REVSTAT – Statistical Journal
Volume 10, Number 2, June 2012, 229–283

NONPARAMETRIC ESTIMATES OF LOW BIAS

Authors: Christopher S. Withers
– Applied Mathematics Group, Industrial Research Limited,
Lower Hutt, New Zealand
c.withers@irl.cri.nz

Saralees Nadarajah
– School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
Saralees.Nadarajah@manchester.ac.uk

Received: October 2011 Revised: February 2012 Accepted: May 2012

Abstract:

- We consider the problem of estimating an arbitrary smooth functional of $k \geq 1$ distribution functions (d.f.s) in terms of random samples from them. The natural estimate replaces the d.f.s by their empirical d.f.s. Its bias is generally $\sim n^{-1}$, where n is the minimum sample size, with a p^{th} *order* iterative estimate of bias $\sim n^{-p}$ for any p. For $p \leq 4$, we give an explicit estimate in terms of the first $2p-2$ von Mises derivatives of the functional evaluated at the empirical d.f.s. These may be used to obtain *unbiased* estimates, where these exist and are of known form in terms of the sample sizes; our form for such unbiased estimates is much simpler than that obtained using polykays and tables of the symmetric functions. Examples include functions of a mean vector (such as the ratio of two means and the inverse of a mean), standard deviation, correlation, return times and exceedances. These p^{th} order estimates require only $\sim n$ calculations. This is in sharp contrast with computationally intensive bias reduction methods such as the p^{th} order bootstrap and jackknife, which require $\sim n^p$ calculations.

Key-Words:

- *bias reduction; correlation; exceedances; multisample; multivariate; nonparametric; ratio of means; return times; standard deviation; von Mises derivatives.*

AMS Subject Classification:

- 62G05, 62G30.

1. INTRODUCTION

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. 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 $T(\cdot)$. For excellent reviews of bias reduction methods, we refer the readers to Gray and Schucany [11], Anderson *et al.* [1], Zacks [30], Efron [8], Hall [12], and Chapter 4 of Beirlant *et al.* [2].

Recently, various analytical methods have been developed for bias reduction in parametric settings. Withers [27] developed methods for bias reduction based on Taylor series expansions. Sen [18] and originally von Mises [22] established asymptotic normality of $\sqrt{n}\,\big\{T(\widehat{F}) - T(F)\big\}$ as $n \to \infty$ under suitable regularity conditions. Cabrera and Fernholz [3], [4] 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 [3], [4] established under suitable regularity conditions that the target estimator has smaller bias and mean squared error than the original estimator. See also Fernholz [9].

The first analytical bias reduction method in a nonparametric setting was proposed by Withers and Nadarajah [29]. The technical tools required for Withers and Nadarajah [29] were contained in an unpublished technical report cited there as Withers (1994a).

This paper is an update of the unpublished technical report. The emphasis of this paper is to describe how to find estimates of low bias for $T(F)$. Because of the material in Withers and Nadarajah [29], the emphasis here will not be on numerical illustrations or applications. In Withers and Nadarajah [29], the estimates proposed here were compared to alternatives. We showed in particular that our estimates consistently outperform bootstrapping, jackknifing and those due to Sen [18] and Cabrera and Fernholz [3], [4]. We also provided computer programs in MAPLE for implementation of the proposed estimates.

Suppose we have $k \geq 1$ independent samples of sizes $n_1, ..., n_k$ from distribution functions (d.f.s) $F = (F_1, ..., F_k)$ on $\mathbb{R}^{s_1}, ..., \mathbb{R}^{s_k}$. Let $\widehat{F} = (\widehat{F}_1, ..., \widehat{F}_k)$ denote their sample d.f.s and let n be the minimum sample size. The problem we consider in this paper is that of finding an estimate of low bias for an arbitrary smooth functional $T(F)$. The natural estimate $T(\widehat{F})$ generally has bias $\sim n^{-1}$, that is, $O(n^{-1})$ as $n \to \infty$.

For the reader's convenience, in Section 2, we repeat the definition of functional derivatives and rules for obtaining them given in Withers [28]. In Section 3, we have a formal asymptotic expansion of the form

$$ET(\widehat{F}) = \sum_{r=0}^{\infty} n^{-r} C_r \ , \tag{1.1}$$

where $C_0 = T(F)$. The coefficient of n^{-r} in $ET(\widehat{F})$, $C_r(F,T) = C_r$ may be written in terms of the (functional or von Mises) derivatives of $T(\widehat{F})$ of order $\leq 2r$, and is given in Section 3 explicitly for $r \leq 4$.

From (1.1) if a functional $T_{(n)}(F)$ can be expanded as

$$T_{(n)} = \sum_{i=0}^{\infty} n^{-i}\, T_i(F)$$

then

$$\begin{aligned} ET_{(n)}(\widehat{F}) &= \sum_{i=0}^{\infty} n^{-i}\, ET_i(\widehat{F}) \\ &= \sum_{i=0}^{\infty} n^{-i} \sum_{r=0}^{\infty} n^{-r} C_r(F, T_i) \\ &= \sum_{j=0}^{\infty} \sum_{r=0}^{j} n^{-j}\, C_r(F, T_{j-r}) \\ &= \sum_{j=0}^{\infty} n^{-j}\, C_j(\mathbf{T}) \ , \end{aligned}$$

where

$$C_j(\mathbf{T}) = \sum_{r=0}^{j} C_r(F, T_{j-r}) \ .$$

Defining T_i iteratively by $T_0 = T$ and

$$T_i(F) = -\sum_{j=1}^{i} C_j(F, T_{i-j}) \tag{1.2}$$

for $i \geq 1$ it follows that for $p \geq 1$

$$T_{n,p}(F) = \sum_{i=0}^{p-1} n^{-i}\, T_i(F) \tag{1.3}$$

satisfies

$$
\begin{aligned}
ET_{n,p}(\widehat{F}) &= \sum_{i=0}^{p-1} n^{-i} ET_i(\widehat{F}) \\
&= \sum_{i=0}^{p-1} n^{-i} \sum_{r=0}^{\infty} n^{-r} C_r(F, T_i) \\
&= \sum_{i=0}^{p-1} n^{-i} \left[\sum_{r=0}^{p-1} n^{-r} C_r(F, T_i) + \sum_{r=p}^{\infty} n^{-r} C_r(F, T_i) \right] \\
&= \sum_{i=0}^{p-1} n^{-i} \sum_{r=0}^{p-1} n^{-r} C_r(F, T_i) + \sum_{i=0}^{p-1} n^{-i} \sum_{r=p}^{\infty} n^{-r} C_r(F, T_i) \\
&= \sum_{j=0}^{p-1} n^{-j} \sum_{r=0}^{j} C_r(F, T_{j-r}) + \sum_{i=0}^{p-1} \sum_{r=p}^{\infty} n^{-i-r} C_r(F, T_i) + O(n^{-p}) \\
&= T_0(F) + \sum_{j=1}^{p-1} n^{-j} T_j(F) + \sum_{j=1}^{p-1} n^{-j} \sum_{r=1}^{j} C_r(F, T_{j-r}) \\
&\quad + \sum_{i=0}^{p-1} \sum_{r=p}^{\infty} n^{-i-r} C_r(F, T_i) + O(n^{-p}) \\
&= T_0(F) + \sum_{i=0}^{p-1} \sum_{r=p}^{\infty} n^{-i-r} C_r(F, T_i) + O(n^{-p}) \\
&= T(F) + O(n^{-p}) \ ,
\end{aligned}
$$

where the two middle terms in the third last step cancel out because of (1.2). So, we can write

$$ET_{n,p}(\widehat{F}) \;=\; T(F) + O(n^{-p}) \ .$$

So, $T_{n,p}(\widehat{F})$ is a p^{th} *order estimate* in the sense that it has bias $O(n^{-p})$. This result was given for the case $k = 1$, $p = 2$ using a different approach in an unpublished technical report by Jaeckel [13].

Note that $T_i(\widehat{F})$ given by (1.2) is the coefficient of n^{-i} in the expansion in powers of n^{-1} of the unbiased estimate (UE) of $T(F)$, if an UE exists.

Section 4 gives $T_i(F)$ explicitly in terms of the first $2\,i$ derivatives of $T(F)$ for $i \leq 3$. So, $T_{n,4}(\widehat{F})$ is an explicit estimate of bias $O(n^{-4})$. Proposition 4.1 shows how to obtain from (1.3) an estimate of bias $O(n^{-p})$ of the form $S_{n,p}(\widehat{F})$, where

$$S_{n,p}(F) \;=\; \sum_{i=0}^{p-1} S_i(F) / \big\{(n-1) \cdots (n-i)\big\} \ .$$

This estimate is *unbiased* for one sample if $T(F)$ is a polynomial in F (such as a moment or cumulant) of degree up to p.

Section 5 gives examples and makes comparisons with the UEs of central moments and cumulants given by James [14] and by Fisher [10]. Our method is demonstrated to be give much simpler results for UEs of products of moments than the polykay system of Wishart [23] as expounded in Section 12.22 of Stuart and Ord [19] using tables of the symmetric functions.

Examples 5.1 to 5.3 estimate an arbitrary function of the vector $\boldsymbol{\mu}(F)$, the mean of one multivariate distribution. Example 5.2 specializes to $T(F) = \mathbf{a}'\boldsymbol{\mu}(F)/\mathbf{b}'\boldsymbol{\mu}(F)$, where $\mathbf{a}, \mathbf{b}$ are given s_1-vectors, in particular for the ratio of means of a bivariate sample,

$$T(F) = \mu_1(F)/\mu_2(F) \ .$$

Examples 5.4 and 5.5 estimate an arbitrary function of the means of k univariate distributions; in particular it considers the case of two univariate samples ($k = 2$, $s_1 = s_2 = 1$) with

$$T(F) = \mu(F_1)/\mu(F_2) \ .$$

Example 5.6 gives an explicit expression for the general derivative of the r^{th} central moment μ_r. Together with the chain rule of Appendix A this enables one to obtain a p^{th} order estimate of any smooth function of moments. In particular, we give fourth order estimates for *any central moment* and UEs for μ_r for $r \leq 7$.

Examples 5.7 to 5.11 extend this to an arbitrary product of moments. An alternative matrix method for obtaining UEs of products of moments is given there. This involves obtaining simultaneously the UEs of all moment products of a given degree. Examples 5.12 to 5.15 give fourth order estimates of the standard deviation and functions of it. Example 5.16 gives third order estimates of the ratio of the mean to the standard deviation.

Examples 5.17 to 5.21 give applications to return times and exceedances. Examples 5.22 and 5.23 illustrate how to obtain UEs for multivariate moments and cumulants from univariate analogs. Finally, Examples 5.24 and 5.25 give second order estimates for the correlation and its square.

The method can also be used to estimate with reduced bias any cumulant of $T(\widehat{F})$. This is illustrated in Section 6 which gives a third order estimate for the covariance of any estimate of the form $\mathbf{T}(\widehat{F})$, where now $\mathbf{T}$ may be a vector. For example, by Example 5.1, if $k = 1$ and $T(F)$ is any function of $\boldsymbol{\mu}(F)$ (such as $\mu_1(F)/\mu_2(F)$) if $s_1 = 2$), this estimate is a function of the mean and covariance of F only, whereas C_1 depends also on the third moment.

Section 7 shows how to estimate the covariance of an estimate of bias.

There are, of course, other p^{th} order estimates of $T(F)$, but they are all *computationally intensive*, requiring $O(n^p)$ calculations (except in special cases), whereas *our method requires only $O(n)$ calculations* for fixed p. The main examples are, firstly, the $(p-1)^{\text{th}}$ iterated bootstrap, $\widehat{\theta}_{p-1}$ of equation (1.35) of Hall [12] in which $(-1)^{i+1}$ should be inserted in the right hand side; and, secondly, the p^{th} order jackknife $\widehat{\theta}^{p-1}$ of equation (4.17) of Schucany *et al.* [17], a ratio of $p \times p$ determinants. To see that this requires $O(n^p)$ calculations note that t_p of their equation (4.19) requires $O(n^p)$ calculations.

The techniques given here can also be applied to quantify their biases. Note that if A and B are two p^{th} order estimates of $T(F)$ then $A - B = O_p(n^{-p})$.

Appendix A gives a very useful chain rule for obtaining the derivatives of a function of a functional. Appendix B gives some results used to obtain $\{T_i\}$ of (1.3). Appendix C shows how to estimate the number of simulated samples needed to estimate the bias to within a given relative error.

[21] by an entirely different method obtained an expansion of the form (1.1) for

$$m(v) \;=\; T(F) \;=\; \prod_{i=1}^{s} E X^{v_i} \;,$$

where $X \sim F$, and so also for $\mu_r(F)$. For these cases he constructs estimates of bias $O(n^{-p})$ given $p \geq 1$. He shows for $T(F) = m(v)$ that the UE $T_{n,\infty}(\widehat{F})$ converges if $E|X|^h < \infty$, where $h = \sum_{i=1}^{s} v_i$ and $n-1 >$ the number of partitions of h. His expression on page 12, Theorem 4, is incorrect. He gives

$$\operatorname{var} \widehat{m}(v) \;=\; n^{-1} V + O(n^{-2}) \;,$$

where

$$V \;=\; m(v)^2\,(A - s^2) \qquad \text{and} \qquad A = \sum_{i=1}^{s} m_{2v_i}\, m_{v_i}^2 \;.$$

Here, A should be

$$\sum_{i,j=1}^{s} m_{v_i+v_j}\, m_{v_i}^{-1}\, m_{v_j}^{-1} \;.$$

For the case $T(F) = \mu^3$ his Table 2 illustrates through simulations for $F = U(0,1)$ and $n = 5,10$ how the bias of $T_{n,p}(\widehat{F})$ falls to zero as p increases.

Throughout the paper, we shall assume that $T(F)$ and all of its relevant derivatives are continuous and bounded, and that (1.1) converges with each term and its relevant derivatives continuous and bounded.

2. FUNCTIONAL PARTIAL DERIVATIVES AND NOTATION

Let $\mathcal{F}_s$ denote the space of d.f.s on $\mathbb{R}^s$. Let $\mathbf{x}, \mathbf{y}, \mathbf{x}_1, ..., \mathbf{x}_r$ be points in $\mathbb{R}^s$, $F \in \mathcal{F}_s$ and $T\colon \mathcal{F}_s \to \mathbb{R}$. In Withers [25] and originally in [22], the r^{th} order functional derivative of $T(F)$ at $(\mathbf{x}_1, ..., \mathbf{x}_r)$

$$T_{\mathbf{x}_1,...,\mathbf{x}_r} = T_F(\mathbf{x}_1, ..., \mathbf{x}_r) ,$$

was defined. It is characterized by the formal functional Taylor series expansion: for G in $\mathcal{F}_s$,

$$T(G) - T(F) \approx \sum_{r=1}^{\infty} \int^r T_F(\mathbf{x}_1, ..., \mathbf{x}_r) \prod_{j=1}^{r} d\big(G(\mathbf{x}_j) - F(\mathbf{x}_j)\big)/r! , \tag{2.1}$$

where $\int^r$ denote r integral signs, and the constraints $T_{\mathbf{x}_1,...,\mathbf{x}_r}$ is symmetric in its r arguments, and

$$\int T_{\mathbf{x}_1,...,\mathbf{x}_r}\, dF(\mathbf{x}_1) = 0 .$$

These imply $F(\mathbf{x}_j)$ in (2.1) can be replaced by zero. In particular, it was shown that, for $0 \le \varepsilon \le 1$,

$$T_x = \partial T\big(F + \varepsilon\,(\delta_x - F)\big)/\partial\varepsilon$$

at $\varepsilon = 0$, where δ_x is the d.f. putting mass 1 at x, that is $\delta_x(y) = I(x \le y) = 1$ if $x \le y$ and 0 otherwise. For example, $T(F) = F(y)$ has first derivative $T_x = T_F(x) = \delta_x(y) - F(y) = F(y)_x$, say.

Also, $T_{\mathbf{x}_1,...,\mathbf{x}_r} = 0$ if $T(F)$ is a 'polynomial in F' of degree less than r (for example, a moment or cumulant of F of order less than r), so that the Taylor series in (2.1) consists of only $r-1$ terms. Note that $T(F)$ is a polynomial in F of degree m if for any G in $\mathcal{F}_s$, $T(F + \varepsilon(G - F))$ is a polynomial in ε of degree m.

Suppose now that $F = (F_1, ..., F_k)$ consists of k distributions on $\mathbb{R}^{s_1}, ..., \mathbb{R}^{s_k}$ and that $T(F)$ is a real functional of F. Then the *functional partial derivative* of $T(F)$ at

$$\begin{pmatrix} a_1, ..., a_r \\ \mathbf{x}_1, ..., \mathbf{x}_r \end{pmatrix}$$

is defined by

$$T^{a_1,...,a_r}_{\mathbf{x}_1,...,\mathbf{x}_r} = T_F\begin{pmatrix} a_1, ..., a_r \\ \mathbf{x}_1, ..., \mathbf{x}_r \end{pmatrix} ,$$

where $\mathbf{x}_i$ in $\mathbb{R}^{s_{a_i}}$ and a_i in $\{1, 2, ..., k\}$, and is obtained by treating the lower order functional partial derivatives and $T(F)$ as functionals of F_a alone for $a = a_1, ..., a_r$.

For example, $T^{a,...,a}_{\mathbf{x}_1,...,\mathbf{x}_r}$ is the ordinary functional derivative of $S(F_a) = T(F)$ at $(\mathbf{x}_1, ..., \mathbf{x}_r)$, and $T^{a,...,a,b,...,b}_{\mathbf{x}_1,...,\mathbf{x}_r\mathbf{y}_1,...,\mathbf{y}_s}$ is the ordinary functional derivative of $S(F_b) = T^{a,...,a}_{\mathbf{x}_1,...,\mathbf{x}_r}$ at $(\mathbf{y}_1, ..., \mathbf{y}_s)$.

Just as $\partial^2 f(x,y)/\partial x\,\partial y = \partial^2 f(x,y)/\partial y\,\partial x$ under mild conditions, swapping columns of $T^{a_1,...,a_r}_{\mathbf{x}_1,...,\mathbf{x}_r}$ (for example, $\substack{a_1\\ \mathbf{x}_1}$ and $\substack{a_2\\ \mathbf{x}_2}$) will not alter its value. So, $T^{a,...,a,b,...,b}_{\mathbf{x}_1,...,\mathbf{x}_r,\mathbf{y}_1,...,\mathbf{y}_s}$ is also the ordinary functional derivative of $S(F_a) = T^{b,...,b}_{\mathbf{y}_1,...,\mathbf{y}_r}$ at $(\mathbf{x}_1, ..., \mathbf{x}_s)$.

The partial derivatives may also be characterized by the formal functional Taylor series expansion: for $G = (G_1, ..., G_k) \in \mathcal{F}_{s_1}\times\cdots\times\mathcal{F}_{s_k}$,

$$T(G) - T(F) \approx \sum_{r=1}^{\infty} \int^r T_F\begin{pmatrix} a_1, ..., a_r \\ \mathbf{x}_1, ..., \mathbf{x}_r \end{pmatrix} \prod_{j=1}^{r} d\big(G_{a_j}(\mathbf{x}_j) - F_{a_j}(\mathbf{x}_j)\big)/r! \tag{2.2}$$

with summation of the repeated subscripts $a_1, ..., a_r$ over their range $1, ..., p$ implicit, together with the constraints

$$T^{a_1,...,a_r}_{\mathbf{x}_1,...,\mathbf{x}_r} \text{ is not altered by swapping columns },$$

and

$$\int T^{a_1,...,a_r}_{\mathbf{x}_1,...,\mathbf{x}_r}\, dF_{a_1}(\mathbf{x}_1) \;=\; 0 \;.$$

These imply $F_{a_j}(\mathbf{x}_j)$ in (2.2) can be replaced by zero. The partial derivatives may also be calculated using

$$T^{a_1,...,a_{r+1}}_{\mathbf{x}_1,...,\mathbf{x}_{r+1}} = \Big(T^{a_1,...,a_r}_{\mathbf{x}_1,...,\mathbf{x}_r}\Big)^{a_{r+1}}_{\mathbf{x}_{r+1}} + \sum_{i=1}^{r} \delta_{a_i,a_{r+1}}\, T\Big\langle{}^{a_1,...,a_{r+1}}_{\mathbf{x}_1,...,\mathbf{x}_{r+1}}\Big\rangle_i \;, \tag{2.3}$$

where $\delta_{i,j} = 1$ or 0 for $i = j$ or $i \neq j$, $\langle\ \rangle_i$ means 'drop the i^{th} column', and $T^a_{\mathbf{x}}$ denotes the ordinary functional derivative of $S(F_a) = T(F)$ at $\mathbf{x}$. The proof of (2.3) is as for equation (2.6) of Withers [25].

3. EXPANSIONS FOR BIAS

Perhaps the easiest method to obtain expressions for the bias coefficients $\{C_r\}$ of (1.1) and the bias reduction coefficients $\{T_i(F)\}$ of (1.3) is from their parametric analogs, given in equation (A.1) and Appendix D (for $i \le 3$) of Withers [27]. The method is to identify $(\theta, \widehat{\theta}, t, \sum)$ with $(F, \widehat{F}, T, \int)$, where the integral is with respect to the appropriate d.f. F_i. This method was used in Withers [28] to derive non-parametric confidence intervals of level $1-\alpha+O(n^{-j/2})$ from their parametric analogs. It is convenient to set

$$T\big(a^i, b^j, ...\big) \;=\; \int\!\cdots\!\int T_F\begin{pmatrix} a^i, b^j & \cdots \\ x^i, y^j & \cdots \end{pmatrix} dF_a(x)\, dF_b(y) \cdots \;, \tag{3.1}$$

where x^i denotes a string of i x's (not a product) and similarly, for a^i. In the notation of Withers [28] this is $[1^i, 2^j, ...]_{a,b,...}$. Setting

$$\lambda_a = n/n_a \qquad \text{with} \quad n = \min n_i \ , \tag{3.2}$$

the above approach yields

$$C_1 \ = \ |2|/2 \ , \qquad C_2 \ = \ |3|/6 + |2^2|/8 \ , \tag{3.3}$$

$$C_3 \ = \ |4|/24 + |2,3|/12 + |2^3|/48, \tag{3.4}$$

$$C_4 \ = \ |5|/120 + |2,4|/48 + |3^2|/72 + |2^2,3|/48 + |2^4|/384 \ , \tag{3.5}$$

where

$$\begin{aligned}
|2| \ &= \sum \lambda_a \, T(a^2) \ , \\
|3| \ &= \sum \lambda_a^2 \, T(a^3) \ , \\
|2^2| \ &= \sum \lambda_{a_1} \lambda_{a_2} \, T(a_1^2, a_2^2) \ , \\
|4| \ &= \sum \lambda_a^3 \Big\{ T(a^4) - 3\, T(a^2, a^2) \Big\} \ , \\
|2,3| \ &= \sum \lambda_a \, \lambda_b^2 \, T(a^2, b^3) \ , \\
|2^3| \ &= \sum \lambda_{a_1} \lambda_{a_2} \lambda_{a_3} \, T(a_1^2, a_2^2, a_3^2) \ , \\
|5| \ &= \sum \lambda_a^4 \Big\{ T(a^5) - 10\, T(a^2, a^3) \Big\} \ , \\
|2,4| \ &= \sum \lambda_a \, \lambda_b^3 \Big\{ T(a^2, b^4) - 3\, T(a^2, b^2, b^2) \Big\} \ , \\
|3^2| \ &= \sum \lambda_{a_1}^2 \lambda_{a_2}^2 \, T(a_1^3, a_2^3) \ , \\
|2^2, 3| \ &= \sum \lambda_{a_1} \lambda_{a_2} \, \lambda_b^2 \, T(a_1^2, a_2^2, b^3) \ , \\
|2^4| \ &= \sum \lambda_{a_1} \lambda_{a_2} \lambda_{a_3} \lambda_{a_4} \, T(a_1^2, a_2^2, a_3^2, a_4^2) \ .
\end{aligned}$$

For example, if $k = 1$ (one sample) then

$$C_1 = T(1^2)/2 \ , \qquad C_2 \ = T(1^3)/6 + T(1^2, 1^2)/8 \ , \qquad \dots \ . \tag{3.6}$$

More generally,

$$\begin{aligned}
|A^i| \ &= \sum \lambda_{a_1}^{A-1} \cdots \lambda_{a_i}^{A-1} \, |A^i|_{a_1,...,a_i} \ , \\
|A^i, B^j| \ &= \sum \lambda_{a_1}^{A-1} \cdots \lambda_{a_i}^{A-1} \lambda_{b_1}^{B-1} \cdots \lambda_{b_j}^{B-1} \, |A^i B^j|_{a_1,...,a_i,b_1,...,b_j}
\end{aligned} \tag{3.7}$$

with each $a_1, ..., b_j$ summed over $1, ..., k$,

$$\begin{aligned}
|A^i, B^j|_{a_1,...,a_i,b_1,...,b_j} &= T\big(a_1^A, ..., a_i^A, b_1^B, ..., b_j^B\big) \qquad \text{if} \quad A \text{ and } B = 2 \text{ or } 3 \ , \\
|4|_a \ &= T(a^4) - 3\, T(a^2, a^2) \ , \\
|5|_a \ &= T(a^5) - 10\, T(a^2, a^3) \ , \\
|2,4|_{a,b} \ &= T(a^2, b^4) - 3\, T(a^2, b^2, b^2) \ .
\end{aligned}$$

For example,

$$|A^2| \;=\; \sum \lambda_{a_1}^{A-1} \lambda_{a_2}^{A-1}\, |A^2|_{a_1,a_2} \;,$$

and

$$\begin{aligned} |A^2|_{a_1,a_2} \;&=\; T(a_1^A, a_2^A) \qquad \text{if} \quad A = 2 \text{ or } 3 \\ &=\; \iint T_F \begin{pmatrix} a_1^A, a_2^A \\ x^A, y^A \end{pmatrix} dF_{a_1}(x)\, dF_{a_2}(y) \;, \end{aligned}$$

so for the one sample case ($k = 1$),

$$\begin{aligned} &|A^i| \;=\; T(1^A, ..., 1^A) \qquad \text{if} \quad A = 2 \text{ or } 3 \;, \\ &|A^i, B^j| \;=\; T\big(1^A, ..., 1^A, 1^B, ..., 1^B\big) \qquad \text{if} \quad A \text{ and } B = 2 \text{ or } 3 \;, \\ &|4| \;=\; T(1^4) - 3\,T(1^2, 1^2) \;, \qquad |5| \;=\; T(1^5) - 10\,T(1^2, 1^3) \;, \\ &|2, 4| \;=\; T(1^2, 1^4) - 3\,T(1^2, 1^2, 1^2) \;. \end{aligned}$$

The general term C_r is given by equation (A.1) of Withers [27], (3.2), (3.7), and

$$|i, j, ...|_{a,b,...} \;=\; \int^i d^i \kappa'_a(\mathbf{x}_1, ..., \mathbf{x}_i) \int^j d\kappa'_b(\mathbf{y}_1, ..., \mathbf{y}_j) \,\cdots\, T_F \begin{pmatrix} a, ..., a, b, ..., b \\ \mathbf{x}_1, ..., \mathbf{x}_i, \mathbf{y}_1, ..., \mathbf{y}_j \end{pmatrix} \cdots \Big) \,,$$

where $\int^i d^i \kappa'_a(\mathbf{x}_1, ..., \mathbf{x}_i)$ is the Lebesgue–Stieltjes integral,

$$\begin{aligned} &\mathbf{x}_1 \wedge \mathbf{x}_2 \wedge \cdots \;=\; \min(\mathbf{x}_1, \mathbf{x}_2, ...) \;\; \text{taken componentwise} \;, \\ &f_{1,2,...} \;=\; F_a(\mathbf{x}_1 \wedge \mathbf{x}_2 \wedge \cdots) \;, \\ &\kappa_a(\mathbf{x}_1, \mathbf{x}_2, ...) \;=\; \kappa(\mathbf{Y}_1, \mathbf{Y}_2, ...) \,, \;\; \text{the joint cumulant at } \mathbf{Y}_j = I(\mathbf{X}_a \le \mathbf{x}_j) \;, \\ &\kappa'_a(\mathbf{x}_1, \mathbf{x}_2, ...) \;=\; \kappa_a(\mathbf{x}_1, \mathbf{x}_2, ...) \quad \text{expressed as a function of } \{f_{i,j,...}\} \text{ at } f_i \equiv 0 \;, \end{aligned}$$

and I is the indicator function and $X_a \sim F_a$. For example, using an obvious summation notation

$$\begin{aligned} \kappa_a(\mathbf{x}_1, \mathbf{x}_2) \;&=\; f_{1,2} - f_1 f_2 \;, \\ \kappa_a(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \;&=\; f_{1,2,3} - \sum^3 f_{1,2}\, f_3 + 2\, f_1 f_2 f_3 \;, \\ \kappa_a(\mathbf{x}_1, ..., \mathbf{x}_4) \;&=\; f_{1,...,4} - \sum^4 f_{1,2,3}\, f_4 - \sum^3 f_{1,2}\, f_{3,4} \;, \end{aligned}$$

imply

$$\begin{aligned} &\kappa'_a(\mathbf{x}_1, \mathbf{x}_2) \;=\; f_{1,2} \;, \qquad \kappa'_a(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \;=\; f_{1,2,3} \;, \\ &\kappa'_a(\mathbf{x}_1, ..., \mathbf{x}_4) \;=\; f_{1,...,4} - \sum^3 f_{1,2}\, f_{3,4} \;. \end{aligned}$$

As a check if $k = 1$, $(C_1, C_2) = (a_{1,1}, a_{1,2})$ on page 580 of Withers [25].

4. ESTIMATES OF BIAS $O(n^{-4})$

Here, we give expressions for $\{T_i,\ i \le 3\}$ of (1.2) and for $\{S_i,\ i \le 3\}$ of Proposition 4.1. Estimates of bias $O(n^{-4})$ are then given by $T_{n,4}(\widehat{F})$ of (1.3) and $S_{n,4}(\widehat{F})$ of (4.5), (4.6).

From their parametric analogs in Appendix D of Withers [27], we obtain (see Appendix B) in the notation of (3.7)

$$T_1(F) = -|2|/2\ , \qquad T_2(F) = |3|/3 + |2^2|/8 - \sum \lambda_a^2\, T(a^2)/2\ , \tag{4.1}$$

and

$$\begin{aligned} T_3(F) = & -\sum \lambda_a^3\, T(a^2)/2 + \sum \lambda_a^3\, T(a^3) - \sum \lambda_a^3\, T(a^4)/4 \\ & + \sum \lambda_a^3\, T(a^2,a^2)/2 + \sum \lambda_a \lambda_b^2\, T(a^2,b^2)/4 - \sum \lambda_a \lambda_b^2\, T(a^2,b^3)/6 \\ & - \sum \lambda_a \lambda_b \lambda_c\, T(a^2,b^2,c^2)/48\ . \end{aligned}$$

For the one sample case ($k = 1$), these reduce to

$$T_1(F) = -T(1^2)/2\ , \tag{4.2}$$

$$T_2(F) = T(1^3)/3 + T(1^2,1^2)/8 - T(1^2)/2\ , \tag{4.3}$$

$$\begin{aligned} T_3(F) = & -T(1^2)/2 + T(1^3) - T(1^4)/4 + 3\,T(1^2,1^2)/4 - T(1^2,1^3)/6 \\ & - T(1^2,1^2,1^2)/48\ . \end{aligned} \tag{4.4}$$

Proposition 4.1. *Let $\{N_i(n),\ i \ge 0\}$ be given functions satisfying $N_i(n) / n^{-i} \to 1$. Then (1.3) may be rewritten as $S_{n,p}(F) + O(n^{-p})$, where*

$$S_{n,p}(F) = \sum_{i=0}^{p-1} N_i(n)\, S_i(F)\ . \tag{4.5}$$

So, $S_{n,p}(\widehat{F})$ is a p^{th} order estimate of $T(F)$.

Suppose now that it is known that there exists an UE and that it has the form $S_{n,p}(\widehat{F})$. Then this gives a method of obtaining it. For example, if $k = 1$ and $T(F)$ is a polynomial of degree p in F (for example, a product of moments or cumulants of total degree p), then the UE of $T(F)$ has the form (4.5) with

$$N_i(n) = 1/(n-1)_i\ , \tag{4.6}$$

where $(r)_i = r!/(r-i)! = r(r-1)\cdots(r-i+1)$. In this case, $\{S_i\}$ are given in terms of $\{T_i\}$ by equation (2.17.2) of Withers [27]:

$$S_0 = T\,, \quad S_1 = T_1\,, \quad S_2 = T_2 - T_1\,, \quad S_3 = T_3 - 3\,T_2 + 2\,T_1\,, \quad \ldots$$

If $k=1$ and we choose $N_i(n)$ as in (4.6), then S_i is generally a simpler expression than T_i:

$$S_0(F) \;=\; T(F)\,, \qquad S_1(F) \;=\; -T(1^2)/2\;,$$

$$S_2(F) \;=\; T(1^3)/3 + T(1^2,1^2)/8\;, \tag{4.7}$$

$$S_3(F) \;=\; -T(1^4)/4 + 3\,T(1^2,1^2)/8 - T(1^2,1^3)/6 - T(1^2,1^2,1^2)/48\;. \tag{4.8}$$

If $k\neq 1$,

$$S_0(F) \;=\; T(F)\,, \qquad S_1(F) = T_1(F) \qquad \textit{of}\;\; (4.2)\;,$$

$$S_2(F) \;=\; T_2(F) - T_1(F) \;=\; |3|/3 + |2^2|/8 + \sum(\lambda_a - \lambda_a^2)\,T(a^2)/2\;,$$

and so on.

For $p\geq 1$, set $e_{n,p}(T,F) = T_{n,p}(F)$ of (1.3) and let $\{U_i(F)\}$ be smooth. Then a p^{th} order estimate of

$$U_n(F) \;=\; \sum_{i=0}^{\infty} n^{-i}\,U_i(F)$$

is

$$U^{\star}_{(n)p}(\widehat{F}) \;=\; \sum_{i=0}^{p-1} n^{-i}\,e_{n,p-i}(U_i,\widehat{F})\;. \tag{4.9}$$

Let $\kappa_r(\mathbf{X})$ denote any r^{th} order cumulant of $\mathbf{X}$, any $q\times 1$ random vector. Then $n^{1-r}\kappa_r(T(\widehat{F}))$ can be expanded in the form (4.9); a method of obtaining $\{U_i\}$ is illustrated in Section 6 for the case $r=2$.

Proposition 4.2. *$ET(\widehat{F})$ may be infinite or may not exist. For example, this is the case if $k=s=1$, $T(F)=\mu(F)^{-I}$, $I\geq 1$ and F has positive density at zero, or $\dot{F}(x)$ approaches zero too slowly as $x\to 0$. So, page 356 in Quenouille [16] is wrong in giving $\overline{X}^{-1}$ finite bias for $X\sim N$ (2,1). In such cases, our method may be salvaged provided we know an upper bound for $|T(F)|$, say $|T(F)|<u<\infty$. By large deviation theory $P\big(|T(\widehat{F})|\geq u\big) = O\big(\exp(-n\lambda)\big)$, where $\lambda>0$. Typically, $\widetilde{T}_{n,p}(\widehat{F})$ is a p^{th} order estimate of $T(F)$, where*

$$\widetilde{T}_{n,p}(F) \;=\; \begin{cases} T_{n,p}(F)\,, & \textit{if}\;\; |T(F)|<u\;,\\ c, & \textit{otherwise}\;,\end{cases} \tag{4.10}$$

and c is any finite constant, for example, u.

The estimates (4.5) and (4.9) can be adapted similarly, to give $\widetilde{S}_{n,p}(\widehat{F})$ and $\widetilde{U}^{\star}_{n,p}(\widehat{F})$ say. Similarly, if $U_{(n)}(F)$ is the formal expansion of $n^{r-1}\kappa_r\big(T_{n,p}(\widehat{F})\big)$ then

$$U^{\star}_{n,q}(\widehat{F})\,I\big(|T(\widehat{F})|<u\big) \quad \textit{is a } q^{\text{th}} \textit{ order estimate of} \quad n^{r-1}\kappa_r\big(\widetilde{T}_{n,p}(\widehat{F})\big)$$

even if $\kappa_r\big(T(\widehat{F})\big)$ is not finite. For example, the variances in equations (10.17)–(10.20) of Kendall and Stuart [15] are infinite if the density at zero is positive.

An alternative estimate of bias $O(n^{-p})$ is $T^+_{n,p}(\widehat{F}) = T_{n,q}(\widehat{F})$, where $q \le p$ is the maximum integer such that $\{n^{-i}\,T_i(\widehat{F}),\ 0 \le i \le q\}$ decreases in absolute value. This may be useful if $T_{n,p}(\widehat{F})$ diverges. Note that $S^+_{n,p}(F)$ and $\widetilde{T}^+_{n,p}(\widehat{F})$ may be defined analogously from (4.5) and (4.10).

5. EXAMPLES

Example 5.1. Suppose $k=1$, $\mathbf{X} \sim F$ on $\mathbb{R}^s$ and $T(F) = g(\mu)$, where $\boldsymbol{\mu} = \boldsymbol{\mu}(F) = E\mathbf{X}$ has dimension $s_1 = s$ and g is a function with finite derivatives at $\boldsymbol{\mu}$. By the chain rule (A.6) or (A.7) of Appendix A,

$$T_F(\mathbf{x}_1, ..., \mathbf{x}_r) \;=\; g_{j_1,...,j_r}\, \mu_{j_1,\mathbf{x}_1} \cdots \mu_{j_r,\mathbf{x}_r}\ ,$$

where

$$\mu_{j,\mathbf{x}} = \mu_{j,F}(\mathbf{x}) = x_j - \mu_j\ , \qquad g_{\cdots} = g_{\cdots}(\boldsymbol{\mu})$$

are the partial derivatives of $g(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$, and summation of the repeated indices $j_1, ..., j_r$ over their range $1, ..., s$ is implicit. So,

$$T(1^{i_1}, 1^{i_2}, ...) \;=\; g_{j_1,...,j_{i_1},k_1,...,k_{i_2},...}\, \mu[j_1, ..., j_{i_1}]\, \mu[k_1, ..., k_{i_2}] \cdots\ ,$$

where

$$\mu[j_1, ..., j_a] \;=\; \int (x_{j_1} - \mu_{j_1}) \cdots (x_{j_a} - \mu_{j_a})\, dF(\mathbf{x})\ , \tag{5.1}$$

the joint central moment. So,

$$\begin{aligned}
T(1^2) &= g_{i,j}\, \mu[i,j] = \sum_{i=1}^{s} g_{i,i}\, \mu[i,i] + 2 \sum_{1\le i<j\le s} g_{i,j}\, \mu[i,j]\ ,\\
T(1^3) &= g_{i,j,k}\, \mu[i,j,k]\ ,\\
T(1^4) &= g_{i,j,k,l}\, \mu[i,j,k,l]\ ,\\
T(1^2,1^2) &= g_{j_1,j_2,k_1,k_2}\, \mu[j_1,j_2]\, \mu[k_1,k_2]\ ,\\
T(1^2,1^3) &= g_{i,j,k,l,m}\, \mu[i,j]\, \mu[k,l,m]\ ,\\
T(1^2,1^2,1^2) &= g_{i,j,k,l,m,n}\, \mu[i,j]\, \mu[k,l]\, \mu[m,n]\ .
\end{aligned}$$

So, by (4.2)–(4.4)

$$\begin{aligned}
T_1(F) &= -C_1 = -g_{i,j}\, \mu[i,j]/2\ ,\\
T_2(F) &= -g_{i,j}\, \mu[i,j]/2 + g_{i,j,k}\, \mu[i,j,k]/3 + g_{i,j,k,l}\, \mu[i,j]\, \mu[k,l]/8\ ,\\
T_3(F) &= -g_{i,j}\, \mu[i,j]/2 + g_{i,j,k}\, \mu[i,j,k] - g_{i,j,k,l} \Big\{\mu[i,j,k,l] - 3\,\mu[i,j]\,\mu[k,l]\Big\}/4\\
&\quad - g_{i,j,k,l,m}\, \mu[i,j]\, \mu[k,l,m]/6 - g_{i,j,k,l,m,n}\, \mu[i,j]\, \mu[k,l]\, \mu[m,n]/48\ .
\end{aligned}$$

A p^{th} order estimate of $T(F)$ is now given in terms of these by $T_{n,p}(\widehat{F})$ of (1.3).

Example 5.2. Consider Example 5.1 with $g(\boldsymbol{\mu}) = \boldsymbol{\alpha}'\boldsymbol{\mu}/\boldsymbol{\beta}'\boldsymbol{\mu} = N/D$, say, where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are given s-vectors. Its i^{th} order partial derivative with respect to $\boldsymbol{\mu}$ is

$$g_{j_1,\dots,j_i} = (-1)^{i-1}(i-1)!\, D^{-i} \sum^{i} \delta_{j_1}\beta_{j_2}\cdots\beta_{j_i} \,, \tag{5.2}$$

where

$$\delta_i = \alpha_i - \beta_i\, T(F) \tag{5.3}$$

and

$$\sum^{m} f_{i_1,\dots,i_m} = f_{i_1,\dots,i_m} + f_{i_2,\dots,i_m,i_1} + \cdots + f_{i_m,i_1,\dots,i_{m-1}} \,.$$

So,

$$\begin{aligned}
T(1^i) &= (-1)^{i-1}\, i!\, D^{-i}\delta_{j_1}\beta_{j_2}\cdots\beta_{j_i}\,\mu[j_1,\dots,j_i] \,,\\
T(1^2,1^2) &= -4!\, D^{-4}\delta_{j_1}\beta_{j_2}\beta_{j_3}\beta_{j_4}\,\mu[j_1,j_2]\,\mu[j_3,j_4] \,,\\
T(1^2,1^3) &= 4!\, D^{-5}\Big\{2\,\delta_{j_1}/\beta_{j_1} + 3\,\delta_{j_3}/\beta_{j_3}\Big\}\beta_{j_1}\cdots\beta_{j_5}\,\mu[j_1,j_2]\,\mu[j_3,j_4,j_5] \,,\\
T(1^2,1^2,1^2) &= -6!\, D^{-6}\delta_{j_1}\beta_{j_2}\cdots\beta_{j_6}\,\mu[j_1,j_2]\,\mu[j_3,j_4]\,\mu[j_5,j_6] \,.
\end{aligned}$$

In particular, for $g(\boldsymbol{\mu}) = \mu_1/\mu_2$ (the ratio of means for one bivariate sample),

$$\begin{aligned}
T(1^i) &= (-1)^{i-1}\, i!\, \mu_2^{-i}\Big\{\mu[1,2^{i-1}] - T(F)\,\mu[2^i]\Big\},\\
T(1^2,1^2) &= -4!\, \mu_2^{-4}\Big\{\mu[1,2]\,\mu[2^2] - T(F)\,\mu[2^2]^2\Big\},\\
T(1^2,1^3) &= 4!\, \mu_2^{-5}\Big\{2\,\mu[1,2]\,\mu[2^3] + 3\,\mu[2^2]\,\mu[1,2^2] - 5\,T(F)\,\mu[2^2]\,\mu[2^3]\Big\},\\
T(1^2,1^2,1^2) &= -6!\, \mu_2^{-6}\Big\{\mu[1,2] - T(F)\,\mu[2^2]\Big\}\,\mu[2^2]^2,
\end{aligned}$$

so

$$\begin{aligned}
S_1(F) = T_1(F) &= -C_1 = \mu_2^{-2}\Big\{\mu[1,2] - T(F)\,\mu[2^2]\Big\},\\
T_2(F) &= 2\,\mu_2^{-3}\Big\{\mu[1,2^2] - T(F)\,\mu[2^3]\Big\} - T_1(F)\Big\{1 + 3\,\mu_2^{-2}\,\mu[2^2]\Big\},
\end{aligned}$$

$S_2(F)$ is the same as $T_2(F)$ with ‘1 +’ deleted,

$$\begin{aligned}
T_3(F) = {} & \mu_2^{-2}\Big\{\mu[1,2] - T(F)\,\mu[2^2]\Big\}\Big\{1 - 18\,\mu_2^{-2}\mu[2^2] - 8\,\mu_2^{-3}\mu[2^3] + 15\,\mu_2^{-4}\mu[2^2]^2\Big\}\\
& + 6\,\mu_2^{-3}\Big\{\mu[1,2^2] - T(F)\,\mu[2^3]\Big\}\Big\{1 - 2\,\mu_2^{-2}\,\mu[2^2]\Big\}\\
& + 6\,\mu_2^{-4}\Big\{\mu[1,2^3] - T(F)\,\mu[2^4]\Big\}
\end{aligned}$$

and

$$\begin{aligned}
S_3(F) = {} & \mu_2^{-2}\Big\{\mu[1,2] - T(F)\,\mu[2^2]\Big\}\Big\{-9\,\mu_2^{-2}\mu[2^2] - 8\,\mu_2^{-3}\mu[2^3] + 15\,\mu_2^{-4}\mu[2^2]^2\Big\}\\
& - 12\,\mu_2^{-5}\Big\{\mu[1,2^2] - T(F)\,\mu[2^3]\Big\}\,\mu[2^2]\\
& + 6\,\mu_2^{-4}\Big\{\mu[1,2^3] - T(F)\,\mu[2^4]\Big\}.
\end{aligned}$$

Example 5.3. Consider Example 5.1 with $g(\boldsymbol{\mu}) = (\boldsymbol{\alpha}'\boldsymbol{\mu})^p = N^p$, say, where $\boldsymbol{\alpha}$ is a given s-vector. The i^{th} order partial derivative of $g(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ is

$$g_{j_1,\ldots,j_i} = (p)_i\, N^{p-i} \alpha_{j_1} \cdots \alpha_{j_i} \ .$$

Set

$$\alpha_{(i)} = N^{-i} \alpha_{j_1} \cdots \alpha_{j_i}\, \mu[j_1, \ldots, j_i] \ .$$

Then

$$\begin{aligned}
T(1^i) &= (p)_i\, N^p \alpha_{(i)} \ , \\
T(1^2, 1^2) &= (p)_4\, N^p \alpha_{(2)}^2 \ , \\
T(1^2, 1^3) &= (p)_5\, N^p \alpha_{(2)}\, \alpha_{(3)} \ , \\
T(1^2, 1^2, 1^2) &= (p)_6\, N^p \alpha_{(2)}^3 \ ,
\end{aligned}$$

$$\begin{aligned}
T_1(F) &= -C_1 = -(p)_2\, N^p \alpha_{(2)}/2 \ , \\
T_2(F) &= N^p \Big\{ -(p)_2\, \alpha_{(2)}/2 + (p)_3\, \alpha_{(3)}/3 + (p)_4\, \alpha_{(2)}^2/8 \Big\} \ , \\
T_3(F) &= N^p \Big\{ -(p)_2\, \alpha_{(2)}/2 + (p)_3\, \alpha_{(3)} - (p)_4 \big[\alpha_{(4)} - 3\,\alpha_{(2)}^2\big]/4 \\
&\quad - (p)_5\, \alpha_{(2)}\, \alpha_{(3)}/6 - (p)_6\, \alpha_{(2)}^3/48 \Big\} \ .
\end{aligned}$$

In particular, for a univariate sample ($s = 1$) with central moments $\{\mu_r\}$ and $g(\mu) = \mu^p$,

$$\begin{aligned}
S_1(F) &= T_1(F) = -(p)_2\, \mu^{p-2} \mu_2/2 \ , \\
T_2(F) &= -(p)_2\, \mu^{p-2} \mu_2/2 + S_2(F) \ , \\
S_2(F) &= (p)_3\, \mu^{p-3} \mu_3/3 + (p)_4\, \mu^{p-4} \mu_2^2/8 \ , \\
T_3(F) &= -(p)_2\, \mu^{p-2} \mu_2/2 + (p)_3\, \mu^{p-3} \mu_3 - (p)_4\, \mu^{p-4} \big(\mu_4 - 3\,\mu_2^2\big)/4 \\
&\quad -(p)_5\, \mu^{p-5} \mu_3\, \mu_2/6 - (p)_6\, \mu^{p-6} \mu_2^3/48
\end{aligned}$$

and

$$S_3(F) = -(p)_4\, \mu^{p-4} \big(2\,\mu_4 - 3\,\mu_2^2\big)/8 - (p)_5\, \mu^{p-5} \mu_3\, \mu_2/6 - (p)_6\, \mu^{p-6} \mu_2^3/48 \ .$$

In particular, for p a positive integer, by Proposition 4.1, an UE for μ^p is

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

where $S_0(F) = \mu^p$, and

for $p = 2$: $S_1(F) = -\mu_2$,

for $p = 3$: $S_1(F) = -3\,\mu\,\mu_2$, $S_2(F) = 2\,\mu_3$,

for $p = 4$: $S_1(F) = -6\,\mu^2 \mu_2$, $S_2(F) = 8\,\mu\,\mu_3 + 3\,\mu_2^2$, $S_3(F) = -6\,\mu_4 + 9\,\mu_2^2$.

These results may be checked by by solving the system of equations given by page 5 in Wishart [23]. For $p = 4$ the system has seven equations. Alternatively, one may follow the method of Section 12.22 of Stuart and Ord [19] using their tables of the symmetric functions. For example, after some labor one obtains for $p=4$ the UE $T_n(\widehat{F})$, where

$$\begin{aligned}(n-1)_3\, T_n(F) \;=\; & \big(N^3-8n^2+23\,n-30\big)\, m_4 \;-\; n\,\big(n^2-7n+4\big)\, m_3 m_1 \\ & -\, n\,\big(n^2-6\,n+6\big)\, m_2^2 \,+\, n^2\,\big(n-9\big)\, m_2 m_1^2 \,+\, n^3 m_1^4 \;,\end{aligned}$$

where $m_i = EX^i$. Clearly, our method gives a much simpler form.

For $p=-1$, that is $T(F)=\mu^{-1}$, the above gives

$$S_{n,p}(F) \;=\; \sum_{i=0}^{p-1} S_i(F)/(n-1)_i \;,$$

where

$$\begin{aligned} S_0(F) \;&=\; \mu^{-1}\;, \qquad S_1(F) \;=\; -\mu^{-3}\mu_2\;, \\ S_2(F) \;&=\; -2\,\mu^{-4}\mu_3 + 3\,\mu^{-5}\mu_2^2\;, \\ S_3(F) \;&=\; -3\,\mu^{-5}\big(2\,\mu_4 - 3\,\mu_2^2\big) + 20\,\mu^{-6}\mu_3\,\mu_2 - 15\,\mu^{-7}\mu_2^3\;, \end{aligned}$$

so setting $\gamma_r = \mu_r\,\mu^{-r}$, $s_i = S_i(F)/T(F)$ is given by

$$\begin{aligned} s_1 \;&=\; -\gamma_2\;, \\ s_2 \;&=\; -2\,\gamma_3 + 3\,\gamma_2^2\;, \\ s_3 \;&=\; -3\big(2\,\gamma_4 - 3\,\gamma_2^2\big) + 20\,\gamma_3\,\gamma_2 - 15\,\gamma_2^3\;. \end{aligned}$$

Some simulations estimating the bias of $\widetilde{S}_{n,i}(\widehat{F})$ of (4.5), (4.6) and Proposition 4.2 with $c=1/u=\mu/10$ for $1\le i\le 4$, for μ^{-1}, are given in Table 1. The estimates present bias even for $n=100$ and bias-corrected estimates of order n^{-2} (i.e. $p=2$): see Appendix C.

Table 1: Relative bias of $\widetilde{S}_{n,p}(\widehat{F})$ for $T(F)=\mu^{-1}$ estimated from two runs of 5000 simulations.

		$n=10$		$n=100$	
		$p=1$	$p=2$	$p=1$	$p=2$
Norm $(1/2,1)$	Run 1	0.0773	−0.0242	0.0089	0.0013
	Run 2	0.0916	−0.0092	0.0087	0.0011
Norm $(1,1)$	Run 1	−0.0780	−0.0105	−0.0149	−0.0094
	Run 2	−0.0660	−0.0040	−0.0141	−0.0087
Norm $(2,1)$	Run 1	0.0208	−0.0048	−0.0046	−0.0070
	Run 2	0.0202	−0.0056	−0.0056	−0.0078
Exp (1)	Run 1	0.1096	0.0120	0.0052	−0.0045
	Run 2	0.1062	0.0184	0.0062	−0.0035

Example 5.1 estimated a smooth function of the mean of one multivariate distribution. We now estimate a smooth function of the means of k univariate distributions.

Example 5.4. Suppose we have k univariate samples (that is $s_1 = \cdots = s_k = 1$) with $T(F) = g(\boldsymbol{\mu})$, where now $\boldsymbol{\mu} = \big(\mu(F_1), ..., \mu(F_k)\big)$. That is, $T(F)$ is a function of the means of k univariate samples. Then

$$T_F\begin{pmatrix} a_1, ..., a_r \\ x_1, ..., x_r \end{pmatrix} = g_{a_1,...,a_r}\, \mu_{a_1,x_1} \cdots \mu_{a_r,x_r} \ ,$$

where $g \cdots$ is the partial derivative with respect to $\boldsymbol{\mu}$ and

$$\mu_{a,x} = \mu_{F_a}(x) = x - \mu(F_a) = x - \mu_a \ .$$

So,

$$T(a^i, b^j, ...) = g_{a^i,b^j,...}\, \mu_i[a]\, \mu_j[b] \cdots \ ,$$

where

$$\mu_i[a] = \mu_i(F_a) = \int (x - \mu_a)^i\, dF_a(x) \ ,$$

the i^{th} central moment of F_a. So, for λ_a of (3.2),

$$C_1 = \sum_a \lambda_a\, g_{a,a}\, \mu_2[a]/2 \ ,$$

$$C_2 = \sum_a \lambda_a^2\, g_{a,a,a}\, \mu_3[a]/6 + \sum_{a,b} \lambda_a \lambda_b\, g_{a,a,b,b}\, \mu_2[a]\, \mu_2[b]/8 \ ,$$

$$\begin{aligned} C_3 = {} & \sum \lambda_a^3\, g_{a,a,a,a} \big\{\mu_4[a] - 3\,\mu_2[a]^2\big\}/24 \\ & + \sum \lambda_a \lambda_b^2\, g_{a,a,b,b,b}\, \mu_2[a]\, \mu_3[b]/12 + \sum \lambda_a \lambda_b \lambda_c\, g_{a,a,b,b,c,c}\, \mu[a]\, \mu_2[b]\, \mu_2[c]/48 \ , \end{aligned}$$

$$T_1(F) = -C_1 \ ,$$

$$T_2(F) = \sum \lambda_a^2\, g_{a,a,a}\, \mu_3[a]/3 + \sum \lambda_a \lambda_b\, g_{a,a,b,b}\, \mu_1[a]\, \mu_2[b]/8 - \sum \lambda_a^2\, g_{a,a}\, \mu_2[a]/2 \ ,$$

$$\begin{aligned} T_3(F) = {} & -\sum \lambda_a^3\, g_{a,a}\, \mu_2[a]/2 + \sum \lambda_a^3\, g_{a,a,a}\, \mu_3[a] \\ & - \sum \lambda_a^3\, g_{a,a,a,a} \big\{\mu_4[a]/4 + \mu_2[a]^2/2\big\} \\ & + \sum \lambda_a^2 \lambda_b\, g_{a,a,b,b}\, \mu_2[a]\, \mu_2[b]/4 - \sum \lambda_a \lambda_b^2\, g_{a,a,b,b,b}\, \mu_2[a]\, \mu_3[b]/6 \\ & - \sum \lambda_a \lambda_b \lambda_c\, g_{a,a,b,b,c,c}\, \mu_2[a]\, \mu_2[b]\, \mu_2[c]/48 \ . \end{aligned}$$

Example 5.5. Consider Example 5.4 with $g(\boldsymbol{\mu}) = \boldsymbol{\alpha}'\boldsymbol{\mu}/\boldsymbol{\beta}'\boldsymbol{\mu} = N/D$, say, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are given k-vectors. Set

$$\begin{aligned}
\gamma_a &= \alpha_a/\beta_a - T(F) \ , \\
A_{i,k,l} &= D^{-kl} \sum_a \lambda_a^{i+kl-1} \beta_a^k \mu_k(a)^l \gamma_a \ , \\
B_{i,k,l} &= \{A_{i,k,l}\} \quad \text{at } \gamma_a \equiv 1 \ , \\
A_k &= A_{0,k,1} \ , \\
B_k &= B_{0,k,1} \ .
\end{aligned}$$

Then, by (5.2),

$$\begin{aligned}
C_1 &= -A_2 \ , \\
C_2 &= A_3 - 6 A_2 B_2 \ , \\
C_3 &= -A_4 + 3 A_{0,2,2} + 6 A_2 B_3 + 9 A_3 B_2 - 15 A_2 B_2^2 \ , \\
T_1(F) &= A_2 \ , \\
T_2(F) &= 2 A_3 - 3 A_2 B_2 + A_{1,2,1}, \\
T_3(F) &= A_{2,2,1} - 9 A_{1,3,1} - 3 A_3 + 6 A_4 - 12 A_{0,2,2} - 3 A_{1,2,1} B_2 - 3 A_2 B_{1,2,1} \\
&\quad - 8 A_2 B_3 - 12 A_3 B_2 + 15 A_2 B_2^2 \ .
\end{aligned}$$

In particular, for $g(\boldsymbol{\mu}) = \mu_1/\mu_2$ (the ratio of means for two univariate samples), setting $\nu_k = \mu_2^{-k}\mu_k[2]$, we obtain

$$\begin{aligned}
C_1 &= \lambda_2 \nu_2 \, \mu_1/\mu_2 \ , \\
C_2 &= \lambda_2^2 \big(-\nu_3 + 6\nu_2^2\big) \, \mu_1/\mu_2 \ , \\
C_3 &= \lambda_2^3 \big(\nu_4 - 3\nu_2^2 - 15 \, \nu_2 \nu_3 + 15 \, \nu_2^3\big) \, \mu_1/\mu_2 \ , \\
T_1(F) &= -\lambda_2 \nu_2 \, \mu_1/\mu_2 \ , \\
T_2(F) &= \lambda_2^2 \big(-2\nu_3 - \nu_2 + 3\nu_2^2\big) \, \mu_1/\mu_2 \ , \\
T_3(F) &= \lambda_2^3 \big(-6\nu_4 - 6\nu_3 - \nu_2 - 15 \, \nu_2^3 + 20 \, \nu_3 \nu_2 + 18 \, \nu_2^2\big) \, \mu_1/\mu_2 \ .
\end{aligned}$$

This may also be derived from (5.2).

Central moments and functions of them may be viewed as functions of noncentral moments and so dealt with using Examples 5.1 and 5.4. However, it is much more convenient to deal with them directly in terms of the derivatives of the central moments. We now give these.

Example 5.6. One univariate sample (that is $k = s_1 = 1$) with $T(F) = \mu_r(F) = \mu_r$, the r^{th} central moment of $X \sim F$. Let $\mu = \mu(F)$ denote the mean of F. Recall that $(r)_i = r!/(r-i)!$ and set $h_i = \mu_{x_i} = x_i - \mu$. The general derivative of $\mu_r(F)$ is

$$(5.4)\qquad \begin{aligned} T_{x_1,\ldots,x_p} &= \mu_{r,F}(x_1, \ldots, x_p) \\ &= (-1)^p \left\{ (r)_p\, \mu_{r-p} - (r)_{p-1} \sum_{i=1}^{p} \bigl(h_i^{r-p} - \mu_{r-p+1} h_i^{-1} \bigr) \right\} \prod_{j=1}^{p} h_j \ . \end{aligned}$$

For example,

$$\begin{aligned} T_x &= -r\,\mu_{r-1}\mu_x + \mu_x^r - \mu_r \ , \\ T_{x,y} &= (r)_2\, \mu_{r-2}\,\mu_x\,\mu_y - r \sum_{x,y}^{2} \bigl(\mu_x^{r-1} - \mu_{r-1} \bigr)\, \mu_y \ , \\ T_{x,y,z} &= -(r)_3\, \mu_{r-3}\,\mu_x\,\mu_y\,\mu_z + (r)_2 \sum_{x,y,z}^{3} \bigl(\mu_x^{r-2} - \mu_{r-2} \bigr)\, \mu_y\,\mu_z \ . \end{aligned}$$

These basic building blocks are written out more explicitly up to $r = 6$ in Appendix D. Setting $q = i_1 + i_2 + \cdots$, this gives

$$(5.5)\qquad \begin{aligned} \mu_r(1^{i_1}, 1^{i_2}, \ldots) &= (-1)^q \Biggl[(r)_q\, \mu_{r-q} \prod_{j=1}^{\infty} \mu_{i_j} \\ &\qquad - (r)_{q-1} \sum_{I=1}^{\infty} i_I \bigl(\mu_{r-q+i_I} - \mu_{r-q+1}\, \mu_{i_I - 1} \bigr) \prod_{j \neq I}^{\infty} \mu_{i_j} \Biggr] \\ &= \begin{cases} 0, & \text{if } q > r \ , \\ (-1)^{r-1} (r-1)! \displaystyle\prod_{j=1}^{\infty} \mu_{i_j}\,, & \text{if } q = r \ . \end{cases} \end{aligned}$$

For example,

$$(5.6)\qquad \mu_r(1^2) = (r)_2\, \mu_{r-2}\,\mu_2 - 2\,r\,\mu_r \ ,$$

$$(5.7)\qquad \mu_r(1^3) = -(r)_3\, \mu_{r-3}\,\mu_3 + 3\,(r)_2 \bigl(\mu_r - \mu_{r-2}\,\mu_2 \bigr) \ ,$$

$$(5.8)\qquad \mu_r(1^4) = (r)_4\, \mu_{r-4}\,\mu_4 - 4\,(r)_3 \bigl(\mu_r - \mu_{r-3}\,\mu_3 \bigr) \ ,$$

$$(5.9)\qquad \mu_r(1^2, 1^2) = (r)_4\, \mu_{r-4}\,\mu_2^2 - 4\,(r)_3\, \mu_{r-2}\,\mu_2 \ ,$$

$$(5.10)\qquad \mu_r(1^2, 1^3) = -(r)_5\, \mu_{r-5}\,\mu_3\,\mu_2 + (r)_4 \bigl(2\,\mu_{r-3}\,\mu_3 + 3\,\mu_{r-2}\,\mu_2 - 3\,\mu_{r-4}\,\mu_2^2 \bigr) \ ,$$

$$(5.11)\qquad \mu_r(1^2, 1^2, 1^2) = (r)_6\, \mu_{r-6}\,\mu_2^3 - 6\,(r)_5\, \mu_{r-4}\,\mu_2^2 \ .$$

Substituting into the expressions of (3.3)–(3.5) for the coefficient C_i of n^{-i} in the expansion of $E\mu_r(\widehat{F})$ gives

$$T_1(F) \;=\; -C_1 \;=\; r\mu_r - (r)_2\,\mu_{r-2}\mu_2/2\;, \tag{5.12}$$

$$\begin{aligned} C_2 \;&=\; (r)_2\,\mu_r/2 - (r)_2\,(r-1)\,\mu_{r-2}\mu_2/2 - (r)_3\,\mu_{r-3}\mu_3/6 \\ &\quad + (r)_4\,\mu_{r-4}\mu_2^2/8\;, \end{aligned} \tag{5.13}$$

$$\begin{aligned} C_3 \;&=\; -(r)_3\,\mu_r/6 + (r)_3\,(r-1)\,\mu_{r-2}\mu_2/4 + (r)_3\,(r-2)\,\mu_{r-3}\mu_3/6 \\ &\quad + (r)_4\,\mu_{r-4}\big(\mu_4 - 3\,(r-1)\,\mu_2^2\big)/24 - (r)_5\,\mu_{r-5}\mu_3\mu_2/12 \\ &\quad + (r)_6\,\mu_{r-6}\mu_2^3/48\;, \end{aligned}$$

$$\begin{aligned} C_4 \;&=\; (r)_4\,\mu_r/24 - (r)_4\,(r-7)\,\mu_{r-2}\mu_2/12 - (r)_6\,\mu_{r-3}\mu_3/2 \\ &\quad + \mu_{r-4}\big\{-(r)_4\,(r-3)\,\mu_4/24 + (r)_4\big(r^2-3r-8\big)\,\mu_2^2/16\big\} \\ &\quad + \mu_{r-5}\big\{-(r)_5\,\mu_5/120 + (r)_6\,(r-2)\,\mu_3\mu_2/12\big\} \\ &\quad + (r)_6\,\mu_{r-6}\big(\mu_4\mu_2/48 + \mu_3^2/72 - r\mu_2^3/48\big) - (r)_7\,\mu_{r-7}\mu_3\mu_2^2/48 \\ &\quad + (r)_8\,\mu_{r-8}\mu_2^4/384\;. \end{aligned}$$

Substituting into the expressions of (4.3)–(4.4) for the coefficient $T_i(\widehat{F})$ of n^{-i} in the expansion for the UE of $\mu_r(F)$ gives

$$T_2(F) \;=\; r^2\mu_r - (r^3-r)\,\mu_{r-2}\mu_2/2 - (r)_3\,\mu_{r-3}\mu_3/3 + (r)_4\,\mu_{r-4}\mu_2^2/8\;,$$

and

$$\begin{aligned} T_3(F) \;&=\; r^3\mu_r - (r^4-r)\,\mu_{r-2}\mu_2/2 - (r)_3\,(r+3)\,\mu_{r-3}\mu_3/3 \\ &\quad + (r)_4\,\mu_{r-4}\big\{-2\,\mu_4 + (r+6)\,\mu_2^2\big\}/8 \\ &\quad + (r)_5\,\mu_{r-5}\mu_3\mu_2/6 - (r)_6\,\mu_{r-6}\mu_2^3/48\;. \end{aligned}$$

Similarly, from (4.7) and (4.8),

$$S_2(F) \;=\; (r)_2\,\mu_r - r^2(r-1)\,\mu_{r-2}\mu_2/2 - (r)_3\,\mu_{r-3}\mu_3/3 + (r)_4\,\mu_{r-4}\mu_2^2/8$$

and

$$\begin{aligned} S_3(F) \;&=\; (r)_3\,\mu_r - r\,(r)_3\,\mu_{r-2}\mu_2/2 - r\,(r)_3\,\mu_{r-3}\mu_3/3 \\ &\quad - (r)_4\,\mu_{r-4}\mu_4/4 + (4r-9)\,(r)_4\,\mu_{r-4}\mu_2^2/8 + (r)_5\,\mu_{r-5}\mu_3\mu_2/6 \\ &\quad - (r)_6\,\mu_{r-6}\mu_2^3/48\;. \end{aligned}$$

Now from page 6 in James [14] the UE for μ_r has the form

$$l_r \;=\; \left\{\sum_{i=0}^{s} a_{i,r}(\widehat{F})\,n^{-i}\right\}\Bigg/\prod_{i=1}^{r-1}\big(1-i/n\big) \tag{5.14}$$

for $r=2s$ or $2s+1$, which can be recovered from $\{T_i,\, i\le s\}$ as in Proposition 4.1. So, the above $\{T_i,\, i\le 3\}$ provide UEs for μ_r for $r\le 7$. These were given for $r\le 6$ on page 6 in James [14] and agree with our results.

For example, for μ_3, $T(1^2) = -2\mu_2$, so $S_1(F) = 3\mu_3$ and $T(1^3) = 12\,\mu_3$, $T(1^2,1^2) = 0$, so $S_2(F) = 4\mu_3$ and so the UE of μ_3 is

$$\mu_3(\widehat{F})\,\big\{1 + 3/(n-1) + 4/(n-1)_2\big\} \;=\; \mu_3(\widehat{F})\,\big\{(1-n^{-1})\,(1-2\,n^{-1})\big\}^{-1}\,.$$

For $r=7$, we obtain in this way $\{a_{i,7}=a_{i,7}(F)\}$ of (5.14) as

$$\begin{aligned} a_{0,7} &= \mu_7\ , \qquad a_{1,7} \;=\; -7\big(2\,\mu_7+3\,\mu_5\,\mu_2\big)\ ,\\ a_{2,7} &= 7\big(11\,\mu_7+39\,\mu_5\,\mu_2-10\,\mu_4\,\mu_3+15\,\mu_3\,\mu_2^2\big)\ ,\\ a_{3,7} &= -7\big(28\,\mu_7+192\,\mu_5\,\mu_2-80\,\mu_4\,\mu_3+60\,\mu_3\,\mu_2^2\big)\ . \end{aligned}$$

Example 5.7. One univariate sample (that is $k=s_1=1$) with $T(F)=\prod_{j=2}^q \mu_j^{p_j}$ for $\{p_j\}$ arbitrary and $\{\mu_j\}$ as in Example 5.6. Set $S_i(\boldsymbol{\mu})=\mu_i$ and $g(\mathbf{S})=\prod S_j^{p_j}$. The ordinary partial derivatives of $g(\mathbf{S})$ are

$$\begin{aligned} g_i &= p_i\,\mu_i^{-1}\,T(F)\ , \qquad g_{i,j} \;=\; p_i\,(p_j-\delta_{i,j})\,(\mu_i\,\mu_j)^{-1}\,T(F)\ ,\\ g_{i,j,k} &= p_i\,(p_j-\delta_{i,j})\,(p_k-\delta_{i,k}-\delta_{j,k})\,(\mu_i\,\mu_j\,\mu_k)^{-1}\,T(F)\ , \end{aligned}$$

and so on, where $\delta_{i,j}=1$ if $i=j$ and 0 otherwise. Set

$$\left[{a,b \atop i,j}\cdots\right] = \int \mu_{i,F}(x^a)\,\mu_{j,F}(x^b)\cdots dF(x)\ .$$

So, $\left[{a \atop i}\right]=\mu_i(1^a)$ of (5.5) and by (5.4), and

$$\left[{1,1 \atop i,j}\right] = i\,j\,\mu_{i-1}\,\mu_{j-1}\,\mu_2-\sum_{i,j}^{2} i\,\mu_{i-1}\,\mu_{j+1}+\mu_{i+j}-\mu_i\,\mu_j\ ,$$

where $\sum_{i_1,\dots,i_m}^m f_{i_1,\dots,i_m}=\sum^m f_{i_1,\dots,i_m}$ is defined in Example 5.2.

By (A.8),

$$\begin{aligned} -2\,T_1(F) &= 2\,C_1\\ &= T(1^2)\\ &= T(F)\,\big\{2\,\langle 1,2\rangle+\langle 1,1\rangle+\langle 1^2\rangle\big\}\,, \end{aligned} \tag{5.15}$$

where

$$\begin{aligned} \langle 1,2\rangle &= \sum_{i<j} p_i\,p_j\left[{1,1 \atop i,j}\right]\mu_i^{-1}\mu_j^{-1}\,,\\ \langle 1,1\rangle &= \sum_i (p_i)_2\left[{1,1 \atop i,i}\right]\mu_i^{-2}\,,\\ \langle 1^2\rangle &= \sum_i p_i\left[{2 \atop i}\right]\mu_i^{-1}\ . \end{aligned}$$

Other terms are calculated similarly. For example, C_2, $T_2(F)$ and $S_2(F)$ are given by (3.6), (4.3), and (4.7) in terms of $T(1^2)$, $T(1^3)$ and $T(1^2,1^2)$. Also by (A.9) to (A.11)

$$\begin{aligned} T(1^3) \;=\; T(F)\Bigg\{&\sum_{i,j,k} p_i\,(p_j-\delta_{i,j})\,(p_k-\delta_{i,k}-\delta_{j,k})\,(\mu_i\,\mu_j\,\mu_k)^{-1}\left[{1,1,1 \atop i,j,k}\right]\\ &+3\sum_{i,j} p_i\,(p_j-\delta_{i,j})\,(\mu_i\,\mu_j)^{-1}\left[{2,1 \atop i,j}\right]+\sum_i p_i\,\mu_i^{-1}\left[{3 \atop i}\right]\Bigg\}\ , \end{aligned} \tag{5.16}$$

and

$$(5.17)\quad \begin{aligned} T(1^2,1^2) \;=\; T(F)\Bigg\{ &\sum_{i,j,k,l} p_i\,(p_j-\delta_{i,j})\,(p_k-\delta_{i,k}-\delta_{j,k})\,(p_l-\delta_{i,l}-\delta_{j,l}-\delta_{k,l}) \\ &\times(\mu_i\,\mu_j\,\mu_k\,\mu_l)^{-1}\left[{}^{1,1}_{k,l}\right] \\ &+\sum_{i,j,k} p_i\,(p_j-\delta_{i,j})\,(p_k-\delta_{i,k}-\delta_{j,k})\,(\mu_i\,\mu_j\,\mu_k)^{-1}\,G_{i,j,k} \\ &+\sum_{i,j} p_i\,(p_j-\delta_{i,j})\,(\mu_i\,\mu_j)^{-1}H_{i,j}\;+\sum_i p_i\,\mu_i^{-1}\mu_i(1^2,1^2)\Bigg\}\;, \end{aligned}$$

where

$$\begin{aligned} G_{i,j,k} \;&=\; 2\left[{}^{1,1}_{i,j}\right]\left[{}^{2}_{k}\right] +\,4\left[1,2_i,1_j,2_k\right]\,, \\ H_{i,j} \;&=\; 4\left[1_i,1,2_j^2\right] + \left[{}^{2}_{i}\right]\left[{}^{2}_{j}\right] + 2\left[1,2_i,1,2_j\right]\,, \\ \left[1^a,2_i^b,1^c,2_j^d,\ldots\right] \;&=\; \iint \mu_i(x^a,y^b)\,\mu_j(x^c,y^d)\cdots\,dF(x)\,dF(y)\;, \end{aligned}$$

so that

$$\begin{aligned} \left[1_i^a,1_j^b,\ldots\right] \;&=\; \left[{}^{a,b}_{i,j}\cdots\right], \\ \left[1,2_i,1_j,2_k\right] \;&=\; (i)_2\,\mu_{i-2}\,A_j\,A_k - i\sum_{j,k}^{2} B_{i,j}\,A_k \end{aligned}$$

for

$$A_j \;=\; \mu_{j+1}-j\,\mu_{j-1}\mu_2\;, \qquad B_{i,j} \;=\; \mu_{i+j-1}-j\,\mu_{j-1}\,\mu_i-\mu_{i-1}\,\mu_j\;.$$

By (5.4),

$$\begin{aligned} \left[{}^{1,1,1}_{i,j,k}\right] \;&=\; -ijk\,\mu_{i-1}\,\mu_{j-1}\,\mu_{k-1}\,\mu_3+\sum^3 ij\,\mu_{i-1}\,\mu_{j-1}\big(\mu_{k+2}-\mu_k\,\mu_2\big) \\ &\quad -\sum^3 i\,\mu_{i-1}\big(\mu_{j+k+1}-\mu_{j+1}\,\mu_k-\mu_{k+1}\,\mu_j\big)+\mu_{i+j+k} \\ &\quad -\sum^3 \mu_i\,\mu_{j+k}+2\,\mu_i\,\mu_j\,\mu_k\;, \\ \left[{}^{2,1}_{i,j}\right] \;&=\; -(i)_2\,j\,\mu_{i-2}\,\mu_{j-1}\,\mu_3+(i)_2\,\mu_{i-2}\big(\mu_{j+2}-\mu_j\,\mu_2\big) \\ &\quad +2\,ij\,\mu_{j-1}\big(\mu_{i+1}-\mu_{i-1}\,\mu_2\big)-2\,i\big(\mu_{i+j}-\mu_i\,\mu_j-\mu_{i-1}\,\mu_{j+1}\big)\;, \\ \left[1_i,1,2_j^2\right] \;&=\; (j)_2\Big\{\big(-3\,i\,\mu_{i-1}\,\mu_{j-1}+\mu_{i+j-2}-\mu_i\,\mu_{j-2}\big)\mu_2+2\,\mu_{i+1}\,\mu_{j-1}\Big\} \\ &\quad +(j)_3\big(i\,\mu_{i-1}\,\mu_{j-3}\,\mu_2^2-\mu_{j-3}\,\mu_{i+1}\,\mu_2\big)\;, \\ \left[1,2_i,1,2_j\right] \;&=\; (i)_2\,(j)_2\,\mu_{i-2}\,\mu_{j-2}\,\mu_2^2-2\sum^2 i(j)_2\,\mu_i\,\mu_{j-2}\,\mu_2 \\ &\quad +2\,ij\,\big(\mu_{i+j-2}\,\mu_2-\mu_{i-1}\,\mu_{j-1}\,\mu_2+\mu_i\,\mu_j\big)\;. \end{aligned}$$

Also $\left[{}^{i}_{r}\right]$ for $2\le i\le 4$ and $\mu_i(1^2,1^2)$ are given by (5.6)–(5.11).

Example 5.8. Consider Example 5.7 with $T(F)=\mu_r^p$. Then

$$\begin{aligned}
T(1^2)/T(F) &= p\begin{bmatrix}2\\r\end{bmatrix}\mu_r^{-1}+(p)_2\,\mu_r^{-2}\begin{bmatrix}1,1\\r,r\end{bmatrix},\\
T(1^3)/T(F) &= p\,\mu_r^{-1}\begin{bmatrix}3\\r\end{bmatrix}+3(p)_2\,\mu_r^{-2}\begin{bmatrix}2,1\\r,r\end{bmatrix}+(p)_3\,\mu_r^{-3}\begin{bmatrix}1,1,1\\r,r,r\end{bmatrix},\\
T(1^2,1^2)/T(F) &= p\,\mu_r^{-1}\mu_r(1^2,1^2)+(p)_2\,\mu_r^{-2}H_{r,r}+(p)_3\,\mu_r^{-3}G_{r,r,r}+(p)_4\,\mu_r^{-4}\begin{bmatrix}1,1\\r,r\end{bmatrix}^2.
\end{aligned}$$

Example 5.9. Consider Example 5.8 with $T(F)=\mu_2^p$. Set $\beta_r=\mu_r\,\mu_2^{-r/2}$. Then

$$\begin{aligned}
T(1^2)/T(F) &= -2\,p+(p)_2\big(\beta_4-1\big)\ ,\\
T(1^3)/T(F) &= -6\,(p)_2\big(\beta_4-1\big)+(p)_3\big(\beta_6-3\,\beta_4+2\big)\ ,\\
T(1^2,1^2)/T(F) &= 12\,(p)_2-4\,(p)_3\big(\beta_4-1+2\,\beta_3^2\big)+(p)_4\big(\beta_4-1\big)^2\,.
\end{aligned}$$

So,

$$\begin{aligned}
-T_1(F)/T(F) &= C_1/T(F) \;=\; -p+(p)_2\big(\beta_4-1\big)/2\ ,\\
C_2/T(F) &= (p)_2\big(5/2-\beta_4\big)+(p)_3\big(\beta_6/6-\beta_4-\beta_3^2+5/6\big)+(p)_4\big(\beta_4-1\big)^2\,,\\
T_2(F)/T(F) &= p+(p)_2\big(4-5\,\beta_4/2\big)+(p)_3\big(2\,\beta_6-9\,\beta_4+7-6\,\beta_3^2\big)/6\\
&\qquad+(p)_4\big(\beta_4-1\big)^2/8 \;=\; \sum_{i=1}^{r}(p)_iA_i \quad \text{say}\,,\\
S_2(F)/T(F) &= (p)_2\big(7/2-2\,\beta_4\big)+\sum_{i=3}^{4}(p)_iA_i\ .
\end{aligned}$$

For $p=2$ this gives $T(F)=\mu_2^2$,

(5.18) $\quad C_1=\mu_4-3\,\mu_2^2\ ,\quad T_1(F)=-\mu_4+3\,\mu_2^2\ ,$

(5.19) $\quad C_2=-2\,\mu_4+5\,\mu_2^2\ ,\quad T_2(F)=-5\,\mu_4+10\,\mu_2^2\ ,\quad S_2(F)=-4\,\mu_4+7\,\mu_2^2\ .$

Note that C_1, C_2 agree with $\mu(2^2)$ of page 368 in Sukhatme [20].

The UE of μ_2^2 has the form

$$l_{2,2} \;=\; \left(\sum_{i=0}^{2}a_{i,2,2}(\widehat{F})\,n^{-i}\right)\Big/\prod_{i=1}^{3}\big(1-i/n\big)\ .$$

So, $\big\{a_i=a_{i,2,2}(F)\big\}$ are given by

$$\begin{aligned}
a_0 &= T(F) \;=\; \mu_2^2\ ,\\
a_1 &= -6\,T(F)+T_1(F) \;=\; -\mu_4-3\,\mu_2^2\ ,\\
a_2 &= 11\,T(F)-6\,T_1(F)+T_2(F) \;=\; \mu_4+3\,\mu_2^2\ .
\end{aligned}$$

We now present a second method for finding an UE of $\prod_i \mu_i^{p_i}$. This method avoids computing $\{T_i(F)\}$, but derives the UE of the vector

$$\mathbf{T}(F)' = \left\{ \prod_i \mu_i^{p_i} : \sum p_i = p \right\} , \tag{5.20}$$

that is, for all products of a given degree p, directly from their first few coefficients $\{\mathbf{C}_i\}$. Suppose $\mathbf{T}(F)$ has dimension $d = d_p$. Then

$$\mathbf{C}_i = \mathbf{A}_i \mathbf{T}(F) ,$$

where $\mathbf{A}_i$ is a $d \times d$ matrix of integers and $\mathbf{A}_0 = \mathbf{I}_d$, the identity matrix. So,

$$\boldsymbol{\alpha}(n)^{-1}\, \mathbf{T}(\widehat{F})$$

is the UE of $\mathbf{T}(F)$, where

$$\boldsymbol{\alpha}(n) = \sum_{i=0}^{\infty} \mathbf{A}_i\, n^{-i} .$$

But this is known to have the form

$$\mathbf{T}_n(\widehat{F}) = \widehat{\boldsymbol{\beta}}_n \Big/ \prod_{i=1}^{p-1} (1 - i/n) , \tag{5.21}$$

where

$$\widehat{\boldsymbol{\beta}}_n = \left\{ \sum_{i=0}^{[p/2]} \mathbf{B}_i\, n^{-i} \right\} \mathbf{T}(\widehat{F}) ,$$

where $\mathbf{B}_i$ is a $d \times d$ matrix of integers with $\mathbf{B}_0 = \mathbf{I}_d$. So,

$$\begin{aligned} \sum_{i=0}^{[p/2]} \mathbf{B}_i\, \varepsilon^i &= \left\{ \prod_{i=1}^{p-1} (1 - i\varepsilon) \right\} \boldsymbol{\alpha}(\varepsilon^{-1}) \\ &= \left\{ 1 - D_1(p)\varepsilon + D_2(p)\varepsilon^2 - \cdots \right\} \\ &\quad \times \left\{ \mathbf{I}_d - \mathbf{A}_1\varepsilon + \left(-\mathbf{A}_2 + \mathbf{A}_1^2\right)\varepsilon^2 + \left(-\mathbf{A}_3 + \mathbf{A}_1\mathbf{A}_2 + \mathbf{A}_2\mathbf{A}_1 - \mathbf{A}_1^3\right)\varepsilon^3 + \cdots \right\}, \end{aligned}$$

where $D_1(p) = (p)_2/2$ and $D_2(p) = (p)_3(p - 1/3)/8$. So, the UE (5.21) is given in terms of $\{A_i,\, i \le p/2\}$:

$$\begin{aligned} \mathbf{B}_0 &= \mathbf{I}_d , \\ \mathbf{B}_1 &= -D_1(p)\mathbf{I}_d - \mathbf{A}_1 , \\ \mathbf{B}_2 &= D_2(p)\mathbf{I}_d + D_1(p)\mathbf{A}_1 - \mathbf{A}_2 + \mathbf{A}_1^2 , \\ \mathbf{B}_3 &= -D_3(p)\mathbf{I}_d - D_2(p)\mathbf{A}_1 - D_1(p)\left(-\mathbf{A}_2 + \mathbf{A}_1^2\right) - \mathbf{A}_3 + \mathbf{A}_1\mathbf{A}_2 + \mathbf{A}_2\mathbf{A}_1 - \mathbf{A}_1^3 , \end{aligned}$$

and so on. The method also applies to obtaining an UE for

$$\mathbf{T}(F)' = \left\{ \mu_1^{p_1} \prod_{i=2}^{q} \mu_i^{p_i} : \sum_{i=1}^{q} p_i = p \right\},$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}(F)$. A third method (for $p \leq 8$) due to Fisher [10] is given in Section 12 of Stuart and Ord [19]. Their Tables 11 and 10, pages 554–555, may be used to verify Examples 5.8 to 5.11 after some labor.

Example 5.10. Consider Example 5.7 with $\mathbf{T}(F) = (\mu_4, \mu_2^2)'$. So, (5.20) holds with $p = 4$ and $d = [p/2] = 2$.

By (5.12), (5.13), for μ_4, $C_1 = -4\mu_4 + 6\mu_2^2$ and $C_2 = 6\mu_4 - 15\mu_2^2$, in agreement with $\mu(4)$ on page 368 in Sukhatme [20]. So, by (5.18), (5.19)

$$\mathbf{A}_1 = \begin{pmatrix} -4 & 6 \\ 1 & -3 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_2 = \begin{pmatrix} 6 & -15 \\ -2 & 5 \end{pmatrix}.$$

So,

$$\mathbf{B}_1 = -6\mathbf{I}_2 - \mathbf{A}_1 = \begin{pmatrix} -2 & -6 \\ -1 & -3 \end{pmatrix}, \qquad \mathbf{B}_2 = 11\mathbf{I}_2 + 6\mathbf{A}_1 - \mathbf{A}_2 + \mathbf{A}_1^2 = \begin{pmatrix} 3 & 9 \\ 1 & 3 \end{pmatrix}.$$

So, UEs of μ_4 and μ_2^2 are $\mu_{4,n}(\widehat{F})$ and $\mu_{2,2,n}(\widehat{F})$, where

$$\mu_{4,n}(F) = \left\{ \mu_4 + \left(-2\mu_4 - 6\mu_2^2\right) n^{-1} + \left(3\mu_4 + 9\mu_2^2\right) n^{-2} \right\} \Big/ \prod_{i=1}^{3} (1 - i/n),$$

and

$$\mu_{2,2,n}(F) = \left\{ \mu_2^2 + \left(-\mu_4 - 3\mu_2^2\right) n^{-1} + \left(\mu_4 + 3\mu_2^2\right) n^{-2} \right\} \Big/ \prod_{i=1}^{3} (1 - i/n).$$

Table 2 gives the relative bias of $S_{n,p}(\widehat{F})$ as estimated from two runs of sixty thousand simulations for $p \leq 2$ and F normal and exponential. The estimates present bias even for $n = 100$ and bias-corrected estimates of order n^{-2} (i.e. $p = 2$): see Example C.3. For $p = 3$ the bias is zero.

Table 2: Relative bias of $S_{n,p}(\widehat{F})$ for $T(F) = \mu_4$ estimated from two runs of 60,000 simulations.

		$n = 5$		$n = 10$		$n = 100$	
		$p = 1$	$p = 2$	$p = 1$	$p = 2$	$p = 1$	$p = 2$
Norm $(0, 1)$	Run 1	−0.3584	−0.1988	−0.1934	−0.0543	−0.0174	0.0021
	Run 2	−0.3572	−0.1947	−0.1871	−0.0460	−0.0206	0.0012
Exp (1)	Run 1	−0.4957	−0.2861	−0.2831	−0.0754	−0.0380	−0.0063
	Run 2	−0.4943	−0.2851	−0.2964	−0.0923	−0.0399	−0.0082

Example 5.11. Consider Example 5.7 with $\mathbf{T}(F) = (\mu_5, \mu_3\mu_2)'$. So, (5.20) holds with $p = 5$ and $d = [p/2] = 2$.

By (5.12), (5.13) for μ_5, $C_1 = -5\mu_5 + 10\mu_3\mu_2$ and $C_2 = 10\mu_5 - 50\mu_3\mu_2$, in agreement with $\mu(5)$ of page 368 in Sukhatme [20]. By (5.15)–(5.17), for $\mu_2\mu_3$,

$$T(1^2) = 2\mu_5 - 16\mu_3\mu_2 \ , \quad T(1^3) = -24\mu_5 + 72\mu_3\mu_2 \ , \quad T(1^2, 1^2) = 96\mu_3\mu_2 \ ,$$

giving $C_1 = \mu_5 - 8\mu_3\mu_2$ and $C_2 = -4\mu_5 + 24\mu_3\mu_2$. So,

$$\mathbf{A}_1 = \begin{pmatrix} -5 & 10 \\ 1 & -8 \end{pmatrix} \qquad \text{and} \qquad \mathbf{A}_2 = \begin{pmatrix} 10 & -50 \\ -4 & 24 \end{pmatrix}.$$

So,

$$\mathbf{B}_1 = -10\mathbf{I}_2 - \mathbf{A}_1 = \begin{pmatrix} -5 & -10 \\ -1 & -2 \end{pmatrix}, \qquad \mathbf{B}_2 = 35\mathbf{I}_2 + 10\mathbf{A}_1 - \mathbf{A}_2 + \mathbf{A}_1^2 = \begin{pmatrix} 10 & 20 \\ 1 & 5 \end{pmatrix}.$$

That is, UEs of μ_5 and $\mu_3\mu_2$ are $\mu_{5,n}(\widehat{F})$, and $\mu_{3,2,n}(\widehat{F})$, where

$$\mu_{5,n}(F) = \Big\{\mu_5 + \big(-5\mu_5 - 10\mu_3\mu_2\big)n^{-1} + \big(10\mu_5 + 20\mu_3\mu_2\big)\, n^{-2}\Big\} \Big/ \prod_{i=1}^{4}(1 - i/n)$$

and

$$\mu_{3,2,n}(F) = \Big\{\mu_3\mu_2 + \big(-\mu_5 - 2\mu_3\mu_2\big)\, n^{-1} + \big(\mu_5 + 5\mu_3\mu_2\big)\, n^{-2}\Big\} \Big/ \prod_{i=1}^{4}(1 - i/n) \ .$$

Example 5.12. Suppose $k = s_1 = 1$ and $T(F) = g(\mu_2)$. Set $g^r = g^{(r)}(\mu_2)$, and $\beta_r = \mu_r\mu_2^{-r/2}$. Then

$$\mu_x = \mu_F(x) = x - \mu \,, \quad \mu_{2,x} = \mu_{2,F}(x) = \mu_x^2 - \mu_2 \,, \quad \mu_{2,x,y} = \mu_{2,F}(x,y) = -2\,\mu_x\mu_y$$

by (5.4). By (A.8),

$$|2| = T(1^2) = g^2\mu_{2,2}(1,1) + g^1\mu_2(1^2) \ ,$$

where

$$\mu_{2,2}(1,1) = \int \mu_{2,x}^2 = \int \mu_{2,x}^2\, dF(x) = \mu_4 - \mu_2^2 \ ,$$

$$\mu_2(1^2) = \int \mu_{2,x,x} = -2\mu_2 \quad \text{by (5.6)} \ .$$

Similarly, by (A.9) to (A.11) and (A.15),

$$T(1^3) = g^3\mu_{2,2,2}(1,1,1) + 3g^2\mu_{2,2}(1,1^2) + g^1\mu_2(1^3) \ ,$$

$$\begin{aligned} T(1^4) = {} & g^4\mu_{2,2,2,2}(1,1,1,1) + 6g^3\mu_{2,2,2}(1,1,1^2) \\ & + g^2\Big\{4\mu_{2,2}(1,1^3) + 3\mu_{2,2}(1^2,1^2)\Big\} \\ & + g^1\mu_2(1^4) \ , \end{aligned}$$

$$\begin{aligned} T(1^2,1^2) \;&=\; g^4\mu_{2,2}(1,1)^2 + g^3\Big\{2\,\mu_{2,2}(1,1)\,\mu_2(1^2) + 4\,\mu_{2,2,2}(ab,a,b)\Big\} \\ &\quad + g^1\mu_2(a^2b^2) \\ &\quad + g^2\Big\{4\,\mu_{2,2}(a,ab^2) + \mu_2(1^2)^2 + 2\,\mu_{2,2}(ab,ab)\Big\} \quad \text{at} \;\; a=b=1 \\ &=\; \sum_{i=2}^{4} g^i a_i \quad \text{say} \;, \\ T(1^2,1^3) \;&=\; g^3A_3 + g^4A_4 + g^5A_5 \;, \end{aligned}$$

and by (A.16)

$$T(1^2,1^2,1^2) \;=\; \sum_{i=3}^{6} g^i B_i \;,$$

where

$$\begin{aligned} &\mu_{2,2,2}(1,1,1) \;=\; \int \mu_{2,x}^3 \;=\; \mu_6 - 3\,\mu_4\mu_2 + 2\,\mu_2^3 \;, \\ &\mu_{2,2}(1,1^2) \;=\; \int \mu_{2,x}\mu_{2,x,x} \;=\; -2\left(\mu_4 - \mu_2^2\right) \;, \\ &\mu_2(1^3) \;=\; \int \mu_{2,x,x,x} \;=\; 0 \;, \\ &\mu_{2,2,2,2}(1,1,1,1) \;=\; \int \mu_{2,x}^4 \;=\; \mu_8 - 4\,\mu_6\mu_2 + 6\,\mu_4\mu_2^2 - 3\,\mu_2^4 \;, \\ &\mu_{2,2,2}(1,1,1^2) \;=\; \int \mu_{2,x}^2\mu_{2,x,x} \;=\; -2\left(\mu_6 - 2\,\mu_4\mu_2 + \mu_2^3\right) \;, \\ &\mu_{2,2}(1,1^3) \;=\; \mu_2(1^4) \;=\; 0 \;, \\ &\mu_{2,2}(1^2,1^2) \;=\; \int \mu_{2,x,x}^2 \;=\; 4\,\mu_4 \;, \\ &\mu_{2,2}(a,ab^2)_{a=b=1} \;=\; \int \mu_{2,x}\mu_{2,x,y,y} \;=\; 0 \;, \\ &\mu_{2,2,2}(ab,a,b)_{a=b=1} \;=\; \iint \mu_{2,x,y}\mu_{2,x}\mu_{2,y} \;=\; -2\,\mu_3^2 \;, \\ &\mu_{2,2}(ab,ab)_{a=b=1} \;=\; \int \mu_{2,x,y}^2 \;=\; 4\,\mu_2^2 \;, \\ &\mu_2(a^2b^2)_{a=b=1} \;=\; \int \mu_{2,x,x,y,y} \;=\; 0 \;, \\ &a_2 \;=\; 12\,\mu_2^2 \;, \qquad a_3 \;=\; -4\left(\mu_4\mu_2 - \mu_2^3 + 2\,\mu_3^2\right) , \qquad a_4 \;=\; \left(\mu_4 - \mu_2^2\right)^2 \;, \end{aligned}$$

and

$$\begin{aligned} A_3 \;&=\; 6\,\mu_{2,2,2}(a,ab,b^2) + 3\,\mu_2(a^2)\mu_{2,2}(b,b^2) + 6\,\mu_{2,2,2}(b,ab,ab) \quad \text{at} \;\; a=b=1 \\ &=\; 3\iint \Big\{2\,\mu_{2,x}\mu_{2,x,y}\mu_{2,y,y} + \mu_{2,y}\mu_{2,y,y}\mu_{2,x,x} + 2\,\mu_{2,y}\mu_{2,x,y}^2\Big\} \\ &=\; 3\iint \Big\{8\left(\mu_x^2 - \mu_2\right)\mu_x\mu_y^3 + 12\left(\mu_y^2 - \mu_2\right)\mu_x^2\mu_y^2\Big\} \\ &=\; 12\,\Big\{2\,\mu_3^2 + 3\left(\mu_2\mu_4 - \mu_2^3\right)\Big\} \\ &=\; 12\,\mu_2^3\,\Big\{2\,\beta_3^2 + 3\,\beta_4 - 3\Big\} \;, \end{aligned}$$

$$\begin{aligned}
A_4 &= \iint \Big\{\mu_{2,x,x}\mu_{2,y}^3 + 6\,\mu_{2,x,y}\mu_{2,y}^2 + 3\,\mu_{2,y,y}\mu_{2,y}\mu_{2,x}^2\Big\} \\
&= -2\iint \Big\{\mu_x^2\big(\mu_y^2-\mu_2\big)^3 + 6\,\mu_x\mu_y\big(\mu_x^2-\mu_2\big)\big(\mu_y^2-\mu_2\big)^2 \\
&\quad + 3\,\mu_y^2\big(\mu_x^2-\mu_2\big)^2\big(\mu_y^2-\mu_2\big)\Big\} \\
&= -2\Big\{\mu_2\big(\mu_6 - 3\,\mu_4\mu_2 + 2\,\mu_2^3\big) + 6\,\mu_3\big(\mu_5 - 2\,\mu_3\mu_2\big) + 3\,\big(\mu_4-\mu_2^2\big)^2\Big\} \\
&= -2\,\mu_2^4\Big\{\beta_6 - 3\,\beta_4 + 2 + 6\,\beta_3\big(\beta_5 - 2\,\beta_3\big) + 3\,\big(\beta_4-1\big)^2\Big\}\ ,
\end{aligned}$$

$$\begin{aligned}
A_5 &= \iint \mu_{2,x}^2\,\mu_{2,y}^3 \;=\; \int\big(\mu_x^2-\mu_2\big)^2\int\big(\mu_y^2-\mu_2\big)^3 \\
&= \big(\mu_4-\mu_2^2\big)\big(\mu_6 - 3\,\mu_4\mu_2 + 2\,\mu_2^3\big) \\
&= \mu_2^5\big(\beta_4-1\big)\big(\beta_6 - 3\,\beta_4 + 2\big)\ ,
\end{aligned}$$

$$\begin{aligned}
B_3 &= B_3^{i,j,k} \qquad \text{at}\ \ \big\{a=b=c=1,\ S=\mu\big\} \\
&= \iiint \Big\{\mu_{2,x,x}\mu_{2,y,y}\mu_{2,z,z} + 6\,\mu_{2,x,x}\mu_{2,y,z}^2 + 8\,\mu_{2,x,y}\mu_{2,y,z}\mu_{2,z,x}\Big\} \\
&= -120\,\mu_2^3\ ,
\end{aligned}$$

$$\begin{aligned}
B_4 &= B_2^{i,j,k,l} \qquad \text{at}\ \ \big\{a=b=c=1, S=\mu_2\big\} \\
&= 3\iiint \Big\{\mu_{2,x}^2\mu_{2,y,y}\mu_{2,z,z} + 2\,\mu_{2,x}^2\mu_{2,y,z}^2 + 4\,\mu_{2,x}\mu_{2,y}\mu_{2,x,y}\mu_{2,z,z} \\
&\quad + 8\,\mu_{2,x}\mu_{2,y}\mu_{2,x,z}\mu_{2,y,z}\Big\} \\
&= 36\,\Big\{\big(\mu_4-\mu_2^2\big)\mu_2^2 + 4\,\mu_3^2\mu_2\Big\} \\
&= 36\,\mu_2^4\Big\{\beta_4 - 1 + 4\,\beta_3^2\Big\}\ ,
\end{aligned}$$

$$\begin{aligned}
B_5 &= 3\iiint \Big\{\mu_{2,x,x}\mu_{2y}^2 + \mu_{2,x,y}\mu_{2,x}\mu_{2,y}\Big\}\,\mu_{2,z}^2 \\
&= -6\,\Big\{\mu_2\big(\mu_4-\mu_2^2\big) + \mu_3^2\Big\}\big(\mu_4-\mu_2^2\big) \\
&= -6\mu_2^5\Big\{\beta_4 - 1 + \beta_3^2\Big\}\big(\beta_4-1\big)\ ,
\end{aligned}$$

$$B_6 = \iiint \mu_{2,x}^2\mu_{2,y}^2\mu_{2,z}^2 \;=\; \big(\mu_4-\mu_2^2\big)^3 \;=\; \mu_2^6\big(\beta_4-1\big)^3 .$$

So,

$$\begin{aligned}
C_1 &= -g^1\mu_2 + g^2\big(\mu_4-\mu_2^2\big)/2\ , \\
C_2 &= g^2\big(5\,\mu_2^2/2 - \mu_4\big) + g^3\big(\mu_6/6 - \mu_3^2 - \mu_4\mu_2 + 5\,\mu_2^3/6\big) + g^4(\mu_4-\mu_2^2)^2/8\ , \\
C_3 &= g^2\mu_4/2 + g^3\big(-\mu_6/2 + 4\,\mu_4\mu_2 + 2\,\mu_3^2 - 6\,\mu_2^3\big) \\
&\quad + g^4\big(\mu_8/24 - \mu_6\mu_2/3 - \mu_5\mu_3 - \mu_4^2/2 + 5\,\mu_4\mu_2^2/2 + 5\,\mu_3^2\mu_2 - 41\,\mu_2^4/24\big) \\
&\quad + g^5\big(\mu_4-\mu_2^2\big)\big(2\,\mu_6 - 9\,\mu_4\mu_2 - 3\,\mu_3^2 + 7\mu_2^3\big)/24 + g^6\big(\mu_4-\mu_2^2\big)^3/48\ ,
\end{aligned}$$

$$\begin{aligned}
T_1(F) &= S_1(F) = -g^2\big(\mu_4-\mu_2^2\big)/2+g^1\mu_2 \ ,\\
T_2(F) &= g^4\big(\mu_4-\mu_2^2\big)^2/8+g^3\big(\mu_6/3-\mu_3^2-3\mu_4\mu_2/2+7\mu_2^3/6\big)\\
&\quad + g^2\big(-5\mu_4/2+4\mu_2^2\big)+g^1\mu_2 \ ,\\
T_3(F) &= \sum_{i=1}^{6} g^i T_{3,i} \ ,\\
S_2(F) &= g^4\big(\mu_4-\mu_2^2\big)^2/8+g^3\big(\mu_6/3-\mu_3^2-3\mu_4\mu_2/2+7\mu_2^3/6\big)\\
&\quad + g^2\big(-2\mu_4+7\mu_2^2/2\big) \ ,\\
S_3(F) &= \sum_{i=2}^{6} g^i S_{3,i} \ ,
\end{aligned}$$

where

$$\begin{aligned}
S_{3,2} &= -3\mu_4+9\mu_2^2/2 \ ,\\
S_{3,3} &= 3\mu_6-27\mu_4\mu_2/2-7\mu_3^2+13\mu_2^3 \ ,\\
S_{3,4} &= -\mu_8/4+4\mu_6\mu_2/3+2\mu_5\mu_3+11\mu_4^2/8-6\mu_4\mu_2^2-7\mu_3^2\mu_2+85\mu_2^4/24 \ ,\\
S_{3,5} &= \big(\mu_4-\mu_2^2\big)\big(-4\mu_6+15\mu_4\mu_2+3\mu_3^2-11\mu_2^3\big)/24 \ ,\\
S_{3,6} &= -B_6/48 \ ,
\end{aligned}$$

$$\begin{aligned}
T_{3,1} &= \mu_2 \ ,\\
T_{3,2} &= -19\mu_4/2+31\mu_2^2/2 \ ,\\
T_{3,3} &= 4\mu_6-18\mu_4\mu_2+33\mu_2^3/2-10\mu_3^2 \ ,\\
T_{3,4} &= -\mu_8/4+4\mu_6\mu_2/3+2\mu_5\mu_3+7\mu_4^2/4-27\mu_4\mu_2^2/4-7\mu_3^2\mu_2+47\mu_2^4/12 \ ,\\
T_{3,5} &= \big(\mu_4-\mu_2^2\big)\big(-4\mu_6+15\mu_4\mu_2+3\mu_3^2-11\mu_2^3\big)/24 \ ,\\
T_{3,6} &= -B_6/48 \ .
\end{aligned}$$

Example 5.13. Consider Example 5.12 with $T(F)=\mu_2^q$. Then

$$\begin{aligned}
g^i &= (q)_i\,\mu_2^{q-i} \ ,\\
T(1^2)/\mu_2^q &= (q)_2\big(\beta_4-1\big)-2q \ ,\\
T(1^3)/\mu_2^q &= (q)_3\big(\beta_6-3\beta_4+2\big)-6(q)_2\big(\beta_4-1\big) \ ,\\
T(1^4)/\mu_2^q &= (q)_4\big(\beta_8-4\beta_6+6\beta_4-3\big)-12(q)_3\big(\beta_6-2\beta_4+1\big)+12(q)_2\beta_4 \ ,\\
T(1^2,1^2)/\mu_2^q &= (q)_4\big(\beta_4-1\big)^2-4(q)_3\big(\beta_4-1+2\mu_3^2\big)+12(q)_2 \ ,\\
T(1^2,1^3)/\mu_2^q &= 12(q)_3\big(2\beta_3^2+3\beta_4-3\big)\\
&\quad -2(q)_4\Big\{\beta_6-3\beta_4+2+6\beta_3\big(\beta_5-2\beta_3\big)+3\big(\beta_4-1\big)^2\Big\}\\
&\quad +(q)_5\big(\beta_4-1\big)\big(\beta_6-3\beta_4+2\big) \ ,\\
T(1^2,1^2,1^2)/\mu_2^q &= -120(q)_3+36(q)_4\big(\beta_4-1+4\beta_3^2\big)\\
&\quad -6(q)_5\big(\beta_4-1+\beta_3^2\big)\big(\beta_4-1\big)+(q)_6\big(\beta_4-1\big)^3 .
\end{aligned}$$

So, $t_i = T_i(F)/T(F)$ and $s_i = S_i(F)/T(F)$ are given by

$$
\begin{aligned}
t_1 &= s_1 = -(q)_2\big(\beta_4 - 1\big)/2 + q \ , \\
t_2 &= (q)_4\big(\beta_4 - 1\big)^2/8 + (q)_3\big(\beta_6/3 - 3\beta_4/2 + 7/6\big) + (q)_2\big(-5\beta_4/2 + 4\big) + q \ , \\
s_2 &= (q)_4\big(\beta_4 - 1\big)^2/8d + (q)_3\big(\beta_6/3 - \beta_3^2 - 3\beta_4/2 + 7/6\big) + (q)_2\big(-2\beta_4 + 7/2\big) \ , \\
t_3 &= \sum_{i=1}^{6}(q)_i\, t_{3,i} \ , \qquad s_3 = \sum_{i=2}^{6}(q)_i\, s_{3,i} \ ,
\end{aligned}
$$

for

$$
\begin{aligned}
t_{3,1} &= 1 \ , \\
t_{3,2} &= \big(31 - 19\beta_4\big)/2 \ , \\
t_{3,3} &= 4\beta_6 - 18\beta_4 - 10\beta_3^2 + 33/2 \ , \\
t_{3,4} &= \big\{-3\beta_8 + 16\beta_6 + 24\beta_5\beta_3 - 84\beta_3^2 + 21\beta_4^2 - 81\beta_4 + 47\big\}/12 \ , \\
t_{3,5} &= s_{3,5} = \big(\beta_4 - 1\big)\big(-4\beta_6 + 15\beta_4 - 11 + 3\beta_3^2\big)/24 \ , \\
t_{3,6} &= s_{3,6} = -\big(\beta_4 - 1\big)^3/48 \ , \\
s_{3,2} &= -3\beta_4 + 9/2 \ , \\
s_{3,3} &= 3\beta_6 - 27\beta_4/2 + 13 - 7\beta_3^2 \ , \\
s_{3,4} &= \big\{-6\beta_8 + 32\beta_6 - 138\beta_4 + 33\beta_4^2 + 85\big\}/24 - 6\beta_4 - 7\beta_3^2 + 2\beta_3\beta_5 \ .
\end{aligned}
$$

Example 5.14. Consider Example 5.13 with $T(F) = \mu_2$, so $ET(\widehat{F}) = (1 - n^{-1})\,T(F)$. As a check $q = 1$ above gives $T(1^2) = -2\mu_2$, $T(1^3) = T(1^4) = T(1^2, 1^2) = T(1^2, 1^3) = T(1^2, 1^2, 1^2) = 0$, so $t_1 = t_2 = t_3 = 1$, $s_1 = 1$, $s_2 = s_3 = 0$.

Example 5.15. Consider Example 5.13 with $T(F) = \mu_2^{1/2} = \sigma(F)$ say. Putting $q = 1/2$ gives $t_1 = s_1 = (\beta_4 + 3)/8$, so an estimate of $\sigma(F)$ of bias $O(n^{-2})$ is

$$
\sigma(\widehat{F})\left\{1 + n^{-1}\big(\beta_4(\widehat{F}) + 3\big)/8\right\} ,
$$

where $\beta_4(F) = \beta_4 = \mu_4\mu_2^{-2}$. To reduce the bias further use

$$
\begin{aligned}
s_2 &= \big(16\beta_6 + 22\beta_4 + 164 - 15\beta_4^2\big)/128 \ , \\
s_3 &= \big(240\beta_8 + 432\beta_6 - 2503\beta_4 + 2817 - 165\beta_4^2 \\
&\quad + 4764\beta_3^2 + 315\beta_4^3 - 560\beta_4\beta_6 + 420\beta_4\beta_3^2 - 1920\beta_3\beta_5\big)/1024 \ .
\end{aligned}
$$

Table 3 gives the relative bias of $S_{n,p}(\widehat{F})$ estimated from simulations for $p \leq 2$ and F normal and exponential. The estimates present bias even for $n = 100$ and bias-corrected estimates of order n^{-2} (i.e. $p = 2$): see Example C.4.

Table 3: Relative bias of $S_{n,p}(\widehat{F})$ for $T(F) = \sigma$.

		$n=5$		$n=10$		$n=100$	
		$p=1$	$p=2$	$p=1$	$p=2$	$p=1$	$p=2$
Norm $(0,1)$	Run 1	-0.1578	-0.0265	-0.0764	-0.0082	0.0281	-0.0045
	Run 2	-0.1592	-0.0277	-0.0745	-0.0080	0.0003	0.0031
Exp (1)	Run 1	-0.2278	-0.1019	-0.1251	-0.0422	-0.0158	-0.0029
	Run 2	-0.2331	-0.1084	-0.1206	-0.0422	-0.0176	-0.0004
Number of simulations/run		10,000		30,000		30,000	

The usual estimator of $\sigma(F)$ is the sample standard deviation, s.d. $= \left\{n\mu_2(\widehat{F})/(n-1)\right\}^{1/2}$, with mean $\sigma\left\{1 - t_1^* n^{-1} + O(n^{-2})\right\}$, where $t_1^* = t_1 - 1/2$. So, bias {s.d.}/bias $\{\sigma(\widehat{F})\} = \lambda_1 + O(n^{-1})$, where $\lambda_1 = (\beta_4 - 1)/(\beta_4 + 3)$.

For the normal, exponential and gamma (γ), $\beta_4 = 3$, 9 and $3 + 6\gamma^{-1}$, so $\lambda_1 = 1/3$, $2/3$ and $(5\gamma + 12)/(6\gamma + 12)$ and the s.d. improves on $\sigma(\widehat{F})$, although both are first order estimates, that is, both have bias $O(n^{-1})$.

To see how $S_{n,2}(\widehat{F})$ improves on the s.d., note that bias $\{S_{n,2}(\widehat{F})\}$/bias {s.d.} $= \lambda_2 n^{-1} + O(n^{-2})$, where $\lambda_2 = s_2/t_1^*$. For the normal, exponential and gamma (γ),

$$\beta_6 = 15, \;\; 265 \;\; \text{and} \;\; 120\gamma^{-2} + 130\gamma^{-1} + 15 \; ,$$

so

$$\begin{aligned} s_2 &= 65/64, \;\; 767/32 \;\; \text{and} \;\; N(\gamma)/64 \; , \\ \lambda_2 &= 65/16 \approx 4.06, \;\; 767/32 \approx 24.1 \;\; \text{and} \;\; N(\lambda)\left(2.5 + 6\lambda^{-1}\right)^{-1}/64 \; , \end{aligned}$$

where $N(\gamma) = 690\gamma^{-2} + 788\gamma^{-1} + 65$.

Example 5.16. Suppose $k = s_1 = 1$, $T(F) = \mu/\sigma = \mu\mu_2^{-1/2} = g(\mu, \mu_2) = \beta$ say. Again set $\beta_r = \mu_r \mu_2^{-r/2}$. Then the partial derivatives of g are $g_1 = \mu_2^{-1/2}$, $g_{1,1} = 0$, $g_2 = -\mu\mu_2^{-3/2}/2$, $g_{1,2} = -\mu_2^{-3/2}/2$, $g_{2,2} = 3\mu\mu_2^{-5/2}/4$, $g_{1,2,2} = 3\mu_2^{-5/2}/4$, $g_{2,2,2} = -15\mu\mu_2^{-7/2}/8$, and so on. Set $U_1(F) = \mu$, $U_2(F) = \mu_2$. Then defining $U_{i,j,\ldots}(1^I, 1^J, \ldots)$ as in (A.12)–(A.14),

$$\begin{aligned} U_{1,1}(1,1) &= \int U_{1,x}^2 = \int \mu_x^2 = \mu_2 \; , \\ U_{1,2}(1,1) &= \int U_{1,x} U_{2,x} = \int \mu_x \mu_{2,x} = \mu_3 \; , \\ U_{2,2}(1,1) &= \int U_{2,x}^2 = \mu_4 - \mu_2^2 \; . \end{aligned}$$

So, by (A.21),

$$T(1^2) \;=\; \beta_3 + \beta\big(3\beta_4 + 1\big)/4\ .$$

Also

$$\begin{aligned}
U_{1,2,2}(1,1,1) \;&=\; \int \mu_x \mu_{2,x}^2 \;=\; \mu_5 - 2\,\mu_2\mu_3\ ,\\
U_{2,2,2}(1,1,1) \;&=\; \int \mu_{2,x}^3 \;=\; \mu_6 - 3\,\mu_4\mu_2 + 2\,\mu_2^3\ ,\\
U_{1,2}(1,1^2) \;&=\; \int \mu_x \mu_{2,x,x} \;=\; -2\,\mu_3\ ,\\
U_{2,1}(1,1^2) \;&=\; \int \mu_{2,x}\mu_{x,x} \;=\; 0\ ,\\
U_{2,2}(1,1^2) \;&=\; \int \mu_{2,x}\mu_{2,x,x} \;=\; -2\left(\mu_4 - \mu_2^2\right)\ ,\\
U_1(1^3) \;&=\; \int \mu_{x,x,x}\ ,\\
U_2(1^3) \;&=\; \int \mu_{2,x,x,x} \;=\; 0\ .
\end{aligned}$$

So, by (A.22)

$$T(1^3)/3 \;=\; \big(3\beta_5 - 2\beta_3\big)/4 + \beta\big(-5\beta_6 + 11\beta_4 - 6\big)/8\ .$$

Similarly, at $(1,1,1^2)$, $U_{2,2,1} = 0$,

$$\begin{aligned}
&U_{1,2,2} \;=\; -2\left(\mu_5 - \mu_3\mu_2\right)\ , \quad U_{2,2,2} \;=\; -2\left(\mu_6 - 2\,\mu_4\mu_2 + \mu_2^3\right)\ ,\\
&U_{i,j}(1,1^3) = U_i(1^4) \;=\; 0\ , \quad U_{1,2}(1^2,1^2) \;=\; 0\ , \quad U_2(1^2,1^2) \;=\; 4\,\mu_4\ ,
\end{aligned}$$

so by (A.23),

$$T(1^4) \;=\; 3\big(-5\beta_7 + 3\beta_5 - 3\beta_3\big)/2 + 3\beta\big(35\beta_8 - 132\beta_6 + 242\beta_4 - 97\big)/16\ .$$

Also at $a = b = 1$,

$$\begin{aligned}
&U_{1,2}(ab,ab) \;=\; U_1(a^2b^2) \;=\; U_2(a^2b^2) \;=\; 0\ , \quad U_{2,2}(ab,ab) \;=\; \int \mu_{2,x,y}^2 \;=\; 4\,\mu_2^2\ ,\\
&U_{1,2,2}(ab,a,b) \;=\; U_{2,2}(a,ab^2) \;=\; U_{1,2}(a,ab^2) \;=\; U_{2,1}(a,ab^2) \;=\; 0\ ,\\
&U_{2,2,2}(ab,a,b) \;=\; \iint \mu_{2,x,y}\,\mu_{2,x}\,\mu_{2,y} \;=\; -2\,\mu_3^2\ .
\end{aligned}$$

So, by (A.24)

$$\begin{aligned}
T(1^2,1^2) \;&=\; 4\,g_{1,2,2,2}\,\mu_3\big(\mu_4 - \mu_2^2\big) + g_{2,2,2,2}\big(\mu_4 - \mu_2^2\big)^2 - 4\,g_{1,2,2}\,\mu_3\,\mu_2\\
&\quad - 4\,g_{2,2,2}\Big\{\big(\mu_4 - \mu_2^2\big)\mu_2 + 2\,\mu_3^2\Big\} + 12\,g_{2,2}\,\mu_2^2\\
&=\; -3\big(5\beta_4 - 43\big)\beta_3/8 + 3\beta\big(35\beta_4^2 + 90\beta_4 + 320\beta_3^2 - 77\big)/16\ .
\end{aligned}$$

So,

$$\begin{aligned} S_1(F) &= T_1(F) = -\beta_3/2 - \beta\big(3\beta_4+1\big)/8 \ , \\ S_2(F) &= \big(48\beta_5 - 15\beta_4\beta_3 - 23\beta_3\big)/64 \\ &\quad + \beta\big(-80\beta_6 + 446\beta_4 - 327 + 105\beta_4^2 + 960\beta_3^2\big)/128 \ . \end{aligned}$$

Note that $T(1^2,1^3)$, $T(1^2,1^2,1^2)$ and $S_3(F)$ may be calculated similarly using (A.7).

In the one sample example above $\boldsymbol{\mu}$ is the mean of $\mathbf{X}\sim F$. In many cases $\mathbf{X}_i = \mathbf{h}(\mathbf{Y}_i)$, where $\mathbf{h}\colon \mathbb{R}^t \to \mathbb{R}^s$ is a given transformation and $\mathbf{Y}_1, ..., \mathbf{Y}_n \sim G$ on $\mathbb{R}^t$ is the original sample. So, $\boldsymbol{\mu}(F) = \int \mathbf{x}\, dF(\mathbf{x}) = \int \mathbf{h}(\mathbf{y})\, dG(\mathbf{y})$. Equivalently, we may replace $\boldsymbol{\mu}(F) = \int \mathbf{x}\, dF(\mathbf{x})$ by $\boldsymbol{\mu}(F) = \int \mathbf{h}(\mathbf{x})\, dF(\mathbf{x})$, so that $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{h}(\mathbf{x}) - \boldsymbol{\mu}$. Similarly, if $s=1$ replace $\mu_r(F) = \int (x-\mu)^r dF(x)$ by $\int \big(h(x)-\mu\big)^r dF(x)$ so that (5.4) holds with $h_i = h_{x_i} = h(x_i) - \mu$. A similar remark holds for several samples.

The next four examples apply this idea to return times and exceedances.

Example 5.17. Take $k=1$, $h(\mathbf{x}) = I(\mathbf{x}\le\mathbf{a})$ for some $\mathbf{a}$ in $\mathbb{R}^s$, and $T(F) = \mu^{-1}$. Since $\mu = F(\mathbf{a})$, $T(F)$ is the return period of the event $\{\mathbf{X}\le\mathbf{a}\}$, where $\mathbf{X}\sim F$. But the case $T(F) = \mu^{-1}$ was dealt with in Example 5.3 in terms of μ_r. In this instance $\mu_r = \mu_r(Bi(1,p))$, where $p = F(\mathbf{a})$, so $\mu_2 = pq$, where $q = 1-p$, $\mu_3 = pq(1-2p)$ and $\mu_4 = pq(1-3pq)$. So, by Examples 5.6, 5.7 and Proposition 4.2 an estimate of the return period p^{-1} of bias $O(n^{-4})$ is $\widetilde{S_{n,4}}[\widehat{p}] = S_{n,4}[\widehat{p}]$ if $\widehat{p} > l$ or l^{-1} if $\widehat{p}\le l$, where $0<l<p$,

$$S_{n,4}[p] \;=\; p^{-1} + \sum_{i=1}^{3} S_i[p]/(n-1)_i \ ,$$

and $S_i[p] = S_i(F)$ is given by $S_1[p] = p^{-1} - p^{-2}$, $S_2[p] = -p^{-1} + p^{-3}$, $S_3[p] = 2p^{-1} + p^{-2} - 2p^{-3} - p^{-4}$.

The same formula with $p = 1 - F(\mathbf{a})$ and $\widehat{p} = 1 - \widehat{F}(\mathbf{a})$ gives an estimate of bias $O(n^{-4})$ for the return time of the event $\{\mathbf{X} > \mathbf{a}\}$. Similarly, for the event $\{\mathbf{x}\in A\}$ with $p = F(A)$ and $\widehat{p} = \widehat{F}(A)$. Similarly, we can apply Example 5.4 to obtain estimates of bias $O(n^{-p})$ for any smooth function $g(p_1, ..., p_k)$ given independent $n_i\widehat{p_i} \sim Bi(n_i, p_i)$, $1\le i\le k$. This problem can also be solved by the parametric method of Withers [27].

Example 5.18. Suppose $k=1$, $\mathbf{X}\sim F$ on $\mathbb{R}^t$ and $T(F) = Er(\mathbf{X}) \mid (\mathbf{X}\in A)$, where $A\subset\mathbb{R}^t$ is a measurable set, $F(A) > 0$ and $r\colon\mathbb{R}^t\to\mathbb{R}$ is a given function. Then $T(F) = \mu_1/\mu_2 = \mu_1(F)/\mu_2(F)$, where $\mu_i(F) = \int h_i(\mathbf{x})\, dF(\mathbf{x})$, $h_1(\mathbf{x}) = r(\mathbf{x}) I(\mathbf{x}\in A)$ and $h_2(\mathbf{x}) = I(\mathbf{x}\in A)$. So, $\{T_i, S_i,\, 1\le i\le 3\}$ are given in Example 5.2 in terms of the moments of (5.1) in which x_{j_i} now needs to be replaced by $h_{j_i}(\mathbf{x})$. Set

$$p = F(A) \ , \qquad q = 1 - p \ , \qquad I_i = \int_A \big(r(\mathbf{x}) - \mu_1\big)^i \, dF(\mathbf{x}) \ .$$

So, $\mu[2^j] = \mu_i(Bi(1,p))$ is given for $2 \le j \le 4$ in Example 5.17 and

$$\mu[1^i, 2^j] \ = \ I_i \, q^j + (-\mu_1)^i \, (-p)^j \, q \ .$$

Using $I_1 = 0$ simplification yields

$$S_{n,4}(F) \ = \ \mu_1 p^{-1} \Big\{ 1 - q^2 p^{-1}/(n-1) + q^3 p^{-2}/(n-1)_2 + q^3 p^{-3}(2p-1)/(n-1)_3 \Big\} \, .$$

Unlike Example 5.17, one does not need to know a lower bound for p, since $\mu_1 = 0$ if $p = 0$; so, if $\widehat{p} = 0$ one interprets $S_{n,4}(\widehat{F})$ as an arbitrary constant. This shows, surprisingly that the bias reduction problem for $T(F) = \mu_1/p$ can be treated as a parametric problem, the parameters being (μ_1, p). The more general problem of $T(F) = g(\mu_1, p)$ does *not* reduce to a finite parameter problem as it involves $\{\int_A r^i dF, \ i \ge 1\}$.

Example 5.19. The conditional distribution of exceedances is

$$F_u(x) \ = \ P\big(X - u < x \mid X - u > 0\big) \\ = \ \big\{F(x+u) - F(u)\big\} / \big\{1 - F(u)\big\} \tag{5.22}$$

for $x \ge 0$. This is μ_1/μ_2 with $A = \{y \colon y > u\} = (u, \infty)$, $B - \{y \colon x + u > y > u\} = (u, x+u)$ and $r(y) = I(y \in B)$. So, Example 5.18 applies with $\mu_1 = F(x+u) - F(u)$, $\mu_2 = 1 - F(u)$.

Example 5.20. The mean conditional exceedance is

$$\mu(F_u) \ = \ \int x \, dF_u(x) \ = \ \mu_1/\mu_2$$

for

$$\mu_1 = \int (x-u)_+ \, dF(x) \ , \qquad \mu_2 \ = \ 1 - F(u) \ ,$$

where

$$x_+ \ = \ \begin{cases} x, & \text{if } x > 0 \, , \\ 0, & \text{if } x \le 0 \, . \end{cases}$$

So, $r(y) = (y-u)_+$ and Example 5.18 applies.

The *central* moments of F_u of (5.22) are *not* covered by Example 5.18 and are probably best dealt with by writing them as functions of the noncentral moments and applying Example 5.1 with $\mu = \big\{\int (x-u)_+^i \, dF(x), \ i \ge 0\big\}$. A more direct approach is given by the following example.

Example 5.21. Suppose $T(F) = S(F_u)$ for F_u of (5.22). Set $C^y(F) = F_u(y)$. Then

$$C^y\big((1-\epsilon)F + \epsilon\,\delta_x\big) \;=\; F_u(y) + \epsilon\, C^y_F(x) + O(\epsilon^2)\ ,$$

and

$$\begin{aligned} T\big((1-\epsilon)F + \epsilon\,\delta_x\big) \;&=\; S\big(F_u(\cdot) + \epsilon\, C^{\cdot}_F(x) + O(\epsilon^2)\big) \\ &=\; S(F) + \epsilon \int S_{F_u}(y)\, C^y_F(x)\, dy \;+\; O(\epsilon^2)\ , \end{aligned}$$

where $C^y_F(x) = \mu_2^{-1} I(u < x < u+y) - \mu_1\,\mu_2^{-2} I(u < x)$. So,

$$T_F(x) \;=\; \int S_{F_u}(y)\, C^y_F(x)\, dy \;=\; \mu_2^{-1} S_{F_u}(x-u)\ . \tag{5.23}$$

Higher derivatives can be calculated from (5.23).

Now let us apply the previous note with $s = 1$, $t = r$, $h(\mathbf{y}) = \mathbf{a}'\mathbf{y}$, where $\mathbf{a}$ lies in $\mathbb{R}^r$. Set $\boldsymbol{\mu} = E\mathbf{Y}$. Then the joint central moment $\mu_{1,\ldots,r} = E(\mathbf{Y} - \boldsymbol{\mu})_1 \cdots \cdots (\mathbf{Y} - \boldsymbol{\mu})_r$ is the coefficient of $a_1 \cdots a_r/r!$ in $\mu_r(\mathbf{a}'\mathbf{Y})$, so the same relation is true of their derivatives. The same is also true of the cumulants. This device allows us to derive results for multivariate moments and cumulants from their univariate analogs.

For example, from Example 5.6, for a univariate random variable, $\mu_2(x) = (x-\mu)^2 - \mu_2$ and $\mu_2(x_1, x_2) = -2\,(x_1 - \mu)\,(x_2 - \mu)$. So, for a bivariate random variable, $\mu_{1,2}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})_1 (\mathbf{x} - \boldsymbol{\mu})_2 - \mu_{1,2}$ and $\mu_{1,2}(\mathbf{x}_1, \mathbf{x}_2) = -2\,(\mathbf{x}_1 - \boldsymbol{\mu})_1 (\mathbf{x}_2 - \boldsymbol{\mu})_2$.

We illustrate this device further with the problems of estimating multivariate moments and the correlation of a bivariate distribution and its square.

Example 5.22. Suppose $k = 1$, $s = 2$ and $T(F) = \mu_{1,2}$. From Example 5.6 and the previous remark, an UE of $\mu_{1,2}$ is $\mu_{1,2}/(1 - n^{-1})$ at $F = \widehat{F}$.

Similarly, we have

Example 5.23. Suppose $k = 1$, $s = 3$ and $T(F) = \mu_{1,2,3}$. An UE of $\mu_{1,2,3}$ is $\mu_{1,2,3}/\{(1 - n^{-1})\,(1 - 2n^{-2})\}$ at $F = \widehat{F}$.

Example 5.24. Suppose $k = 1$, $s = 2$, and $T(F) = \mu_{1,2}\{\mu_{1,1}\,\mu_{2,2}\}^{-1/2}$, the correlation of a bivariate sample. So, (A.1) of Appendix A holds with $\mathbf{S}(F) = (\mu_{1,2}, \mu_{1,1}, \mu_{2,2})$ and $g(\mathbf{S}) = S_1(S_2 S_3)^{-1/2}$. We shall apply (A.8). Set $\nu_{i,j,\ldots} = \mu_{i,j,\ldots}(\mu_{i,i}\,\mu_{j,j} \cdots)^{-1/2}$. So, $T(F) = \nu_{1,2}$. Now $S_1(1^2) = \int S_{1,\mathbf{x},\mathbf{x}} = -2\,\mu_{1,2}$, $S_2(1^2) = \int S_{2\mathbf{x},\mathbf{x}} = -2\,\mu_{1,1}$ and $S_3(1^2) = \int S_{3,\mathbf{x},\mathbf{x}} = -2\,\mu_{2,2}$. Also $g_1 = (\mu_{1,1}\,\mu_{2,2})^{-1/2}$, $g_2 = -\nu_{1,2}/\mu_{1,1}$, $g_3 = -\nu_{1,2}/\mu_{2,2}$. So, $g_i S_i(1^2) = T(F)\,(-2+1+1) = 0$. Similarly, $S_{1,\mathbf{x}} = (\mathbf{x} - \boldsymbol{\mu})_1 (\mathbf{x} - \boldsymbol{\mu})_2 - \mu_{1,2}$, so $S_{1,1}(1,1) = \int S^2_{1,\mathbf{x}} = \mu_{1,1,2,2} - \mu^2_{1,2}$, and similarly $S_{1,2}(1,1) = \mu_{1,1,1,2} - \mu_{1,1}\,\mu_{1,2}$, $S_{1,3}(1,1) = \mu_{1,2,2,2} - \mu_{1,2}\,\mu_{2,2}$, $S_{2,2}(1,1) = \mu_{1,1,1,1} - \mu_{1,1}^2$, $S_{3,3}(1,1) = \mu_{2,2,2,2} - \mu_{2,2}^2$, and $S_{2,3}(1,1) = \mu_{1,1,2,2} - \mu_{1,1}\,\mu_{2,2}$. So, an estimate of bias $O(n^{-2})$ is $T(F) - T(1^2)/(2n)$ or $T(F) - T(1^2)/(2n-2)$ at $F = \widehat{F}$, where by (A.8), $T(1^2) = \nu_{1,2}(3\nu_{1,1,1,1} + 3\nu_{2,2,2,2} + 2\nu_{1,1,2,2})/4 - \nu_{1,1,1,2} - \nu_{1,2,2,2}$.

Example 5.25. Suppose $k=1$, $s=2$ and $T(F) = \mu_{1,2}^2\{\mu_{1,1}\,\mu_{2,2}\}^{-1} = \nu_{1,2}^2$, the square of the correlation of a bivariate sample. Again (A.1) holds with $\mathbf{S}(F) = (\mu_{1,2}, \mu_{1,1}, \mu_{2,2})$ but now $g(\mathbf{S}) = S_1^2(S_2 S_3)^{-1}$, so $g_1 = 2\,T(F)\,S_1^{-1}$, $g_2 = -T(F)\,S_2^{-1}$, $g_3 = -T(F)\,S_3^{-1}$, $g_{i,i} = 2\,T(F)\,S_i^{-2}$, $g_{1,2} = -2\,T(F)\,(S_1 S_2)^{-1}$, $g_{1,3} = -2\,T(F)\,(S_1 S_3)^{-1}$, and $g_{2,3} = T(F)\,(S_2 S_3)^{-1}$. Again $g_i\,S_i(1^2) = T(F)(-4+2+2) = 0$. So, an estimate of bias $O(n^{-2})$ is $T(F) - T(1^2)/(2n)$ or $T(F) - T(1^2)/(2n-2)$ at $F = \widehat{F}$, where by (A.8), $T(1^2) = 2\,\nu_{1,2}^2(\nu_{1,1,1,1} + \nu_{2,2,2,2} + 2\,\nu_{1,1,2,2} - 2\,\nu_{1,1,1,2} - 2\,\nu_{1,2,2,2})$.

6. ESTIMATING COVARIANCES OF ESTIMATES

In this section, we give an estimate of bias $O(n^{-3})$ for $\mathbf{V}_n(F)$, the covariance of $\mathbf{T}(\widehat{F})$, where now $\mathbf{T}(F)$ is a $q\times 1$ *vector* with *components* $\{T^\alpha(F),\ 1 \le \alpha \le q\}$. After Example 6.1, we estimate the covariance of more general estimates of $\mathbf{T}(F)$.

From the formulas for $\{K_i^{a,b}\}$ on pages 66 and 67 in Withers [24],

$$V_n^{\alpha,\beta}(F) \;=\; \operatorname{covar}\bigl(T^\alpha(\widehat{F}), T^\beta(\widehat{F})\bigr) \;=\; \sum_{i=1}^{\infty} n^{-i} K_i^{\alpha,\beta}(F)\ , \tag{6.1}$$

where

$$\begin{aligned} K_1^{\alpha,\beta}(F) \;&=\; t_i^\alpha\, t_j^\beta\, k^{i,j} \;=\; \sum \lambda_a \iint T_F^\alpha\binom{a}{x}\, T_F^\beta\binom{a}{y}\, d\kappa_a(x,y) \\ &=\; \sum \lambda_a T^{\alpha,\beta}(a,a)\ , \end{aligned} \tag{6.2}$$

$$\begin{aligned} K_2^{\alpha,\beta}(F) \;&=\; \sum^2 t_{i,j}^\alpha\, t_k^\beta k^{i,j,k}/2 \;+\; \left(\sum^2 t_{i,j,k}^\alpha\, t_l^\beta + t_{i,k}^\alpha\, t_{j,l}^\beta\right) k^{i,j}\, k^{k,l}/2 \\ &=\; \sum \lambda_a \sum^2 \int^3 T_F^\alpha\binom{a,a}{x,y}\, T_F^\beta\binom{a}{z}\, d\kappa_a(x,y,z)/2 \\ &\quad+ \sum \lambda_a \lambda_b \int^4 \left\{\sum^2 T_F^\alpha\binom{a,a,b}{w,x,y}\, T_F^\beta\binom{b}{z} \right. \\ &\quad\left. + T_F^\alpha\binom{a,b}{w,x}\, T_F^\beta\binom{a,b}{y,z}\right\}\, d\kappa_a(w,x)\, d\kappa_b(y,z)/2 \\ &=\; \sum \lambda_a \sum^2 T^{\alpha,\beta}(a^2,a)/2 \\ &\quad+ \sum \lambda_a \lambda_a \left\{\sum^2 T^{\alpha,\beta}(a^2 b, b) + T^{\alpha,\beta}(ab,ab)\right\}/2\ , \end{aligned} \tag{6.3}$$

$$\sum^{2} f_{\alpha,\beta} \;=\; f_{\alpha,\beta}+f_{\beta,\alpha}\ ,$$

$$T^{\alpha,\beta}(a,a) \;=\; \int T_F^{\alpha}\binom{a}{x}\,T_F^{\beta}\binom{a}{x}\,dF_a(x)\ ,$$

$$T^{\alpha,\beta}(a^2,a) \;=\; \int T_F^{\alpha}\binom{a,a}{x,x}\,T_F^{\beta}\binom{a}{x}\,dF_a(x)\ , \tag{6.4}$$

$$T^{\alpha,\beta}(a^2b,b) \;=\; \iint T_F^{\alpha}\binom{a,a,b}{x,x,y}\,T_F^{\beta}\binom{b}{y}\,dF_a(x)\,dF_b(y)\ , \tag{6.5}$$

and

$$T^{\alpha,\beta}(ab,ab) \;=\; \iint T_F^{\alpha}\binom{a,b}{y,x}\,T_F^{\beta}\binom{a,b}{x,y}\,dF_a(x)\,dF_b(y)\ . \tag{6.6}$$

Also, setting $V^{\alpha,\beta}(F)=K_1^{\alpha,\beta}(F)$ and differentiating, we have

$$V_F^{\alpha,\beta}\binom{a}{x}\Big/\lambda_a \;=\; T_F^{\alpha}\binom{a}{x}\,T_F^{\beta}\binom{a}{x}-T^{\alpha,\beta}(a,a)\;+\sum^{2}\int T_F^{\alpha}\binom{a,a}{y,x}\,T_F^{\beta}\binom{a}{y}\,dF_a(y)\ ,$$

and

$$\begin{aligned}
V_F^{\alpha,\beta}\binom{a,a}{x,x}\Big/\lambda_a \;=\; \sum^{2}\Bigg[&\left\{T_F^{\alpha}\binom{a,a}{x,x}-T_F^{\alpha}\binom{a}{x}\right\}T_F^{\beta}\binom{a}{x}\;+\;T_F^{\alpha}\binom{a,a}{x,x}\,T_F^{\beta}\binom{a}{x}\\
&-\int T_F^{\alpha}\binom{a,a}{x,y}\,T_F^{\beta}\binom{a}{y}\,dF_a(y)\;+\int T_F^{\alpha}\binom{a,a}{x,y}\,T_F^{\beta}\binom{a,a}{x,y}\,dF_a(y)\\
&+\int\left\{T_F^{\alpha}\binom{a,a,a}{x,x,y}-T_F^{\alpha}\binom{a,a}{x,y}\right\}T_F^{\beta}\binom{a}{y}\,dF_a(y)\Bigg]\ ,
\end{aligned}$$

so that

$$\begin{aligned}
C_1(V^{\alpha,\beta},F) \;&=\; \sum\lambda_a V^{\alpha,\beta}(a^2)\\
&=\; \sum\lambda_a^2\Bigg[\sum^{2}\Big\{T^{\alpha,\beta}(a^2,a)+T^{\alpha,\beta}(a^2b,b)/2\Big\}\\
&\qquad +2\,T^{\alpha,\beta}(ab,ab)-T^{\alpha,\beta}(a,a)\Bigg]_{b=a}\ .
\end{aligned}$$

So, $n^{-1}K_1^{\alpha,\beta}(\widehat{F})$ given by (6.2) estimates $V_n^{\alpha,\beta}(F)$ with bias $O(n^{-2})$ and $n^{-1}K_1^{\alpha,\beta}(\widehat{F})$ $+\,n^{-2}L^{\alpha,\beta}(\widehat{F})$ estimates $V_n^{\alpha,\beta}(F)$ with bias $O(n^{-3})$, where

$$\begin{aligned}
L^{\alpha,\beta}(F) \;&=\; K_2^{\alpha,\beta}(F)-C_1(V^{\alpha,\beta},F)\\
&=\; \sum\big(\lambda_a-\lambda_a^2\big)\ \sum^{2}T^{\alpha,\beta}(a^2,a)/2\\
&\quad +\sum\lambda_a\lambda_b\ \left\{\sum^{2}T^{\alpha,\beta}(a^2b,b)+T^{\alpha,\beta}(ab,ab)\right\}\Big/2\\
&\quad -\sum\lambda_a^2\ \left\{\sum^{2}T^{\alpha,\beta}(a^2b,b)/2+2\,T^{\alpha,\beta}(ab,ab)-T^{\alpha,\beta}(a,a)\right\}_{b=a}\ .
\end{aligned}$$

If $k=1$ this reduces to

$$L^{\alpha,\beta}(F) \;=\; T^{\alpha,\beta}(a,a) - 3\,T^{\alpha,\beta}(ab,ab)/2 \tag{6.7}$$

at $a=b=1$, so that

$$(n-1)^{-1}\,T^{\alpha,\beta}(a,a) - 3\,n^{-2}\,T^{\alpha,\beta}(ab,ab)/2 \tag{6.8}$$

at $\{F=\widehat{F},\ a=b=1\}$ estimates $V_n^{\alpha,\beta}(F)$ with bias $O(n^{-3})$, where at $a=b=1$,

$$T^{\alpha,\beta}(a,a) \;=\; \int T_F^{\alpha}(x)\,T_F^{\beta}(x)\,dF(x)\ ,$$

and

$$T^{\alpha,\beta}(ab,ab) \;=\; \iint T_F^{\alpha}(x,y)\,T_F^{\beta}(x,y)\,dF(x)\,dF(y)\ .$$

One may prefer to use $n^{-1}-n^{-2}$ instead of $(n-1)^{-1}$ in (6.8). Remarkably, unlike the case $k>1$, the estimate (6.8) does not depend on $T^{\alpha,\beta}(a^2,a)$ or $T^{\alpha,\beta}(a^2b,b)$ at $a=b=1$.

We now show how to estimate

$$\mathbf{W}_n(F) \;=\; \mathrm{covar}\,\mathbf{T}_{(n)}(\widehat{F})\ , \tag{6.9}$$

where

$$\mathbf{T}_{(n)} \;=\; \sum_{i=0}^{\infty} n^{-i}\,\mathbf{T}_i$$

is $q\times 1$ and $\mathbf{T}_0=\mathbf{T}$. Clearly, $\mathbf{T}_{(n)}(\widehat{F})$ estimates $\mathbf{T}(F)$. Now

$$\mathbf{W}_n(F) \;=\; \sum_{i,j\ge 0} n^{-i-j}\,\mathbf{W}_n(\mathbf{T}_i,\mathbf{T}_j)\ ,$$

where

$$\mathbf{W}_n(\mathbf{T}_i,\mathbf{T}_j) \;=\; \mathrm{covar}\big(\mathbf{T}_i(\widehat{F}),\mathbf{T}_j(\widehat{F})\big)$$

has (α,β) element

$$W_n^{\alpha,\beta}(\mathbf{T}_i,\mathbf{T}_j) \;=\; \mathbf{W}_n(T_i^{\alpha},T_j^{\beta}) \;=\; V_n^{1,2}(F)$$

of (6.1) with $(T^1,T^2)=(T_i^{\alpha},T_j^{\beta})$. So,

$$W_n^{\alpha,\beta}(F) \;=\; \sum_{l=1}^{\infty} n^{-l} K_l^{\alpha,\beta}[F]\ ,$$

where

$$K_l^{\alpha,\beta}[F] \;=\; \sum_{i+j+k=l} K_k(T_i^{\alpha},T_j^{\beta})\ ,$$

and

$$K_k(T^1,T^2) \;=\; K_k^{1,2}(F) \quad \text{of (6.1)}\;.$$

So,

$$K_1^{\alpha,\beta}[F] \;=\; K_1(T^\alpha,T^\beta) \;=\; K_1^{\alpha,\beta}(F)$$

of (6.2), and

$$K_2^{\alpha,\beta}[F] \;=\; K_2^{\alpha,\beta}(F) + \triangle^{\alpha,\beta}\;,$$

where

$$\triangle^{\alpha,\beta} \;=\; \sum^{2} K_1(T^\alpha,T_1^\beta)\;,$$

and

$$K_1(T^\alpha,T_1^\beta) \;=\; K_1^{\alpha,\beta}(F)$$

of (6.2) at $T^\beta = T_1^\beta$.

So, $n^{-1}K_1^{\alpha,\beta}(\widehat{F})$ and $n^{-1}K_1^{\alpha,\beta}(\widehat{F})+n^{-2}L^{\alpha,\beta}(\widehat{F})$ estimate $W_n^{\alpha,\beta}(F)$ with bias $O(n^{-2})$ and $O(n^{-3})$, respectively, where

$$L^{\alpha,\beta}[F] \;=\; K_2^{\alpha,\beta}[F] - C_1(V^{\alpha,\beta},F) \;=\; L^{\alpha,\beta}(F) + \triangle^{\alpha,\beta}\;. \tag{6.10}$$

Alternatively, for $k=1$, the sum of (6.8) and $n^{-2}\triangle^{\alpha,\beta}$ at $F=\widehat{F}$ estimates $W_n^{\alpha,\beta}(F)$ with bias $O(n^{-3})$. Now for $p\geq 2$, $T_{n,p}$ of (1.3) has the form $\mathbf{T}_{(n)}$ of (6.9) with T_1 given by (4.1), so that

$$T_{1,F}^{\beta}\binom{a}{x} \;=\; -\lambda_a\left\{T_F^\beta\binom{a^2}{x^2} - T^\beta(a^2) + \int T_F^\beta\binom{a^2,a}{y^2,x}\,dF_a(y)\right\}/2\;,$$

and so

$$K_1(T^\alpha,T_1^\beta) \;=\; -\sum \lambda_a^2\left\{T^{\beta,\alpha}(a^2,a) + T^{\beta,\alpha}(a^3,a)\right\}/2\;,$$

$$\triangle^{\alpha,\beta} \;=\; -\sum \lambda_a^2 \sum^{2}\left\{T^{\alpha,\beta}(a^2,a) + T^{\alpha,\beta}(a^2b,b)\right\}_{b=a}/2\;, \tag{6.11}$$

$$\begin{aligned}
K_2^{\alpha,\beta}[F] \;&=\; \sum\left(\lambda_a-\lambda_a^2\right)\sum^{2} T^{\alpha,\beta}(a^2b,b)/2 \;-\sum\lambda_a^2\sum T^{\alpha,\beta}(a^2b,b)_{b=a}/2\\
&\quad +\sum\lambda_a\lambda_b\left\{\sum^{2} T^{\alpha,\beta}(a^2b,b) + T^{\alpha,\beta}(ab,ab)\right\}/2\;,
\end{aligned}$$

$$\begin{aligned}
L^{\alpha,\beta}[F] \;&=\; \sum\left(\lambda_a/2-\lambda_a^2\right)\sum^{2} T^{\alpha,\beta}(a^2,a)\\
&\quad +\sum\lambda_a\lambda_b\left\{\sum^{2} T^{\alpha,\beta}(a^2b,b) + T^{\alpha,\beta}(ab,ab)\right\}/2\\
&\quad -\sum\lambda_a^2\left\{\sum^{2} T^{\alpha,\beta}(a^2b,b) + 2\,T^{\alpha,\beta}(ab,ab) - T^{\alpha,\beta}(a,a)\right\}_{b=a}\;.
\end{aligned}$$

For $k=1$, at $a=b=1$, this gives

$$\begin{aligned}
&\triangle^{\alpha,\beta} \;=\; -\sum^{2}\Big\{T^{\alpha,\beta}(a^2,a)+T^{\alpha,\beta}(a^2b,b)\Big\}/2\ ,\\
(6.12)\quad &K_2^{\alpha,\beta}[F] \;=\; T^{\alpha,\beta}(ab,ab)/2\ ,\\
&\mathrm{covar}\big(T^{\alpha}_{n,p}(\widehat{F}),T^{\beta}_{n,p}(\widehat{F})\big) \;=\; n^{-1}T^{\alpha,\beta}(a,a)+n^{-2}T^{\alpha,\beta}(ab,ab)/2+O(n^{-3})
\end{aligned}$$

which, remarkably, does not depend on $T(a^2,a)$ or $T(a^2b,b)$ to this accuracy — whereas $L^{\alpha,\beta}[F]$ does.

Example 6.1. Consider again Example 5.1, that is $k=1$, $\mathbf{T}(F)=\mathbf{g}(\boldsymbol{\mu})$, where now $\mathbf{g}$ may be a vector $\{g^\alpha\}$. By (A.17)–(A.20) at $a=b=1$

$$\begin{aligned}
K_1^{\alpha,\beta}(F) \;&=\; T^{\alpha,\beta}(a,a) \;=\; g_i^\alpha\, g_j^\beta\, \mu[i,j]\ ,\\
T^{\alpha,\beta}(ab,ab) \;&=\; g_{i,j}^\alpha\, g_{k,l}^\beta\, \mu[i,k]\,\mu[j,l]\ ,\\
T^{\alpha,\beta}(a^2,a) \;&=\; g_{i,j}^\alpha\, g_k^\beta\, \mu[i,j,k]\ ,\\
T^{\alpha,\beta}(a^2b,b) \;&=\; g_{i,j,k}^\alpha\, g_l^\beta\, \mu[i,j]\,\mu[k,l]\ ,
\end{aligned}$$

and $K_2^{\alpha,\beta}(F)$, $L^{\alpha,\beta}(F)$, $K_2^{\alpha,\beta}[F]$, $L^{\alpha,\beta}[F]$ are given by (6.3), (6.7), (6.10), (6.11), (6.12). Note that $L^{\alpha,\beta}$ depends only on the first and second moments of F, even though $K_2^{\alpha,\beta}$ depends on the third moments!

Example 6.2. Consider Example 6.1 with $g(\boldsymbol{\mu})=\boldsymbol{\alpha}'\boldsymbol{\mu}/\boldsymbol{\beta}'\boldsymbol{\mu}=N/D$, say, — that is, Example 5.2. Since $q=1$ we drop suffixes α, β. Define $\mu[\cdot]$ and δ_i as in (5.1) and (5.3). Then at $a=b=1$

$$\begin{aligned}
K_1(F) \;&=\; T(a,a) \;=\; D^{-2}\mu_2[\delta,\delta]\ ,\\
T(ab,ab) \;&=\; 2\,\mu_2[\delta,\beta]^2+2\,\mu_2[\delta,\delta]\,\mu_2[\beta,\beta]\ ,\\
T(a^2,a) \;&=\; -2\,D^{-3}\mu_3[\delta,\delta,\beta]\ ,\\
T(a^2b,b) \;&=\; 2\,D^{-4}\big\{2\,\mu_2[\delta,\beta]^2+\mu_2[\delta,\delta]\,\mu_2[\beta,\beta]\big\}\,,
\end{aligned}$$

where $\mu_2[\delta,\beta]=\delta_i\,\beta_j\,\mu[i,j]$ and $\mu_3[\alpha,\beta,\gamma]=\alpha_i\,\beta_j\,\gamma_k\,\mu[i,j,k]$. In particular, for $g(\boldsymbol{\mu})=\mu_1/\mu_2$, at $a=b=1$ setting $\gamma_{i,j,\ldots}=\mu(i,j,\ldots)\,\mu_i^{-1}\mu_j^{-1}\cdots$, we have

$$\begin{aligned}
(6.13)\qquad &K_1(F) \;=\; T(a,a) \;=\; \big(\mu_1/\mu_2\big)^2\big(\gamma_{1,1}-2\,\gamma_{1,2}+\gamma_{2,2}\big)\,,\\
&T(ab,ab) \;=\; 2\,\big(\mu_1/\mu_2\big)^2\big(\gamma_{1,1}\,\gamma_{2,2}-4\,\gamma_{1,2}\,\gamma_{2,2}+2\,\gamma_{2,2}^2\big)\,,\\
(6.14)\qquad &T(a^2,a) \;=\; -2\big(\mu_1/\mu_2\big)^2\big(\gamma_{1,1,2}-2\,\gamma_{1,2,2}+\gamma_{2,2,2}\big)\,,\\
&T(a^2b,b) \;=\; 2\,\big(\mu_1/\mu_2\big)^2\big(2\,\gamma_{1,2}^2-5\,\gamma_{1,2}\,\gamma_{2,2}+3\,\gamma_{2,2}^2+\gamma_{1,1}\,\gamma_{2,2}\big)\,.
\end{aligned}$$

Note that (6.13) is in agreement with equation (10.17) of Kendall and Stuart [15].

Example 6.3. Consider Example 6.1 with $g(\mu)=N^p$, where $N=\boldsymbol{\alpha}'\boldsymbol{\mu}$, that is, we consider Example 5.3. In the notation there, with $a=b=1$

$$\begin{aligned} K_1(F) &= T(a,a) = p^2 N^{2p}\alpha_{(2)} \ , \\ T(ab,ab) &= p^2\big(p-1\big)^2 N^{2p}\alpha_{(2)}^2 \ , \\ T(a^2,a) &= p^2\big(p-1\big)\,N^{2p}\alpha_{(3)} \ , \\ T(a^2b,b) &= (p)_3\,p\,N^{2p}\alpha_{(2)}^2 \ . \end{aligned}$$

In particular, for $s=1$ and $g(\mu)=\mu^p$, with $a=b=1$

$$\begin{aligned} T(a,a) &= p^2\mu^{2p-2}\mu_2 \ , \qquad T(ab,ab) = p^2\big(p-1\big)^2\mu^{2p-4}\mu_2^2 \ , \\ T(a^2,a) &= p^2\big(p-1\big)\,\mu^{2p-3}\mu_3 \ , \qquad T(a^2b,b) = (p)_3\,p\,\mu^{2p-4}\mu_2^2 \ . \end{aligned}$$

For example, $\mathrm{var}\{\widehat{\mu}^{-1}\}$ or (if Proposition 4.2 needs to be applied), $\mathrm{var}\{\widehat{\mu}^{-1}I(|\widehat{\mu}|>l)\}$, where $l>0$ is a known lower bound for $|\mu|$, can be estimated by

$$\widehat{T}_{n,2} = (n-1)^{-1}\widehat{\mu}^{-4}\,\widehat{\mu}_2 - 6\,n^{-2}\widehat{\mu}^{-6}\,\widehat{\mu}_2^2$$

or by

$$\widehat{T}_{n,2}\,I\big(|\widehat{\mu}|>l\big)$$

with bias $O(n^{-3})$, where $(\widehat{\mu},\widehat{\mu}_2)$ is (μ,μ_2) at $F=\widehat{F}$. Alternatively, replacing n^{-2} in $\widehat{T}_{n,2}$ by $(n-1)^{-2}$ and setting $s^2=\widehat{\mu}_2\,n/(n-1)$, the UE of μ_2, we obtain

$$T^\star_{n,2} = n^{-1}\,\widehat{\mu}^{-4}s^2 - 6\,n^{-2}\,\widehat{\mu}^{-6}s^4 \ , \qquad T^\star_{n,2}\,I\big(|\widehat{\mu}|>l\big)$$

as estimates with bias $O(n^{-3})$.

7. ESTIMATING THE COVARIANCE OF AN ESTIMATE OF BIAS

The emphasis of this paper has been to reduce bias, not estimate it. However, a number of papers have given methods for estimating the variance of an estimate of bias for the case $k=1$. See, for example, Efron [7] and Davison *et al.* [6]. These papers provide bootstrap and jackknife methods of an order of magnitude less efficient computationally than the Taylor series method (also called the delta method or the infinitesimal jackknife when $p=2$) used here.

Suppose then $\mathbf{T}(F)$ is a $q\times 1$ functional. Note that $\mathbf{T}(\widehat{F})$ has bias $n^{-1}\mathbf{B}(F)/2 + O(n^{-2})$, where $\mathbf{B}(F)=|2|=\sum\lambda_a T(a^2)$. Its estimate $n^{-1}\mathbf{B}(\widehat{F})/2$ has covariance $n^{-2}\mathbf{V}(F)/4+O(n^{-3})$, where

$$V^{\alpha,\beta}(F) = \sum\lambda_a\int B_F^\alpha\Big(\begin{smallmatrix}a\\x\end{smallmatrix}\Big)\,B_F^\beta\Big(\begin{smallmatrix}a\\x\end{smallmatrix}\Big)\,dF_a(x) =$$

$$= \sum \lambda_a^3 \left\{ \int T^\alpha\!\binom{a,a}{x,x} T^\beta\!\binom{a,a}{x,x} - T^\alpha(a^2)\, T^\beta(a^2) \right.$$
$$\left. + \sum^2 \iint T^\alpha\!\binom{a,a,a}{x,x,y} T^\beta\!\binom{a,a}{y,y} + \iiint T^\alpha\!\binom{a,a,a}{x,x,z} T^\beta\!\binom{a,a,a}{y,y,z} \right\}$$

and $dF_a(x)$, $dF_a(y)$, $dF_a(z)$ are implicit in the integrals. Finally, $n^{-2}\mathbf{V}(\widehat{F})/4$ estimates $\operatorname{covar}\{n^{-1}\mathbf{B}(\widehat{F})/2\}$ with bias $O(n^{-3})$.

The same is true if we replace $\mathbf{B}(\widehat{F})$ by $\mathbf{B}_{n,p}(\widehat{F})$. If desired, one could apply Section 6 to reduce this bias to $O(n^{-4})$.

In equation (2.6) of Davison *et al.* [6] and the following line a factor 1/2 should be inserted. So, the usual bootstrap and the usual jackknife estimates of bias as well as our estimate $n^{-1}\mathbf{B}(F)/2$, all have bias $O(n^{-2})$.

APPENDIX A

Here, we note and illustrate the following chain rule for the partial derivatives of

(A.1) $$T(F) = g\big(\mathbf{S}(F)\big) ,$$

where $\mathbf{S}(F)$ is $q\times 1$ and $g\colon \mathbb{R}^q \to \mathbb{R}$.

First, suppose $k=1$, that is, F is a single d.f. Given $r\geq 1$, let $\mathbf{s}(\mathbf{y})\colon \mathbb{R}^r \to \mathbb{R}^q$ be an arbitrary function. Set $\partial_i = \partial/\partial y_i$. Then

(A.2) $$T_F(\mathbf{x}_1, ..., \mathbf{x}_r) = \partial_1 \cdots \partial_r\, g\big(\mathbf{s}(\mathbf{y})\big) ,$$

evaluated with $\mathbf{s}(\mathbf{y})$ replaced by $\mathbf{S}(F)$, and $\partial_1\cdots\partial_r\,\mathbf{s}(\mathbf{y})$ replaced by $\mathbf{S}_F(\mathbf{x}_1, ..., \mathbf{x}_r)$. So, setting

$$\begin{aligned} T_{1,...,r} &= T_F(\mathbf{x}_1, ..., \mathbf{x}_r) , \\ S_{i,1,...,r} &= S_{i,F}(\mathbf{x}_1, ..., \mathbf{x}_r) , \\ g_{i,j,...} &= \partial_i \partial_j \cdots g(\mathbf{s}) \end{aligned}$$

with $\partial_i = \partial/\partial s_i$ at $\mathbf{s} = \mathbf{S}(F)$, we have

(A.3) $$T_1 = g_i\, S_{i,1} , \qquad T_{1,2} = g_{i,j}\, S_{i,1}\, S_{j,2} + g_i\, S_{i,1,2} ,$$

(A.4) $$T_{1,2,3} = g_{i,j,k}\, S_{i,1}\, S_{j,2}\, S_{k,3} + g_{i,j} \sum^3 S_{i,1,2}\, S_{j,3} + g_i\, S_{i,1,2,3} ,$$

(A.5) $$\begin{aligned} T_{1,2,3,4} &= g_{i,j,k,l}\, S_{i,1}\, S_{j,2}\, S_{k,3}\, S_{l,4} + g_{i,j,k} \sum^6 S_{i,1}\, S_{j,2}\, S_{k,3,4} \\ &\quad + g_{i,j} \left(\sum^4 S_{i,1}\, S_{j,2,3,4} + \sum^3 S_{i,1,2}\, S_{j,3,4} \right) + g_i\, S_{i,1,2,3,4} , \end{aligned}$$

where summation over repeated suffixes $i, j, \ldots$ is implicit, and by the multivariate version of Faa de Bruno's chain rule given in Withers [26], for $r \geq 1$,

$$T_{1,\ldots,r} = \sum_{k=1}^{r} g_{i_1,\ldots,i_k}\big(\mathbf{S}(F)\big) \sum_{\mathbf{n}} \sum^{m(\mathbf{n})} S_{i_1,\pi_1} \cdots S_{i_k,\pi_k} \ , \tag{A.6}$$

where $\sum^{m(\mathbf{n})}$ sums over all $m(\mathbf{n}) = r!/\prod_{i=1}^{r}(i!^{n_i}\, n_i!)$ partitions $(\pi_1, \ldots, \pi_k)$ of $1, \ldots, r$ giving distinct terms with n_i of the π's of length i, and $\sum_{\mathbf{n}}$ sums over $\big\{\mathbf{n} \in N^r,\ \sum_{i=1}^{r} n_i = k,\ \sum_{i=1}^{r} i\, n_i = r\big\}$. For example,

$$\sum^{3} S_{i,1,2}\, S_{j,3,4} = S_{i,1,2}\, S_{j,3,4} + S_{i,1,3}\, S_{j,2,4} + S_{i,1,4}\, S_{j,2,3} \ .$$

The reader can derive $T_{1,2,3}$ from $T_{1,2}$ using equation (2.6) of Withers [25] to appreciate the labor-saving this rule gives.

By equation [4c] of Comtet [5] the general term can be written in terms of the multivariate exponential Bell polynomials, $\{B_{r,k}(\mathbf{S})_{i_1,\ldots,i_k}\}$:

$$T_{1,\ldots,r} = \sum_{k=1}^{r} g_{i_1,\ldots,i_k}\, B_{r,k}(\mathbf{S})_{i_1,\ldots,i_k} \ . \tag{A.7}$$

This is a much easier form to use than (A.6) as these polynomials are immediately derived from the univariate polynomials $B_{r_k}(\mathbf{S})$ tabled on pages 307–308 of Comtet [5]. For example, the table gives

$$\begin{aligned} B_{4,1}(\mathbf{S}) &= S_4 \ , \\ B_{4,2}(\mathbf{S}) &= 4\, S_1 S_3 + 3\, S_2^2 \ , \\ B_{4,3}(\mathbf{S}) &= 6\, S_1^2 S_2 \ , \\ B_{4,4}(\mathbf{S}) &= S_1^4 \ , \end{aligned}$$

so

$$\begin{aligned} B_{4,1}(\mathbf{S})_{i_1} &= S_{i_1,1,2,3,4} \ , \\ B_{4,2}(\mathbf{S})_{i_1,i_2} &= \sum^{4} S_{i_1,1}\, S_{i_2,2,3,4} + \sum^{3} S_{i_1,1,2}\, S_{i_2,3,4} \ , \\ B_{4,3}(\mathbf{S})_{i_1,i_2,i_3} &= \sum^{6} S_{i_1,1}\, S_{i_2,2}\, S_{i_3,3,4} \ , \\ B_{4,4}(\mathbf{S})_{i_1,\ldots,i_4} &= S_{i_1,1} \cdots S_{i_4,4} \ , \end{aligned}$$

and (A.7) for $r \leq 4$ reduces to (A.3)–(A.5).

Now suppose F consists of k d.f.s: the only change is to replace $(\mathbf{x}_1, \ldots, \mathbf{x}_r)$ by $\binom{a_1,\ldots,a_r}{\mathbf{x}_1,\ldots,\mathbf{x}_r}$ wherever it occurs. So, in the notation of (3.1), (A.3)–(A.5) imply

$$T(a^2) = g_{i,j}\, S_{i,j}(a,a) + g_i\, S_i(a^2) \ , \tag{A.8}$$

$$T(a^3) = g_{i,j,k}\, S_{i,j,k}(a,a,a) + 3\, g_{i,j}\, S_{i,j}(a,a^2) + g_i\, S_i(a^3) \ , \tag{A.9}$$

$$\text{(A.10)}\quad \begin{aligned} T(a^4) \;=\;& g_{i,j,k,l}\, S_{i,j,k,l}(a,a,a,a) + 6\, g_{i,j,k}\, S_{i,j,k}(a,a,a^2) \\ &+ g_{i,j}\left\{4\, S_{i,j}(a,a^3) + 3\, S_{i,j}(a^2,a^2)\right\} + g_i\, S_i(a^4), \end{aligned}$$

$$\text{(A.11)}\quad \begin{aligned} T(a^2,b^2) \;=\;& g_{i,j,k,l}\, S_{i,j}(a,a)\, S_{k,l}(b,b) \\ &+ g_{i,j,k}\left\{S_{i,j}(a,a)\, S_k(b^2) + S_{i,j}(b,b)\, S_k(a^2) + 4\, S_{i,j,k}(ab,a,b)\right\} \\ &+ g_{i,j}\left\{2\, S_{i,j}(a,ab^2) + 2\, S_{i,j}(b,a^2b) \right. \\ &\qquad \left. + S_i(a^2)\, S_j(b^2) + 2\, S_{i,j}(ab,ab)\right\} \\ &+ g_i\, S_i(a^2b^2)\ , \end{aligned}$$

where

$$\text{(A.12)}\qquad S_{i,j,\ldots}(a^I,a^J,\ldots) \;=\; \int S_{i_F}\binom{a^I}{x^I} S_{j_F}\binom{a^J}{x^J} \cdots\, dF_a(x)\ ,$$

$$\text{(A.13)}\qquad \binom{a^I}{x^I} = {}^{a,\ldots,a}_{x,\ldots,x} \qquad \text{with } I \text{ columns}\ ,$$

$$\text{(A.14)}\qquad \begin{aligned} &S_{i,j}(a^I,b^J,\ldots,a^K,b^L,\ldots) \;= \\ &\quad = \int\!\cdots\!\int S_{i,F}\left(\begin{matrix} a^I \\ x^I \end{matrix}, \begin{matrix} b^J \\ y^J \end{matrix}, \ldots\right) S_{j,F}\left(\begin{matrix} a^K \\ x^K \end{matrix}, \begin{matrix} b^L \\ y^L \end{matrix}, \ldots\right) dF_a(x)\, dF_b(y)\ , \end{aligned}$$

and so on. Similarly, from (A.7) at $r=5$ we obtain

$$\text{(A.15)}\qquad T(a^2,b^3) \;=\; \sum_{k=1}^{5} g_{i_1,\ldots,i_k}\, A^{i_1,\ldots,i_k}\ ,$$

where

$$\begin{aligned} A^i \;=\;& S_i(a^2b^3)\ , \\ A^{i,j} \;=\;& 2\, S_{i,j}(a,ab^3) + 3\, S_{i,j}(b,a^2b^2) + S_i(a^2)\, S_j(b^3) \\ &+ 6\, S_{i,j}(ab,ab^2) + 3\, S_{i,j}(b^2,a^2b)\ , \\ A^{i,j,k} \;=\;& S_{i,j}(a,a)\, S_k(b^3) + 3\, S_{i,j,k}(b,b,a^2b) \\ &+ 6\, S_{i,j,k}(a,b,ab^2) + 6\, S_{i,j,k}(a,ab,b^2) \\ &+ 3\, S_{i,k}(b,b^2)\, S_j(a^2) + 6\, S_{i,j,k}(b,ab,ab)\ , \\ A^{i,j,k,l} \;=\;& S_i(a^2)\, S_{j,k,l}(b,b,b) + 6\, S_{i,j,k,l}(ab,a,b,b) + 3\, S_{i,l}(b^2,b) S_{j,k}(a,a)\ , \\ A^{i_1,\ldots,i_5} \;=\;& S_{i_1,i_2}(a,a)\, S_{i_3,i_4,i_5}(b,b,b)\ , \end{aligned}$$

and from (A.7) at $r=6$ we obtain

$$\text{(A.16)}\qquad T(a^2,b^2,c^2) \;=\; \sum_{k=1}^{6} g_{i_1,\ldots,i_k}\, B^{i_1,\ldots,i_k}\ ,$$

where

$$
\begin{aligned}
B^i &= S_i(a^2b^2c^2) \ , \\
B^{i,j} &= B_1^{i,j} + B_2^{i,j} + B_3^{i,j} \ , \\
B_1^{i,j} &= 2 \sum^3 S_{i,j}(a, ab^2c^2) \ , \\
B_2^{i,j} &= \sum^3 S_i(a^2)\, S_j(b^2c^2) \ + \ 4 \sum^3 S_{i,j}(ab, abc^2) \ , \\
B_3^{i,j} &= 2 \sum^3 S_{i,j}(a^2b, bc^2) \ + \ 4\, S_{i,j}(abc, abc) \ ,
\end{aligned}
$$

$$
\begin{aligned}
B^{i,j,k} &= B_1^{i,j,k} + B_2^{i,j,k} + B_3^{i,j,k} \ , \\
B_1^{i,j,k} &= \sum^3 S_{i,j}(a,a)\, S_k(b^2c^2) \ + \ 4 \sum^3 S_{i,j,k}(a, b, abc^2) \ , \\
B_2^{i,j,k} &= 2 \sum^6 S_{i,k}(a, ac^2)\, S_j(b^2) \ + \ 4 \sum^6 S_{i,j,k}(a, ab, bc^2) \\
&\quad + 8 \sum^3 S_{i,j,k}(a, bc, abc) \ , \\
B_3^{i,j,k} &= S_i(a^2)\, S_j(b^2)\, S_k(c^2) + 2 \sum^3 S_i(a^2)\, S_{j,k}(bc, bc) + 8\, S_{i,j,k}(ab, bc, ca) \ ,
\end{aligned}
$$

$$
\begin{aligned}
B^{i,j,k,l} &= B_1^{i,j,k,l} + B_2^{i,j,k,l} \ , \\
B_1^{i,j,k,l} &= 2 \sum^6 S_{i,j}(a^2b, b)\, S_{k,l}(c,c) + 8\, S_{i,j,k,l}(abc, a, b, c) \ , \\
B_2^{i,j,k,l} &= \sum^3 \Big\{ S_{i,j}(a,a)\, S_k(b^2)\, S_l(c^2) + 2\, S_{i,j}(a,a)\, S_{k,l}(bc, bc) \\
&\quad + 4\, S_{i,j,k}(a, b, ab)\, S_l(c^2) + 8\, S_{i,j,k,l}(a, b, ac, bc) \Big\} ,
\end{aligned}
$$

$$
\begin{aligned}
B^{i_1,\ldots,i_5} &= \sum^3 \Big\{ S_{i_1}(a^2)\, S_{i_2,i_3}(b,b)\, S_{i_4,i_5}(c,c) + S_{i_1,i_2,i_3}(ab, a, b)\, S_{i_4,i_5}(c,c) \Big\} , \\
B^{i_1,\ldots,i_6} &= S_{i_1,i_2}(a,a)\, S_{i_3,i_4}(b,b)\, S_{i_5,i_6}(c,c) \ ,
\end{aligned}
$$

and $\sum^m$ is interpreted in the obvious manner by permuting a, b, c. For example,

$$\sum^3 S_{i,j}(a, ab^2c^2) \ = \ S_{i,j}(a, ab^2c^2) + S_{i,j}(b, bc^2a^2) + S_{i,j}(c, ca^2b^2) \ .$$

Similarly, if we now allow $\mathbf{T}$ and $\mathbf{g}$ to be r-vectors with components $\{T^\alpha\}$ and $\{g^\alpha\}$, then by (A.3), $T^{\alpha,\beta}(a,a)$ of (6.2) is given by

(A.17) $$T^{\alpha,\beta,\cdots}(a, a, \ldots) \ = \ g_i^\alpha g_i^\beta \cdots \ S_{i,j,\ldots}(a, a, \ldots)$$

and $T^{\alpha,\beta}(ab, ab)$ of (6.6) satisfies

(A.18) $$\begin{aligned} T^{\alpha,\beta}(ab, ab) \ &= \ g_{i,j}^\alpha\, g_{k,l}^\beta\, S_{i,k}(a,a)\, S_{j,l}(b,b) + \sum_{\alpha,\beta}^2 g_i^\alpha\, g_{j,k}^\beta\, S_{i,j,k}(ab, a, b) \\ &\quad + g_i^\alpha\, g_j^\beta\, S_{i,j}(ab, ab) \ , \end{aligned}$$

where

$$S_{i,j,k}(ab,a,b) \;=\; \iint S_{i,F}\binom{a,b}{x,y}\, S_{j,F}\binom{a}{x}\, S_{k,F}\binom{b}{y}\, dF_a(x)\, dF_b(y) \ .$$

Similarly, (6.4), (6.5) yield

$$T^{\alpha,\beta}(a^2,a) \;=\; \Big\{ g^{\alpha}_{i,j}\, S_{i,j,k}(a,a,a) + g^{\alpha}_{i}\, S_{i,k}(a^2,a) \Big\}\, g^{\beta}_{k} \ , \tag{A.19}$$

and

$$\begin{aligned} T^{\alpha,\beta}(a^2b,b) \;=\; &\Big\{ g^{\alpha}_{i,j,k}\, S_{i,j}(a,a)\, S_{k,l}(b,b) + g^{\alpha}_{i,j}\, \big[S_i(a^2)\, S_{j,l}(b,b) + 2\, S_{i,j,l}(ab,a,b)\big] \\ &+ g^{\alpha}_{i}\, S_{i,l}(a^2b,b) \Big\}\, g^{\beta}_{l} \ . \end{aligned} \tag{A.20}$$

Similarly,

$$T^{\alpha,\beta,\delta}(ab,a,b) \;=\; \Big\{ g^{\alpha}_{i,j}\, S_{i,j,k,l}(a,b,a,b) + g^{\alpha}_{i}\, S_{i,k,l}(ab,a,b) \Big\}\, g^{\beta}_{k}\, g^{\delta}_{l} \ .$$

We now consider the case, where $\mathbf{S}(F)$ is bivariate, that is $q=2$. Since $S_{i,j}(a^I, a^J) = S_{j,i}(a^J, a^I)$, (A.8)–(A.11) can be written as

$$T(a^2) \;=\; \Big\{ g_{1,1}\, S_{1,1} + 2\, g_{1,2}\, S_{1,2} + g_{2,2}\, S_{2,2} \Big\}(a,a) \;+\; \Big\{ g_1 S_1 + g_2\, S_2 \Big\}(a^2)\, , \tag{A.21}$$

$$\begin{aligned} T(a^3) \;=\; &\Big\{ g_{1,1,1}\, S_{1,1,1} + 3\, g_{1,1,2}\, S_{1,1,2} + 3\, g_{1,2,2}\, S_{1,2,2} + g_{2,2,2}\, S_{2,2,2} \Big\}(a,a,a) \\ &+ 3 \Big\{ g_{1,1}\, S_{1,1} + g_{1,2}\big(S_{1,2} + S_{2,1}\big) + g_{2,2}\, S_{2,2} \Big\}\, (a,a^2) \\ &+ \Big\{ g_1 S_1 + g_2\, S_2 \Big\}\, (a^3) \ , \end{aligned} \tag{A.22}$$

$$\begin{aligned} T(a^4) \;=\; &\Big\{ g_{1,1,1,1}\, S_{1,1,1,1} + 4\, g_{1,1,1,2}\, S_{1,1,1,2} + 6\, g_{1,1,2,2}\, S_{1,1,2,2} \\ &\quad + 4\, g_{1,2,2,2}\, S_{1,2,2,2} + g_{2,2,2,2}\, S_{2,2,2,2} \Big\}\, (a,a,a,a) \\ &+ 6 \Big\{ g_{1,1,1}\, S_{1,1,1} + g_{1,1,2}\, S_{1,1,2} + 2\, g_{1,2,1}\, S_{1,2,1} \\ &\quad + g_{2,2,1}\, S_{2,2,1} + 2\, g_{1,2,2}\, S_{1,2,2} + g_{2,2,2}\, S_{2,2,2} \Big\}\, (a,a,a^2) \\ &+ 4 \Big\{ g_{1,1}\, S_{1,1} + g_{1,2}\big(S_{1,2} + S_{2,1}\big) + g_{2,2}\, S_{2,2} \Big\}\, (a,a^3) \\ &+ 3 \Big\{ g_{1,1}\, S_{1,1} + 2\, g_{1,2}\, S_{1,2} + g_{2,2}\, S_{2,2} \Big\}\, (a^2,a^2) \\ &+ \Big\{ g_1 S_1 + g_2\, S_2 \Big\}\, (a^4) \ , \end{aligned} \tag{A.23}$$

$$
\begin{aligned}
T(a^2,b^2) = {} & \Big\{g_{1,1,1,1}\,S_{1,1}\,S_{1,1} + 2\,g_{1,1,1,2}\,S_{1,1}\,S_{1,2} + g_{1,1,2,2}\,S_{1,1}\,S_{2,2} \\
& + 2\,g_{1,2,1,1}\,S_{1,2}\,S_{1,1} + 4\,g_{1,2,1,2}\,S_{1,2}\,S_{1,2} + 2\,g_{1,2,2,2}\,S_{1,2}\,S_{2,2} \\
& + g_{2,2,1,1}\,S_{2,2}\,S_{1,1} + 2\,g_{2,2,1,2}\,S_{2,2}\,S_{1,2} + g_{2,2,2,2}\,S_{2,2}\,S_{2,2}\Big\}\,(a,a)\,(b,b) \\
& + \Big\{g_{1,1,1}\,S_{1,1}\,S_1 + 2\,g_{1,2,1}\,S_{1,2}\,S_1 + g_{2,2,1}\,S_{2,2}\,S_1 + g_{1,1,2}\,S_{1,1}\,S_2 \\
& + 2\,g_{1,2,2}\,S_{1,2}\,S_2 + g_{2,2,2}\,S_{2,2}\,S_2\Big\}\Big\{(a,a)(b^2) + (b,b)(a^2)\Big\} \\
& + 4\Big\{g_{1,1,1}\,S_{1,1,1} + 3\,g_{1,1,2}\,S_{1,1,2} + 3\,g_{1,2,2}\,S_{1,2,2} + g_{2,2,2}\,S_{2,2,2}\Big\}(ab,a,b) \\
& + 2\Big\{g_{1,1}\,S_{1,1} + g_{1,2}\big(S_{1,2} + S_{2,1}\big) + g_{2,2}\,S_{2,2}\Big\}\Big\{(a,ab^2) + (b,a^2b)\Big\} \\
& + \Big\{g_{1,1}\,S_1 S_1 + g_{1,2}\big(S_1 S_2 + S_2\,S_1\big) + g_{2,2}\,S_2\,S_2\Big\}\,(a^2)(b^2) \\
& + 2\Big\{g_{1,1}\,S_{1,1} + 2\,g_{1,2}\,S_{1,2} + g_{2,2}\,S_{2,2}\Big\}\,(ab,ab) \\
& + \Big\{g_1 S_1 + g_2\,S_2\Big\}\,(a^2b^2)\;.
\end{aligned}
\tag{A.24}
$$

The convention here is that

$$
\begin{aligned}
\big(g_{\pi_1} S_{\pi_2} + \cdots\big)\,\big(a^I, ...\big) &= g_{\pi_1} S_{\pi_2}\big(a^I, ...\big)\;, \\
\big(g_{\pi_1} S_{\pi_2}\,S_{\pi_3} + \cdots\big)\,\big(a^I, ...\big)\,\big(b^J, ...\big) &= g_{\pi_1} S_{\pi_2}\big(a^I, ...\big)\,S_{\pi_3}\big(b^J, ...\big)\;.
\end{aligned}
$$

Similarly, for $q = 2$, splitting the third term in (A.15), $g_{i,j,k}\,A^{i,j,k}$, into the six components corresponding to $A^{i,j,k}$, the first is

$$
g_{i,j,k}\,S_{i,j,k} = \Big\{g_{1,1,k}\,S_{1,1,k} + 2\,g_{1,2,k}\,S_{1,2,k} + g_{2,2,k}\,S_{2,2,k}\Big\}
$$

at (a, a, b^3) and similarly for the second and sixth components. Similarly, for the three components of the fourth term, the first being

$$
g_{i,...,l}\,S_{i,...,l} = \left\{\sum_{j=1}^{2} g_{i,j,j,j}\,S_{i,j,j,j} + 3\,g_{i,1,1,2}\,S_{i,1,1,2} + g_{i,1,2,2}\,S_{i,1,2,2}\right\}
$$

at (a^2, b, b, b), and for the fifth term

$$
\begin{aligned}
g_{i_1,...,i_5}\,S_{i_1,...,i_5} &= \\
&= \big(g_{1,1-}\,S_{1,1-} + 2\,g_{1,2-}\,S_{1,2-} + g_{2,2-}\,S_{2,2-}\big) \\
&\quad \times \big(g_{-1,1,1}\,S_{-1,1,1} + 3\,g_{-1,1,2}\,S_{-1,1,2} + 3\,g_{-1,2,2}\,S_{-1,2,2} + g_{-2,2,2}\,S_{-2,2,2}\big)
\end{aligned}
$$

at (a, a, b, b, b), where $g_{\pi-}\,S_{\pi-}\,g_{-\pi'}\,S_{-\pi'}$ is interpreted as $g_{\pi,\pi'} S_{\pi,\pi'}$.

Similarly, for $q = 2$, the term B_3^i in (A.16) has the component

$$
4\,g_{i,j}\,S_{i,j} = 4\sum_{i=1}^{2} g_{i,i}\,S_{i,i} + 8\,g_{1,2}\,S_{1,2}
$$

at (abc, abc). The sixth component is

$$\left(g_{1,1-}\,S_{1,1-} + 2\,g_{1,2-}\,S_{1,2-} + g_{2,2-}\,S_{2,2-}\right) \times \\ \times \left(g_{-1,1-}\,S_{-1,1-} + 2\,g_{-1,2-}\,S_{-1,2-} + g_{-2,2-}\,S_{-2,2-}\right) \times \\ \times \left(g_{-1,1}\,S_{-1,1} + 2\,g_{-1,2}\,S_{-1,2} + g_{-2,2}\,S_{-2,2}\right)$$

at (a, a, b, b, c, c), where $g_{\pi_1-}\,S_{\pi_1-}\,g_{-\pi_2-}\,S_{-\pi_2-}\,g_{-\pi_3}\,S_{-\pi_3}$ interpreted as g_{π_1,π_2,π_3} S_{π_1,π_2,π_3}, and so on.

APPENDIX B

The nonparametric analogs of the terms for t_2 and equation (D.1) of Withers [27] needed for T_2 and T_3 — apart from those given in (3.3)–(3.5) are as follows. Summation over a, b, c is implicit, where they occur. These terms are listed both for the purpose of checking and for application to other problems. Note that T_2 requires

$$\begin{vmatrix} 22 \\ 10 \end{vmatrix} = |3| \qquad \text{and} \qquad \begin{vmatrix} 22 \\ 20 \end{vmatrix} = -2\,\lambda_a^2\,|2|_a$$

and that T_3 requires

$$\begin{vmatrix} 23 \\ 10 \end{vmatrix} = \begin{vmatrix} 222 \\ 110 \end{vmatrix} = \lambda_a^3 \left\{ T(a^4) - T(a^2, a^2) \right\},$$

$$\begin{vmatrix} 23 \\ 20 \end{vmatrix} = -2\,\lambda_a^3\,T(a^3)\ ,$$

$$\begin{vmatrix} 2\ 2\ 2 \\ \\ 1\ 0\ 0 \end{vmatrix} = |23|\ , \qquad \begin{vmatrix} 2\ 2\ 2 \\ \\ 2\ 0\ 0 \end{vmatrix}_2 = -2\,\lambda_a^2 \lambda_b\,T(a^2, b^2)\ ,$$

$$\begin{vmatrix} 2\ 2\ 2 \\ \\ 0\ 2\ 0 \end{vmatrix}_1 = -2\,\lambda_a^3\,T(a^2, a^2)\,, \qquad \begin{vmatrix} 2\ 2\ 2 \\ 1\ 2\ 0 \end{vmatrix}_i = \begin{vmatrix} 2\ 2\ 2 \\ \\ 2\ 1\ 0 \end{vmatrix} - 2\,\lambda_a^3\,T(a^3) \quad \text{for } 1 \le i \le 3\,,$$

$$\begin{vmatrix} 2\ 2\ 2 \\ \\ 2\ 2\ 0 \end{vmatrix} = 4\,\lambda_a^3\,T(a^2)\ , \qquad \begin{vmatrix} 32 \\ 10 \end{vmatrix} = \lambda_a^3 \left\{ T(a^4) - 3\,T(a^2, a^2) \right\},$$

$$\begin{vmatrix} 32 \\ 20 \end{vmatrix} = -6\,|3|\ .$$

Also,

$$\begin{vmatrix} 23 \\ 30 \end{vmatrix} = \begin{vmatrix} 2\ 2\ 2 \\ \\ 0\ 3\ 0 \end{vmatrix} = \begin{vmatrix} 2\ 2\ 2 \\ \\ 0\ 4\ 0 \end{vmatrix} = 0$$

since $\kappa_a(x_1,x_2)$, being quadratic in F_a, has functional derivatives higher than two equal to zero. To illustrate the proof,

$$\left|\begin{matrix}222\\210\end{matrix}\right|_1 = \kappa^{x_1,x_2}_{y_1,z_1}\,\kappa^{y_1,y_2}_{z_2}\,\kappa^{z_1,z_2}\,t_{x_1,x_2,y_2}$$
$$= \int^6 \lambda_a\lambda_b\lambda_c\; d_{\mathbf{x}}U_F\left({}^{b,c}_{y_1,z_1}\right) d_{\mathbf{y}}V_F\left({}^{c}_{z_2}\right) d\kappa_c(z_1,z_2)T_F\left({}^{a,a,b}_{x_1,x_2,y_2}\right),$$

where $U(F)=\kappa^{x_1,x_2}=\kappa_a(x_1,x_2)$ and $V(F)=\kappa^{y_1,y_2}=\kappa_b(y_1,y_2)$. Note that

$$V_F\left({}^{c}_{z_2}\right) = 0$$

unless $c=b$ and

$$U_F\left({}^{b,c}_{y_1,z_1}\right) = 0$$

unless $b=c=a$. Also

$$U_F\left({}^{a,a}_{y_1,z_1}\right) = -\sum_{x_1,x_2}^{2}\triangle_{y_1}(x_1)\,\triangle_{z_1}(x_2)\,,$$

and

$$V_F\left({}^{a}_{z}\right) = \triangle_z(y_1\wedge y_2) - \sum_{y_1,y_2}^{2}\triangle_z(y_1)\,F_a(y_2)\,,$$

where $\triangle_y(x)=(F_a(x))_y=I(y\le x)-F_a(x)$. Integrate first with respect to $\mathbf{x}=(x_1,x_2)$: since columns in $T_F\left({}^{\cdot,\cdot,\cdot}_{\cdot,\cdot,\cdot}\right)$ are interchangeable we may replace $\sum^2_{x_1,x_2}$ by 2. Since

$$\text{(B.1)}\qquad \int T_F\left({}^{a,a,a}_{x_1,x_2,y_2}\right) dF_a(x_i) = 0$$

for $i=1,2$, and

$$d_{\mathbf{x}}\Big\{I(y_1\le x_1)\,I(z_1\le x_2)\Big\} = \delta(x_1-y_1)\,\delta(x_2-z_1)\,dx_1\,dx_2$$

with δ the Dirac delta function,

$$\int^2 d_{\mathbf{x}}\,U_F\left({}^{a,a}_{y_1,z_1}\right)T_F\left({}^{a,a,a}_{x_1,x_2,y_2}\right) = -2\,T_F\left({}^{a,a,a}_{y_1,z_1,y_2}\right).$$

So,

$$\left|\begin{matrix}222\\210\end{matrix}\right|_1 = -2\,\lambda_a^3\int^4 d\kappa_a(z_1,z_2)\,T_F\left({}^{a,a,a}_{y_1,y_2,z_1}\right)d_{\mathbf{y}}V_F\left({}^{a}_{z_2}\right).$$

Integrate with respect to $\mathbf{y}=(y_1,y_2)$: (B.1) implies the contribution from the last two out of the three terms in $V_F\left({}^{a}_{z}\right)$ is zero. Also,

$$\triangle_z(y_1\wedge y_2) = I(z\le y_1)\,I(z\le y_2) - F_a(y_1\wedge y_2)\,,$$

so

$$d_{\mathbf{y}} \triangle_z(y_1 \wedge y_2) \;=\; \delta(y_1 - z)\,\delta(y_2 - z)\,dy_1\,dy_2 - \delta(y_1 - y_2)\,dy_2\,dF_a(y_2)\ .$$

So,

$$\int^2 T_F\Big(\,{}^{a,a,a}_{y_1,y_2,z_1}\Big)\,d_{\mathbf{y}}V_F\Big(\,{}^{a}_{z_2}\Big) \;=\; T_F\Big(\,{}^{a,a,a}_{z_2,z_2,z_1}\Big) - \int T_F\Big(\,{}^{a,a,a}_{y_1,y_1,z_1}\Big)\,dF_a(y_1)\ .$$

Now integrate with respect to $\mathbf{z} = (z_1, z_2)$: by (B.1) the second out of two terms from $d\kappa_a(z_1, z_2)$ contributes zero. So, putting

$$L \;=\; \int dF_a(z_2)\,T_F\Big(\,{}^{a,a,a}_{y_1,y_1,z_2}\Big) \;=\; 0\ ,$$

we obtain

$$\begin{aligned}\left|\begin{matrix}222\\210\end{matrix}\right|_1 &= -2\,\lambda_a^3 \int^2 dF_a(z_1 \wedge z_2)\,\Big\{T_F\Big(\,{}^{a,a,a}_{z_2,z_2,z_1}\Big) - \int T_F\Big(\,{}^{a,a,a}_{y_1,y_1,z_1}\Big)\,dF_a(y_1)\Big\}\\ &= -2\,\lambda_a^3\Big\{\int T_F\Big(\,{}^{a,a,a}_{z,z,z}\Big)\,dF_a(z) - \int dF_a(y_1)\,L\Big\} \;=\; -2\,\lambda_a^3\,T(a^3)\ .\end{aligned}$$

APPENDIX C

Here, we show how to estimate N, the number of simulated samples needed to estimate the bias to within a given relative error ϵ.

Note that $T_{n,p}(\widehat{F})$ has bias $-n^{-p}\,T_p(F) + O(n^{-p-1})$ and that $S_{n,p}(\widehat{F})$ has bias $-(n-1)_p^{-1}S_p(F) + O(n^{-p-1}) = -n^{-p}\,S_p(F) + O(n^{-p-1})$. Suppose we estimate the bias of $Y = S_{n,p}(\widehat{F})$ by $Z = \overline{Y} - T(F)$, where $\overline{Y} = N^{-1}\sum_{j=1}^N Y_j$, $Y_j = S_{n,p}(\widehat{F}_j)$ and $\widehat{F}_j$ is the empirical d.f. of the j^{th} simulated sample. Then $EZ = ES_{n,p}(\widehat{F}) - T(F)$ is the true bias of Y and we can write $Z = EZ + (v_n/N)^{1/2}\,\{\mathcal{N}(0,1) + o_p(1))\}$ as $N \to \infty$, where $v_n = \operatorname{var} Y_1 = V_T\,n^{-1} + O(n^{-2})$ as $n \to \infty$, and $V_T = V_T(F) = \sum \lambda_a T(a,a)$ with $T(a,a) = \int T_F(\,{}^a_x)^2\,dF_a(x)$. So, if $S_p = S_p(F) \neq 0$, the relative error in the estimate of bias,

$$\begin{aligned}\text{(bias estimate} - \text{bias)/bias} \;&\approx\; -(v_n/N)^{1/2}\mathcal{N}(0,1)\,n^p\,S_p(F)\\ &\approx\; -V_T(F)^{1/2}\,S_p(F)^{-1}\,n^{p-1/2}N^{-1/2}\mathcal{N}(0,1)\end{aligned}$$

is bounded by a given number ϵ with probability greater than $0.975 + O_p(n^{-1/2})$ if

$$2\,V_T(F)^{1/2}\,S_p(F)^{-1}\,n^{p-1/2}N^{-1/2} \;\leq\; \epsilon\ ,$$

that is, if

$$N \geq N_{\epsilon,p,n} = \epsilon^{-2} n^{2p-1} \phi_p ,$$

where $\phi_p = 4\, V_T(F) S_p(F)^{-2}$. This implies that for $\epsilon = 0.1$ and n large, say $n = 100$, it is not practical to carry out enough simulations to give meaningful estimates of bias unless $p = 1$. This is reflected by the poor estimates of bias in the tables for the case $p = 2$ obtained for $n = 100$ using $N = 10{,}000$.

Consider the following one sample examples. Set $\beta_r = \mu_r \mu_2^{-2/2}$. For $F = \mathcal{N}(0,1)$, $\mu_4 = 3$, $\mu_6 = 15$, $\mu_8 = 105$ and for $F = \exp(1)$, $\mu_2 = 1$, $\mu_3 = 2$, $\mu_4 = 9$, $\mu_5 = 44$, $\mu_6 = 305$, $\mu_8 = 14{,}833$.

Example C.1. Consider $T(F) = \mu_2$. Then $V_T = \mu_4 - \mu_2^2$, $S_1 = \mu_2$, $\phi_1 = 4(\beta_4 - 1)$. So, for a normal sample $\phi_1 = 8$ and $\widehat{\mu}_2 = \mu_2(\widehat{F})$ needs

$$N \geq N_{\epsilon,1,n} = 8\,\epsilon^{-2}\, n = \begin{cases} 80{,}000\, n \text{ simulations} & \text{for } \epsilon = 0.01\,, \\ 800\, n \text{ simulations} & \text{for } \epsilon = 0.1\,. \end{cases}$$

For an exponential sample $\phi_1 = 32$, so one needs four times as many simulations. Since $S_2(F) = 0$, ϕ_2 is not defined.

Example C.2. Consider $T(F) = \mu_2^2$. Then $V_T = 4\,\mu_2^2(\mu_4 - \mu_2^2)$ and by Example 5.8, $S_1 = -\mu_4 + \mu_2^2$, $S_2 = -4\,\mu_4 + 7\mu_2^2$ so for a unit normal, $V_T = 8$, $S_1 = -2$, $\phi_1 = 8$, $S_2 = -29$, $\phi_2 = 0.1522$ so $N_{0.1,1,n} = 800\, n$ and $N_{0.1,2,n} = 152\, n^3$ and for exp(1), $V_T = 14{,}048$, $S_1 = 30$, $\phi_1 = 62.44$, $S_2 = 87$, $\phi_2 = 7.424$, so $N_{0.1,1,n} = 6{,}244\, n$ and $N_{0.1,2,n} = 74.24\, n^3$.

Example C.3. Consider $T(F) = \mu_4$. Then $V_T = \mu_8 - \mu_4^2 - 8\,\mu_5\mu_3$, and by Example 5.6 or 5.10, $S_1 = 2(2\,\mu_4 - 3\,\mu_2^2)$, $S_2 = 3(4\,\mu_4 - 7\mu_2^2)$, so for a unit normal, $V_T = 96$, $S_1 = 6$, $\phi_1 = 32/3$, $S_2 = 15$, $\phi_2 = 128/75$, so $N_{0.1,1,n} = 1067\, n$ and $N_{0.1,2,n} = 171\, n^3$ and for exp(1), $V_T = 14{,}048$, $S_1 = 30$, $\phi_1 = 62.44$, $S_2 = 87$, $\phi_2 = 7.424$, so $N_{0.1,1,n} = 6{,}244\, n$ and $N_{0.1,2,n} = 74.24\, n^3$.

Example C.4. Consider $T(F) = \sigma = \mu_2^{1/2}$. Then $V_T = \mu_2(\beta_4 - 1)/4$, so by Example 5.15, for a unit normal, $V_T = 1/2$, $S_1 = 3/4$, $\phi_1 = 32/9$, $S_2 = 1/32$, $\phi_2 = 2048$, so $N_{0.1,1,n} = 356\, n$ and $N_{0.1,2,n} = 204{,}800\, n^3$ and for exp(1), $V_T = 2$, $S_1 = 3/2$, $\phi_1 = 32/9$, $S_2 = 213/8 = 26.625$, $\phi_2 = 0.01129$, so $N_{0.1,1,n} = 356\, n$ and $N_{0.1,2,n} = 1.129\, n^3$.

APPENDIX D

Here, we list the non-zero derivatives $\mu_{r\cdot 1,2,\ldots,p} = \mu_{r,F}(x_1,\ldots,x_p)$ for $2 \le p \le r \le 6$. They are obtained from (5.4) in terms of $h_i = \mu_{x_i}$, where $\mu_x = x - \mu$, the first derivative of μ:

$$\mu_{2\cdot 1} = h_1^2 - \mu_2 \ ,$$
$$\mu_{2\cdot 1,2} = -2\,h_1 h_2 \ ,$$

$$\mu_{3\cdot 1} = h_1^3 - \mu_3 - 3\,h_1\mu_2 \ ,$$
$$\mu_{3\cdot 1,2} = -3(h_1^2 - \mu_2)h_2 - 3\,h_1(h_2^2 - \mu_2) \ ,$$
$$\mu_{3\cdot 1,2,3} = 12\,h_1 h_2 h_3 \ ,$$

$$\mu_{4\cdot 1} = h_1^4 - \mu_4 - 4\,h_1\mu_3 \ ,$$
$$\mu_{4\cdot 1,2} = 12\,h_1h_2\mu_2 - 4(h_1^3 - \mu_3)h_2 - 4\,h_1(h_2^3 - \mu_3) \ ,$$
$$\mu_{4\cdot 1,2,3} = 12(h_1^2 - \mu_2)h_2h_3 + 12\,h_1(h_2^2 - \mu_2)h_3 + 12\,h_1h_2(h_3^2 - \mu_2) \ ,$$
$$\mu_{4\cdot 1,2,3,4} = -72\,h_1h_2h_3h_4 \ ,$$

$$\mu_{5\cdot 1} = h_1^5 - \mu_5 - 5\,h_1\mu_4 \ ,$$
$$\mu_{5\cdot 1,2} = 20h_1h_2\mu_3 - 5(h_1^4 - \mu_4)h_2 - 5\,h_1(h_2^4 - \mu_4) \ ,$$
$$\mu_{5\cdot 1,2,3} = -60h_1h_2h_3\mu_2 + 20(h_1^3 - \mu_3)h_2h_3 + 20h_1(h_2^3 - \mu_3)h_3 + 20h_1h_2(h_3^3 - \mu_3) \ ,$$
$$\mu_{5\cdot 1,2,3,4} = -60(h_1^2 - \mu_2)h_2h_3h_4 - 60h_1(h_2^2 - \mu_2)h_3h_4 - 60h_1h_2(h_3^2 - \mu_2)h_4 - 60h_1h_2h_3(h_4^2 - \mu_2) \ ,$$
$$\mu_{5\cdot 1,2,3,4,5} = 480h_1h_2h_3h_4h_5 \ ,$$

$$\mu_{6\cdot 1} = h_1^6 - \mu_6 - 6\,h_1\mu_5 \ ,$$
$$\mu_{6\cdot 1,2} = 30h_1h_2\mu_4 - 6(h_1^5 - \mu_5)h_2 - 6\,h_1(h_2^5 - \mu_5) \ ,$$
$$\mu_{6\cdot 1,2,3} = -120h_1h_2h_3\mu_3 + 30(h_1^4 - \mu_4)h_2h_3 + 30h_1(h_2^4 - \mu_4)h_3 + 30h_1h_2(h_3^4 - \mu_4) \ ,$$
$$\mu_{6\cdot 1,2,3,4}/120 = 3h_1h_2h_3h_4\mu_2 - (h_1^3 - \mu_3)h_2h_3h_4 - h_1(h_2^3 - \mu_3)h_3h_4 - h_1h_2(h_3^3 - \mu_3)h_4 - h_1h_2h_3(h_4^3 - \mu_3) \ ,$$
$$\mu_{6\cdot 1,2,3,4,5}/360 = (h_1^2 - \mu_2)h_2h_3h_4h_5 + h_1(h_2^2 - \mu_2)h_3h_4h_5 + h_1h_2(h_3^2 - \mu_2)h_4h_5 + h_1h_2h_3(h_4^2 - \mu_2)h_5 + h_1h_2h_3h_4(h_5^2 - \mu_2) \ .$$

Note that

$$\mu_{r\cdot 1,2,\ldots,r} = (-1)^{r-1}(r-1)\,r!\prod_{j=1}^{r} h_j \ ,$$

and

$$\mu_{r\cdot 1,2,\ldots,r-1} = (-1)^r\,(r!/2)\sum^{r-1}(h_1^2 - \mu_2)\,h_2\cdots h_{r-1} \ ,$$

where $\sum^{r-1}$ sums over all $r-1$ like terms.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the referee for careful reading and for their comments which improved the paper.

REFERENCES

[1] Anderson, S.; Auquier, A.; Hauck, W.W.; Oakes, D.; Vandaele, W. and Weisberg, H.I. (1980). *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*, John Wiley and Sons, New York.

[2] Beirlant, J.; Goegebeur, Y.; Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*, John Wiley and Sons, Chichester.

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

[4] 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.

[5] Comtet, L. (1974). *Advanced Combinatorics*, Reidel, Dordrecht.

[6] Davison, A.C.; Hinkley, D.V. and Schechtman, E. (1986). Efficient bootstrap simulation, *Biometrika*, **73**, 555–566.

[7] Efron, B. (1981). Nonparametric estimates of standard errors: The jackknife, the bootstrap and other methods, *Biometrika*, **68**, 589–599.

[8] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia.

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

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

[11] Gray, H.L. and Schucany, W.R. (1972). *The Generalized Jackknife Statistic*, Marcel Dekker, New York.

[12] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer Verlag, New York.

[13] Jaeckel, L.A. (1972). The infinitesimal jackknife, *Bell Laboratories Technical Report MM 72-1215-11*, June 30, 1972, New Jersey.

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

[15] Kendall, M.G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, volume 1, fourth edition, Griffin, London.

[16] QUENOUILLE, M.H. (1956). Notes on bias in estimation, *Biometrika*, **43**, 353–360.

[17] SCHUCANY, W.R.; GRAY, H. and OWEN, D.B. (1971). On bias reduction in estimation, *Journal of the American Statistical Association*, **66**, 524–533.

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

[19] STUART, A. and ORD, J.K. (1987). *Kendall's Advanced Theory of Statistics*, volume 1, fifth edition, Griffin, London.

[20] SUKHATME, P.V. (1944). Moments and product moments of moment-statistics for samples of the finite and infinite populations, *Sanhkyā*, **6**, 363–382.

[21] TIIT, E. (1988). *Unbiased and n^{-k}-estimations of entire rational functions of moments*. In: "Statistical and Probabilistic Models", Report 798, Tartu State University, Estonia, pp. 3–17.

[22] VON MISES, R. (1947). Differentiable statistical functions, *Annals of Mathematical Statistics*, **18**, 309–348.

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

[24] WITHERS, C.S. (1982). The distribution and quantiles of a function of parameter estimates, *Annals of the Institute of Statistical Mathematics*, A, **34**, 55–68.

[25] 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.

[26] WITHERS, C.S. (1984). A chain rule for differentiation with applications to multivariate Hermite polynomials, *Bulletin of the Australian Mathematical Society*, **30**, 247–250.

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

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

[29] WITHERS, C.S. and NADARAJAH, S. (2008). Analytic bias reduction for k sample functionals, *Sankhyā*, A, **70**, 186–222.

[30] ZACKS, S. (1981). *Parametric Statistical Inference: Basic Theory and Modern Approaches*, Pergamon Press, New York.