.$$

7

If $T(F)$ is a polynomial of degree p_1 in $F_1, \ldots, p_k$ in F_k, then the UE reduces to the finite sum $V_{\boldsymbol{n:p}}(\widehat{F})$, containing $(p_1 + 1) \cdots (p_k + 1)$ terms. This is because $\widehat{V}_{\boldsymbol{r}} = 0$ if $r_1 > p_1$ or $\cdots$ or $r_k > p_k$. Set

$$T_F(i^r j^s \cdots) = \int \int \cdots T_F \left(\begin{array}{ccc} i^r & j^s & \\ x^r & y^s & \cdots \end{array} \right) dF_j(x) dF_j(y) \cdots ,$$

where the first column in the integrand stands for r repeated columns of $\binom{i}{x}$, and similarly for the other columns. Set

$$T_{FG}[\pi_1, \ldots, \pi_k] = T_{FG}(1^{\pi_{11}} \cdots 1^{\pi_{m_1 1}} \cdots 1^{\pi_{1k}} \cdots 1^{\pi_{m_k k}})$$

for $\pi_1 = (\pi_{11} \cdots \pi_{m_1 1}), \ldots, \pi_k = (\pi_{1k} \cdots \pi_{m_k k})$. (So, m_k is the length of the vector π_k.) Set

$$\widehat{T}[\pi_1 \cdots \pi_k] = T_{\widehat{F}\widehat{F}}[\pi_1 \cdots \pi_k], \ \widehat{T}(i^r j^s \cdots) = T_{\widehat{F}\widehat{F}}(i^r j^s \cdots), \ T(i^r j^s \cdots) = T_{FF}(i^r j^s \cdots).$$

Analogous to (2.5) we have

$$\widehat{V}_{\boldsymbol{r}} = (\boldsymbol{n})_{\boldsymbol{r}}^{-1} \sum_{\pi_1}^{\sim r_1} \cdots \sum_{\pi_k}^{\sim r_k} c(\pi_1) \cdots c(\pi_k) \widehat{T}[\pi_1, \ldots, \pi_k] n_1^{m(\pi_1)} \cdots n_k^{m(\pi_k)}. \tag{2.9}$$

The MAPLE procedures Vestsum(...) and Vest(...) in Appendix B calculate (2.9) for any $\boldsymbol{r}$ and hence (2.8) for any p_0, so estimates of bias of any order can be obtained. Let e_i be the ith unit vector in R^k. Then the first six $\widehat{V}_{r_0}$ of (2.8) are

$$\widehat{V}_0 = T(\widehat{F}),$$

$$\widehat{V}_1 = \sum_{i=1}^{k} \widehat{V}_{e_i} = 0,$$

$$\widehat{V}_2 = \sum_{i=1}^{k} \widehat{V}_{2e_i} = -\sum_{i=1}^{k} T_{\widehat{F}}(i^2)/(n_i - 1),$$

since $\widehat{V}_{e_i + e_j} = 0$ for $i \neq j$,

$$\widehat{V}_3 = \sum_{i=1}^{k} \widehat{V}_{3e_i} = 2\sum_{i=1}^{k} \widehat{T}(i^3)/(n_i - 1)_2,$$

$$\begin{aligned}
\widehat{V}_4 &= \sum_{i=1}^{k} \widehat{V}_{4e_i} + \binom{4}{2} \sum_{1 \leq i < j \leq k} \widehat{V}_{2e_i + 2e_j} \\
&= \sum_{i=1}^{k} \{-6\widehat{T}(i^4) + 3\widehat{T}(i^2 i^2) n_i\}/(n_i - 1)_3 + 6\sum_{1 \leq i < j \leq k} \widehat{T}(i^2 j^2)/\{(n_i - 1)(n_j - 1)\},
\end{aligned}$$

$$\begin{aligned}
\widehat{V}_5 &= \sum_{i=1}^{k} \widehat{V}_{5e_i} + \binom{5}{2} \sum_{i \neq j} \widehat{V}_{3e_i + 2e_j} \\
&= \sum_{i=1}^{k} \{24\widehat{T}(i^5) - 20\widehat{T}(i^3 i^2) n_i\}/(n_i - 1)_4 - 20\sum_{i \neq j} \widehat{T}(i^3 j^2)/\{(n_i - 1)_2 (n_j - 1)\}
\end{aligned}$$

8

and

$$
\begin{aligned}
\widehat{V}_6 &= \sum_{i=1}^{k} \widehat{V}_{6e_i} + \binom{6}{2} \sum_{i \neq j} \widehat{V}_{4e_i+2e_j} + \binom{6}{3} \sum_{1 \leq i < j \leq k} \widehat{V}_{3e_i+3e_j} + \binom{6}{222} \sum_{1 \leq i < j < l \leq k} \widehat{V}_{2e_i+2e_j+2e_l} \\
&= \sum_{i=1}^{k} \left\{ -120\widehat{T}(i^6) + 90\widehat{T}(i^4 i^2)n_i + 40\widehat{T}(i^3 i^3)n_i - 15\widehat{T}(i^2 i^2 i^2)n_i^2 \right\} / (n_i - 1)_5 \\
&\quad -15 \sum_{i \neq j} \{ -6\widehat{T}(i^4 j^2) + 3\widehat{T}(i^2 i^2 j^2)n_i / \} \{ (n_i - 1)_3 (n_j - 1) \} \\
&\quad +80 \sum_{1 \leq i < j \leq k} \widehat{T}(i^3 j^3) / \{ (n_i - 1)_2 (n_j - 1)_2 \} \\
&\quad -90 \sum_{1 \leq i < j < l \leq k} \widehat{T}(i^2 j^2 l^2) / \{ (n_i - 1)(n_j - 1)(n_l - 1) \}.
\end{aligned}
$$

By (2.8) the above formulas give $V_{n4}(\widehat{F}) = \sum_{r=0}^{6} \widehat{V}_r / r!$ as an estimate of $T(F)$ with bias $O(n_0^{-4})$, where $n_0 = \min_{1 \leq i \leq k} n_i$.

3   The S Estimate

Here we derive the S form of UE, (1.1), from the V estimate, (1.3).

Suppose first that $k = 1$, (that is, univariate data). Then for $r \geq 2$, $(n)_r \widehat{V}_r$ is a polynomial in $n = n_1$ of degree $[r/2]$, the integral part of $r/2$, say:

$$
\widehat{V}_r = (n)_r^{-1} \sum_{1 \leq i \leq r/2} U_{r,i}(\widehat{F}) n^i = (n-1)_{r-1}^{-1} \sum_{0 \leq j \leq r/2-1} U_{r,j+1}(\widehat{F}) n^j
$$

for $r \geq 2$. Writing

$$
n^j = \sum_{k=0}^{j} [n - r + 1]_k\, c_{kjr}, \tag{3.1}
$$

where $[a]_k = a(a + 1) \cdots (a + k - 1)$, we obtain

$$
\widehat{V}_r = \sum_{r/2 \leq i < r} V_{ri}(\widehat{F}) / (n - 1)_i,
$$

where

$$
V_{ri}(F) = \sum_{0 \leq j \leq r/2-1} c_{r-i-1,jr} U_{r,j+1}(F).
$$

The MAPLE procedure cc(...) in Appendix B calculates the coefficients c_{kjr} in (3.1) for any given k, j and r. It follows that the UE is

$$
\widehat{V}_0 + \sum_{r=2}^{\infty} \widehat{V}_r / r! = \sum_{i=0}^{\infty} S_i(\widehat{F}) / (n - 1)_i,
$$

9

where $S_0(\widehat{F}) = \widehat{V}_0 = T(\widehat{F})$ and

$$S_i(F) = \sum_{r=i+1}^{2i} V_{ri}(F)/r! \tag{3.2}$$

for $i \geq 1$. So,

$$S_{np}(\widehat{F}) = \sum_{i=0}^{p-1} S_i(\widehat{F})/(n-1)_i \tag{3.3}$$

has bias $O(n^{-p})$. The MAPLE procedure Sest(...) in Appendix B can be used to calculate (3.3) to obtain a bias of any order. Using (3.2), the first few $S_i(F)$ can be shown to be

$$\begin{aligned}
S_1(F) &= -T[2]/2, &\text{(3.4)}\\
S_2(F) &= T[3]/3 + T[2^2]/8, &\text{(3.5)}\\
S_3(F) &= -T[4]/4 + 3T[2^2]/8 - T[32]/6 - T[2^3]/48, &\text{(3.6)}\\
S_4(F) &= T[5]/5 - 2T[32]/3 - 3T[2^3]/16 + T[42]/8 + T[3^2]/18 \\
&\quad + T[32^2]/24 + T[2^4]/384, &\text{(3.7)}\\
S_5(F) &= -T[6]/6 + 5T[42]/8 + 5T[3^2]/18 - T[52]/10 - T[43]/12 + 3T[2^4]/64 \\
&\quad - T[42^2]/32 - T[3^2 2]/36 - T[32^3]/144 - T[2^5]/3840, &\text{(3.8)}\\
S_6(F) &= T[7]/7 - 3T[52]/5 - T[43]/2 + T[32^2]/140 + 127T[2^4]/64 \\
&\quad - 13T[42^2]/32 - 377T[3^2 2]/1008 + T[62]/12 + T[53]/15 + T[4^2]/32 \\
&\quad - T[32^3]/48 + T[52^5]/40 + T[432]/24 + T[3^3]/324 - T[2^5]/160 \\
&\quad + T[3^2 2^2]/144 + T[42^3]/192 + T[32^4]/1152 + T[2^6]/46080. &\text{(3.9)}
\end{aligned}$$

So, $S_{n2}(\widehat{F}) = T(\widehat{F}) - (n-1)^{-1}T[2]/2$ has bias $O(n^{-2})$, $S_{n3}(\widehat{F}) = S_{n2}(\widehat{F}) + (n-1)_2^{-1}\{T[3]/3 + T[2^2]/8\}$ has bias $O(n^{-3})$, and so on.

Now suppose $k \geq 1$, (that is, multivariate data). The UE

$$\sum_{\boldsymbol{r}=\boldsymbol{0}}^{\infty} \widehat{V}_{\boldsymbol{r}}/\boldsymbol{r}! = \sum_{r_0=0}^{\infty} \widehat{V}_{r_0}/r_0!$$

can be written

$$\sum_{\boldsymbol{i}=\boldsymbol{0}}^{\infty} S_{\boldsymbol{i}}(\widehat{F})/(\boldsymbol{n}-\boldsymbol{1})_{\boldsymbol{i}} = \sum_{i_0=0}^{\infty} S_{i_0}^{\boldsymbol{n}}(\widehat{F}),$$

where

$$S_{i_0}^{\boldsymbol{n}}(F) = \sum_{\boldsymbol{i'1}=i_0} S_{\boldsymbol{i}}(F)/(\boldsymbol{n}-\boldsymbol{1})_{\boldsymbol{i}} = O(n_0^{-i_0})$$

for $n_0 = \min_{1 \leq i \leq k} n_i$. An estimate of $T(F)$ of bias $O(n_0^{-p_0})$ is

$$S_{\boldsymbol{n}p_0}(\widehat{F}) = \sum_{i_0=0}^{p_0-1} S_{i_0}^{\boldsymbol{n}}(\widehat{F}). \tag{3.10}$$

10

The MAPLE procedure Sest(...) in Appendix B can be used to calculate (3.10) to obtain a bias of any order. For example, to obtain an estimate of bias $O(n_0^{-4})$ one can show that

$$S_0^{\boldsymbol{n}}(F) \quad = \quad T(F), \tag{3.11}$$

$$S_1^{\boldsymbol{n}}(F) \quad = \quad -(1/2)\sum_{i=1}^{k} T(i^2)/(n_i - 1), \tag{3.12}$$

$$S_2^{\boldsymbol{n}}(F) \quad = \quad S_{\boldsymbol{n}}^2 + S_{\boldsymbol{n}}^{11}, \tag{3.13}$$

where

$$S_{\boldsymbol{n}}^2 \quad = \quad \sum_{i=1}^{k} \{T(i^3)/3 + T(i^2 i^2)/8\}(n_i - 1)_2,$$

$$S_{\boldsymbol{n}}^{11} \quad = \quad (1/4) \sum_{1 \le i < j \le k} T(i^2 j^2)\{(n_i - 1)(n_j - 1)\}$$

and

$$S_3^{\boldsymbol{n}}(F) \quad = \quad S_{\boldsymbol{n}}^3 + S_{\boldsymbol{n}}^{21} + S_{\boldsymbol{n}}^{111}, \tag{3.14}$$

where

$$S_{\boldsymbol{n}}^3 \quad = \quad \sum_{i=1}^{k} \{-T(i^4)/4 + 3T(i^2 i^2)/8 - T(i^3 i^2)/6 - T(i^2 i^2 i^2)/48\}/(n_i - 1)_3,$$

$$S_{\boldsymbol{n}}^{21} \quad = \quad (-1/2)\sum_{i \ne j} \{T(i^3 j^2)/3 + T(i^2 i^2 j^2)/8\}/\{(n_i - 1)_2(n_j - 1)\},$$

$$S_{\boldsymbol{n}}^{111} \quad = \quad (-1/2)^3 \sum_{1 \le i < j < l \le k} T(i^2 j^2 l^2)/\{(n_i - 1)(n_j - 1)(n_l - 1)\}.$$

In general

$$S_r^{\boldsymbol{n}}(F) = \sum_{\pi}^{r} S_{\boldsymbol{n}}^{\pi}, \tag{3.15}$$

where for $\pi = (i_1, i_2, \ldots)$, $S_{\boldsymbol{n}}^{i_1 i_2 \cdots}$ can be written down from the formulas for $S_{i_1}(F), S_{i_2}(F), \ldots$: let us write

$$S_i(F) = \sum_{\pi}^{*i} d_{i\pi} T[\pi], \tag{3.16}$$

where $\sum_{\pi}^{*i}$ sums over partitions π of $\{i + 1, \ldots, 2i\}$. (Therefore many $d_{i\pi}$ are zero.) Then

$$S_{\boldsymbol{n}}^{i_1 \cdots i_m}(F) = \sum_{\boldsymbol{\pi}}^{*i_1} \cdots \sum_{\boldsymbol{\pi}}^{*i_m} d_{i_1 \pi_1} \cdots d_{i_m \pi_m} \sum_{j_1 \cdots j_m}^{k} T(j_1^{\pi_1} \cdots j_m^{\pi_m})/\{(n_{j_1} - 1)_{i_1} \cdots (n_{j_m} - 1)_{i_m}\},$$

where $j^{\pi} = j^{\pi_1} j^{\pi_2} \cdots$ and $\sum_{j_1 \cdots j_m}^{k}$ sums over $j_1, \ldots, j_m$ distinct in $1, \ldots, k$ with $j_i < j_{i+1}$ if $\pi_i = \pi_{i+1}$. For example,

$$S_4^{\boldsymbol{n}}(F) = S_{\boldsymbol{n}}^4 + S_{\boldsymbol{n}}^{31} + S_{\boldsymbol{n}}^{22} + S_{\boldsymbol{n}}^{211} + S_{\boldsymbol{n}}^{1111}, \tag{3.17}$$

11

where by (3.7),

$$S_{\boldsymbol{n}}^4 = \sum_{i=1}^{k} \{ T(i^5)/5 - 2T(i^3 i^2)/3 - 3T(i^2 i^2 i^2)/16$$
$$+ T(i^4 i^4)/8 + T(i^3 i^3)/18 + T(i^3 i^2 i^2)/24 + T(i^2 i^2 i^2 i^2)/384 \} /(n_i - 1)_4,$$

and by (3.4), (3.6),

$$S_{\boldsymbol{n}}^{31} = (-1/2) \sum_{i \neq j} \{ -T(i^3 j^2)/4 - 3T(i^2 i^2 j^2)/18 - T(i^3 i^2 j^2)/6$$
$$- T(i^2 i^2 i^2 j^2)/48 \}/\{ (n_i - 1)_3 (n_j - 1) \},$$

and by (3.5),

$$S_{\boldsymbol{n}}^{22} = \sum_{i \leq i < j \leq k} \{ T(i^3 j^3)/9 + T(i^2 i^2 j^2 j^2)/64 \}/\{ (n_i - 1)_2 (n_j - 1)_2 \}$$
$$+ 2(1/3)(1/8) \sum_{i \neq j} T(i^3 j^2 j^2)/\{ (n_i - 1)_2 (n_j - 1) \},$$

and by (3.4), (3.5),

$$S_{\boldsymbol{n}}^{211} = (-1/2)^2 \sum_{i \neq j, i \neq l, j < l} \{ -T(i^3 j^2 l^2)/3 + T(i^2 i^2 j^2 l^2)/8 \}/\{ (n_i - 1)_2 (n_j - 1)(n_l - 1) \}$$

and by (3.4),

$$S_{\boldsymbol{n}}^{1111} = (-1/2)^4 \sum_{1 \leq i < j < l < m \leq k} T(i^2 j^2 l^2 m^2)/\{ (n_i - 1)(n_j - 1)(n_l - 1)(n_m - 1) \}.$$

So, (3.10), (3.4)–(3.9) provide the S-estimate of bias $O(n_0^{-6})$.

4 The T Estimate

The T form of the UE, $\sum_{\boldsymbol{i}=0}^{\infty} T_{\boldsymbol{i}}(\widehat{F}) \boldsymbol{n}^{-\boldsymbol{i}}$, is easily derived from the S estimate, but is less useful: its truncated form cannot be an UE for $T(F)$ a polynomial in F, and the number of components in $T_{\boldsymbol{i}}(F)$ rapidly increases over the number in $S_{\boldsymbol{i}}(F)$ as $\boldsymbol{i}$ increases. Since

$$\prod_{j=1}^{i} (1 - j\epsilon)^{-1} = \sum_{\beta=0}^{\infty} \epsilon^\beta G_{\beta i},$$

we have

$$G_{\beta i} = \sum_{\alpha_1 + \cdots + \alpha_i = \beta} 1^{\alpha_1} \cdots i^{\alpha_i}, \tag{4.1}$$

and

$$(n - 1)_i^{-1} = \sum_{\alpha=i}^{\infty} n^{-\alpha} D_{\alpha i},$$

12

where

$$D_{\alpha i} = G_{\alpha - i, i}. \tag{4.2}$$

The MAPLE procedures gbetai(...) and dalphai(...) in Appendix B calculate (4.1) and (4.2), respectively. Set $D_{\boldsymbol{\alpha i}} = D_{\alpha_1 i_1} \cdots D_{\alpha_k i_k}$, so $D_{\boldsymbol{i+\beta, i}} = G_{\beta_1 i_1} \cdots G_{\beta_k i_k}$. Then

$$(\boldsymbol{n} - \boldsymbol{1})_{\boldsymbol{i}}^{-1} = \sum_{\boldsymbol{\alpha} = \boldsymbol{i}}^{\infty} D_{\boldsymbol{\alpha i}} \boldsymbol{n}^{-\boldsymbol{\alpha}}.$$

So, the UE is

$$\sum_{\boldsymbol{i}=\boldsymbol{0}}^{\infty} S_{\boldsymbol{i}}(\widehat{F})/(\boldsymbol{n} - \boldsymbol{1})_{\boldsymbol{i}} = \sum_{\boldsymbol{\alpha}=\boldsymbol{0}}^{\infty} T_{\boldsymbol{\alpha}}(\widehat{F}) \boldsymbol{n}^{-\boldsymbol{\alpha}} = \sum_{r=0}^{\infty} T_r^{\boldsymbol{n}}(\widehat{F}),$$

where

$$T_{\boldsymbol{\alpha}}(F) = \sum_{\boldsymbol{i}=\boldsymbol{0}}^{\boldsymbol{\alpha}} D_{\boldsymbol{\alpha i}} S_{\boldsymbol{i}}(F)$$

and

$$T_r^{\boldsymbol{n}}(F) = \sum_{\{\boldsymbol{\alpha'1} = r\}} T_{\boldsymbol{\alpha}}(F) \boldsymbol{n}^{-\boldsymbol{\alpha}} = O(n_0^{-r}).$$

Note that

$$T_{\boldsymbol{n}p}(\widehat{F}) = \sum_{r=0}^{p-1} T_r^{\boldsymbol{n}}(\widehat{F}) \tag{4.3}$$

has bias $O_p(n_0^{-p})$. The MAPLE procedure Test(...) in Appendix B can be used to calculate (4.3) to obtain a bias of any order. The first few $\{T_r^{\boldsymbol{n}}(F)\}$ are

$$T_0^{\boldsymbol{n}}(F) = T(F),$$

$$T_1^{\boldsymbol{n}}(F) = -(1/2) \sum_{i=1}^{k} T(i^2) n_i^{-1},$$

$$
\begin{aligned}
T_2^{\boldsymbol{n}}(F) &= \sum_{i=1}^{k} \{-T(i^2)/2 + T(i^3)/3 + T(i^2 i^2)/8\}/(n_i - 1)_2 \\
&\quad + (1/4) \sum_{i<j} T(i^2 j^2)/\{(n_i - 1)(n_j - 1)\},
\end{aligned}
$$

$$
\begin{aligned}
T_3^{\boldsymbol{n}}(F) &= \sum_{i=1}^{k} \{-T(i^2)/2 + T(i^3) + 3T(i^2 i^2)/4 - T(i^4)/4 \\
&\quad -T(i^3 i^2)/6 - T(i^2 i^2 i^2)/48\} n_i^{-3} \\
&\quad + \sum_{i \neq j} \{T(i^2 j^2)/4 - T(i^3 j^2)/6 - T(i^2 i^2 j^2)/16\} n_i^{-2} n_j^{-1} \\
&\quad - \sum_{i<j<l} T(i^2 j^2 l^2) n_i^{-1} n_j^{-1} n_l^{-1}/8
\end{aligned}
$$

and

$$
\begin{aligned}
T_4^{\boldsymbol{n}}(F) \;=\; & \sum_{i=1}^{k}\{-T(i^2)/2 + 7T(i^3)/3 + 25T(i^2i^2)/8 - 3T(i^4)/2 \\
& -5T(i^3i^2)/3 - 5T(i^2i^2i^2)/16 + T(i^4i^4)/8 \\
& +T(i^3i^3)/18 + T(i^2i^2i^2)/24 + T(i^2i^2i^2i^2)/384\}n_i^{-4} \\
& -(1/2)\sum_{i\neq j}\{-T(i^2j^2)/2 + T(i^3j^2) \\
& -T(i^4j^2)/4 - T(i^3i^2j^2)/6 - T(i^2i^2i^2j^2)/48\}n_i^{-3}n_j^{-1} \\
& +\sum_{i<j}\{T(i^2j^2)/4 + T(i^3j^3)/9 + T(i^3j^2j^2)/12 + T(i^2i^2j^2j^2)/64\}n_i^{-2}n_j^{-2} \\
& +\sum_{i\neq j}\{-T(i^3j^2)/6 - T(i^2i^2j^2)/6\}n_i^{-2}n_j^{-2} \\
& +\sum_{i\neq j,i\neq l,j<l}\{-T(i^2j^2l^2)/8 + T(i^3j^2l^2)/12 + T(i^2i^2j^2l^2)/32\}n_i^{-2}n_j^{-1}n_l^{-1} \\
& +\sum_{i<j<l<m}T(i^2j^2l^2m^2)n_i^{-1}n_j^{-1}n_l^{-1}n_m^{-1}/16.
\end{aligned}
$$

5   Number of Computations Required

As $n_0 = \min_{1\leq i\leq k} n_i \to \infty$, for fixed p the estimators of bias $O(n_0^{-p})$, $V_{\boldsymbol{n}p}(\widehat{F})$ of (2.8), $S_{\boldsymbol{n}p}(\widehat{F})$ of (3.10), and $T_{\boldsymbol{n}p}(\widehat{F})$ of (4.3), all require $O(n_0)$ calculations. This is in sharp contrast with bootstrap estimates of bias $O(n_0^{-p})$, such as that of (1.35) of Hall (1992), which requires $O(n_0^p)$ calculations for the case $k = 1$ given there, and so at least this if generalized to $k > 1$. (Note that Hall's (1.35) should have a factor $(-1)^{i+1}$ inserted.) If one allows p to increase and counts the number of terms needed, then as shown by Table 5.1, $S_{\boldsymbol{n}p}(\widehat{F})$ requires increasingly fewer terms than do $V_{\boldsymbol{n}p}(\widehat{F})$ and $T_{\boldsymbol{n}p}(\widehat{F})$.

[Tables 5.1–5.3 about here.]

The number of terms in an expression of the form $\sum_{1\leq i_1<\cdots<i_m\leq k} a_{\boldsymbol{n},i_1\cdots i_m}$ is $s_m = (k)_m/m!$. So, for general k one obtains the results in Table 5.2. For $k = 2, 3$ this gives the results in Table 5.3.

6   Examples

For $k = 1$ and $p \leq 3$, $S_{np}(F)$ was given in Withers (1994a). For $k > 1$, $S_{n_0p}(\widehat{F})$ and $T_{n_0p}(\widehat{F})$ in Withers (1994a) differ from $S_{\boldsymbol{n}p}(\widehat{F})$ and $T_{\boldsymbol{n}p}(\widehat{F})$ given here. All have bias $O(n_0^{-p})$, but the forms given here are more natural.

In this section we go over some of the examples of Withers (1994a) to obtain S estimates, $S_{\boldsymbol{n}p}(\widehat{F})$, of bias $O(n_0^{-p})$ up to $p = 7$. Recall that for $k = 1$ this is given by

$$
S_{np}(\widehat{F}) = \sum_{i=0}^{p-1} S_i(\widehat{F})/(n-1)_i \tag{6.1}
$$

for $S_i(F)$ of (3.4)–(3.9), and for $k > 1$ by (3.10)–(3.17). We begin with two problems estimating a general function of the means of the distributions after initial given transformations of the data. When this function of the means is a polynomial of total degree p_0 we show that for both problems $S_{np_0}(\widehat{F})$ is an UE. We then give examples estimating functions of central moments. We use the convention that repeated indices are summed over their range. For example, in (6.2) below, $i_1 \cdots j_1 \cdots$ are implicitly summed over $1, \ldots, t$.

Example 6.1 $k = 1, T(F) = g(\mu)$, where $\mu = \int h dF = Eh(X)$ for $X \sim F$, where $h : R^s \to R^t$ and $g : R^t \to R$ are given functions.

Assume that the partial derivatives of $g(\mu)$ are finite: $g_{j_1 \cdots j_r} = \partial_{j_1} \cdots \partial_{j_r} g(\mu)$, where $\partial_j = \partial/\partial_{\mu_j}$. Then the general derivative of $T(F)$ is $T_F(x_1 \cdots x_r) = g_{j_1 \cdots j_r} \mu_{j_1 x_1} \cdots \mu_{j_r x_r}$, where $\mu_{jx} = \mu_{jF}(x) = h_j(x) - \mu_j$. So,

$$T[\pi] = g_{i_1 \cdots i_{\pi_1} j_1 \cdots j_{\pi_2} \cdots} \mu[i_1 \cdots i_{\pi_1}] \mu[j_1 \cdots j_{\pi_2}] \cdots , \qquad (6.2)$$

where $\mu[i_1 i_2 \cdots] = \int (h(x) - \mu)_{i_1} (h(x) - \mu)_{i_2} \cdots dF(x)$, the joint central moment of $h(X)$. Substituting into (3.4)–(3.9) gives $S_i(F)$ for $i \leq 6$ so (6.1) gives an estimate of bias $O(n^{-7})$. Now for $i \geq 1, S_i(F)$ has the form $\sum_{r=i+1}^{2i} g_{j_1 \cdots j_r} c_{ij_1 \cdots j_r}$. So, if $g(\mu)$ is a polynomial of degree $\boldsymbol{p}$ in μ (that is of degree p_i in μ_i for $1 \leq i \leq t$), then summation may be restricted to $j_1, \ldots, j_r$ with $\sum_{i=1}^{r} I(j_i = a) \leq p_a$ for $1 \leq a \leq t$, and $S_i(F) = 0$ for $i \geq |\boldsymbol{p}| = p_1 + \cdots + p_t$ so $S_{n,|\boldsymbol{p}|}(\widehat{F})$ is an UE. So, (6.1), (3.4)–(3.9) gives an UE for all polynomials $g(\mu)$ of total degree seven or less. $\square$

Example 6.1.1 $g(\mu) = N^q$, where $N = \alpha' \mu$ for some t-vector α. If $\pi = (\pi_1, \ldots, \pi_m)$ and $r = \pi_1 + \cdots + \pi_m$ then $T[\pi] = (q)_r N^q \alpha(\pi_1) \cdots \alpha(\pi_m)$, where $\alpha(i) = N^{-i} \alpha_{j_1} \cdots \alpha_{j_i} \mu[j_1 \cdots j_i]$. For example, $\alpha(2) = N^{-2} \alpha' \widehat{V} \alpha$, where $V = covar\{h(X)\}$. So, a second order estimate of $(\alpha' \mu)^q$ is $(\alpha' \widehat{\mu})^q \{1 - \binom{q}{2}(\alpha' \widehat{\mu})^{-2} \alpha' \widehat{V} \alpha/(n-1)\}$, where $\widehat{\mu}, \widehat{V}$ are the sample estimates of $\mu = h(X)$, and V.

[Figures 6.1 and 6.2 about here.]

Now suppose $q = t = 2$, $\alpha = (\alpha_1, 1 - \alpha_1)$ and $V = I_2$. The second order estimate reduces to $\{\alpha_1 \widehat{\mu_1} + (1 - \alpha_1)\widehat{\mu_2}\}^2 - \{\alpha_1^2 + (1 - \alpha_1)^2\}/(n-1)$. If $h(X)$ is bivariate normal with unit means and the given V then the target estimator (Cabrera and Fernholz, 1999, 2004) of $(\alpha' \mu)^2$ is $\{\alpha_1 \widehat{\mu_1} + (1 - \alpha_1)\widehat{\mu_2}\}^2 - \{\alpha_1^2 + (1 - \alpha_1)^2\}/n$. Sen (1988)'s asymptotic normality provides $\widehat{\omega}$, the maximum likelihood estimate of ω given $(\alpha' \widehat{\mu})^2 \sim N(\omega, (3/n)\{\alpha_1^2 + (1 - \alpha_1)^2\}^2 + (4\omega/n)\{\alpha_1^2 + (1 - \alpha_1)^2\})$. Figures 6.1 and 6.2 compare the performance of these three estimates and those obtained by bootstrapping and jackknifing. Our estimate has the lowest absolute bias and the lowest mean squared error. The estimates due to Cabrera and Fernholz (1999, 2004) are so close to ours that they are indistinguishable.

Next suppose $t = \alpha = 1$ so $T(F) = \mu^q$, $\mu = Eh(X)$, $\alpha(i) = \mu^{-i} \mu_i$, $\mu_i = E(h(X) - \mu)^i$ and $T[\pi] = (q)_r \mu^{q-r} \mu_{\pi_1} \cdots \mu_{\pi_m}$. Also for $i \geq 1$,

$$S_i(F) = \sum_{r=i+1}^{2i} (q)_r \mu^{q-r} S_{ir}, \qquad (6.3)$$

where $S_{12} = -\mu_2/2$, $S_{23} = \mu_3/3$, $S_{24} = \mu_2^2/8$, $S_{34} = -\mu_4 + 3\mu_2^2/8$, $S_{35} = -\mu_3\mu_2/6$, $S_{36} = -\mu_2^3/48$, $S_{45} = \mu_5/5 - 2\mu_3\mu_2/3$, $S_{46} = -3\mu_3^2/16 + \mu_4\mu_2/8 + \mu_3^2/18$, $S_{47} = \mu_3\mu_2^2/48$, $S_{48} = \mu_2^4/384$, $S_{56} = -\mu_6/6 + 5\mu_4\mu_2/8 + 5\mu_3^2/18$, $S_{57} = -\mu_5\mu_2/10 - \mu_4\mu_3/12$, $S_{58} = 3\mu_2 4\mu_2^2/32 - \mu_3^2\mu_2/36$, $S_{59} = -\mu_3\mu_2^3/144$, $S_{5,10} = -\mu_2^5/3840$, $S_{67} = \mu_7/7 - 3\mu_5\mu_2/5 - \mu_4\mu_3/2 + \mu_3\mu_2^2/140$, $S_{68} = 127\mu_2^4/64 - 13\mu_4\mu_2^2/32 - 327\mu_3^2\mu_2/100 + \mu_6\mu_2/12 + \mu_5\mu_3/15 + \mu_4^2/32$, $S_{69} = -7\mu_3\mu_2^3/48 + \mu_5\mu_2^2/40 + \mu_4\mu_3\mu_2/24 + \mu_3^3/324$, $S_{6,10} = -\mu_2^5/160 + \mu_3^2\mu_2^2/144 + \mu_4\mu_3^3/192$, $S_{6,11} = \mu_3\mu_2^4/1152$ and $S_{6,12} = \mu_2^6/46080$. This

gives $S_i(F)$ for $i \leq 6$ and so $S_{n7}(\widehat{F})$ of bias $O(n^{-7})$. If q is a positive integer, then $S_i(F) = 0$ for $i \geq q$ so $S_{nq}(\widehat{F})$ of (6.1), (6.3) is an UE of μ^q. For example, an UE of μ^4 is given by (6.1) with $S_1(F) = -6\mu^2\mu_2$, $S_2(F) = 8\mu\mu_3 + 3\mu_2^2$, $S_3(F) = -6\mu_4 + 9\mu_2^2$. This result may be checked by solving the system of seven equations given by Wishart (1952, page 5). Alternatively one may follow the method of Stuart and Ord (1987, Section 12.22) using their tables of the symmetric polynomials. For example, one obtains the UE $T_n(\widehat{F})$ for μ^4, where

$$
\begin{aligned}
(n-1)_3 T_n(F) &= (n^3 - 8n^2 + 23n - 30)m_4 - n(n^2 - 7n + 4)m_3m_1 - n(n^2 - 6n + 6)m_2^2 \\
&\quad + n^2(n-9)m_2m_1^2 + n^3 m_1^4,
\end{aligned}
$$

where $m_i = EX^i$. Clearly our method gives the simpler form. $\square$

Example 6.1.2 $t = 2$, $g(\mu) = \mu_1\mu_2^{-1}$. Then $g_{12^j} = (-1)^j j! \, \mu_2^{-j-1} = a_j$ say, $g_{2^j} = \mu_1 a_j$, and $g_\pi = 0$ unless π is a permutation of 2^j or 12^j for some j. For $\pi = (\pi_1, \pi_2, \ldots)$ set $|\pi| = \pi_1 + \pi_2 + \cdots$ and

$$
\alpha[\pi] = \mu[\pi]/(\mu_{\pi_1}\mu_{\pi_2}\cdots) = E(X_{\pi_1}/\mu_{\pi_1} - 1)(X_{\pi_2}/\mu_{\pi_2} - 1)\cdots .
$$

Then

$$
T[\pi] = \mu_1\mu_2^{-1}(-1)^{|\pi|}\alpha[2^{\pi_1}]\alpha[2^{\pi_2}]\cdots\{|\pi|! - (|\pi| - 1)!\sum_i \pi_i\alpha[12^{\pi_i-1}]/\alpha[2^{\pi_i}]\}.
$$

So,

$$
\begin{aligned}
S_1(F)/T(F) &= \alpha[12] - \alpha[22] = EX_1X_2/(\mu_1\mu_2) - EX_2^2/\mu_2^2, \\
S_2(F)/T(F) &= -2\alpha[222] + 2\alpha[122] + 3\alpha[22]^2 - 3\alpha[12]\alpha[22], \\
S_3(F)/T(F) &= 6\alpha[12^3] - 6\alpha[2^4] - 9\alpha[12]\alpha[22] + 9\alpha[22^2] \\
&\quad + 20\alpha[2^3]\alpha[2^2] - 12\alpha[12^2]\alpha[2^2] - 8\alpha[2^3])\alpha[12] \\
&\quad + 15\alpha[12]\alpha[22]^2 - 15\alpha[22]^3, \\
S_4(F)/T(F) &= 24\alpha[12^4] - 24\alpha[2^5] + 80\alpha[2^3]\alpha[2^2] - 48\alpha[12^2]\alpha[2^2] \\
&\quad - 32\alpha[2^3]\alpha[12] + 45\alpha[22^2](\alpha[12] - \alpha[22]) \\
&\quad + 90\alpha[2^4]\alpha[2^2] - 60\alpha[12^3]\alpha[2^2] - 30\alpha[2^4]\alpha[12] \\
&\quad + 40\alpha[2^3](\alpha[2^3] - \alpha[12^2]) - 210\alpha[2^3]\alpha[2^2]^2 + 90\alpha[12^2]\alpha[2^2]^2 \\
&\quad + 120\alpha[2^3]\alpha[12]\alpha[22] + 105\alpha[22]^3(\alpha[22] - \alpha[12]), \\
S_5(F)/T(F) &= 120\alpha[12^5] - 120\alpha[2^6] + 150\alpha[2^4](3\alpha[22] - \alpha[12]) \\
&\quad - 300\alpha[12^3]\alpha[2^2] + 200\alpha[2^3](\alpha[2^3] - \alpha[12^2]) \\
&\quad + 72\alpha[2^5](7\alpha[2^3] - 3\alpha[12^3]) - 360\alpha[12^4]\alpha[2^3] \\
&\quad + 60\alpha[2^4](7\alpha[2^3] - 3\alpha[12^2]) - 240\alpha[12^3]\alpha[2^3] \\
&\quad + 1890\alpha[2^2]^3(\alpha[22] - \alpha[12]) + 630\alpha[24]\alpha[22](\alpha[12] - 2\alpha[22]) \\
&\quad + 630\alpha[12^3]\alpha[2^2]^2 + 210\alpha[2^3]^2(\alpha[12] - 4\alpha[22]) \\
&\quad - 630\alpha[12^2]\alpha[2^3]\alpha[22] + 945\alpha[2^2]^4(\alpha[12] - \alpha[22]) \\
&\quad + 840\alpha[2^2]^2(3\alpha[2^3]\alpha[2^2] - \alpha[12^2]\alpha[2^2] - 2\alpha[2^3]\alpha[12]),
\end{aligned}
$$

and similarly for $S_6(F)$.

[Figures 6.3 and 6.4 about here.]

It follows that a second order estimate of μ_1/μ_2 is $(\overline{X_1}/\overline{X_2})\{1 + n^{-1}[\overline{X_1X_2}/(\overline{X_1}\,\overline{X_2}) - \overline{X_2^2}/\overline{X_2}^2]\}$, where $\overline{X_1}$ and $\overline{X_2}$ are the sample means for X_1 and X_2, respectively. If X_i, $i = 1, 2$ are independent

16

exponential random variables with means μ_i, $i = 1, 2$ then a target estimator (Cabrera and Fernholz, 1999, 2004) of μ_1/μ_2 is $((n_2-1)/n_2)(\overline{X_1}/\overline{X_2})$, where n_2 denotes the sample size for $\overline{X_2}$. Sen (1988)'s conditions for asymptotic normality do not hold here. Figures 6.3 and 6.4 compare the performance of the two estimates and those obtained by bootstrapping and jackknifing. Our estimate again provides the lowest absolute bias and the lowest mean squared error, but the differences with respect to the target estimator do not appear to be significant. $\square$

Example 6.2 $T(F) = g(\mu)$, where $\mu = (\mu_1, \ldots, \mu_k)$, $\mu_i = \int h_i dF_i = Eh_i(X_i)$ for $X_i \sim F_i$, $h_i : R^{s_i} \to R$ is a given function, and $g : R^k \to R$ is a given function with finite partial derivatives $g_{j_1 \cdots j_r}$. Then

$$T_F \begin{pmatrix} i_1 & \cdots & i_r \\ x_1 & \cdots & x_r \end{pmatrix} = g_{i_1 \cdots i_r} \mu_{i_1 x_1} \cdots \mu_{i_r x_r},$$

where $\mu_{ix} = \mu_{iF_i}(x) = h_i(x) - \mu_i$. So,

$$T(i^a j^b \cdots) = g_{i^a j^b \ldots} \mu_a[i] \mu_b[j] \cdots ,$$

where $\mu_a[i] = \mu_a(h_i(X_i)) = \int (h_i(x) - \mu_i)^a dF_i(x)$. An estimate of bias $O(n_0^{-p})$ is $S_{\boldsymbol{n}p}(\widehat{F}) = \sum_{i=0}^{p-1} S_i^{\boldsymbol{n}}(\widehat{F})$, where by (3.11)–(3.17),

$$S_0^{\boldsymbol{n}}(F) = g(\mu),$$

$$S_1^{\boldsymbol{n}}(F) = -(1/2) \sum_{i=1}^{k} g_{i^2} \mu_2[i](n_i - 1)^{-1},$$

$$\begin{aligned}
S_2^{\boldsymbol{n}}(F) &= \sum_{i=1}^{k} \{g_{i^3} \mu_3[i]/3 + g_{i^4} \mu_2[i]^2/8\}(n_i - 1)_2^{-1} \\
&\quad + \sum_{1 \leq i < j \leq k} g_{i^2 j^2} \mu_2[i] \mu_2[j](n_i - 1)^{-1}(n_j - 1)^{-1}/4,
\end{aligned}$$

$$\begin{aligned}
S_3^{\boldsymbol{n}}(F) &= \sum_{i=1}^{k} \{g_{i^4}(-\mu_4[i]/4 + 3\mu_2[i]^2/8) - g_{i^5} \mu_3[i] \mu_2[i]/6 - g_{i^6} \mu_2[i]^3/48\}(n_i - 1)_3^{-1} \\
&\quad - \sum_{i \neq j} \{g_{i^3 j^3} \mu_3[i] \mu_3[j]/6 + g_{i^4 j^2} \mu_2[i]^2 \mu_2[j]/16\}(n_i - 1)_2^{-1}(n_j - 1)^{-1} \\
&\quad - (1/8) \sum_{i < j < l} g_{i^2 j^2 l^2} \mu_2[i] \mu_2[j] \mu_2[l](n_i - 1)^{-1}(n_j - 1)^{-1}(n_l - 1)^{-1}
\end{aligned}$$

and

$$
\begin{aligned}
S_4^{\boldsymbol{n}}(F) \;=\; & \sum_{i=1}^{k}\{g_{i^5}(\mu_5[i]/5 - 2\mu_3[i]\mu_2[i]/3) \\
& + g_{i^6}(\mu_4[i]^2/8 + \mu_3[i]^2/18 - 3\mu_2[i]^3/16) \\
& + g_{i^7}\mu_3[i]\mu_2[i]^2/24 + g_{i^8}\mu_2[i]^4/384\}(n_i-1)_4^{-1} \\
& -(1/2)\sum_{i\neq j}\{g_{i^4 j^2}(-\mu_4[i]/4 + 3\mu_2[i]^2/18)\mu_2[j] \\
& - g_{i^5 j^2}\mu_3[i]\mu_2[i]\mu_2[j]/6 + g_{i^6 j^2}\mu_2[i]^3\mu_2[j]/48\}(n_i-1)_3^{-1}(n_j-1)^{-1} \\
& + \sum_{i<j}\{g_{i^3 j^3}\mu_3[i]\mu_3[j]/9 + g_{i^4 j^4}\mu_2[i]^2\mu_2[j]^2/64\}(n_i-1)_2^{-1}(n_j-1)_2^{-1} \\
& + \sum_{i\neq j} g_{i^3 j^3}\mu_3[i]\mu_3[j](n_i-1)_2^{-1}(n_j-1)_2^{-1}/12 \\
& + \sum_{i\neq j, i\neq l, j<l}\{g_{i^3 j^2 l^2}\mu_3[i]\mu_2[j]\mu_2[l]/12 \\
& + g_{i^4 j^2 l^2}\mu_4[i]^2\mu_2[j]\mu_2[l]/32\}(n_i-1)_2^{-1}(n_j-1)^{-1}(n_l-1)^{-1} \\
& + \sum_{i<j<l<m} g_{i^2 j^2 l^2 m^2}\mu_2[i]\mu_2[j]\mu_2[l]\mu_2[m](n_i-1)^{-1}(n_j-1)^{-1}(n_l-1)^{-1}(n_m-1)^{-1}/16.
\end{aligned}
$$

As in Example 6.1, for $i \geq 1$, $S_i^{\boldsymbol{n}}(F)$ has the form $\sum_{r=i+1}^{2i} g_{j_1\cdots j_r} c_{ij_1\cdots j_r}^{\boldsymbol{n}}$. So, if $g(\mu)$ is a polynomial of degree $\boldsymbol{p}$ in μ (that is degree p_i in μ_i for $1 \leq i \leq k$), then $S_{\boldsymbol{n}|\boldsymbol{p}|}(\widehat{F})$ is an UE of $g(\mu)$ for $|\boldsymbol{p}| = \boldsymbol{p}'\mathbf{1}$.
$\square$

Example 6.2.1 $g(\mu) = N^q$, where $N = \alpha'\mu$ for some k-vector α. Then

$$
g_{i_1\cdots i_r} = (q)_r N^{q-r}\alpha_{i_1}\cdots\alpha_{i_r}.
$$

So,

$$
T(i^a j^b \cdots) = (q)_r N^{q-r}\alpha_a[i]\alpha_b[j]\cdots,
$$

where $r = a + b + \cdots$, $\alpha_a[i] = \alpha_i^a \mu_a[i]$, and for $i \geq 1$,

$$
S_i^{\boldsymbol{n}}(F) = \sum_{r=i+1}^{2i} (q)_r N^{q-r} S_{ir}^{\boldsymbol{n}},
$$

where

$$
S_{12}^{\boldsymbol{n}} = -\sum_{i=1}^{k} \alpha_2[i](n_i-1)^{-1}/2,
$$

$$
S_{23}^{\boldsymbol{n}} = \sum_{i=1}^{k} \alpha_3[i](n_i-1)_2^{-1}/3,
$$

$$
S_{24}^{\boldsymbol{n}} = \sum_{i=1}^{k} \alpha_2[i]^2(n_i-1)_2^{-1}/8 + \sum_{i<j} \alpha_2[i]\alpha_2[j](n_i-1)^{-1}(n_j-1)^{-1}/4,
$$

18

$$S_{34}^{\boldsymbol{n}} = \sum_{i=1}^{k}(-\alpha_4[i]/4 + 3\alpha_2[i]^2/8)(n_i - 1)_3^{-1},$$

$$S_{35}^{\boldsymbol{n}} = -\sum_{i=1}^{k}\alpha_3[i]\alpha_2[i](n_i - 1)_3^{-1}/6,$$

$$
\begin{aligned}
S_{36}^{\boldsymbol{n}} \;=\; & -\sum_{i=1}^{k}\alpha_2[i]^3(n_i - 1)_3^{-1}/48 \\
& -\sum_{i\neq j}(\alpha_3[i]\alpha_3[j]/6 + \alpha_2[i]^2\alpha_2[j]/16)(n_i - 1)_2^{-1}(n_j - 1)^{-1} \\
& -\sum_{i<j<l}\alpha_2[i]\alpha_2[j]\alpha_2[l](n_i - 1)^{-1}(n_j - 1)^{-1}(n_l - 1)^{-1}/8,
\end{aligned}
$$

$$S_{45}^{\boldsymbol{n}} \;=\; \sum_{i=1}^{k}(\alpha_5[i]/5 - 2\alpha_3[i]\alpha_2[i]/3)(n_i - 1)_4^{-1},$$

$$
\begin{aligned}
S_{46}^{\boldsymbol{n}} \;=\; & \sum_{i=1}^{k}(\alpha_4[i]^2/18 + \alpha_3[i]^2/18 - 3\alpha_2[i]^3/16)(n_i - 1)_4^{-1} \\
& -\sum_{i\neq j}(-\alpha_4[i]/4 + 3\alpha_2[i]^2/18)\alpha_2[j](n_i - 1)_3^{-1}(n_j - 1)^{-1}/2 \\
& +\sum_{i<j}\alpha_3[i]\alpha_3[j](n_i - 1)_2^{-1}(n_j - 1)_2^{-1}/9,
\end{aligned}
$$

$$
\begin{aligned}
S_{47}^{\boldsymbol{n}} \;=\; & \sum_{i=1}^{k}\alpha_3[i]\alpha_2[i]^2(n_i - 1)_4^{-1}/384 \\
& +\sum_{i\neq j}\alpha_3[i]\{\alpha_2[i]\alpha_2[j](n_i - 1)_3^{-1}(n_j - 1)^{-1} + \alpha_2[j]^2(n_i - 1)_2^{-1}(n_j - 1)_2^{-1}\}/12 \\
& +\sum_{i\neq j, i\neq l, j<l}\alpha_3[i]\alpha_2[j]\alpha_1[l](n_i - 1)_2^{-1}(n_j - 1)_2^{-1}(n_l - 1)_2^{-1}/12
\end{aligned}
$$

and

$$
\begin{aligned}
S_{48}^{\boldsymbol{n}} \;=\; & \sum_{i=1}^{k}\alpha_3[i]^4(n_i - 1)_4^{-1}/384 \\
& +\sum_{i\neq j}\alpha_2[i]^3\alpha_2[j](n_i - 1)_3^{-1}(n_j - 1)^{-1}/96 \\
& +\sum_{i<j}\alpha_2[i]^2\alpha_2[j]^2(n_i - 1)_2^{-1}(n_j - 1)_2^{-1}/64 \\
& +\sum_{i\neq j, i\neq l, j<l}\alpha_4[i]\alpha_2[j]\alpha_2[l](n_i - 1)_2^{-1}(n_j - 1)^{-1}(n_l - 1)^{-1}/32 \\
& +\sum_{i<j<l<m}\alpha_2[i]\alpha_2[j]\alpha_2[l]\alpha_2[m](n_i - 1)^{-1}(n_j - 1)^{-1}(n_l - 1)^{-1}(n_m - 1)^{-1}/16.
\end{aligned}
$$

For example, a second order estimate of $(\alpha'\mu)^q$ is

$$(\alpha'\widehat{\mu})^q\{1 - \binom{q}{2}(\alpha'\widehat{\mu})^{-2}\sum_{i=1}^k \alpha_i^2 \widehat{v}_i/(n_i - 1)\},$$

where $\widehat{\mu}_i, \widehat{v}_i$ are the sample estimates of $\mu_i, v_i = \mu_2[i]$, the mean and variance of $h_i(X_i)$. $\square$.

Example 6.2.2 $k = 2$, $g(\mu) = \mu_1\mu_2/(\mu_1 + \mu_2)$. Set $D = \mu_1 + \mu_2$. Then $g(\mu) = 1 - \mu_2^2 D^{-1}$ so $g_{1^a}\mu_2^{-2} = g_{2^a}\mu_1^{-2} = (-D)^{-a-1}a!$. Also since $\mu_2^2 = (D - \mu_1)^2$,

$$g_{1^a 2^b} = -(-D)^{-a-b-1}(a + b - 2)!\{(b^2 - b)\mu_1^2 - 2ba\mu_1\mu_2 + (a^2 - a)\mu_2^2\}.$$

So,

$$S_1^{\mathbf{n}}(F) \;=\; T(F)^2 D^{-1}\sum_{i=1}^2 \mu_i^{-2}\mu_2[i](n_i - 1)^{-1},$$

$$\begin{aligned} S_2^{\mathbf{n}}(F) \;=\; & T(F)^2 \sum_{i=1}^2 (2D^{-2}\mu_3[i] - 3D^{-3}\mu_2[i]^2)\mu_i^{-2}(n_i - 1)_2^{-1} \\ & + D^{-5}(\mu_1^2 + \mu_2^2 - 4\mu_1\mu_2)\mu_2[1]\mu_2[2](n_1 - 1)^{-1}(n_2 - 1)^{-1}, \end{aligned}$$

$$\begin{aligned} S_3^{\mathbf{n}}(F) \;=\; & T(F)^2 \sum_{i=1}^2 \mu_i^{-2}\{D^{-3}(6\mu_4[i] - 9\mu_2[i]^2) \\ & + 20D^{-4}\mu_3[i]\mu_2[i] + 15D^{-5}\mu_2[i]^3\}(n_i - 1)_3^{-1} \\ & - 48D^{-7}(\mu_1^2 + \mu_2^2 - 3\mu_1\mu_2)\mu_3[1]\mu_3[2]\{(n_1 - 1)_2^{-1}(n_2 - 1)^{-1} \\ & + (n_1 - 1)^{-1}(n_2 - 1)_2^{-1}\} \\ & - 3D^{-7}(\mu_1^2 - 8\mu_1\mu_2 + 6\mu_2^2)\mu_2[1]^2\mu_2[2](n_1 - 1)_2^{-1}(n_2 - 1)^{-1} \\ & - 3D^{-7}(6\mu_1^2 - 8\mu_1\mu_2 + \mu_2^2)\mu_2[2]^2\mu_2[1](n_2 - 1)_2^{-1}(n_1 - 1)^{-1}, \end{aligned}$$

and so on. $\square$

Example 6.3 $k = 1$, $T(F) = \mu_r = \mu_r(h(X)) = E(h(X) - \mu)^r$, where $\mu = Eh(X)$ for $X \sim F$ and $h : R^s \to R$ a given function. By (5.6) of Withers (1994a), its general pth order derivative is

$$\mu_{rF}(x_1\cdots x_p) = (-1)^p\{(r)_p\mu_{r-p} - (r)_{p-1}\sum_{i=1}^p (h_i^{r-p} - \mu_{r-p+1}h_i^{-1})\}\prod_{j=1}^p h_j,$$

where $h_i = \mu_F(x_i) = h(x_i) - \mu$. So, $T[i_1 i_2 \cdots] = T(1^{i_1}1^{i_2}\cdots)$ and $S_i(F)$ for $i \le 3$ are as given there. For example,

$$\begin{aligned} T[i] \;&=\; (-1)^i\{(r)_i\mu_{r-i}\mu_i - i(r)_{i-1}(\mu_r - \mu_{r-i+1}\mu_{i-1})\}, \\ T[4] \;&=\; (r)_4\mu_{r-4}\mu_4 - 4(r)_3(\mu_r - \mu_{r-3}\mu_3), \\ T[5] \;&=\; -(r)_5\mu_{r-5}\mu_5 + 5(r)_4(\mu_r - \mu_{r-3}\mu_3), \\ T[42] \;&=\; (r)_6\mu_{r-6}\mu_4\mu_2 - 2(r)_5\{-2\mu_{r-5}\mu_3\mu_2 + \mu_{r-4}\mu_4 + 2\mu_{r-2}\mu_2\}, \\ T[3^2] \;&=\; (r)_6\mu_{r-6}\mu_3^2 - 6(r)_5(\mu_{r-3}\mu_{r-5}\mu_2)\mu_3, \\ T[32^2] \;&=\; -(r)_7\mu_{r-7}\mu_3\mu_2^2 + (r)_6\{-3\mu_{r-6}\mu_2^3 + 4\mu_{r-5}\mu_3\mu_2 + 3\mu_{r-4}\mu_2^2\}, \\ T[2^4] \;&=\; (r)_8\mu_{r-8}\mu_2^4 - 8(r)_7\mu_{r-6}\mu_2^3. \end{aligned}$$

So,

$$S_4[F] = \sum_{i=0}^{8} \mu_{r-i}(r)_i S_{4-i},$$

where $S_{4-0} = (r)_4$, $S_{4-1} = 0$, $S_{4-2} = -r(r-2)_2\mu_2/2$, $S_{4-3} = -r(r-3)\mu_3/3$, $S_{4-4} = -r\mu_4 + r^2\mu_2^2/8$, $S_{4-5} = -\mu_5/5 + (r+4)\mu_3\mu_2/6$, $S_{4-6} = -3\mu_2^3/16 + \mu_4\mu_2/8 + \mu_3^2/18 - r\mu_2^3/48$, $S_{4-7} = -\mu_3\mu_2^2/24$ and $S_{4-8} = \mu_2^4/384$. The fifth order S estimate is now given by (6.1). Also by Example 5.3 of Withers (1994a) one can now obtain an UE for μ_r for $r \le 9$. Obtaining S_5, S_6 from (3.8)–(3.9) one can obtain an UE for μ_r for $r \le 12$. Our results agree with those of James (1958, page 6): by a different method he obtained an UE for μ_r for $r \le 6$. $\square$

For $T(F)$ a function of a central moment, one may apply the chain rule, as illustrated in the next example for the case $T(F) = \sigma = \mu_2^{1/2}$.

Example 6.4 Here we specialise an application of the chain rule given in Appendix A of Withers (1994a). Suppose $k = 1$, $g : R \to R$ is a function with finite derivatives and $T(F) = g(S(F))$, where $S(F)$ is a smooth univariate functional. Set $g_r = g^{(r)}(S(F))$. The third order S estimate of $T(F)$ is given by (6.1), (3.4), (3.5) in terms of

$$
\begin{aligned}
T[2] &= g_2 S(1,1) + g_1\, S[2], \\
T[3] &= g_3 S(1,1,1) + 3g_2 S(1,1^2) + g_1\, S[3], \\
T[2^2] &= g_4 S(1,1)^2 + g_3\{2S(1,1)S[2] + 4S(ab,a,b)\} \\
&\quad + g_2\{4S(a,ab^2) + S[2]^2 + 2S(ab,ab)\} + g_1\, S[2^2],
\end{aligned}
$$

where

$$S(1,1) = \int S_x^2, \quad S(1,1,1) = \int S_x^3, \quad S(1,1^2) = \int S_x\, S_{xx}, \quad S(ab,a,b) = \int\int S_{xy}\, S_x\, S_y,$$

$$S(a,ab^2) = \int\int S_x\, S_{xyy}, \quad S(ab,ab) = \int\int S_{xy}^2, \quad S_{xy\cdots} = S_F\,(x,y,\ldots)$$

and integration is with respect to $F(x)$ and $F(y)$.

Take $T(F) = \sigma$, that is $S(F) = \mu_2$ and $g(s) = s^{1/2}$. Then $g_r = (1/2)_r\mu_2^{1/2-r}$, where $(r)_i = r!/(r-i)!$, $\mu_x = x - \mu$, $\mu_{2x} = \mu_x^2 - \mu_2$, $\mu_{2xy} = -2\mu_x\mu_y$, and higher derivatives vanish. So, $S(1,1) = \mu_4 - \mu_2^2$, $S[2] = -2\mu_2$, $S(1,1,1) = \mu_6 - 3\mu_4\mu_2 + 2\mu_2^3$, $S(1,1^2) = -2(\mu_4 - \mu_2^2)$, $S[3] = S(a,ab^2) = S[2^2] = 0$, $S(ab,a,b) = -2\mu_3^2$ and $S(ab,ab) = 4\mu_2^2$. Setting $\beta_r = \mu_2\,\mu_2^{-r/2}$, we obtain $S_1(F)/\mu_2^{1/2} = (\beta_4 + 3)/8$ and $S_2(F)/\mu_2^{1/2} = (16\beta_6 + 22\,\beta_4 + 164 - 15\beta_4^2)/128$.

[Figures 6.5 and 6.6 about here.]

It follows that a second order estimate of σ is $\widehat{\sigma}\{1 + (1/(n-1))(1/8)(\widehat{\beta}_4 + 3)\}$. Suppose now F is exponential with mean σ. Then a target estimator (Cabrera and Fernholz, 1999, 2004) of σ is the sample mean $\overline{X}$. Sen (1988)'s asymptotic normality provides $\widehat{\omega}$, the maximum likelihood estimate of ω given $\widehat{\sigma} \sim N(\omega, \omega^2/n)$. Figures 6.5 and 6.6 compare the performance of these three estimates and those obtained by bootstrapping and jackknifing. Our estimate again gives the lowest absolute bias and the lowest mean squared error. $\square$

We end this section with examples of functions of more than one moment. These appeal to some results given in Withers (1994a).

Example 6.5 $k = 1, T(F) = g(\mu, \mu_2)$ for $\mu = EY, \mu_2 = E(Y - \mu)^2, Y = h(X)$, given $h : R^s \to R$ and $g : R^2 \to R$ with finite derivatives $g_{j_1 \cdots j_r} = \partial_{j_1} \cdots \partial_{j_r} g(U_1, U_2)$ at $U_1 = \mu$, $U_2 = \mu_2$, where $\partial_i = \partial/\partial U_i$. We give the fourth order S estimate for general g, then specialise. Note that $S_1(F)$, $S_2(F)$, $S_3(F)$ are given by (3.4)–(3.6) and (A14)–(A15), (A20)–(A23) of Withers (1994a). Set

$$U_{ij..}(a^I b^J .., a^K b^L .., ..) = \int \int ..U_{iF}(a^I b^J ..)U_{jF}(a^K b^L ..)..dF(a)dF(b)...$$

Allowing for permutations the nonzero terms needed are as follows:

- for $T[2]$: at $(a^2), U_2 = -2\mu_2$; at $(a, a), U_{11} = \mu_2, U_{12} = \mu_3, U_{22} = \mu_4 - \mu_2^2$;

- for $T[3]$: at $(a, a^2), U_{12} = -2\mu_3, U_{22} = -2(\mu_4 - \mu_2^2)$; at (a, a, a), $U_{111} = \mu_3$, $U_{112} = \mu_4 - \mu_2^2$, $U_{122} = \mu_5 - 2\mu_3\mu_2$, $U_{222} = \mu_6 - 3\mu_4\mu_2 + 2\mu_2^3$;

- for $T[2^2]$: $U_{22}(ab, ab) = 4\mu_2^2$; at (ab, a, b), $U_{211} = -2\mu_2^2, U_{212} = U_{221} = -2\mu_2\mu_3, U_{222} = -2\mu_3^2$;

- for $T[4]$: $U_{22}(a^2, a^2) = 4\mu_4$; at (a, a, a^2), $U_{112} = \mu_4 - \mu_2^2$, $U_{122} = -2(\mu_5 - \mu_3\mu_2)$, $U_{222} = -2(\mu_6 - 2\mu_4\mu_2 + \mu_2^2)$; at (a, a, a, a), $U_{1111} = \mu_4$, $U_{1112} = \mu_5 - \mu_3\mu_2$, $U_{1122} = \mu_6 - 2\mu_4\mu_2 + \mu_2^2$, $U_{1222} = \mu_7 - 3\mu_5\mu_2 + 3\mu_3\mu_2^2$, $U_{2222} = \mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 - 3\mu_2^4$.

- By (A14), $T[32] = T(a^2 b^3) = \sum_{k=1}^5 \nu_{23.k}$ say, where $\nu_{23.1} = \nu_{23.2} = 0$. For $\nu_{23.3}$: at (a, ab, b^2), $U_{122} = 4\mu_3\mu_2, U_{222} = 4\mu_3^2$; at (b, ab, ab), $U_{122} = 4\mu_3\mu_2, U_{222} = 4(\mu_4 - \mu_2^2)\mu_2$. For $\nu_{23.4}$: at (ab, a, b, b), $U_{2111} = -2\mu_3\mu_2$, $U_{2112} = -2\mu_2(\mu_4 - \mu_2^2)$, $U_{2122} = -2\mu_2(\mu_5 - 2\mu_3\mu_2)$, $U_{2212} = -2\mu_3(\mu_4 - \mu_2^2)$, $U_{2222} = -2\mu_3(\mu_5 - 2\mu_3\mu_2)$.

- By (A15), $T[2^3] = T(a^2 b^2 c^2) = \sum_{k=1}^6 \nu_{222.k}$, where $\nu_{222.1} = \nu_{222.2} = 0$. For $\nu_{222.3} = g_{ijk}B_3^{ijk}$: $U_{222}(ab, bc, ca) = -8\mu_2^3$. For $\nu_{222.4} = g_{ijkl}B_3^{ijkl}$: at (a, b, ab), $U_{112} = -2\mu_2^2$, $U_{122} = -2\mu_3\mu_2$, $U_{222} = -2\mu_3^2$; at (a, b, ac, bc), $U_{1122} = 4\mu_3^3, U_{1222} = 4\mu_3\mu_2^2, U_{2222} = 4\mu_3^2\mu_2$.

Other components are covered by the above $U_{ij....}$ □

Example 6.5.1 $g(\mu, \mu_2) = \mu\mu_2^{-1/2} = \beta$ say. Let $\beta_r = \mu_2^{-r/2}\mu_r$. Then $S_1(F), S_2(F), T[4] = T(1^4)$ and $T[2^2] = T(1^2 1^2)$ are given by Example 5.6 of Withers (1994a). So, $S_3(F)$ is given by (3.6) - and hence the fourth order S estimate - once we specify $T[32] = \sum_{k=3}^5 \nu_{23.k}$ and $T[2^3] = \sum_{k=3}^6 \nu_{222.k}$. Since $g_1 = \mu_2^{-1/2}$, $g_2/\mu = g_{12} = -\mu_2^{-3/2}/2$, $g_{22}/\mu = g_{122} = 2\mu_2^{-5/2}/4$, and so on, these components are given by

$$
\begin{aligned}
\nu_{23.3} &= 45\beta_3 - 15\beta(2\beta_3^2 + 3\beta_4 - 3)/2, \\
\nu_{23.4} &= 45(3\beta_5 - 6\beta_3 + 7\beta_4 - 7)/4 \\
&\quad -105\beta\{\beta_6 - 3\beta_4 + 2 + 12\beta_3(\beta_5 - 2\beta_3) + 6(\beta_4 - 1)^2\}/16, \\
\nu_{23.5} &= 105\{2\beta_3(\beta_6 - 3\beta_4 + 2) + 3(\beta_4 - 1)(\beta_5 - 2\beta_3)\}/16 \\
&\quad -945\beta(\beta_4 - 1)(\beta_6 - 3\beta_4 + 2)/32, \\
\nu_{222.3} &= -775\beta, \\
\nu_{222.4} &= 45\{-y5\beta_3 + 7\beta(\beta_4 - 1 + 4\beta_3^2)/4\}, \\
\nu_{222.5} &= 315\{-3(\beta_4 - 1)\beta_3 + \beta_3^3 + 3\beta(\beta_4 - 1)(\beta_4 - 1 + \beta_3^2)\}/16, \\
\nu_{222.6} &= 945\{-12(\beta_4 - 1)^2\beta_3 + 11\beta(\beta_4 - 1)^3\}/64.
\end{aligned}
$$

[Figures 6.7 and 6.8 about here.]

It follows that a second order estimate of $\mu/\sqrt{\mu_2}$ is $\widehat{\mu}/\sqrt{\widehat{\mu_2}} - (1/n)\{\widehat{\beta}_3/2 + (1/8)(\widehat{\mu}/\sqrt{\widehat{\mu_2}})(3\widehat{\beta}_4 + 1)\}$. Suppose now $Y = h(X)$ is normally distributed with mean μ and variance μ_2. Then a target estimator (Cabrera and Fernholz, 1999, 2004) of $\mu/\sqrt{\mu_2}$ is $\sqrt{2}\Gamma((n-1)/2)\bar{Y}/\{\sqrt{n-1}\Gamma(n/2-1)S_Y\}$, where $\bar{Y}$ and S_Y denote the sample mean and the sample standard deviation, respectively. Sen (1988)'s conditions for asymptotic normality do not hold here. Figures 6.7 and 6.8 compare the performance of the two estimates and those obtained by bootstrapping and jackknifing. Our estimate again provides the lowest absolute bias and the lowest mean squared error, but the differences with respect to the target estimator do not appear to be significant. $\square$

Example 6.6 $k = s = 1$, $T(F) = g(\mu_2, \mu_3)$ for $\mu_r = \mu_r(X)$, $X \sim F$ and g having finite derivatives

$$g_{ij\cdots} = (\partial/\partial\mu_i)(\partial/\partial\mu_j)\cdots g(\mu_2, \mu_3).$$

By Example 6.3, $\mu_{3x} = \mu_x^3 - \mu_3 - 3\mu_2\mu_x$, $\mu_{3xy} = -3(\mu_x + \mu_y)(\mu_x\mu_y - \mu_2)$, $\mu_{3xyz} = 12\mu_x\mu_y\mu_z$. Now use (A8–A11) of Withers (1994a) with $U_i(F) = \mu_i$: for $T[2]$ we need $U_3(a^2) = -6\mu_3$, $U_{23}(a, a) = \mu_5 - 4\mu_3\mu_2$, $U_{33}(a, a) = \mu_6 - 6\mu_4\mu_2 - \mu_3^2 + 9\mu_2^3$. This gives

$$T[2] = -2g_2\mu_2 - 6g_3\mu_3 + g_{22}(\mu_4 - \mu_2^2) + 2g_{23}(\mu_5 - 4\mu_3\mu_2) + g_{33}(\mu_6 - 6\mu_4\mu_2 - \mu_3^2) + 9\mu_2^3.$$

For $T[3]$ we need

$$
\begin{aligned}
U_3(a^3) &= 12\mu_3, \\
U_{23}(a, a^2) &= -6(\mu_5 - 2\mu_3\mu_2), \\
U_{32}(a, a^2) &= -2(\mu_5 - 4\mu_3\mu_2), \\
U_{33}(a, a^2) &= -6(\mu_7 - 4\mu_5\mu_2) - \mu_4\mu_3 + 4\mu_3\mu_2^2, \\
U_{223}(a, a, a) &= \mu_7 - 5\mu_5\mu_2 - \mu_4\mu_3 + 8\mu_3\mu_2^2, \\
U_{233}(a, a, a) &= \mu_8 - 3\mu_6\mu_2 - 2\mu_5\mu_3 + 15\mu_4\mu_2^2 + 4\mu_3^2\mu_2, \\
U_{333}(a, a, a) &= \mu_9 - 9\mu_7\mu_2 - 3\mu_6\mu_3 + 27\mu_5\mu_2^2 + 18\mu_4\mu_3\mu_2 + 2\mu_3^3 - 135\mu_3\mu_2^3.
\end{aligned}
$$

This gives

$$
\begin{aligned}
T[3] &= 12g_3\mu_3 - 6g_{22}(\mu_4 - \mu_2^2) - 12g_{23}(2\mu_5 - 5\mu_3\mu_2) \\
&\quad -18g_{33}(\mu_7 - 4\mu_5\mu_2 - \mu_4\mu_3 + 4\mu_3\mu_2^2) \\
&\quad +g_{222}(\mu_6 - 3\mu_4\mu_2 + 2\mu_2^3) + 3g_{223}(\mu_7 - 5\mu_5\mu_2 - \mu_4\mu_3 + 8\mu_3\mu_2^2) \\
&\quad +g_{233}U_{233}(a, a, a) + g_{333}U_{333}(a, a, a).
\end{aligned}
$$

For $T[2^2]$ we need

$$
\begin{aligned}
U_{23}(a, ab^2) &= 12\mu_3\mu_2, \\
U_{33}(a, ab^2) &= 12(\mu_4 - 3\mu_2^2)\mu_2, \\
U_{23}(ab, ab) &= 12\mu_3\mu_2, \\
U_{33}(ab, ab) &= 18(\mu_4\mu_2 + \mu_3^2 - \mu_2^3),
\end{aligned}
$$

and at (ab, a, b):

$$
\begin{aligned}
U_{222} &= -2\mu_3^2, \\
U_{223} &= U_{232} = -2\mu_3\kappa_4, \\
U_{233} &= -2\kappa_4^2, \\
U_{322} &= -6\mu_3(\mu_4 - \mu_2^2), \\
U_{323} &= U_{332} = -3\{(\mu_4 - \mu_2^2)\kappa_4 + \mu_3(\mu_5 - 4\mu_3\mu_2)\}, \\
U_{333} &= -6\kappa_4(\mu_5 - 4\mu_3\mu_2).
\end{aligned}
$$

This gives

$$
\begin{aligned}
T[2^2] &= 12g_{22}\sigma^4 + 48g_{23}\sigma^5\beta_3 + 12g_{33}\sigma^6(7\beta_4 + 6\beta_3^2 - 15) \\
&\quad -8g_{222}\sigma^6(\beta_4 - 1 + \beta_3^2) - 4g_{223}\sigma^7(2\beta_5 + 13\beta_4\beta_3 - 29\beta_3) \\
&\quad -4g_{233}\sigma^8(\beta_6 + 12\beta_5\beta_3 + 2\beta_4^2 - 42\beta_4 - 49\beta_3^2 + 9) \\
&\quad -12g_{333}\sigma^9\{(\beta_6 - 14\beta_4 - \beta_3^2 + 33)\beta_3 + 2(\beta_4 - 3)\beta_5\} \\
&\quad +g_{2222}\sigma^8(\beta_4 - 1)^2 + 4g_{2223}\sigma^9(\beta_5 - 4\beta_3)(\beta_4 - 1) \\
&\quad +2g_{2233}\sigma^{10}\{(\beta_6 - 6\beta_4 - \beta_3^2)(\beta_4 - 1) + 2(\beta_5 - 4\beta_3)^2\} \\
&\quad +4g_{2333}\sigma^{11}(\beta_5 - 4\beta_3)(\beta_6 - 6\beta_4 - \beta_3^2) + g_{3333}\sigma^{12}(\beta_6 - 6\beta_4 - \beta_3^2)^2,
\end{aligned}
$$

where again $\beta_r = \sigma^{-r}\mu_r$. By (3.4), (3.5) this gives $S_1(F)$ and $S_2(F)$ so that (6.1) gives the third order S estimate. $\square$

Example 6.6.1 $g(\mu_2, \mu_3) = \mu_3\mu_2^{-3/2}$, the skewness of F. One obtains

$$
\begin{aligned}
T[2] &= -3\beta_5 + 15(\beta_4 - 1)\beta_3/4 + 9\beta_3, \\
T[3] &= -105(\beta_6 - 3\beta_4 + 2)\beta_3/8 + 36\beta_5 - 45\beta_4\beta_3/2 - 111\beta_3/2, \\
T[2^2] &= -105(\beta_5 - 4\beta_3)(\beta_4 - 1)/2 + 945(\beta_4 - 1)^2\beta_3/16 + \beta_3\delta,
\end{aligned}
$$

where $\delta = 78 - 105(\beta_4 + \beta_3^2) - 15(2\beta_5 + 13\beta_4\beta_3 - 29\beta_3)$. Note that T[2] for both skewness and kurtosis were first given in Withers (1994b): β_i in (3.1) there should be β_{i+r}. $\square$

Acknowledgments

References

[1] Cabrera, J. and Fernholz, L. T. (1999). Target estimation for bias and mean square error reduction. *Annals of Statistics*, **27**, 1080–1104.

[2] Cabrera, J. and Fernholz, L. T. (2004). Multivariate targeting with applications to ellipse estimation. Contemporary data analysis: theory and methods. *Journal of Statistical Planning and Inference*, **122**, 79–94.

[3] David, F. N. and Kendall, M. G. (1949). Tables of symmetric functions. *Biometrika*, **36**, 431–449.

[4] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.

[5] Fernholz, L. T. (2001). On multivariate higher order von Mises expansions. *Metrika*, **53**, 123–140.

[6] Fisher, R. A. (1929). Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society*, 2, **30**, 199–238.

[7] Hall, P. (1992). *The Bootstrap and Edgeworth Expansions.* Springer, New York.

[8] Jaeckel, L. A. (1972). The infinitesimal jackknife. *Bell Laboratories Technical Report*, MM 72–1215–11, New Jersey.

[9] James, G. S. (1958). On moments and cumulants of systems of statistics. *Sankhyā*, **20**, 1–30.

[10] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *Annals of Mathematical Statistics*, **14**, 217–226.

[11] Sen, P. K. (1988). Functional jackknifing: rationality and general asymptotics. *Annals of Statistics*, **16**, 450–469.

[12] Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, volume 1. Hafner, New York.

[13] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, **18**, 309–348.

[14] Wishart, J. (1952). Moment coefficients of the k-statistics in samples from a finite population. *Biometrika*, **39**, 1–13.

[15] Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Annals of Statistics*, **11**, 577–587.

[16] Withers, C. S. (1987). Bias reduction by Taylor series. *Communications in Statistics—Theory and Methods*, **16**, 2369–2384.

[17] Withers, C. S. (1988). Nonparametric confidence intervals for functions of several distributions. *Annals of the Institute of Statistical Mathematics*, **40**, 727–746.

[18] Withers, C. S. (1994a). Estimates of low bias for functionals of distribution functions. *Technical Report*, Industrial Research Ltd., New Zealand.

[19] Withers, C. S. (1994b). Bias reduced estimates of skewness and kurtosis. *Procedings of the 1994 New Zealand Statistics Association Conference*, Massey University, Palmerston North, pp. 345–349.

Appendix A

Here we explain the infinitesimal jackknife given in the unpublished report Jaeckel (1972) and its relation to our case $p = 2, k = 1$. In fact he offers two forms and these are equivalent to our second order S and T estimates, but are arrived at without using functional derivatives.

He considers the case $k = 1$ for $T(F)$ but with $\widehat{F(x)}$ the *weighted* empirical distribution, $\widehat{F(x)} = \sum_{i=1}^{n} w_i I(X_i \leq x)$. His first form is equivalent to our T estimate with

$$T_1(\widehat{F}) = -(1/2)n^{-1} \sum_{i=1}^{n} \partial^2 T(\widehat{F})/\partial w_i^2$$

evaluated at $w_1 = \cdots = w_n = 1/n$. His second form is essentially our S estimate and is defined in the same way except that in the definition of the weighted empirical distribution w_i is replaced by $w_i/(\sum_{j=1}^{n} w_j)$.

Appendix B

The following procedures in MAPLE calculate the V, S and T estimates described in Sections 2 to 4 to obtain bias of any given order. The electronic versions of these procedures can be obtained from the corresponding author.

```
#this procedure expresses the first u symmetric polynomials in terms of power
#sum functions the output from this procedure is needed as input for
#the Vest() and Vestsum() procedures
symmpoly := proc(u)
local m,e;
e[0]:=1;
for m from 1 to u do
e[m]:=(1/m)*sum((-1)**(i-1)*e[m-i]*p[i],i=1..m);
od;
return([seq(e[mm],mm=1..u)]);
end proc;


#this procedure computes the G function given by (4.1)
gbetai := proc (beta,i)
local tt,ttt,ii,j,k,p;
tt:=0;
for ii from 1 to i do
p:=composition(beta,ii);
for j from 1 to nops(p) do
ttt:=1;
for k from 1 to ii do
ttt:=ttt*k**(p[j][k]);
od;
tt:=tt+ttt;
od;
od;
return(tt);
end proc;


#this procedure computes the D function given by (4.2)
dalphai := proc (alpha,i)
gbetai(alpha-i,i);
end proc;


#this procedure computes the coefficients c[k,j,r] given by equation (3.1)
cc := proc (k, j, r)
local d,tt,i,n;
d[j]:=1/Amkr(j,j,r);
tt:=d[j];
```

```
if (j>k) then
for i from 1 to (j-k) do
d[j-i]:=0;
for n from (j-i+1) to j do
d[j-i]:=d[j-i]+d[n]*Amkr(j-i,n,r);
od;
d[j-i]:=-d[j-i]/(Amkr(j-i,j-i,r));
od;
tt:=d[k];
end if;
return(tt);
end proc;


#the following procedure is needed as part of the above cc (k, j, r) procedure
Amkr := proc (m, k, r)
sum(((-1)**(l+k))*stirling1(k,l)*binomial(l,m)*(1-r)**(l-m),l=m..k);
end proc;


#this procedure computes the k (ndim) fold summation in equation (2.9)
#here it is assumed implicitly that k=5
#but the procedure can be applied for any k by making appropriate changes to
#the five do-loops and the five statements following it
#this procedure requires as input: n[1],n[2],...,n[k],r[1],r[2],...,r[k],
#ndim=the value of k,the output from symmpoly() and the T(.) function
#in equation (2.4) the output from this procedure is needed as input for
#the Vest() and coefficients() procedures
Vestsum := proc (ndim,n::list,r::list,e::list,T)
local kk,pa,ss,i1,i2,i3,i4,i5,paa,bb,tt,kkk,j,nn,ttt,mmm;
for kk from 1 to ndim do
pa[kk]:=partition(r[kk]);
od;
ss:=0;
for i1 from 1 to nops(pa[1]) do
for i2 from 1 to nops(pa[2]) do
for i3 from 1 to nops(pa[3]) do
for i4 from 1 to nops(pa[4]) do
for i5 from 1 to nops(pa[5]) do
paa[1]:=pa[1][i1];
paa[2]:=pa[2][i2];
paa[3]:=pa[3][i3];
paa[4]:=pa[4][i4];
paa[5]:=pa[5][i5];
bb:=member(1,paa[1])=false;
for kk from 2 to ndim do
bb:=bb and member(1,paa[kk])=false;
od;
```

```
if (bb) then
for kk from 1 to ndim do
tt[kk]:=expand(e[r[kk]]);
od;
for kkk from 1 to ndim do
for j from 1 to r[kkk] do
nn:=0;
for kk from 1 to nops(paa[kkk]) do
if (j=paa[kkk][kk]) then nn:=nn+1 end if;
od;
if (nn>0) then tt[kkk]:=coeff(tt[kkk],p[j],nn) end if;
od;
od;
ttt:=1;
for mmm from 1 to ndim do
ttt:=ttt*tt[mmm]*(n[mmm])**(nops(paa[mmm]));
od;
ss:=ss+ttt*T([seq(paa[mm],mm=1..ndim)]);
end if;
od;
od;
od;
od;
od;
ss:=expand(ss);
return(ss);
end proc;


#this procedure computes of the coefficients of the output from
#with respect to powers of n[1],n[2],...,n[k]
#It is assumed implicitly that k is 5
#the procedure can be applied to any k by changing the number
#of do loops and making other appropriate changes
#this procedure requires as input: n[1],n[2],...,n[k],r[1],r[2],...,r[k]
#and the output from Vestsum()
#the output from this procedure is needed as input for the Sest()
#and Test() procedures
coefficients := proc(n::list,r::list,ss)
local i1,i2,i3,i4,i5,sss,A;
A:=array(1..r[1]/2,1..r[2]/2,1..r[3]/2,1..r[4]/2,1..r[5]/2);
for i1 from 1 to r[1]/2 do
for i2 from 1 to r[2]/2 do
for i3 from 1 to r[3]/2 do
for i4 from 1 to r[4]/2 do
for i5 from 1 to r[5]/2 do
sss:=coeff(ss,n[1],i1);
sss:=coeff(sss,n[2],i2);
```

```
sss:=coeff(sss,n[3],i3);
sss:=coeff(sss,n[4],i4);
sss:=coeff(sss,n[5],i5);
A[i1,i2,i3,i4,i5]:=sss;
od;
od;
od;
od;
od;
return(A);
end proc;


#this procedure computes V_r for the V estimate in Section 2
#this procedure requires as input: n[1],n[2],...,n[k],r[1],r[2],...,r[k],
#ndim=the value of k,the output from symmpoly() and the T(.) function
#in equation (2.4)
Vest := proc(ndim,n::list,r::list,e::list,T)
local ss,mm;
ss:=Vestsum(ndim,n,r,e,T);
ss:=ss*(product(factorial(n[mm]-r[mm]),mm=1..ndim));
ss:=ss/(product(factorial(n[mm]),mm=1..ndim));
ss:=simplify(ss);
return(ss);
end proc;


#this procedure computes V_r for the S estimate in Section 3
#It is assumed implicitly that k is 5
#the procedure can be applied to any k by changing the number
#of do loops and making other appropriate changes
#this procedure requires as input: n[1],n[2],...,n[k],
#r[1],r[2],...,r[k] and the output from coefficients()
Sest := proc(n::list,r::list,A::array)
local i1,i2,i3,i4,i5,j1,j2,j3,j4,j5,tt,ttt,ss;
ss:=0;
for i1 from r[1]/2 to r[1]-1 do
for i2 from r[2]/2 to r[2]-1 do
for i3 from r[3]/2 to r[3]-1 do
for i4 from r[4]/2 to r[4]-1 do
for i5 from r[5]/2 to r[5]-1 do
tt:=0;
for j1 from r[1]-i1-1 to r[1]/2-1 do
for j2 from r[2]-i2-1 to r[2]/2-1 do
for j3 from r[3]-i3-1 to r[3]/2-1 do
for j4 from r[4]-i4-1 to r[4]/2-1 do
for j5 from r[5]-i5-1 to r[5]/2-1 do
ttt:=A[j1+1,j2+1,j3+1,j4+1,j5+1]*cc(r[1]-i1-1,j1,r[1])*cc(r[2]-i2-1,j2,r[2]);
```

30

```
ttt:=ttt*cc(r[3]-i3-1,j3,r[3])*cc(r[4]-i4-1,j4,r[4])*cc(r[5]-i5-1,j5,r[5]);
tt:=tt+ttt;
od;
od;
od;
od;
od;
tt:=tt*factorial(n[1]-i1-1)*factorial(n[2]-i2-1)*factorial(n[3]-i3-1);
tt:=tt*factorial(n[4]-i4-1)*factorial(n[5]-i5-1);
tt:=tt/(factorial(n[1]-1)*factorial(n[2]-1));
tt:=tt/(factorial(n[3]-1)*factorial(n[4]-1)*factorial(n[5]-1));
ss:=ss+tt;
print(i1,i2,i3,i4,i5,ss);
od;
od;
od;
od;
od;
ss:=simplify(ss);
return(ss);
end proc;


#this procedure computes V_r for the T estimate in Section 4
#It is assumed implicitly that k is 5
#the procedure can be applied to any k (ndim) by changing the number
#of do loops and making other appropriate changes
#this procedure requires as input: n[1],n[2],...,n[k],r[1],r[2],...,r[k],
#ndim=the value of k and the output from coefficients()
Test := proc(ndim,n::list,r::list,A::array)
local i1,i2,i3,i4,i5,j1,j2,j3,j4,j5,nn,tt,ttt,tttt,ss,a,kk;
ss:=0;
for i1 from r[1]/2 to r[1]-1 do
for i2 from r[2]/2 to r[2]-1 do
for i3 from r[3]/2 to r[3]-1 do
for i4 from r[4]/2 to r[4]-1 do
for i5 from r[5]/2 to r[5]-1 do
tt:=0;
for j1 from r[1]-i1-1 to r[1]/2-1 do
for j2 from r[2]-i2-1 to r[2]/2-1 do
for j3 from r[3]-i3-1 to r[3]/2-1 do
for j4 from r[4]-i4-1 to r[4]/2-1 do
for j5 from r[5]-i5-1 to r[5]/2-1 do
ttt:=A[j1+1,j2+1,j3+1,j4+1,j5+1]*cc(r[1]-i1-1,j1,r[1])*cc(r[2]-i2-1,j2,r[2]);
ttt:=ttt*cc(r[3]-i3-1,j3,r[3])*cc(r[4]-i4-1,j4,r[4])*cc(r[5]-i5-1,j5,r[5]);
tt:=tt+ttt;
od;
od;
```

```
od;
od;
od;
nn[1]:=i1;
nn[2]:=i2;
nn[3]:=i3;
nn[4]:=i4;
nn[5]:=i5;
ttt:=1;
for kk from 1 to ndim do
tttt:=0;
for a from nn[kk] to 100 do
tttt:=tttt+(n[kk])**(-a)*dalphai(a,nn[kk]);
od;
ttt:=ttt*tttt;
od;
ss:=ss+tt*ttt;
print(i1,i2,i3,i4,i5,ss);
od;
od;
od;
od;
od;
ss:=simplify(ss);
return(ss);
end proc;
```

Table 5.1. Number of terms for estimates of bias $O(n^{-p})$ for $k = 1$.

$p =$	1	2	3	4	5	6	7
$V_{\boldsymbol{n}p}(\widehat{F})$	1	2	5	11	22	42	77
$S_{\boldsymbol{n}p}(\widehat{F})$	1	2	4	8	15	25	44
$T_{\boldsymbol{n}p}(\widehat{F})$	1	2	5	11	21	.	.

Table 5.2. Number of terms for estimates of bias $O(n_0^{-p})$, where $n_0 = \min n_i$.

$p =$	1	2	3	4	5
$V_{\boldsymbol{n}p}(\widehat{F})$	1	$k+1$	$s_2 + 4k + 1$ $= (k^2 + 7k + 2)/2$	$s_3 + 8s_2 + 10k + 1$ $= (k^3 + 21k^2 + 38k + 6)/6$	$s_4 + 14s_3 + 31s_2 + 21k + 1$ $= (k^4 + 40k^3 + 305k^2 + 230k + 24)/24$
$S_{\boldsymbol{n}p}(\widehat{F})$	1	$k+1$	$s_2 + 3k + 1$ $= (k^2 + 5k + 2)/2$	$s_3 + 3s_2 + 7k + 1$ $= (k^3 + 6k^2 + 35k + 6)/6$	$s_4 + 9s_3 + 15s_2 + 14k + 1$ $= (k^4 + 30k^3 + 143k^2 + 234k + 24)/24$
$T_{\boldsymbol{n}p}(\widehat{F})$	1	$k+1$	$s_2 + 4k + 1$ $= (k^2 + 7k + 2)/2$	$s_3 + 7s_2 + 10k + 1$ $= (k^3 + 18k^2 + 20k + 6)/6$	$s_4 + 13s_3 + 25s_2 + 20k + 1$ $= (k^4 + 36k^3 + 245k + 270k + 24)/24$

Table 5.3. Number of terms for estimates of bias $O(n^{-p})$ for $k = 2$ and 3.

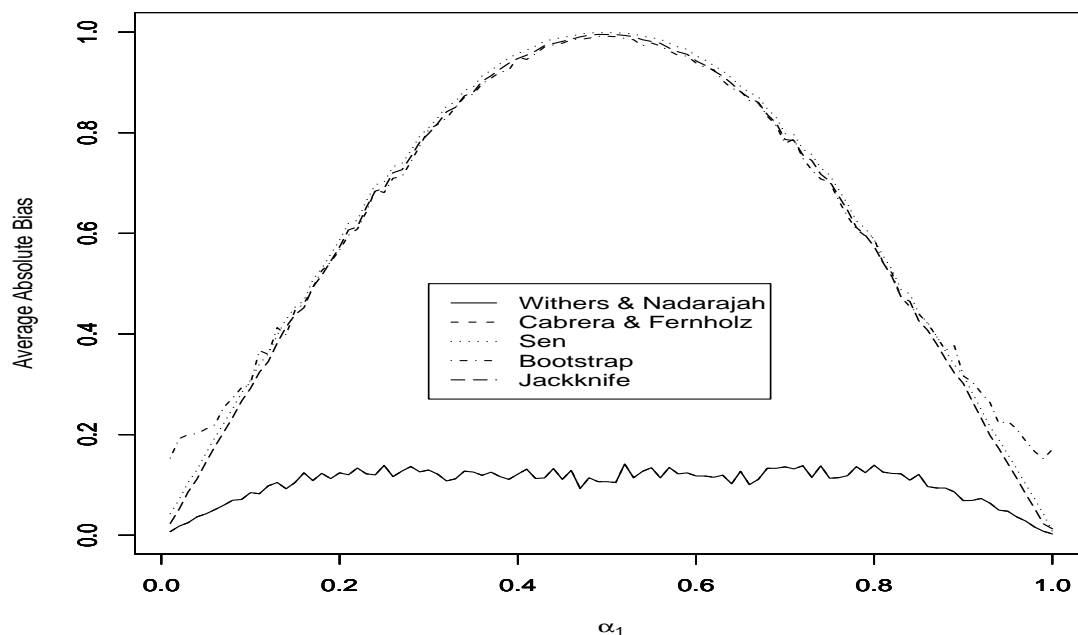|	$k = 2$					$k = 3$				
$p =$	1	2	3	4	5	1	2	3	4	5
$V_{\boldsymbol{n}p}(\widehat{F})$	1	3	10	29	74	1	4	16	56	171
$S_{\boldsymbol{n}p}(\widehat{F})$	1	3	8	18	44	1	4	13	32	97
$T_{\boldsymbol{n}p}(\widehat{F})$	1	3	10	28	66	1	4	16	53	149

Figure 6.1. The average absolute bias of the estimator of $\{\alpha_1\mu_1 + (1 - \alpha_1)\mu_2\}^2$ when μ_1 and μ_2 are the means of two independent normal distributions with unit standard deviations. The average is based on 100 random samples each of size 100.
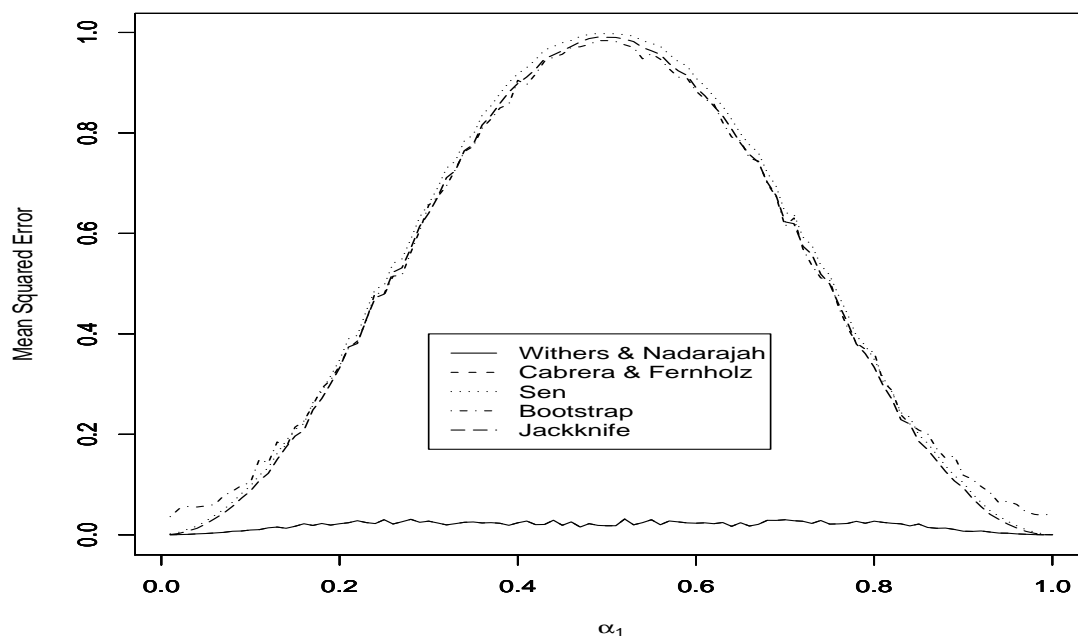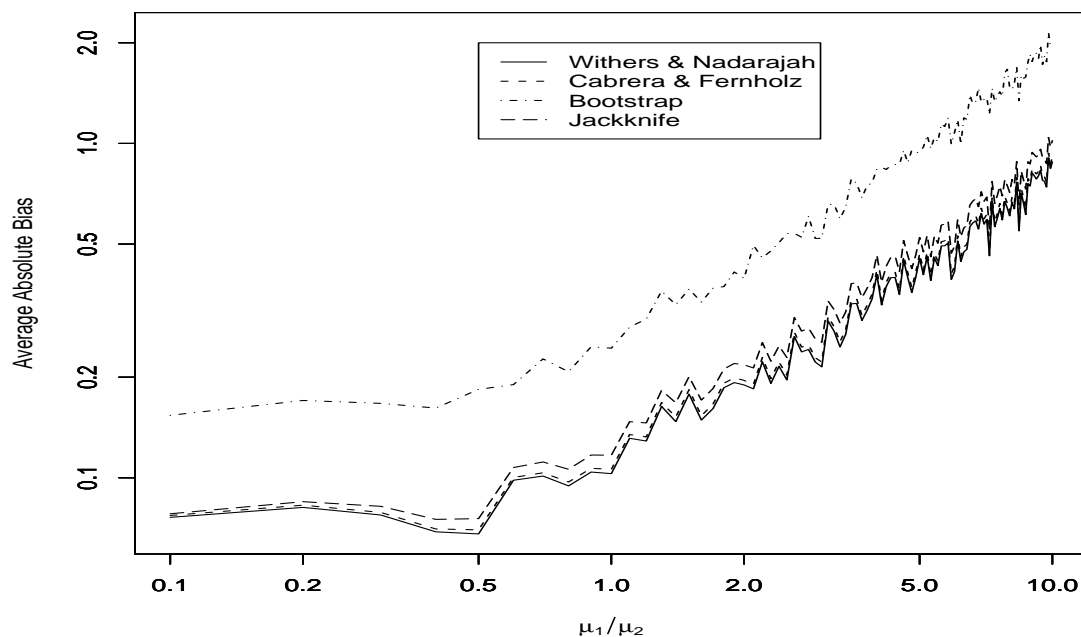

Figure 6.2. Mean squared error of the estimator of $\{\alpha_1\mu_1 + (1 - \alpha_1)\mu_2\}^2$ when μ_1 and μ_2 are the means of two independent normal distributions with unit standard deviations. The MSE is based on 100 random samples each of size 100.

Figure 6.3. The average absolute bias of the estimator of μ_1/μ_2 when μ_1 and μ_2 are the means of two independent exponential distributions. The average is based on 100 random samples each of size 100. The x axis is in log scale.
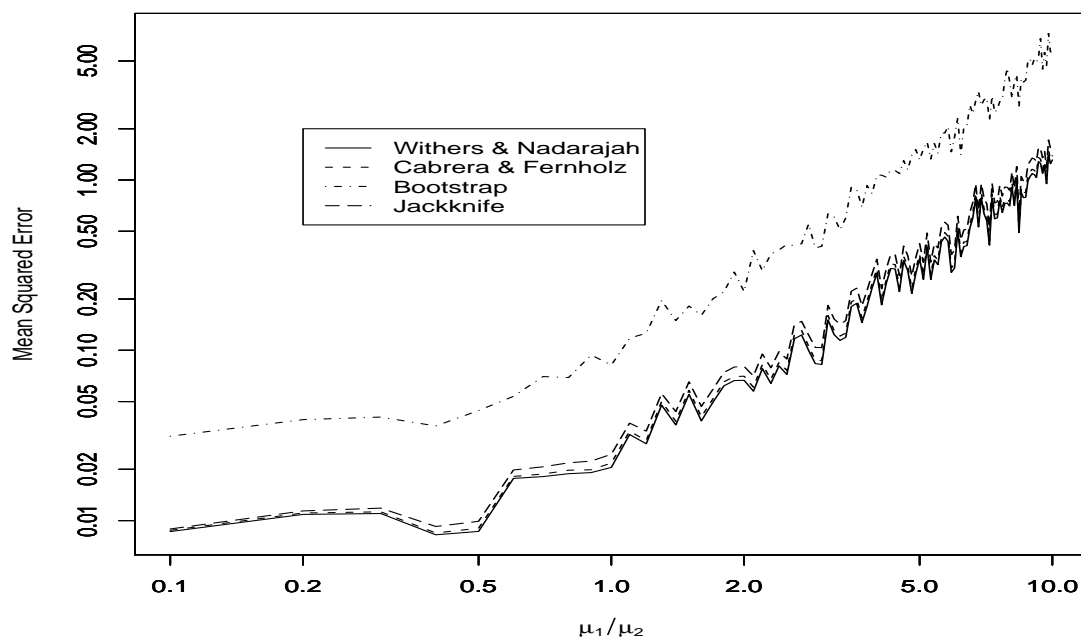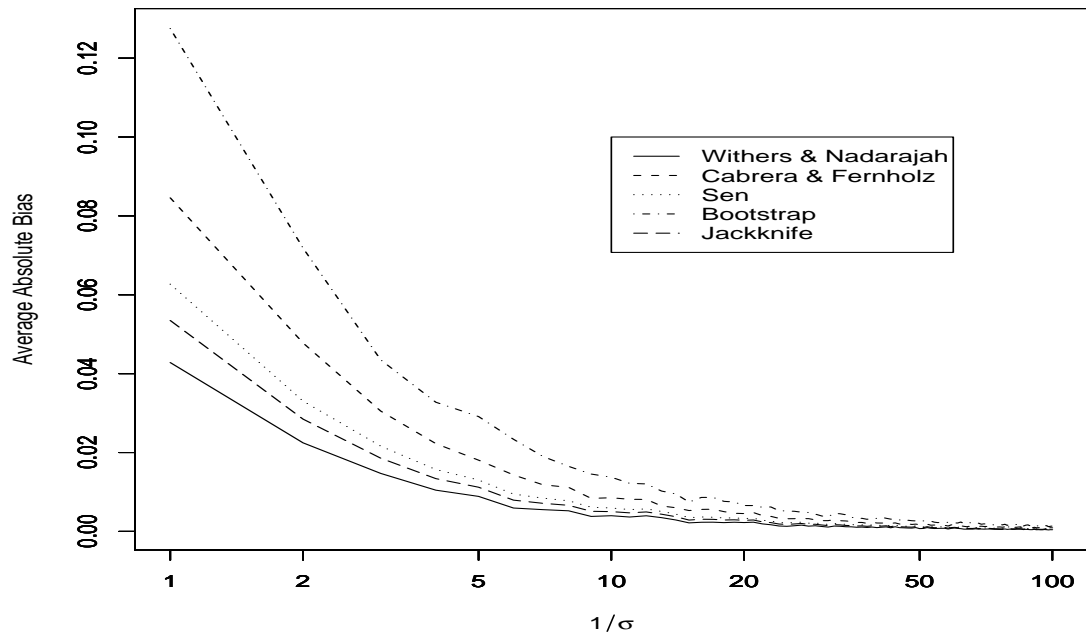

Figure 6.4. Mean squared error of the estimator of μ_1/μ_2 when μ_1 and μ_2 are the means of two independent exponential distributions. The MSE is based on 100 random samples each of size 100. The x axis is in log scale.

Figure 6.5. The average absolute bias of the estimator of σ when σ is the mean of an exponential distribution. The average is based on 100 random samples each of size 100. The x axis is in log scale.
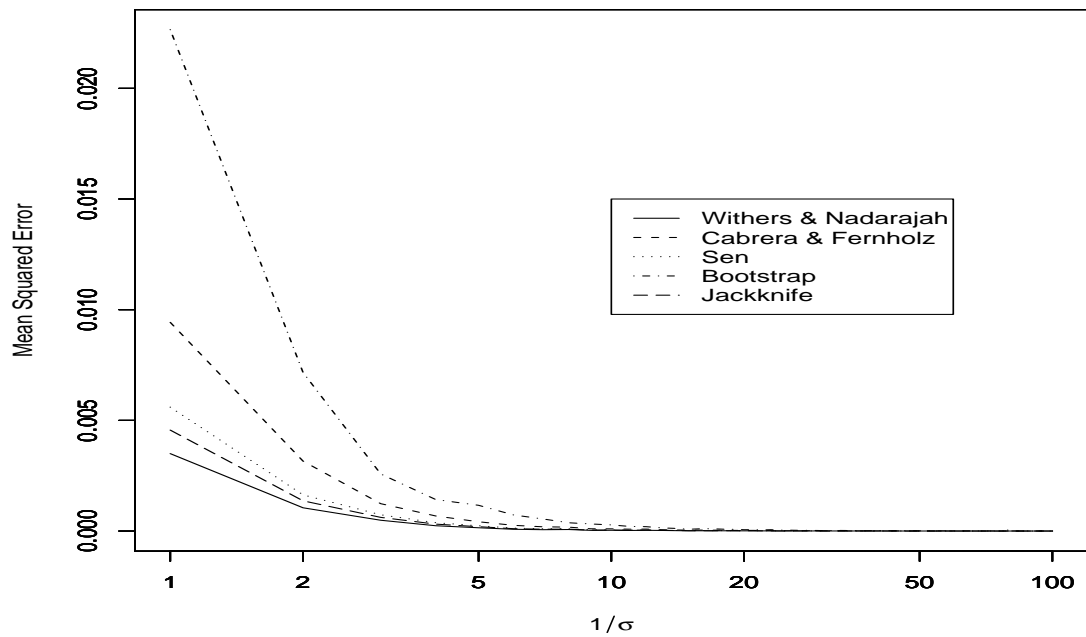

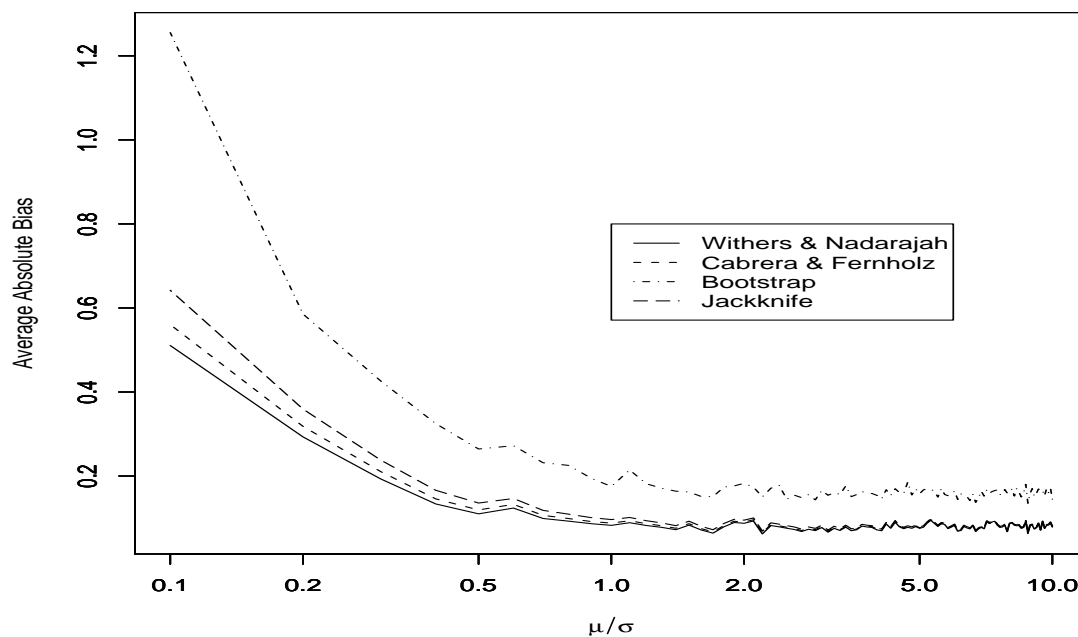**Figure 6.6.** Mean squared error of the estimator of σ when σ is the mean of an exponential distribution. The average is based on 100 random samples each of size 100. The x axis is in log scale.

Figure 6.7. The average absolute bias of the estimator of μ/σ when μ and σ are the mean and the standard deviation of a normal distribution. The average is based on 100 random samples each of size 100. The x axis is in log scale.
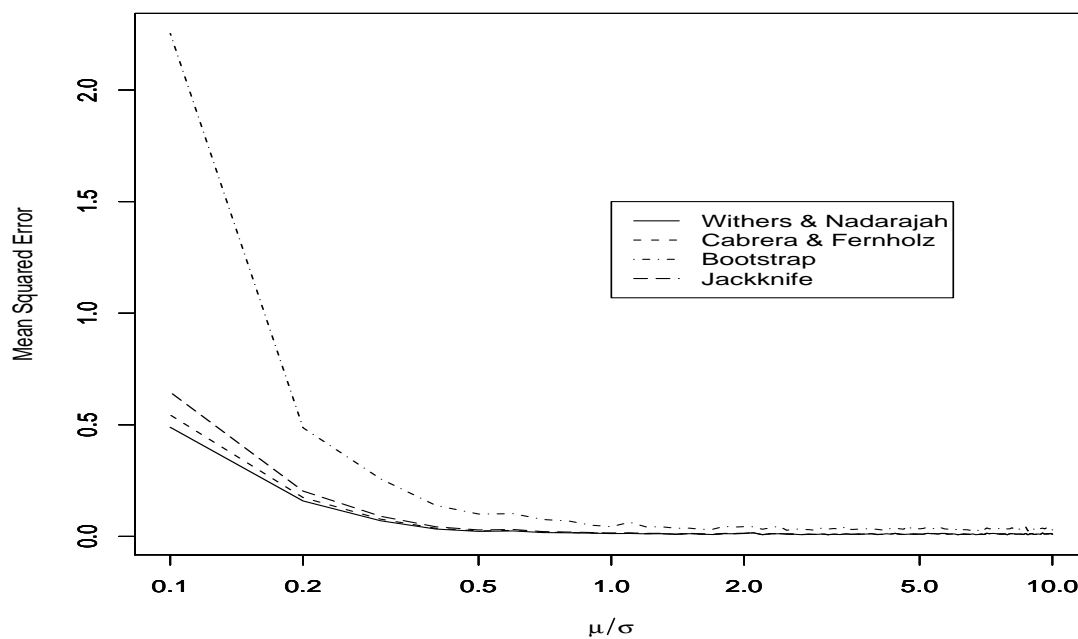

Figure 6.8. The mean squared error of the estimator of μ/σ when μ and σ are the mean and the standard deviation of a normal distribution. The MSE is based on 100 random samples each of size 100. The x axis is in log scale.