The chain rule for functionals
with applications to functions of moments

by

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

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

Abstract: The chain rule for derivatives of a function of a function is extended to a function of a statistical functional, and applied to obtain approximations to the cumulants, distribution and quantiles of functions of sample moments, and so to obtain third order confidence intervals and estimates of reduced bias for functions of moments. As an example we give the distribution of the standardized skewness for a normal sample to magnitude $O(n^{-2})$, where $n$ is the sample size.

AMS 2000 subject classification: Primary 62E20; Secondary 62G30.

Keywords and phrases: Asymptotic expansions; Bias; Chain rule; Confidence intervals; Cumulants; Moments; Nonparametric.

1 Introduction

The derivatives introduced by von Mises (1947) and their subsequent versions have wide ranging applications in statistics. Two prominent application areas are the construction of nonparametric confidence intervals and analytic bias reduction.

Suppose we want to construct a nonparametric confidence interval of level $\alpha + O(n^{-3/2})$ for a smooth functional $T(F)$ based on $\hat{F}$ say, the sample or empirical distribution for a sample of size $n$ from $F$. Withers (1983) showed that the limit can be given in terms of integrals of products of von Mises derivatives evaluated at $\hat{F}$. First one Studentizes using the asymptotic variance of $n^{1/2}\{T(\hat{F}) - T(F)\}$, $a_{21}(F) = [1^2]_T = \int_{-\infty}^{\infty} T_F(x_1)^2 dF(x_1)$, where $T_F(x)$ is the first derivative or influence function of $T(F)$.

Similarly, it is known that for smooth $T(F)$, an estimate of $T(F)$ of bias $O(n^{-2})$ is $T(\hat{F}) - n^{-1}a_{11}(\hat{F})$, where $a_{11}(F) = [11]_T = \int_{-\infty}^{\infty} T_F(x_1, x_1) dF(x_1)$ and $T_F(x_1, x_2)$ is the second derivative of $T(F)$. For convenience we refer to these integrals of products of derivatives like $[1^2]_T$ and $[11]_T$ as bracket functions.

Other recent application areas of von Mises derivatives include: least squares support vector regression filtering methods, bootstrapping, functional principal components analysis, linearization and composite estimation, dimension reduction, quantile regressions,
machine learning, cusum statistics, methods of sieves and penalization, change point estimation, Hadamard differentiability, change-of-variance function, measuring and testing dependence by correlation of distances, empirical finite-time ruin probabilities (Loisel et al., 2009), nonparametric maximum likelihood estimators (Nickl, 2007), estimating mean dimensionality of analysis of variance decompositions, monotonicity of information in the Central Limit Theorem, generalizations of the Anderson-Darling statistic, $M$-estimation, $U$-statistics (Volodko, 2011), information criteria in model selection, goodness-of-fit tests for kernel regression, empirical Bayes estimation, and estimation of Kendall's tau.

The aim of this paper is to develop tools for extending the use of von Mises derivatives. In Section 2, we extend Faa di Bruno's chain rule for the derivative of a function of a univariate function to functions of a multivariate function and show how it can be applied to a function of a function of $F$, say $T(F) = g(S(F))$, where $g : \mathbb{R}^a \to \mathbb{R}$ is a smooth function and $S(F)$ a smooth functional.

In Section 3, we apply it to obtain derivatives and bracket functions for powers, products, quotients, standardized and Studentized functionals.

Section 4 gives the general derivative for a moment and applies previous results to obtain expansions up to $O(n^{-2})$ for the distribution and quantiles of functions of sample moments. As an example we give the distribution of the standardized skewness for a normal sample to magnitude $O(n^{-2})$, where $n$ is the sample size. Also we give confidence intervals and bias reduction methods for functions of moments.

Some of the results in the paper follow easily from Withers (1983, 1987), see Theorems 3.1 to 3.3. But these results are not the main contributions of this paper. The main contributions are: 1) the tools developed to compute von Mises type derivatives, see Theorems 2.1 and 2.2; 2) their applications to obtain bracket functions for general functionals, see Examples 3.1 to 3.4 and Appendix A. The functionals considered by these examples include $T(F) = g(S(F))$, where $g$ is a univariate function, $T(F) = S_1(F)S_2(F)$, a product of two functionals, Studentized forms of $T(F)$ and $T(F) = U(F)g(S(F))$, where $S(F)$ is real valued; 3) also the applications of Theorems 2.1 and 2.2 to obtain derivatives of central moments, see Theorem 4.1, Corollary 4.1, Corollary 4.2 and Appendix B.

Fisher and Wishart gave unbiased estimates only for cumulants and their products: see, for example, Stuart and Ord (1987). Our two methods for bias reduction apply to any smooth functional - and our second estimate reduces to their results for the cases they consider. Also our method does not need to use unbiased estimates of cumulants to reduce the bias of functions of cumulants.

Analogous to Fisher's tables for his $k$-statistics and their cumulants, Appendix B gives the terms needed for bias reduction of any smooth function of one or more moments.

2 Chain rules for functions and functionals

Let $s$ and $g$ be real functions on $\mathbb{R}$ with finite derivatives. Comtet (1974, page 137) gives Faa di Bruno's chain rule for the $r$th derivative of

$$t(x) = g(s(x))$$
for \( r = 1, 2, \ldots \) in the form
\[
i^{(r)}(x) = \sum_{h=1}^{r} g^{(h)}(s(x)) B_{r-h}(s) \tag{2.1}
\]
evaluated at \( s = (s_1, s_2, \ldots) \), \( s_i = s^{(i)}(x) \), where \( B_{r-h} \) is the partial exponential Bell polynomial defined by the coefficients in the formal expansion in powers of real \( \varepsilon \),
\[
\left( \sum_{i=1}^{\infty} \varepsilon^{i} s_{i}/i! \right)^{j}/j! = \sum_{r=j}^{\infty} \varepsilon^{r} B_{rj}(s)/r!
\]
for \( j \geq 0 \). Comtet (1974) shows they are given by
\[
B_{rj}(s)/j! = \sum_{n \text{ in } \mathbb{N}^{r}} \left\{ \frac{s_{1}^{n_1} \cdots s_{r}^{n_r}}{n_1! \cdots n_r!} : n_1 + \cdots + n_r = j, \, 1 \cdot n_1 + \cdots + r \cdot n_r = r \right\},
\]
where \( \mathbb{N} = \{0, 1, 2, \ldots\} \). Comtet (1974, page 307) tables them for \( r \leq 12 \). For example,
\[
B_{r1}(s) = s_r, \quad B_{rt}(s) = s_1^{r}, \quad B_{32}(s) = 3s_1s_2, \quad B_{42}(s) = 4s_1s_3 + 3s_2^2, \quad B_{43}(s) = 6s_1^2s_2.
\tag{2.2}
\tag{2.3}
\]

Theorem 2.1 provides an extension of (2.1) to the case \( s : \mathbb{R}^{a} \to \mathbb{R}^{b} \) and \( g : \mathbb{R}^{b} \to \mathbb{R} \).

**Theorem 2.1** Define the partial derivatives
\[
t_{i_{1}, \ldots, i_{r}}(x) = \partial^{r} t(x)/\partial x_{i_{1}} \cdots \partial x_{i_{r}},
\]
\[
s_{i_{1}, \ldots, i_{r}}(x) = \partial^{r} s_{i_{1}}(x)/\partial x_{i_{1}} \cdots \partial x_{i_{r}},
\]
\[
g_{i_{1}, \ldots, i_{r}}(s) = \partial^{r} g(s)/\partial s_{i_{1}} \cdots \partial s_{i_{r}}.
\]

The extension of (2.1) is
\[
t_{1, \ldots, r}(x) = \sum_{h=1}^{r} g_{i_{1}, \ldots, i_{h}}(s(x)) B_{r}^{i_{1}, \ldots, i_{h}}(s).
\tag{2.4}
\]

In (2.4) and throughout, we use the tensor sum convention that repeated indices \( i_{1}, i_{2}, \ldots \) are implicitly summed over their range \( (1, \ldots, b \text{ in the case of (2.4)}) \).

Note that \( B_{r}^{i_{1}, \ldots, i_{h}}(s) \) can be written down on sight from \( B_{r-h} \). Some particular cases of \( B_{r}^{i_{1}, \ldots, i_{h}}(s) \) can be obtained from (2.2) and (2.3):
\[
B_{r}^{i}(s) = s_{i, 1, \ldots, r}, \quad B_{r}^{i_{1}, \ldots, i_{r}}(s) = s_{i_{1}, 1} \cdots s_{i_{r}, r},
\]
\[
B_{3}^{i_{1}i_{2}}(s) = \sum_{3} s_{i_{1}, 1}s_{i_{2}, 23},
\]
\[
B_{4}^{i_{1}i_{2}}(s) = \sum_{3} s_{i_{1}, 1}s_{i_{2}, 234} + \sum_{3} s_{i_{1}, 12}s_{i_{2}, 34}, \quad B_{4}^{i_{1}i_{2}i_{3}}(s) = \sum_{6} s_{i_{1}, 1}s_{i_{2}, 2}s_{i_{3}, 34},
\]
where
\[
\sum_{r} h_{1, \ldots, r} = h_{1, \ldots, r} + h_{2, \ldots, r} + \cdots + h_{r, \ldots, 1}.
\]

3
A form of the multivariate chain rule (2.4) was given in Withers (1984).

Let $\mathcal{F}$ be a convex set of probability measures on a measurable space $(\Omega, A)$. Suppose for $x \in \Omega$, that $\delta_x$ lies in $\mathcal{F}$, where $\delta_x$ is the measure putting mass 1 at $x$ and 0 elsewhere. Let $x, \{x_i\}$ be points in $\Omega$. Let $F$ lie in $\mathcal{F}$, and $T : F \to \mathbb{R}$ be some functional. Define the $r$th derivative of $T(F)$ at $(x_1, \ldots, x_r)$,

$$T_{1 \cdots r} = T_F(x_1 \cdots x_r) = T_F^{(r)}(x_1, \ldots, x_r),$$

as in Withers (1983). The only derivative we need give here is the first, also known as the influence function:

$$T_1(x_1) = T_F(x_1) = \lim_{\epsilon \to 0} \{T((1 - \epsilon)F + \epsilon \delta_{x_1}) - T(F)\} / \epsilon.$$ 

For example, $T(F) = \int_{-\infty}^{\infty} g(x) dF(x)$ has first derivative $T_1 = T_1(x) = g(x) - T(F)$. The results stated in Withers (1983) for $\Omega = \mathbb{R}$ generalize immediately to general $\Omega$. In particular, the rule (2.11) for the derivative of the $r$th derivative may be stated as

$$(T_{1 \cdots r})_{r+1} = T_{1 \cdots r+1} - \sum_{i=1}^{r} [T_{1 \cdots r+1}]_i,$$  

(2.5)

where $[T_{1 \cdots r+1}]_i = T_{1 \cdots r+1}$ with the $i$th argument dropped. So,

$$(T_1)_2 = T_{12} - T_2,$$

$$(T_{12})_3 = T_{123} - T_{23} - T_{13}.$$ 

In this way higher derivatives may be calculated from successive first derivatives. For example, the second derivative of $\int_{-\infty}^{\infty} g(x) dF(x)$ is zero. Now suppose for some function $g : \mathbb{R}^k \to \mathbb{R},$

$$T(F) = g(S(F))$$

where $S(F)$ is a real functional in $\mathbb{R}^d$. (2.6)

Applying (2.5) gives

$$T_1 = g_i S_{i1},$$

(2.7)

$$T_{12} = g_i S_{i12} + g_{ij} S_{i1} S_{j2},$$

(2.8)

$$T_{123} = g_{ij} S_{i123} + g_{ijk} S_{i1} S_{j2} S_{k3},$$

(2.9)

and so on, where $S_{a12\ldots}$ is the $r$th derivative of $S_a(F)$. Despite the fact that by (2.5) the derivative of a derivative is not a second derivative, the expressions (2.7)-(2.9) are precisely those for the derivatives of a function of a vector function of a vector given in (2.4). That is,

$$T_{1 \cdots r} = \sum_{h=1}^{r} g_{i_1 \cdots i_h} (S(F)) B_{r}^{i_1 \cdots i_h}(S),$$

(2.10)

where $S = (S_1, S_2, \ldots)$, $S_i = S^{(i)}(F)$. A proof that (2.10) holds for general $r$ follows using (2.5) and induction. The result can be formally stated as follows.
Theorem 2.2 If (2.6) holds, $T_{1...r}$ is given by the chain rule for

$$T_{1...r} = T_{1...r}(x) = \partial^r T(x)/\partial x_1 \cdots \partial x_r$$

when $T(x) = g(S(x))$ for $S(x) : \mathbb{R}^r \to \mathbb{R}^d$ with $S_{i1...r} = S_{i1...r}(x) = \partial^r S_i(x)/\partial x_1 \cdots \partial x_r$ re-interpreted as $S_i(F(x_1 \cdot \cdot \cdot x_r)$ and $S(x)$ as $S(F)$. So,

$$T_{1...r} = \sum_{h=1}^{r} g_{i_1...i_h} \sum_{n} \bigg( \prod S_{i_1 \cdot \Pi_1} \cdots S_{i_h \cdot \Pi_h} \bigg),$$

(2.11)

where $g_{i_1...i_h}(y) = \partial_{i_1} \cdots \partial_{i_h} g(y)$ at $y = S(F)$ for $\partial_i = \partial/\partial y_i$, $\sum_n$ sums over $n = (n_1 \cdots n_r) \in \mathbb{N}^r$ satisfying $\sum_{i=1}^{r} n_i = k$, $\sum_{i=1}^{r} in_i = r$, $m(n) = r! \prod_{i=1}^{r} i^{n_i} n_i !$, the partition function, and $\sum_{\Pi} m(n)$ sums over all partitions $(\Pi_1 \cdots \Pi_k)$ of $(1 \cdots r)$ with $i \Pi$’s of length $n_i$. 

Corollary 2.1 applies Theorem 2.2 to obtain the next two derivatives.

Corollary 2.1 We

$$T_{1234} = g_{i_4} S_{i_1 1234} + g_{i_1 i_2} \left( \sum_{i_3} S_{i_1 12 i_2 3} + \sum_{i_3} S_{i_1 12 i_2 4} \right)$$

$$+ g_{i_1 i_2 i_3} \sum_{i_4} S_{i_1 12 i_2 3 i_4} + g_{i_1 ... i_5} S_{i_1 12 i_2 3 i_4 i_5},$$

$$T_{12345} = g_{i_5} S_{i_1 1235} + g_{i_1 i_2} \left( \sum_{i_3} S_{i_1 12 i_2 3 4} + \sum_{i_3} S_{i_1 12 i_2 3 4 5} \right)$$

$$+ g_{i_1 i_2 i_3} \sum_{i_4} S_{i_1 12 i_2 3 4 i_4} + g_{i_1 ... i_6} S_{i_1 12 i_2 3 i_4 i_5 i_6},$$

so

$$\begin{bmatrix} ^1 \varepsilon \end{bmatrix} = g_{i_1} \cdots g_{i_k} \left[ S_{i_1 \cdots i_k} \right],$$

$$\begin{bmatrix} [1] \varepsilon \end{bmatrix} = g_i [11] S_i + g_i g_j \left[ S_{i 1 j} \right],$$

$$\begin{bmatrix} [1, 2, 12] \varepsilon \end{bmatrix} = g_i g_{i_2} g_{i_3} \left[ S_{i_1 12 i_2 3 i_3} \right] + g_{i_1 i_2 i_3 i_4 i_5} \left[ S_{i_1 1 i_2 3 i_3 i_4} \right] \left[ S_{i_1 i_2 2 i_3 i_4} \right],$$

$$\begin{bmatrix} [111] \varepsilon \end{bmatrix} = g_i [111] S_i + 3 g_{i j} \left[ S_{i 1 j 11} \right] + g_{i j k} \sum_{i_1 i_2} \left[ S_{i 1 j 1 k 1} \right],$$

$$\begin{bmatrix} [1122] \varepsilon \end{bmatrix} = g_i [1122] S_i + g_{i_1 i_2} \left( 4 \left[ S_{i_1 1 i_2 122} \right] + \sum \left[ S_{i_1 12 i_2 12} \right] + [11] S_i [11] S_i \right)$$

$$+ g_{i_1 i_2 i_3} \left( 2 \left[ S_{i_1 12 i_2 2 i_3 12} \right] + S_{i_1 12 i_2 13} \left[ 11 \right] S_i \right),$$

$$\begin{bmatrix} [1, 122] \varepsilon \end{bmatrix} = g_i g_{i_2} \left[ S_{i_1 1 i_2 122} \right] + 2 g_{i j} g_{j k} \left[ S_{i_1 1 i_2 2 j k} \right] + g_{i j k l} \left[ S_{i 1 j 1 k l} \right] \left[ S_{i j 1 k 1} \right],$$

$$\begin{bmatrix} [12^2] \varepsilon \end{bmatrix} = g_i g_{i_2} \left[ S_{i_1 12 i_2 12} \right] + 2 g_{i j} g_{j k} \left[ S_{i_1 12 i_2 12 k} \right] + g_{i j k l} \left[ S_{i 1 i j k l} \right] \left[ S_{i j 1 i 1} \right],$$

5
and so on, where

\[ [S_{i,1}S_{j,1}] = \int_{-\infty}^{\infty} S_{i,1}S_{j,1}dF(x_1), \]

\[ [S_{i,1}S_{j,1}S_{k,12}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_{i,1}S_{j,1}S_{k,12}dF(x_1)dF(x_2), \]

and so on.

### 3 Some applications

Let \( \hat{F} \) be the empirical distribution of a random sample of size \( n \) from \( F \). By Withers (1983), for a broad class of \( T \), the cumulants of \( T(\hat{F}) \) satisfy

\[ \kappa_r \left( T(\hat{F}) \right) \approx \sum_{i=r-1}^{\infty} n^{-i} a_{ri} \]

for \( r \geq 1 \), where the cumulant coefficient \( a_{ri}(T) = a_{ri} \) is a certain function of the derivatives of \( T(F) \). The most important are \( a_{10} = T(F) \),

\[ a_{21} = [1^2]_T = [T_1^2], \quad (3.1) \]
\[ a_{11} = [1]_T/2 = [T_{11}]/2, \quad (3.2) \]
\[ a_{32} = [1^3]_T + 3[1,2,12]_T = [T_3^3] + 3[T_1T_2T_{12}], \quad (3.3) \]

where

\[ [f(T_1,T_{11},\ldots)] = \int_{-\infty}^{\infty} f(T_1,T_{11},\ldots) dF_1(x_1), \]
\[ [f(T_1,T_2,T_{11},T_{12},T_{22},T_{122},\ldots)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(T_1,T_2,T_{11},T_{12},T_{22},T_{122},\ldots) dF_1(x_1)dF_2(x_2), \]

and so on, for \( F_i = F(x_i) \), and

\[ [1^i,2^j,11^k,12^d,22^m,\ldots]_T = [T_1^i,T_2^j,T_{11}^k,T_{12}^d,T_{22}^m,\ldots], \]

and so on. We refer to the functionals \([\cdots]\) as bracket functions. They are the building blocks for the cumulant coefficients \( a_{ri} \) and the cumulant coefficients of the Studentized statistics, and hence for the Edgeworth-Cornish-Fisher expansions of the standardized form of \( T(\hat{F}) \),

\[ Y_n = (n/a_{21})^{1/2} \left\{ T(\hat{F}) - T(F) \right\}, \quad (3.4) \]

and its Studentized form. They are also the building blocks for obtaining nonparametric confidence intervals and estimates of low bias for \( T(F) \).

As a start we have these approximations to the bias, variance, and skewness of \( T(\hat{F}) \):

\[ \mathbb{E} \left[ T(\hat{F}) \right] = T(F) + n^{-1} a_{11} + O\left(n^{-2}\right), \]
\[ \text{var} \left[ T(\hat{F}) \right] = n^{-1} a_{21} + O\left(n^{-2}\right), \]
\[ \mu_3 \left[ T(\hat{F}) \right] = n^{-2} a_{32} + O\left(n^{-3}\right), \]
where \( \mu_3[X] = \mathbb{E}[(X - \mathbb{E}[X])^3] \).

Theorem 3.1 lists the bracket functions needed for bias and bias reduction. Theorem 3.2 lists the bracket functions needed for Edgeworth-Cornish-Fisher expansions. Theorem 3.3 lists the bracket functions needed for nonparametric confidence intervals.

**Theorem 3.1** Under regularity conditions,

\[
\mathbb{E}\left[T\left(\hat{F}\right)\right] = T(F) + \sum_{i=1}^{j} n^{-i}a_{1i} + O\left(n^{-j-1}\right),
\]

\[
a_{11} = [11]T/2, \quad a_{12} = [111]T/6 + [1122]T/8, \quad a_{13} = [1111]T/24 + [1122]T/12 + [112233]T/48,
\]

and so on. The estimates of \( T(F) \) of bias \( O(n^{-j-1}) \) are

\[
T\left(\hat{F}\right) + \sum_{i=1}^{j} n^{-i}T_i\left(\hat{F}\right) \quad \text{and} \quad T\left(\hat{F}\right) + \sum_{i=1}^{j} S_i\left(\hat{F}\right) / (n-1)_i, \quad (3.5)
\]

where \((m)_i = m!/(m-i)! = m(m-1) \cdots (m-i+1) \) and

- \( T_1(F) = S_1(F) = -[11]T/2, \quad (3.6) \)
- \( T_2(F) = [111]T/3 + [1122]T/8 - [11]T/2, \quad (3.7) \)
- \( T_3(F) = -[11]T/2 + [111]T - [1111]T/4 + 3[1122]T/4 - [11122]T/6 - [112233]T/48, \)
- \( S_2(F) = [111]T/3 + [1122]T/8, \)
- \( S_3(F) = -[1111]T/4 + 3[1122]T/8 - [11122]T/6 - [112233]T/48. \)

**Proof:** Follows by equation (2.4) of Withers (1987). □

**Theorem 3.2** The ‘reduced’ Edgeworth and Cornish-Fisher expansions of \( T(\hat{F}) \) to \( O\left(n^{-j+1/2}\right) \) needs

- for \( j = 0 \): \( a_{21} \),
- for \( j = 1 \): \( a_{11} \) and \( a_{32} \),
- for \( j = 2 \): \( a_{22} = [1,11]T + \left[12^2\right]T/2 + [1,122]T, \)
- \( a_{43} = [1^4]T - 3\left[12^2\right]T + 12[1,2^2,12]T + 12[1,2,13,23]T + 4[1,2,3,123]T, \) \( \quad (3.8) \)

and so on. In particular, for \( Y_n \) of \((3.4)\), under regularity conditions,

\[
P\left(Y_n \leq x\right) = \Phi(x) - \phi(x) \left[n^{-1/2}h_1(x) + n^{-1}h_2(x)\right] + O\left(n^{-3/2}\right)
\]

for \( h_1 = A_{11} + A_{32}He_2/6 \) and \( h_1 = (A_{22} + A_{11}^2)He_1/2 + (A_{43} + 4A_{11}A_{32})He_3/24 + A_{32}^2He_5/72, \) where \( \Phi, \phi \) are the distribution and density of a unit normal random variable, \( He_r \) is the \( r \)th Hermite polynomial, and \( A_{ri} = a_{ri}/a_{21}^r \), the standardized cumulant coefficient.

**Proof:** Follows by Withers (1983). □

The regularity conditions needed for Theorems 3.1 and 3.2 are the same as those given in Withers (1983, 1987). So, they are not stated here.
Theorem 3.3 A confidence interval for $T(F)$ of level $1 - \alpha + O(n^{-(j+1)/2})$ requires the bracket functions

for $j = 0: a_{21} = \left[1^2\right]_T$,
for $j = 1: [11]_T, \left[1^3\right]_T, [1, 2, 12]_T$,
for $j = 2: [1, 11]_T, [12^2]_T, [1, 122]_T, [1^4]_T, [1, 2^2, 12]_T, [1, 2, 13, 23]_T, [1, 2, 3, 123]_T.

(3.9)

Proof: Follows by Theorem 5.1 in Withers (1983). □

By Withers (1989), the bracket functions in Theorem 3.3 are also the terms needed for the distribution and quantiles of the Studentized form of $T(F)$ to $O(n^{-(j+1)/2})$.

For the distribution of $|T(\hat{F}) - T(F)|$ to $O(n^{-j-1})$ or for a symmetric confidence interval for $T(F)$ of level $1 - \alpha + O(n^{-j-1})$ one needs, by equations (2.4) and (2.5) of Withers (1982), $a_{21}$ for $j = 0$ and $a_{11}, a_{32}, a_{22}, a_{43}$ for $j = 1$.

For convenience, set $T = T(F)$ and $g_i = g_i^{(i)}(S(F))$ for $S(F)$ in $\mathbb{R}$.

Example 3.1 This example gives bracket functions for a function of a univariate functional. Suppose (2.7) holds with $b = 1$. Then

\begin{align*}
T_1 &= g_1 S_1, \\
T_{12} &= g_1 S_{12} + g_2 S_1 S_2, \\
T_{123} &= g_1 S_{123} + g_2 \sum_{3} S_1 S_{23} + g_3 S_1 S_2 S_3, \\
T_{1234} &= g_1 S_{1234} + g_2 \left( \sum_{4} S_1 S_{234} + \sum_{3} S_1 S_{24} \right) + g_3 \sum_{6} S_1 S_2 S_{34} + g_4 S_1 S_2 S_3 S_4, \\
T_{12345} &= g_1 S_{12345} + g_2 \left( \sum_{10} S_1 S_{2345} + \sum_{15} S_1 S_{2345} \right) + g_3 \sum_{10} S_1 S_2 S_{345} + g_4 \sum_{15} S_1 S_2 S_3 S_{45} + g_5 S_1 \cdots S_5.
\end{align*}

So,

\begin{align*}
\left[1^k\right]_T &= g_1^k \left[1^k\right]_S, \\
[11]_T &= g_1[11]_S + g_2 \left[1^2\right]_S, \\
[1, 2, 12]_T &= g_1^2[1, 2, 12]_S + g_1^2 g_2 \left[1^2\right]_S^2, \\
[111]_T &= g_1[111]_S + 3g_2[1, 11]_S + g_3 \left[1^3\right]_S, \\
[1122]_T &= g_1[1122]_S + g_2 \left( 4[1, 122]_S + [11]_S^2 + 2 \left[12^2\right]_S \right) + 2g_3 \left( \left[1^2\right]_S [11]_S + 2[1, 12]_S \right) + g_4 \left[1^2\right]_S^2, \\
[1, 122]_T &= g_1^2[1, 122]_S + g_1 g_2 \left( \left[1^2\right]_S [111]_S + 2[1, 2, 12]_S \right) + g_1 g_3 \left[1^2\right]_S, \\
[12^2]_T &= g_1^2 \left[12^2\right]_S + 2g_1 g_2 \left[1, 2, 12\right]_S + g_2^2 \left[1^2\right]_S^2.
\end{align*}
Example 3.2 This example gives bracket functions for a product. Suppose that \( T(F) = S_1(F)S_2(F) \). Then

\[
\begin{align*}
T_1 &= S_2S_{1.1} + S_1S_{2.1} = \sum S_1S_{1.1} \text{ say,} \\
T_{12} &= (S_2S_{1.12} + S_1S_{2.12}) + (S_{1.1}S_{2.2} + S_{2.1}S_{1.2}) \\
&= \sum (S_2S_{1.12} + S_1S_{2.12}) \text{ say,} \\
T_{123} &= (S_2S_{1.123} + S_1S_{2.123}) + \sum^3 (S_{1.1}S_{2.23} + S_{2.1}S_{1.23}) \\
&= \sum S_2S_{1.123} + \sum S_{1.1}S_{2.23} \text{ say,} \\
T_{1234} &= (S_2S_{1.1234} + S_1S_{2.1234}) + \sum^4 (S_{1.1}S_{2.234} + S_{2.1}S_{1.234}) \\
&\quad + \sum^3 (S_{1.12}S_{2.34} + S_{2.12}S_{1.34}) \\
&= \sum S_2S_{1.1234} + \sum S_{1.1}S_{2.234} + \sum S_{1.2}S_{2.34} \text{ say.}
\end{align*}
\]

So,

\[
\begin{align*}
[1^2]_T &= \sum^2 S_2^2 [1^2]_{S_1} + 2S_1S_2 [S_1S_{1.2.1}], \\
[11]_T &= \sum^2 S_2 [11]_{S_1} + 2 [S_1S_{1.2.1}], \\
[1^3]_T &= \sum^2 (S_2^2 [1^3]_{S_1} + 3S_2^2 [S_2S_{1.1}^2]), \\
[1, 2, 12]_T &= \sum^2 \left\{ S_2^3 [1, 2, 12]_{S_1} + S_2^2S_1 \left\{ [S_1S_{1.2}S_{2.12}] \\
&\quad + 2[S_1S_{2.2}S_{1.12}] \right\} + 2S_2^2 [1^2]_{S_1} [S_1S_{1.2.1}] \right\} \\
&\quad + 2S_1S_2 \left\{ [S_1S_{1.2}^2] + [1^2]_{S_1} [1^2]_{S_2} \right\}, \\
[111]_T &= \sum^2 (S_2 [111]_{S_1} + 3 [S_2S_{1.1}]), \\
[1122]_T &= \sum^2 (S_2 [1122]_{S_1} + 4 [S_1S_{1.2.1.2}]) + 2 [11]_{S_1} [11]_{S_2} + 4 [S_1S_{1.2}S_{2.1}].
\end{align*}
\]

Example 3.3 This example gives bracket functions for a Studentized function. The Studentized form of \( T(F) \) is

\[
T_0 \left( \hat{F} \right) = V \left( \hat{F} \right)^{-1/2} \left\{ T \left( \hat{F} \right) - T(F) \right\}
\]

for \( V(F) = a_{21} \). Its bracket functions \([\cdots]_{T_0} \) (and so also its cumulant coefficients) may be expressed in terms of the bracket functions \([\cdots]_{T} \). For details, see Appendix A of Withers (1989).
If one makes other assumptions such as symmetry of $F$ or a parametric form for $F$, then $V(F) = a_{21}$ will generally take a simpler form. Similarly, in some circumstances one is interested in standardizing a functional in a different way, for example, replacing $\mu_r$ by $\mu_r/\mu_r^{3/2}$. The next example covers this situation for the special case of a $T(F)$ a function of a univariate functional.

**Example 3.4** Suppose that $T(F) = U(F) g(S(F))$ with $S(F)$ in $\mathbb{R}$. Then

$$T_1 = g_0 U_1 + g_1 U S_1,$$
$$T_{12} = g_0 U_{12} + g_1 \left( U S_{12} + \sum_{i=1}^{2} U_1 S_{2i} \right) + g_2 U S_1 S_2,$$
$$T_{123} = g_0 U_{123} + g_1 \left( U S_{123} + \sum_{i=1}^{6} U_1 S_{2i} \right) + g_2 \sum (U S_1 S_{23} + U_1 S_{2} S_{3}) + g_3 U S_1 S_2 S_3,$$
$$T_{1234} = g_0 U_{1234} + g_1 \left( U S_{1234} + \sum_{i=1}^{8} U_1 S_{2i} \right) + g_2 \left( U \sum S_{1} S_{2i} + U \sum S_{1} S_{2} S_{3} + U \sum U_1 S_{2} S_{4} + U \sum U_1 S_{2} S_{3} S_{4} \right) + g_3 \left( U \sum S_{1} S_{2} S_{3} S_{4} + U \sum U_1 S_{2} S_{3} S_{4} \right) + g_4 U S_1 S_2 S_3 S_4.$$

So, the cumulant coefficients $a_{21}$, $a_{11}$, $a_{32}$ needed for third order inference are given by (3.1)-(3.3) in terms of the bracket functions

$$[1^2]_T = g_0^2 [1^2]_U + 2 g_0 g_1 U [U_1 S_1] + g_1^2 U^2 [1^2]_S,$$
$$[11]_T = g_0^2 [1]_U + g_1 (U [11]_S + 2 [U_1 S_1]) + g_2 U [1^2]_S,$$
$$[1^3]_T = g_3^3 [1^3]_U + 3 g_0 g_1 U [U_1 S_1] + 3 g_0 U^2 [U_1 S_1] + g_3 U [3^3]_S,$$
$$[1,2,12]_T = g_0^3 [1,2,12]_U + g_0 g_1 (U [U_1 U_2 S_{12}] + 2 [1^2]_U [U_1 S_1] + 2 U [U_1 S_1 U_{12}]) + g_0 g_2 U [U_1 S_1]^2 + g_0 g_1 U \left( 2 [U_1 S_1]^2 + 2 [1^2]_U [1^2]_S + 2 U [U_1 S_1 S_{12}] \right) + U [S_{1} S_{2} U_{12}] + 2 g_0 g_1 g_2 U^2 [U_1 S_1] [1^2]_S + g_3 U^2 \left( U [1,2,12]_S + 2 [U_1 S_1] [1^2]_S \right),$$

Similarly, the cumulant coefficients $a_{22}$, $a_{43}$ needed for third order inference are given by (3.5) in terms of the bracket functions given in Appendix A. The bracket functions needed
Example 4.1 Suppose for estimates of $T(F)$ of bias $O(n^{-3})$ are
\[
\begin{align*}
[111]_T &= g_0 [111]_U + g_1 (U [111]_S + 3 [U_1 S_{11}] + 3 [U_{11} S_1]) \\
&+ 3g_2 (U [1, 11]_S + [U_1 S^2_1]) + g_3 U [1^3]_S,
\end{align*}
\]
\[
\begin{align*}
[1122]_T &= g_0 [1122]_U + g_1 \left(U [1122]_S + 4 [U_1 S_{122}] + 4 [U_{122} S_1]\right) \\
&+ 2 [11]_U [11]_S + 4 [U_{12} S_{12}] + g_2 \left(4U [1, 122]_S + U [11]_S^2 + 2U [12^2]_S\right) \\
&+ 8 [U_1 S_{2} S_{12}] + 4 [S_1 S_2 U_{12}] + 4 [U_1 S_{1}] [11]_S + 2 [11]_U [1^2]_S \\
&+ 2g_3 \left(U [1^2]_S [11]_S + 2U [1, 2, 12]_S + 2 [U_1 S_{1}] [1^2]_S\right) + g_4 U [1^2]_S^2.
\end{align*}
\]
Further terms are given in Appendix A.

If $g(s) = s^r$ then $g_i = (r)_i s^{r-i}$. Putting $r = -1$ gives the derivatives of a quotient $(-1)_i = (-1)^i i!$.

4 Applications to moments

Suppose $X \sim F$ on $\mathbb{R}$. Set $\mu = \mathbb{E}[X]$, $\mu'_r = \mathbb{E}[X^r]$ and let $\{\mu_r, \kappa_r\}$ be the central moments and cumulants of $F$. Set $\mu(F) = \mu$ and so on. Let $\hat{F}$ be the empirical distribution of a random sample of size $n$ from $F$.

Many authors have studied problems of moments and cumulants: see, for example, Stuart and Ord (1987). Fisher’s $k$-statistic $\kappa_r$, the unbiased estimate of $\kappa_r$, is given there in Section 12.9 for $r \leq 8$ in terms of $\{s_i = n \mu'_i(\hat{F}) = \sum^n_{j=1} X_j^i\}$. Fisher’s expressions for unbiased estimates of the joint cumulants of $k$-statistics are given there in Section 12.16. Wishart’s unbiased estimates of products of cumulants are given there in Section 12.16 in terms of symmetric functions, which can be converted to $\{s_i\}$ using Appendix Table 10.

Generally one only wants approximations. (Indeed without making parametric assumptions on $F$ only approximations are possible except for estimating polynomials in moments). One problem with these “traditional” approaches is that it is not an easy task to separate out terms beyond the first in decreasing order of importance in order to make such approximations. As noted in Section 3 the present approach does not suffer from this disadvantage.

For $S(F)$ a polynomial in $F$ of degree $r$ (for example, $\mu'_r$, $\mu_r$ or $\kappa_r$), derivatives of order beyond $r$ vanish.

Example 4.1 Suppose $T(F)$ is a function of a univariate mean, say $T(F) = g(\mu(F))$. Setting $g_k = g^{(k)}(\mu)$, Example 3.1 implies
\[
\begin{align*}
a_{21} &= g_1^2 \mu_2, \\
a_{11} &= g_2 \mu_2^2, \\
a_{32} &= g_1^3 \mu_3 + 3g_1^2 g_2 \mu_3^2, \\
a_{22} &= g_1 g_2 \mu_3 + \left(g_2^3/2 + g_1 g_3\right) \mu_3^2, \\
a_{43} &= g_1^4 \left(\mu_4 - 3\mu_2^2\right) + 12g_1^3 g_2 \mu_3 \mu_2 + 4 \left(3g_1^2 g_2^2 + g_1^3 g_3\right) \mu_3^2.
\end{align*}
\]
For,
\[ T_{1:1:p} = g_p h_1 \cdots h_p, \]
where \( h_i = x_i - \mu \). So,
\[
\begin{align*}
[1^k] &= g_1^k \mu_k, \\
[1 \cdots 1] &= g_k \mu_k \text{ if } 1 \cdots 1 \text{ contains } k \text{'s,}
[1, 2, 12] &= g_1^2 g_2 \mu_3^2, \\
[1, 11] &= g_1 g_2 \mu_3, \\
[12^2] &= g_2^2 \mu_2^2, \\
[1, 122] &= g_1 g_3 \mu_2^2, \\
[1, 2^2, 12] &= g_1^2 g_2 \mu_3 \mu_2, \\
[1, 2, 13, 23] &= g_1 g_2 \mu_3^3, \\
[1, 2, 3, 123] &= g_1^3 g_3 \mu_3^2.
\end{align*}
\]
So, an estimate of \( g(\mu(F)) \) of bias \( O(n^{-4}) \) is given by (3.3) with \( j = 3 \) in terms of
\[
\begin{align*}
S_1(F) &= -g_2 \mu_2/2, \\
S_2(F) &= g_3 \mu_3/3 + g_4 \mu_2^2/8, \\
S_3(F) &= -g_4 \mu_4/4 + 3g_4 \mu_2^2/8 - g_5 \mu_3 \mu_2/6 - g_6 \mu_2^3/48.
\end{align*}
\]
For example, an estimate of \( \mu^r \) of bias \( O(n^{-4}) \) is given by substituting \( g_i = (r)_i \mu^{r-i} \). If \( \mu \geq 0 \), \( r \) need not be an integer. However, regularity conditions generally breakdown if \( r < 0 \) and \( F(0) \neq 0 \).

Functions of non-central moments can be handled with similar ease. We now present an important result which was stated without proof in equation (4.1) of Withers (1987), the derivatives of a central moment.

**Theorem 4.1** For \( r, p \) in \( \{1, 2, \ldots \} \), the \( p \)th derivative of \( \mu_r(F) \) is
\[
\mu_{r,1:1:p} = (-1)^p \left\{ (r)_p \mu_{r-p} - (r)_{p-1} \sum_{i=1}^{p} (h_i^{r-p} - \mu_{r-p+1} h_i^{-1}) \right\} \prod_{i=1}^{p} h_i,
\]
where \( h_i = x_i - \mu \).

**Proof:** As in Example 3.1, \( T(F) = \mu'_k \mu^j \) has derivatives
\[
T_{1:1:p} = \left( g_{p-1} \sum_{i=1}^{p} U_i S_i^{-1} + g_p U \right) S_1 \cdots S_p,
\]
where
\[
g_p = (j)_p \mu^{j-p}, \ U = \mu'_k, \ U_i = x_i^k - \mu'_k, \ S_i = x_i - \mu.
\]
But $\mu_r = \sum_{k=0}^{r} (\binom{r}{k})(-1)^{r-k} \mu_k \mu^{r-k}$. So,

$$
\mu_{r-1:p} = \sum_{k=0}^{r} \left( \binom{r}{k}(-1)^{r-k}\left\{ (r-k)p-1\mu^{r-k-p+1}\sum_{i=1}^{p} (x_i^k - \mu_k) h_i^{-1} \\
+ (r-k)p\mu^{r-k-p}\mu_k \right\} \right) \prod_{j=1}^{p} h_j.
$$

Now simplify. □

Some particular cases of the theorem are given by the following corollaries.

**Corollary 4.1** We have

$$
\begin{align*}
\mu_{r-1} &= h_1^r - \mu_r - rh_1\mu_{r-1}, \\
\mu_{r-12} &= -r \sum_{i=3}^{2} (h_1^{r-1} - \mu_{r-1}) h_2 + (r)h_1 h_2\mu_{r-2}, \\
\mu_{r-123} &= (r)h_2 \sum_{i=3}^{2} (h_1^{r-2} - \mu_{r-2}) h_3 - (r)h_1 h_2 h_3\mu_{r-3}, \\
\mu_{r-12-\cdots-1} &= (-1)^{r-1} (r! / 2) \sum_{i=3}^{r-1} (h_1^2 - \mu_2) h_2 \cdots h_{r-1}, \\
\mu_{r-12-\cdots-r} &= (-1)^{r}(r-1)! h_1 \cdots h_r, \\
\mu_{r-1p} &= (-1)^p \left\{ (r)p\mu_{r-p} h_1^p - p(r)p-1h_1^p - \mu_{r-p+1}h_1^{p-1} \right\}.
\end{align*}
$$

**Corollary 4.2** We have

$$
\begin{align*}
\mu_{2-1} &= h_1^2 - \mu_2, \\
\mu_{3-1} &= h_1^3 - \mu_3 - 3h_1\mu_2, \\
\mu_{4-1} &= h_1^4 - \mu_4 - 4h_1\mu_3, \\
\mu_{5-1} &= h_1^5 - \mu_5 - 5h_1\mu_4, \\
\mu_{2-12} &= -2h_1 h_2, \\
\mu_{3-12} &= -2\sum_{i=3}^{2} (h_1^2 - \mu_2) h_2, \\
\mu_{4-12} &= 12h_1 h_2\mu_2 - 4\sum_{i=3}^{2} (h_1^3 - \mu_3) h_2, \\
\mu_{5-12} &= 20h_1 h_2\mu_3 - 5\sum_{i=3}^{2} (h_1^4 - \mu_4) h_2, \\
\mu_{3-123} &= 12h_1 h_2 h_3, \\
\mu_{4-123} &= 12 \sum_{i=3}^{3} (h_1^2 - \mu_2) h_2 h_3, \\
\mu_{5-123} &= 20 \sum_{i=3}^{3} (h_1^3 - \mu_3) h_2 h_3 - 60h_1 h_2 h_3, \\
\mu_{4-1234} &= 72h_1 h_2 h_3 h_4, \\
\mu_{5-1234} &= 60 \sum_{i=4}^{4} (h_1^2 - \mu_2) h_2 h_3 h_4, \\
\mu_{5-12345} &= -480h_1 h_2 h_3 h_4 h_5.
\end{align*}
$$
So, for example, for $T(F) = \mu_r$,
\[
\begin{align*}
[1^3]_T &= \mu_{3r} - 3r\mu_{2r+1}\mu_{r-1} - 3\mu_{2r}\mu_r + 3r^2\mu_{r+2}\mu_{r-1} + 6r\mu_{r+1}\mu_{r-1} \\
&\quad + 2\mu_r^3 - 3r^2\mu_r^2\mu_{r-1} + r^3\mu_{r-1}^3,
\end{align*}
\]

\[
[1, 2, 12]_T = -\mu_{2r-1}(2r\mu_{r+1} + r^2\mu_{r-1}\mu_2) + (r)_{2r+1}\mu_{r-2} \\
- r^2\mu_{r+1}\mu_{r-1} + 2r^2(r - 1)\mu_{r+1}\mu_{r-2}\mu_2 \\
- (2r^3 - r^2)\mu_r^2\mu_{r-1}\mu_2 + r^3(r - 1)\mu_{r-1}\mu_{r-2}\mu_2^2,
\]
giving
\[
\begin{align*}
a_{21} &= [1^2]_T = r^2\mu_{r-1}\mu_2 - 2r\mu_{r-1}\mu_{r+1} + \mu_{2r} - \mu_r^2, \\
a_{11} &= [11]_T/2 = (r)_{2r-2}\mu_2/2 - r\mu_r, \\
a_{32} &= \mu_{3r} - 3r\mu_{2r+1}\mu_{r-1} - 3\mu_{2r}\mu_r - 3\mu_{2r-1}(2r\mu_{r+1} + r^2\mu_{r-1}\mu_2) \\
&\quad + 3r^2\mu_{r+2}\mu_{r-1} + 3(r)_{2r+1}\mu_{r-2} - 3r(r - 2)\mu_{r+1}\mu_{r-1} \\
&\quad + 6r^2(r - 1)\mu_{r+1}\mu_{r-1}\mu_{r-2}\mu_2 + 2\mu_r^3 - 3r^2\mu_r^2\mu_{r-1}\mu_2 - r^3\mu_{r-1}^3.
\end{align*}
\]

Similarly, estimates of $\mu_r$ for general $r$ of bias $O(n^{-3})$ are given by \([13, 5]\) in terms of
\[
[111]_T = -(r)_{3r-3}\mu_3 + 3(r)_{2r} (\mu_r - \mu_{r-1}\mu_2), \\
[1122]_T = (r)_{4r-4}\mu_2^2.
\]

For $r = 2$ this gives
\[
a_{21} = \mu_4 - \mu_2^2, \quad a_{11} = -\mu_2, \quad a_{32} = \mu_6 - 3\mu_4\mu_2 + 2\mu_2^3, \quad [111]_T = [1122]_T = 0,
\]
and for $r = 3$ this gives
\[
a_{21} = \mu_6 - 4\mu_4\mu_2 - \mu_2^3 + 9\mu_2^3, \quad a_{11} = -3\mu_3, \\
a_{32} = \mu_6 - 9\mu_4\mu_2 - 3\mu_6\mu_3 - 18\mu_5\mu_4 - 9\mu_4\mu_3\mu_2 + 2\mu_3^3, \\
[111]_T = 12\mu_3, \quad [1122]_T = 0.
\]

**Example 4.2** Suppose that $r$ is an odd integer and $F$ is symmetric. So, odd cumulants of $\mu_r(F)$ are zero so that $a_{11} = a_{32} = 0$ and the Edgeworth-Cornish-Fisher expansions are in powers of $n^{-1}$, not just $n^{-1/2}$. Taking $r = 3$ gives for $T(F) = \mu_3$,
\[
\begin{align*}
[12^2]_T &= 2\mu_2 (\mu_4 - \mu_2^2), \\
[1, 11]_T &= -6 (\mu_6 - 4\mu_4\mu_2 + 3\mu_2^3), \\
[1, 122]_T &= 12\mu_2 (\mu_4 - 3\mu_2^2), \\
[1^4]_T &= \mu_{12} - 12\mu_{10}\mu_2 + 5\mu_8\mu_2^2 - 108\mu_6\mu_2^3 + 81\mu_4\mu_2^4, \\
[1, 2^2, 12]_T &= -3\mu_4 (\mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 - 3\mu_2^4), \\
[1, 2, 13, 23]_T &= 9 (\mu_4 - \mu_2^2) (\mu_4 - 3\mu_2^2)^2.
\end{align*}
\]

So,
\[
\begin{align*}
a_{21} &= \mu_6 - 6\mu_4\mu_2 + 9\mu_2^3, \\
a_{22} &= -6 (\mu_6 - 7\mu_4\mu_2 + 10\mu_2^3), \\
a_{43} &= \mu_{12} - 12\mu_{10}\mu_2 - \mu_8 (72\mu_4 - 5\mu_2^2) - 3\mu_6 (\mu_6 - 108\mu_4\mu_2 + 54\mu_2^3) \\
&\quad + 3\mu_4 (52\mu_2^2 - 576\mu_4\mu_2^2 + 1179\mu_2^4) - 2511\mu_2^6.
\end{align*}
\]

For $F$ normal this gives $a_{21} = 6\mu_2^3, \quad a_{22} = -24\mu_2^3, \quad a_{43} = -11625\mu_2^6$. 

14
Example 4.3 This example is about standardized central moments. Suppose $T(F) = \nu_r$, where $\nu_r = \mu_r / \mu_r^2$. Then the $[\cdot]_T$ needed for third order inference and bias reduction, are given by Example 3.4 with $S = \mu_2$ and $U = \mu_r$ and
\[
g_j = (-r/2) j \mu_2^{-r/2-j} = r(r+2)(r+4) \cdots (r+2j-2)(-2\mu_2)^{-j} \mu_2^{-r/2}
\]
in terms of $[1^2]_U$, $[11]_U$, $[1^3]_U$, $[1^4]_U$, $[1^2]_S$, $[11]_S$, $[1^3]_S$, $[1^4]_S$ given by Example 4.2 and the bracket functions
\[
\begin{align*}
[U_1 S_1] &= \mu_r + 2 - 2\mu_r \mu_2 - r \mu_r \mu_3, \\
[U_2 S_1] &= \mu_2 + 2 - 2\mu_r \mu_2 - 2r \mu_{r+1} (\mu_2 - \mu_r \mu_3) - 2\mu_r - 2\mu_r \\
&+ 2r \mu_2 + r^2 \mu_{r-1} (\mu_3 - \mu_2) , \\
[U_1 S_2] &= \mu_r + 4 - 2\mu_r \mu_2 + \mu_r (-\mu_4 + 2\mu_2^2) - r \mu_{r-1} (\mu_5 + 2\mu_3 \mu_2), \\
[U_1 U_2 S_1] &= -2 (\mu_r + 1 - r \mu_{r-1} \mu_2), \\
[U_1 S_2 U_1] &= -r \mu_{r-1} \mu_3 - r^2 \mu_{r-1} + (r^2 + r) \mu_{r+1} \mu_{r-1} \mu_2 + (r) \mu_{r+1} \mu_r \mu_{r-2} \mu_3 \\
&+ (r^2 + r) \mu_r \mu_{r-1} \mu_3 - r^2 \mu_r \mu_{r-1} \mu_2 - r^2 (r-1) \mu_{r-1} \mu_r \mu_{r-2} \mu_3 \mu_2, \\
[U_1 S_2 U_1] &= 2 (\mu_r + 1) + 2 \mu_r \mu_2 + r \mu_{r-1} \mu_3, \\
[S_1 S_2 U_1] &= -2r \mu_{r+1} \mu_3 - 4 \mu_{r+1} \mu_3 \mu_2 + (r) \mu_{r+2} \mu_3, \\
[U_1 S_1] &= 2 (\mu_r + 2 + r \mu_{r-2} \mu_3), \\
[U_1 S_1] &= 0, \\
[U_1 S_1] &= 2 (\mu_r + 2 + r \mu_{r-2} \mu_3), \\
[U_1 S_1] &= 4 \mu_r \mu_2 - 2 (r) \mu_{r+2} \mu_2.
\end{align*}
\]
For example, suppose that $r = 3$ and $F$ is symmetric. Then $a_{ri} = 0$ for $r$ odd and
\[
\begin{align*}
a_{21} &= \nu_6 - 6\nu_4 + 9, \\
a_{22} &= -3 (\nu_8 - 5\nu_6 + 7\nu_4 - 3) + 12 \nu_6 (2\nu_4 - 1) / 4 \\
&+ 2\nu_4 (107\nu_4 - 489) / 4 + 9 (4\nu_4 - 11), \\
a_{43} &= \nu_{12} - 12 \nu_{10} + 54 \nu_8 - 108 \nu_6 + 81 \nu_4 - 3a_{21} \\
&- 18 (\nu_8 - 4\nu_6 + 3\nu_4 - 3) (\nu_6 - 4\nu_4 + 9) \\
&+ 27 (\nu_4 - 1) (a_{21}^2 + 4a_{21} (\nu_4 - 3) + 4 (\nu_4 (\nu_4 - 1) + 9)) \\
&+ 12 (\nu_4 - 3)^2 (3\nu_6 - 14\nu_4 + 15).
\end{align*}
\]
For $F$ normal this is in agreement with Fisher (1931) who gave the result
\[
\mu (-r, a_3, a_4, \ldots) = \mu (a_3, a_4, \ldots) (n-1)^r / \{(n-1)(n+1) \cdots (n+2r-3)\mu_r^2 \}
\]
for $\mu(a_2, a_3, a_4, \ldots) = E[k_2^{a_2} k_3^{a_3} \cdots]$. See Agostino and Pearson (1973) for a simulation approach.

Figure 4.1 compares the bias reduced estimator of $\nu_3$ versus the usual one by means of simulation. The biases of the estimators are computed by simulating ten thousand replications of samples of size $n$ from the following distributions: standard normal, Student’s $t$ with two degrees of freedom, Student’s $t$ with five degrees of freedom, Student’s $t$ with ten
degrees of freedom, standard logistic, standard Laplace. As expected, the bias reduced estimators give substantially smaller biases for each \( n \) and for each of the six distributions. The biases appear largest for the Student’s \( t \) distribution with two degrees of freedom. The biases appear smallest for the normal distribution, the Student’s \( t \) distribution with ten degrees of freedom, and the logistic distribution.

As noted \( k_r \) is the unbiased estimate of \( \kappa_r \) so \( k_2 = \mu_2(\hat{F})n/(n-1) \), and \( k_3 = \mu_3(\hat{F})n^2/(n-1)(n-2) \).

**Example 4.4** Suppose \( T(F) = \mu_r/\mu^r \), where \( \mu \neq 0 \). Then the \([\cdot]_r \) needed for third order inference and bias reduction are given by Example 3.4 with \( g_j = (-r)_j\mu^{-r-j} \), \( S = \mu \), \( U = \mu_r \), \([1^2]_U \), \([1]_U \), \([1^3]_U \), \ldots given by Example 4.2, \([1^r] \) \( S = \mu \), the other non-zero leading terms needed for Example 3.3 being

\[
[U_1S_1] = \mu_{r+1} - r\mu_{r-1}\mu_2,
[U^2_1S_1] = \mu_{2r+1} - 2r\mu_{r+2}\mu_{r-1} - 2\mu_{r+1}\mu_r + 2r\mu_r\mu_{r-1}\mu_2 + r^2\mu_{r-1}^2\mu_3,
[U_1S_2U_{12}] = -r\mu_{2r-1}\mu_2 - r\mu_{r+1}\mu_r + r(2r+1)\mu_r\mu_{r-1}\mu_2
+ (r)_{2r-2} (\mu_{r+1}\mu_2 - r\mu_{r-1}\mu_2^2),
[S_1S_2U_{12}] = -2r\mu_r\mu_2 + (r)_{2r-2}\mu_2^2,
[U_{11}S_1] = -2r (\mu_{r+1} - \mu_{r-1}\mu_2) + (r)_{2r-2}\mu_{r-2}\mu_3,
[U_{122}S_1] = 3(r)_{2r-1}\mu_2 - (r)_{3r-3}\mu_2^2.
\]

For example, the asymptotic variance of \( n^{1/2}(T(\hat{F}) - T(F)) \) is

\[
\mu^{-2r} (r^2\mu_{r-1}^2 - 2r\mu_{r-1}\mu_{r+1} + \mu_{2r} - \mu_r^2)
- 2r\mu^{-2r-1} (\mu_{r+1} - r\mu_{r-1}\mu_2) + r^2\mu^{-2r-2}\mu_r^2 (\mu_4 - \mu_2^2).
\]

For \( r = 2 \), this reduces to \( T(F)^2(\mu_4\mu_2^{-2} - 1 - 4\mu_3\mu_2^{-1}\mu_1^{-1} + 4\mu_2\mu_1^{-2}) \).

**Example 4.5** This example is about the coefficient of variation. Suppose \( T(F) = \mu_2^{1/2}/\mu \). Then the \([\cdot]_r \) needed for third order inference and bias reduction are given by Example 3.4 with \( g_j = (1/2)_j\mu_2^{1/2-j} \), \( S = \mu_2 \), \( U = \mu^{-1} \). By Example 4.1, \( U_{1\ldots p} = (-1)_{p}\mu^{-1-p}h_1 \ldots h_p \)
so the terms needed in Example 3.4 are

\[
\begin{align*}
[1^2]_U &= \mu^{-4} \mu_2, \\
[11]_U &= 2 \mu^{-3} \mu_2, \\
[11]_S &= -2 \mu_2, \\
[1^3]_U &= -\mu^{-6} \mu_3, \\
[U_1^2 S_1] &= \mu^{-4} (\mu_4 - \mu_2^2), \\
[U_1 S_1^2] &= -\mu^{-2} (\mu_5 - 2 \mu_3 \mu_2), \\
[1^3]_S &= \mu_6 - 3 \mu_4 \mu_2 + 2 \mu_3^3, \\
[1, 2, 12]_U &= 2 \mu^{-7} \mu_2^2, \\
[U_1 U_2 S_{12}] &= -2 \mu^{-4} \mu_2^2, \\
[U_1 S_2 U_{12}] &= -2 \mu^{-5} \mu_3 \mu_2, \\
[U_1 S_2 S_{12}] &= 2 \mu^{-2} \mu_3 \mu_2, \\
[S_1 S_2 U_{12}] &= 2 \mu^{-3} \mu_3^2, \\
[1, 2, 12]_S &= -2 \mu_3^2, \\
[111]_U &= -6 \mu^{-4} \mu_3, \\
[111]_S &= 0, \\
[U_1 S_1 11] &= 2 \mu^{-2} \mu_3, \\
[U_1 S_1 11] &= 2 \mu^{-3} (\mu_4 - \mu_2^2), \\
[S_1 S_1 11] &= -2 (\mu_4 - \mu_2^2), \\
[1122]_U &= 24 \mu^{-5} \mu_2^2, \\
[1122]_S &= 0, \\
[U_1 S_{1122}] &= 0, \\
[U_1 S_{1221}] &= -6 \mu^{-4} \mu_2^2, \\
[U_1 S_{1221}] &= -4 \mu^{-3} \mu_2^2, \\
[1, 122]_S &= 0, \\
[112]_S &= 4 \mu_4, \\
[12^2]_S &= 4 \mu_2^2, \\
[U_1 S_1 S_{12}] &= 2 \mu^{-2} \mu_3 \mu_2.
\end{align*}
\]

For example,

\[
a_{21} = [1^2] = T(F)^2 \left( \mu_2 \mu^{-2} - \mu_3 \mu^{-1} \mu_2^{-1} + \mu_4 \mu_2^{-2} \mu_4^{-1} - 4^{-1} \right)
\]

as given by Section 10.6 of Stuart and Ord (1987). Also

\[
\begin{align*}
[1^3] &= T(F)^3 \sum_{i=0}^{3} \mu^{-i} A_i, \\
[1, 2, 12] &= T(F)^3 \sum_{i=0}^{4} \mu^{-i} B_i, \\
[a_{32}] &= T(F)^3 \sum_{i=0}^{4} \mu^{-i} C_i,
\end{align*}
\]

where

\[
\begin{align*}
A_0 &= 1/4 - 3 \mu_2^{-2} \mu_4 / 8 + \mu_2^{-3} \mu_6 / 8, \\
A_1 &= 3 \left( 2 \mu_2^{-1} \mu_3 - \mu_2^{-2} \mu_5 / 4 \right) / 4, \\
A_2 &= 3 \left( -\mu_2 + \mu_2^{-1} \mu_4 / 2 \right) / 2, \\
A_3 &= -\mu_3, \\
B_0 &= - (1 - \mu_2^{-2} \mu_4^2) / 16 - \mu_2^{-2} \mu_4^2 / 4, \\
B_1 &= \mu_2^{-1} \mu_3 / 2, \\
B_2 &= -3 \mu_2 / 2 + \mu_2^{-1} \mu_4 / 2 + 3 \mu_2^{-2} \mu_3^2 / 4, \\
B_3 &= -3 \mu_3, \\
B_4 &= 2 \mu_2^2, \\
C_0 &= 1/16 - 3 \mu_2^{-2} \mu_3^2 / 4 + \mu_2^{-3} \mu_6 / 8 - 3 \mu_2^{-4} \mu_4^2 / 16, \\
C_1 &= 3 \left( \mu_2^{-1} \mu_3 - \mu_2^{-2} \mu_5 / 4 \right) / 4, \\
C_2 &= 3 \left( -\mu_2 + \mu_2^{-1} \mu_4 + 3 \mu_2^{-2} \mu_3^2 / 4 \right), \\
C_3 &= -10 \mu_3, \\
C_4 &= 6 \mu_2^2.
\end{align*}
\]


Appendix A

Continuing Example 3.4, the terms needed for (3.8)-(3.9) are:

\[
[1, 11]_T = g_0^2 [1, 11]_U + g_0 g_1 \left( U [S_1 U_{11}] + U [U_1 S_{11}] \right) + 2 [U^2 S_1] + g_0 g_2 U [U_1 S^2_1] + g_1 g_2 U^2 \left[ 1^2 \right]_S,
\]

\[
[12^2]_T = g_0^2 [12^2]_U + 2 g_0 g_1 \left( U [U_1 S_{12}] + 2 [U_1 S_2 U_{12}] \right) + g_1^2 \left( U^2 [12^2]_s + 4 U [U_1 S_{2 S_{12}}] + 2 \left[ 1^2 \right]_U [1^2]_S + 2 [U_1 S_1]^2 \right) + 2 g_0 g_2 U [S_1 S_2 U_{12}] + 2 g_1 g_2 U \left( U [1, 2, 12] + 2 [U_1 S_1] [1^2]_S \right) + g_2^2 U^2 \left[ 1^2 \right]_S,
\]

\[
[1, 122]_T = g_0^2 [1, 122]_U + g_0 g_1 \left( U [U_1 S_{122}] + [1^2]_U [11]_S + 2 [U_1 U_2 S_{12}] \right) + 2 [U_1 S_2 U_{12}] + [U_1 S_1] [11]_U + U [S_1 U_{122}] + g_0 g_2 \left( U [U_1 S_1] [11]_S + 2 U [U_1 S_{2 S_{12}}] + [1^2]_U [1^2]_S + 2 [U_1 S_1]^2 \right) + g_0 g_3 U [U_1 S_1] [1^2]_S + g_1^2 U \left( U [1, 122] + [S_1 U_1] [11]_S \right) + 2 [U_1 S_2 S_{12}] + [1^2]_S [11]_U + 2 [S_1 S_2 U_{12}] + g_1 g_2 U \left( U [1^2]_S [11]_S \right) + 2 U [1, 2, 12] + [S_1 U_1] [1^2]_S + 2 [1^2]_S [U_1 S_1] + 2 g_1 g_3 U^2 \left[ 1^2 \right]_S,
\]

\[
[1^4]_T = g_0^4 [1^4]_U + 4 g_0^3 g_1 U [U^3 S_{1}] + 6 g_0^2 g_1^2 U^2 \left[ U^2 S^2_1 \right] + 4 g_0 g_1^3 U^3 \left[ U_1 S^3_1 \right] + g_1^4 U^4 [1^4]_S,
\]
\[ [1, 2^2, 12]_T = g_0^4 [1, 2^2, 12]_U + g_0^3 g_1 \left( U [U_1 U_2^2 S_{12}] + [1^2]_U [U_3 S_{13}] \right) \]

\[ + [1^3]_U [U_1 S_{13}] + 2 U [U_1 U_2 S_{2} U_{12}] + U [U_3^2 S_{2} U_{12}] \]

\[ + g_0^3 g_2 U [U_1 S_{13}] [U_2 S_{12}] \]

\[ + g_0^2 g_1^2 U \left( 2 U [U_1 U_2 S_{2} U_{12}] + 2 [1^2]_U [U_1 S_{12}] + 2 [U_1 S_{13}] [U_2 S_{12}] + [1^3]_U [1^2]_S \right) \]

\[ + g_0^2 g_1 g_2 U^2 \left( 2 [U_1 S_{13}] [U_2 S_{12}] + [1^2]_S [1^2]_S \right) \]

\[ + g_0^2 g_1^2 U^2 \left( U [U_1 S_{2} U_{12}] + [1^2]_U [1^3]_S + [U_1 S_{13}] [U_2 S_{2} U_{12}] + 2 U [S_1 S_{2} S_{2} U_{12}] \right) \]

\[ + g_0 g_1^2 g_2 U^3 [U_1 S_{13}] [1^3]_S \]

\[ + 2 g_1^4 U^3 \left( U [1, 2^2, 12]_S + [U_1 S_{13}] [1^3]_S + [U_1 S_{2} S_{13}] [1^2]_S \right) \]

\[ + 2 g_1^3 g_2 U^4 [1^2]_S [1^3]_S, \]
\[ [1, 2, 13, 23]_T = g_0^4 [1, 2, 13, 23]_U \\
+ g_0^3 g_1 \left( 2U [U_1 S_2 U_{13} U_{23}] + 2U [U_1 U_2 U_{13} S_{23}] + 2 [U_1 S_2 U_{12}] [U_1^2] \\
+ 2 [U_1 S_1] [1, 2, 12]_U \right) \\
+ g_0^3 g_1^2 \left( U^2 [S_1 S_2 U_{13} U_{23}] + 2U^2 [U_1 S_2 U_{13} S_{23}] + 2U^2 [U_1 S_2 U_{13} S_{23}] \\
+ 2U \left[ 1^2 \right]_U [S_1 S_2 U_{12}] + 4U \left[ U_1 S_1 \right] [U_1 S_2 U_{12}] + 2U \left[ 1, 2, 12 \right]_U \left[ 1^2 \right]_S \\
+ U^2 [U_1 U_2 S_{13} S_{23}] + 2U \left[ U_1 S_{12} S_2 \right] [1^2]_U + 2U \left[ U_1 U_2 S_{12} \right] [U_1 S_1] \\
+ \left[ 1^2 \right] U \left[ 1^2 \right]_S + 3 [U_1 S_1] [1^2]_S + [U_1 S_1]^3 \right) \\
+ 2g_0 g_2 U [U_1 S_2 U_{12}] [U_1 S_1] \\
+ 2g_0 g_1 g_2 \left( U^2 [U_1 S_1] [S_1 S_2 U_{12}] + U^2 [U_1 S_2 U_{12}] [1^2]_S + 2U [U_1 S_2 S_{12}] [U_1 S_1] \\
+ U \left[ 1^2 \right] U [U_1 S_1] [1^2]_S + U [U_1 S_1]^3 \right) \\
+ 2g_0 g_1^2 g_2 U^2 \left( U \left[ S_1 S_2 U_{12} \right] [1^2]_S + U \left[ U_1 S_1 \right] [1, 2, 12]_S + U \left[ U_1 S_1 S_{12} \right] [1^2]_S \\
+ 3 [U_1 S_1] [1^2]_S + \left[ 1^2 \right] U \left[ 1^2 \right]_S \right) \\
+ 2g_1^3 g_2 U^3 \left( U \left[ 1^2 \right]_S [1, 2, 12]_S + 2 [U_1 S_1] [1^2]_S \right) \\
+ g_0^2 g_2 U^2 [U_1 S_1] [1^2]_S \\
+ 2g_0 g_1 g_2^2 U^3 [U_1 S_1] [1^2]^2_S \\
+ g_1^2 g_2^3 U^{4} [1^2]^3_S \right).
\[ \begin{align*}
[1, 2, 3, 123]_T & = g_0^4 [1, 2, 3, 123]_U \\
& + g_0^3 g_1 \left( U [U_1 U_2 U_3 S_{123}] + 3 \left[ 1^2 \right]_U [U_1 U_2 S_{12}] + 3 [1, 2, 12]_U [U_1 S_1] \\
& + 3U [U_1 U_2 S_3 U_{123}] \right) \\
& + g_0^3 g_2 \left( 3U [U_1 S_1] [U_1 U_2 S_{12}] + 3 \left[ 1^2 \right]_U [U_1 S_1]^2 \right) \\
& + g_0^3 g_3 U [U_1 S_1]^3 \\
& + 3g_0^2 g_1^2 U \left( U [U_1 U_2 S_3 S_{123}] + 2 \left[ 1^2 \right]_U [U_1 S_2 S_{12}] + [U_1 S_1] [U_1 U_2 S_{12}] \\
& + 2 [U_1 S_1] [U_1 S_2 U_{12}] + [1, 2, 12]_U \left[ 1^2 \right]_S + U [U_1 S_1 S_3 U_{123}] \right) \\
& + 3g_0^2 g_1 g_2 U \left( 2U [U_1 S_1] [U_1 S_2 S_{12}] + U [U_1 U_2 S_{12}] \left[ 1^2 \right]_S \\
& + 2 \left[ 1^2 \right]_U [U_1 S_1] \left[ 1^2 \right]_S + [U_1 S_1]^3 \right) \\
& + 3g_0^2 g_1 g_3 U^2 [U_1 S_1]^2 \left[ 1^2 \right]_S \\
& + g_0^3 U^2 \left( 3U [U_1 S_2 S_3 S_{123}] + 3 \left[ 1^2 \right]_U [S_1 S_2 S_{12}] + 6 [U_1 S_1] [U_1 S_2 S_{12}] \\
& + 3 [U_1 S_1] [S_1 S_2 U_{12}] + 6 [U_1 S_2 U_{12}] \left[ S_1^2 \right] + U [S_1 S_2 S_3 U_{123}] \right) \\
& + 3g_0^2 g_1 g_2 U^2 \left( U [U_1 S_1] [1, 2, 12]_S + 2U [U_1 S_2 S_{12}] \left[ 1^2 \right]_S \\
& + \left[ 1^2 \right]_U \left[ 1^2 \right]_S + 2 \left[ U_1 S_1 \right]^2 \left[ 1^2 \right]_S \right) \\
& + 3g_0^2 g_1 g_3 U^3 [U_1 S_1] \left[ 1^2 \right]_S^2 \\
& + g_1^4 U^3 \left( U [1, 2, 3, 123]_S + 3 [U_1 S_1] [1, 2, 12]_S + 3 [S_1 S_2 U_{12}] \left[ 1^2 \right]_S \right) \\
& + 3g_1^2 g_2 U^3 \left( U [1]^3 [1, 2, 12]_S + [U_1 S_1] \left[ 1^2 \right]_S^2 \right) \\
& + g_3^3 g_4 U^4 \left[ 1^2 \right]_S^3. 
\end{align*} \]
Appendix B

Here, we give bracket functions for moments to order seven, that is expressions for

\[ \left( 1^n s^b t^a u^d \right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mu_{r-1}^{a} \mu_{s-1}^{b} \mu_{t-2}^{c} \mu_{u-12}^{d} \cdots dF_1 (x_1) dF_2 (x_2) \cdots \]

up to \( N = ar + bs + ct + \cdots = 7 \), where now \( \mu_1 \) means \( \mu \). This enables one to obtain the Edgeworth-Cornish-Fisher expansions and bias reduction for any smooth function of \((\mu, \mu_2, \mu_3, \ldots)\). We exclude separable terms like \((t_1^2 t_2^2) = (t_1^2) (t_2^2)\). For each \( N \), we order the terms by its partition functions.

\[
\begin{align*}
N & = 2 \text{ (2 terms) : } \\
1^2 & : (1^1_1) = \mu_2, \\
2 & : (11_2) = -2 \mu_2. \\
N & = 3 \text{ (5 terms) : } \\
1^3 & : (1^1_3) = \mu_3, \\
12 & : (1_1 1_2) = \mu_3, (1_1 1_1 2) = -2 \mu_3, \\
3 & : (1_3) = -6 \mu_3, (1_1 1_1 1_3) = -12 \mu_3. \\
N & = 4 \text{ (13 terms) : } \\
1^4 & : (1^1_4) = \mu_4, \\
1^2 2 & : (1^2_1 1_2) = \mu_4 - \mu_2^2, (1^2_1 1_1 1_2) = -2 \mu_4, \\
13 & : (1_1 1_3) = \mu_4 - \mu_2^2, (1_1 1_1 1_3) = -6 (\mu_4 - \mu_2^2), (1_1 1_1 1_1 1_3) = 12 \mu_4, \\
& \quad (1_1 1_2 2_3) = 12 \mu_2^2, \\
2^2 & : (1^2_2) = \mu_4 - \mu_2^2, (1_2 1_1 1_2) = -2 (\mu_4 - \mu_2^2), (1^2_1 1_2) = 4 \mu_4, (1^2_2) = 4 \mu_2^2, \\
4 & : (1_1 1_1 1_4) = -72 \mu_4, (1_1 1_2 2_4) = -72 \mu_2^2. \\
N & = 5 \text{ (43 terms) : } \\
1^5 & : (1^1_5) = \mu_5, \\
1^3 2 & : (1^3_1 1_2) = \mu_5 - \mu_3 \mu_2, (1^3_1 1_1 1_2) = \mu_5 - 2 \mu_3 \mu_2, (1^3_1 2_1 1_2) = -2 \mu_3 \mu_2, \\
1^2 3 & : (1^2_1 1_3) = \mu_3 \mu_2, (1^2_1 1_1 1_3) = -6 (\mu_5 - \mu_3 \mu_2), (1^2_1 1_1 1_1 1_3) = 12 \mu_5, \\
& \quad (1^2_1 1_2 2_3) = 12 \mu_3 \mu_2, (1_1 2_1 1_2 3) = -2 \mu_4^2, (1_1 1_2 1_1 1_3) = 12 \mu_3 \mu_2, \\
12^2 & : (1^2_1 1_2) = \mu_5 - 2 \mu_3 \mu_2, (1_1 1_2 1_1 2) = -2 (\mu_5 - \mu_3 \mu_2), (1_1 1_2 1_2 2) = 4 \mu_3 \mu_2, \\
& \quad (1^1_1 1_2^2) = 4 \mu_5, (1^1_1 1_2^2) = 4 \mu_3 \mu_2, \\
14 & : (1_1 1_1 1_4) = \mu_5 - 4 \mu_3 \mu_2, (1_1 1_1 1_4) = -8 \mu_5 + 20 \mu_3 \mu_2, \\
& \quad (1_1 1_1 1_1 1_4) = 36 (\mu_5 - \mu_3 \mu_2), (1_1 1_1 1_1 1_1 1_4) = -72 \mu_5, \\
& \quad (1_1 1_2 2_4) = 6 \mu_3 \mu_2, (1_1 1_1 2_2 4) = -72 \mu_3 \mu_2, \\
& \quad (1_1 1_2 2_2 4) = -72 \mu_3 \mu_2, \\
23 & : (1_2 1_3) = \mu_5 - 4 \mu_3 \mu_2, (1_2 1_1 1_3) = -6 (\mu_5 - \mu_3 \mu_2), (1_2 1_1 1_1 1_3) = 12 (\mu_5 - \mu_3 \mu_2), \\
& \quad (1_1 1_1 1_3) = -2 (\mu_5 - 4 \mu_3 \mu_2), (1_1 1_1 1_3) = 12 (\mu_5 - \mu_3 \mu_2), (1_1 1_1 1_1 1_3) = -24 \mu_5, \\
& \quad (1_1 1_2 2_3) = -24 \mu_3 \mu_2, (1_2 1_1 1_2 3) = -24 \mu_3 \mu_2, \\
5 & : (1_1^5) = -10 (\mu_5 - 2 \mu_3 \mu_2), (1_1^5) = 60 (\mu_5 - 2 \mu_3 \mu_2), (1_1^5) = -240 (\mu_5 - \mu_3 \mu_2), \\
& \quad (1_1 1_1 1_1 1_5) = 480 \mu_5, (1_1 1_2 2_5) = 0, (1_1 1_2 2_2 5) = 480 \mu_3 \mu_2.
\end{align*}
\]
\[ N = 6 \text{ (85 terms):} \]

\[ \begin{align*}
1^6 & : \quad (1^6) = \mu_6, \\
1^4 & : \quad (1^4_1) = \mu_6 - \mu_4 \mu_2, \quad (1^4_1 12) = -2 \mu_6, \quad (1^4_1 21 12) = -2 \mu_4 \mu_2, \\
& \quad (1^4_1 2^2 12) = -2 \mu_5^2, \\
1^3 & : \quad (1^3_1) = \mu_6 - 3 \mu_4 \mu_2 - \mu_3^2, \quad (1^3_1 11) = -6 (\mu_6 - \mu_4 \mu_2), \\
& \quad (1^3_1 111) = 12 \mu_6, \quad (1^3_1 122) = 12 \mu_4 \mu_2, \quad (1^3_1 21 12) = -3 (\mu_4 \mu_2 + \mu_3^2 - \mu_2^3), \\
& \quad (1^3_2 111) = 12 \mu_4 \mu_2, \quad (1^3_2 122) = 12 \mu_3^2, \quad (1^3_2 21 12) = 12 \mu_2^3, \\
1^2 & : \quad (1^2_1) = \mu_6 - 4 \mu_4 \mu_2 - 4 \mu_3^2, \quad (1^2_1 11) = -4 (2 \mu_6 - 3 \mu_4 \mu_2 - 2 \mu_2^3), \\
& \quad (1^2_1 111) = -12 (\mu_6 - \mu_4 \mu_2), \quad (1^2_1 1111) = -72 \mu_6, \\
& \quad (1^2_1 122) = -12 (\mu_4 \mu_2 + 2 \mu_3^2 - \mu_2^3), \quad (1^2_1 1122) = -72 \mu_4 \mu_2, \\
& \quad (1^2_2 111) = -72 \mu_3^2, \quad (1^2_2 1111) = -4 (2 \mu_4 \mu_2 - 3 \mu_2^3), \\
& \quad (1^2_2 112) = -12 (2 \mu_4 \mu_2 + \mu_3^2 - 2 \mu_2^3), \quad (1^2_2 1112) = 480 \mu_4 \mu_2, \\
& \quad (1^2_2 1222) = 480 \mu_3^2, \quad (1^2_2 1233) = 480 \mu_2^3, \\
1^2^2 & : \quad (1^2^2_1 12) = 4 \mu_6, \quad (1^2^2_1 12^2) = 4 \mu_4 \mu_2, \quad (1^2^2_1 12 122) = 4 \mu_3^2, \\
& \quad (1^2^2_2 12) = -2 \mu_5^2, \\
123 & : \quad (1^1 112 111) = -24 \mu_6, \quad (1^1 112 122) = -24 \mu_4 \mu_2, \\
& \quad (1^1 122 112) = -24 \mu_4 \mu_2, \quad (1^1 122 122) = -24 \mu_3^2, \\
& \quad (1^1 122 222) = -24 \mu_4 \mu_2, \\
15 & : \quad (1^1_1) = -10 (\mu_6 - \mu_4 \mu_2 - 2 \mu_2^3), \quad (1^1_1 11) = 60 (\mu_6 - \mu_4 \mu_2 - \mu_3^2), \quad (1^1_1 111) = 60 \mu_4 \mu_2, \\
& \quad (1^1_1 1111) = 240 (\mu_6 - \mu_4 \mu_2), \quad (1^1_1 1112) = 120 (\mu_4 \mu_2 + \mu_3^2 - \mu_2^3), \\
& \quad (1^1_1 122) = 60 (3 \mu_4 \mu_2 + \mu_3^2 - 3 \mu_2^3), \quad (1^1_1 1111) = 480 \mu_6, \\
& \quad (1^1_1 1122) = 480 \mu_4 \mu_2, \quad (1^1_1 11222) = 480 \mu_3^2, \\
& \quad (1^1_1 12222) = 480 \mu_4 \mu_2, \quad (1^1_1 12233) = 480 \mu_2^3, \\
2^3 & : \quad (1^3_1) = 2 \mu_6 - 3 \mu_4 \mu_2 - 2 \mu_3^2, \quad (1^3_1 12) = -2 (\mu_6 - 2 \mu_4 \mu_2 + \mu_2^3), \\
& \quad (1^3_2 12) = 2 \mu_6, \quad (1^3_2 12^2) = 2 \mu_6 - \mu_4 \mu_2, \\
& \quad (1^3_2 12 12) = 4 (\mu_6 - \mu_4 \mu_2), \\
& \quad (1^3_2 2^2) = 4 (\mu_4 \mu_2 - \mu_2^3), \quad (1^3_2 12 1) = -8 \mu_6, \\
& \quad (1^3_2 12 2) = -8 \mu_4 \mu_2, \quad (1^3_2 12 2 12) = -8 \mu_2^3, \\
24 & : \quad (1^1 11) = 2 \mu_6 - \mu_4 \mu_2 - 2 \mu_3^2, \quad (1^1 11) = -4 (2 \mu_6 - 3 \mu_4 \mu_2 - 2 \mu_3^3 + 3 \mu_2^3), \\
& \quad (1^2 11) = -36 (\mu_6 - 2 \mu_4 \mu_2 + 2 \mu_2^3), \quad (1^2 111) = 72 (\mu_6 - \mu_4 \mu_2), \\
& \quad (1^2 112) = -72 (\mu_4 \mu_2 - \mu_2^3), \quad (1^2 12) = -2 (\mu_6 - 4 \mu_4 \mu_2 - 2 \mu_2^3), \\
& \quad (1^2 112) = 8 (2 \mu_6 - 3 \mu_4 \mu_2 - 2 \mu_3^2), \quad (1^2 121) = 72 (\mu_6 - \mu_4 \mu_2), \\
& \quad (1^2 121) = 144 \mu_6, \quad (1^2 1221) = 144 \mu_4 \mu_2, \\
& \quad (1^2 122) = 8 (2 \mu_4 \mu_2 - 3 \mu_3^3), \quad (1^2 1221) = -4 (2 \mu_4 \mu_2 + \mu_3^3 - 2 \mu_2^3), \\
& \quad (1^2 1221) = -144 \mu_3^3, \\
3^2 & : \quad (1^3_1) = 2 \mu_6 - 6 \mu_4 \mu_2 - \mu_3^2 + 9 \mu_2^3, \quad (1^3_1 11) = -6 (\mu_6 - 4 \mu_4 \mu_2 - \mu_3^2 + 3 \mu_2^3), \\
& \quad (1^3_1 111) = 12 (\mu_6 - 3 \mu_4 \mu_2 - \mu_2^3), \quad (1^3_1 122) = 12 (\mu_4 \mu_2 - 3 \mu_2^3), \\
& \quad (1^3_1 1222) = 72 (\mu_4 \mu_2 - \mu_3^2), \quad (1^3_1 1222) = 18 (\mu_4 \mu_2 + \mu_3^2 - \mu_2^3), \\
& \quad (1^3_1 1222) = -36 (\mu_4 \mu_2 + \mu_3^2 - \mu_2^3), \\
& \quad (1^3_1 1222) = 144 \mu_3^3.
\end{align*} \]
6: \( (11_6) = -6 (2\mu_6 - 5\mu_4\mu_2), \quad (111_6) = 30 (3\mu_6 - 3\mu_4\mu_2 - 4\mu_3^2), \)
\( (1111_6) = -120 (\mu_6 - \mu_4\mu_2 + \mu_2^3), \quad (11111_6) = -1800 (\mu_6 - \mu_4\mu_2), \)
\( (112_6) = -120 (4\mu_4\mu_2 - 3\mu_3^2), \quad (11112_6) = -360 (3\mu_4\mu_2 + 2\mu_3^2 - 3\mu_2^3), \)
\( (111111_6) = -3600\mu_6, \quad (111122_6) = -3600\mu_4\mu_2, \)
\( (11122_6) = -3600\mu_2^2, \quad (11223_6) = -3600\mu_3^2. \)
As noted for $N = 2, \ldots, 6$ there are 2, 5, 13, 43 and 85 terms, or without $\mu$, 1, 2, 6, 20 and 39 terms. We now give the 90 terms for $N = 7$ without $\mu$.

\[ N = 7 : \]

\[ 2^3 : \]
\[
(1^2_3) = \mu_5 - 4\mu_3\mu_2, \\
(1^2_3) = -6(\mu_7 - 3\mu_5\mu_2 + 3\mu_3\mu_2^2), \\
(1^2_3) = -24\mu_5, \\
(1^2_3) = (12_2) = -24\mu_3\mu_2, \\
(1^2_3) = 2(3\mu_5\mu_2 - \mu_4\mu_3 + \mu_3^2), \\
(1^2_3) = 12(\mu_7 - 2\mu_5\mu_2 + 3\mu_3\mu_2^2), \\
(1^2_3) = -24(\mu_7 - \mu_5\mu_2), \\
(1^2_3) = -12(\mu_5 - 3\mu_3\mu_2)\mu_2, \\
(1^2_3) = 4(\mu_4 - \mu_2^3)\mu_3, \\
(1^2_3) = 6(\mu_5\mu_2 - 2\mu_4\mu_3), \\
(1^2_3) = -24(\mu_5 - \mu_3\mu_2)\mu_2, \\
(1^2_3) = -24(\mu_4 - \mu_2^3)\mu_3, \\
(1^2_3) = -24\mu_4\mu_3, \\
(1^2_3) = -24\mu_3\mu_2^2.
\]

\[ 25 : \]
\[
(1^2_5) = \mu_7 - 5\mu_5\mu_2 - 5\mu_4\mu_3, \\
(1^2_5) = -10(\mu_7 - 5\mu_5\mu_2 + 3\mu_4\mu_3 - 2\mu_3\mu_2^2), \\
(1^2_5) = 60(\mu_7 - 2\mu_5\mu_2 - \mu_4\mu_3 + 2\mu_3\mu_2^2), \\
(1^2_5) = 20(\mu_5\mu_2 + 2\mu_4\mu_3 - 4\mu_3\mu_2^2), \\
(1^2_5) = (12_1) = 480(\mu_7 - \mu_5\mu_2), \\
(1^2_5) = 120(\mu_5\mu_2 + \mu_4\mu_3 - 3\mu_3\mu_2^2), \\
(1^2_5) = (11^2_2) = 480(\mu_5 - \mu_3\mu_2)\mu_2, \\
(1^2_5) = (12_2) = -(12_2) = 480(\mu_4 - \mu_2^2)\mu_3, \\
(1^2_5) = (12_2) = (11^2_2) = 480\mu_4\mu_3, \\
(1^2_5) = (12_2) = (11^2_2) = 480\mu_3\mu_2^2, \\
(1^2_5) = -2(\mu_7 - 5\mu_5\mu_2 - 5\mu_4\mu_3), \\
(1^2_5) = 20(\mu_7 - 3\mu_4\mu_3), \\
(1^2_5) = -120(\mu_7 - \mu_5\mu_2 - \mu_4\mu_3), \\
(1^2_5) = -40(\mu_5\mu_2 + 2\mu_4\mu_3 - 4\mu_3\mu_2^2), \\
(1^2_5) = 240(\mu_7 - 2\mu_5\mu_2 + \mu_3\mu_2^2), \\
(1^2_5) = 120(\mu_5\mu_2 + \mu_4\mu_3 - 3\mu_3\mu_2^2), \\
(1^2_5) = 480(\mu_7 - \mu_5\mu_2), \\
(1^2_5) = -40(2\mu_5\mu_2 + \mu_4\mu_3 - 5\mu_3\mu_2^2), \\
(1^2_5) = -120(3\mu_5\mu_2 + \mu_4\mu_3 - 3\mu_3\mu_2^2), \\
(1^2_5) = -120(3\mu_5\mu_2 + \mu_4\mu_3 - 3\mu_3\mu_2^2), \\
(1^2_5) = -960\mu_5\mu_2, \\
(1^2_5) = -960\mu_4\mu_3, \\
(1^2_5) = -960\mu_3\mu_2^2.
\]
34:  \[(1314) = \mu_7 - 3\mu_5\mu_2 - 5\mu_4\mu_3 + 12\mu_3\mu_2^2,\]
\[(13114) = -4 (2\mu_7 - 9\mu_5\mu_2 - 4\mu_4\mu_3 - 18\mu_3\mu_2^2),\]
\[(131114) = -36 (\mu_7 - 4\mu_5\mu_2 - \mu_4\mu_3 + 4\mu_3\mu_2^2),\]
\[(1311114) = 72 (\mu_7 - 3\mu_5\mu_2 - \mu_4\mu_3),\]
\[(1311224) = 72 (\mu_5 - 4\mu_3\mu_2) \mu_2,\]
\[(1131) = (113114) / 4 = -6 (\mu_7 - \mu_5\mu_2 - 4\mu_4\mu_3 + 4\mu_3\mu_2^2),\]
\[(1131114) = 216 (\mu_7 - 2\mu_5\mu_2 + \mu_3\mu_2^2),\]
\[(11311224) = 432 (\mu_7 - \mu_5\mu_2),\]
\[(113112224) = 432 (\mu_5 - \mu_3\mu_2) \mu_2,\]
\[(12312) = 24 (5\mu_5\mu_2 + \mu_4\mu_3 - 5\mu_3\mu_2^2),\]
\[(123112) = 24 (5\mu_5\mu_2 + \mu_4\mu_3 - 4\mu_3\mu_2^2),\]
\[(12311124) = (12311224) = 216 (5\mu_5\mu_2 + \mu_4\mu_3 - 3\mu_3\mu_2^2),\]
\[(12311224) = 432 (\mu_4 - \mu_2^2) \mu_3,\]
\[(12312334) = 432\mu_3\mu_2^2,\]

7:  \[(117) = -14 (\mu_7 - 3\mu_5\mu_2),\]
\[(11117) = 42 (3\mu_7 - 3\mu_5\mu_2 - 5\mu_4\mu_3),\]
\[(111117) = -840 (\mu_7 - 2\mu_4\mu_3),\]
\[(11227) = -140 (4\mu_5\mu_2 - 3\mu_3\mu_2^2),\]
\[(1111117) = 840 (5\mu_7 - 3\mu_5\mu_2 - 5\mu_4\mu_3),\]
\[(1111227) = 420 (3\mu_5\mu_2 + 2\mu_4\mu_3 - 6\mu_3\mu_2^2),\]
\[(11111117) = 15120 (\mu_7 - \mu_5\mu_2),\]
\[(11111227) = 7! (2\mu_5\mu_2 + \mu_4\mu_3 - 2\mu_3\mu_2^2),\]
\[(11112227) = 15120 (\mu_4 - \mu_2^2) \mu_3,\]
\[(1122337) = 15120\mu_3\mu_2^2,\]
\[(111111117) = 30240\mu_7,\]
\[(111111227) = 30240\mu_5\mu_2,\]
\[(111112227) = 30240\mu_4\mu_3,\]
\[(111122337) = 30240\mu_3\mu_2^2.\]

For $F$ symmetric terms corresponding to odd $N$ reduce to zero.

References

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

[2] D’Agostino, R. and Pearson, E. S. (1973). Tests for departure from normality, empirical results for the distribution of $b_2$ and $\sqrt{a_1}$. *Biometrika*, 60, 613-622.

[3] Fisher, R. A. (1931). The moments of the distribution for normal samples of measures of departure from normality. *Proceedings of the Royal Society London*, A, 130, 16-28.
[4] Loisel, S., Mazza, C. and Rulliere, D. (2009). Convergence and asymptotic variance of bootstrapped finite-time ruin probabilities with partly shifted risk processes. *Insurance: Mathematics and Economics*, **45**, 374-381.

[5] Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probability Theory and Related Fields*, **138**, 411-449.

[6] Stuart, A. and Ord, J. K. (1987). *Kendall’s Advanced Theory of Statistics*, volume 1, fifth edition. Griffin, London.

[7] Volodko, N. V. (2011). Limit theorems for canonical von Mises and $U$-statistics of $m$-dependent observations. *Theory of Probability and Its Applications*, **55**, 271-290.

[8] von Mises, R. (1947). Differentiable statistical functions. *Annals of Mathematical Statistics*, **18**, 309-348.

[9] Withers, C. S. (1982). Second order inference for asymptotically normal random variables. *Sankhyā, B.*, **44**, 19-27.

[10] 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.

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

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

[13] Withers, C. S. (1989). The distribution and cumulants of a Studentised statistic. *Communications in Statistics—Theory and Methods*, **16**, 295-318.
Figure 4.1 Biases of the usual (black) and bias reduced (red) estimators of skewness versus $n = 2, 3, \ldots, 100$. 