ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 10, No.2
June 2012

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 426 364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *300 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT included]

- Single issue ………………………………………………………….. **€ 11**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)………. **€ 26**
- Annual subscription (No. 2, No. 3) ……………………………….. **€ 18**

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 10, Number 2, June 2012, 167–179

MONTE CARLO TEST FOR POLYNOMIAL COVARIATES

Authors: Abdeljelil Farhat
– Research Unit: Applied Economics and Simulation, Mahdia, Tunisia
Abdeljelil.Farhat@fsegma.rnu.tn

Sami Mestiri
– Research Unit: Applied Economics and Simulation, Mahdia, Tunisia
mestirisami2007@gmail.com

Received: December 2009 Revised: November 2011 Accepted: November 2011

Abstract:

- In this paper, we review the score test procedure used for testing polynomial covariate effects in a semi parametric additive mixed model. This test is based on the mixed model representation of the smoothing spline estimator of the nonparametric function and treating the inverse of the smoothing parameter as an extra variance component. Zhang and Lin (2003) found that the score test of polynomial test for non Gaussian responses follows a scaled chi-squared distribution. Simulation studies showed that their approximation is less satisfactory for binary data. To overcome this deficiency, we apply the technique of Monte Carlo in order to obtain provably exact procedures. Derivation and performance of each testing procedure are discussed throughout the simulations that we conducted.

Key-Words:

- *semi parametric additive mixed models; polynomial test; score test; Monte Carlo test.*

AMS Subject Classification:

- 62G08, 62J12, 62H15.

1. INTRODUCTION

Linear mixed models [Laird and Ware (1982)] and their extension, generalized linear mixed models (GLMMs) [Breslow and Clayton (1993); Zeger and Karim (1991)] are popular statistical models for analyzing correlated data. An important feature of these models is that the conditional mean of the response given covariates and random effects, after transformed by a link function, is linearly related to the fixed covariate effects and random effects. Since this parametric assumption in GLMMs is strong and may not be appropriate for data with complex covariate effects, Lin and Zhang (1999) proposed generalized additive mixed models (GAMMs) that allow for flexible modeling of the covariate effects by replacing the linear predictor in GLMMs with an additive combination of nonparametric functions of covariates and random effects. Therefore, it is of practical importance to check the adequacy of the assumption for the parametric linear covariate effects.

In order to evaluate the adequacy of a parametric covariate effect in a regression model, one common approach is to cast the problem in the hypothesis testing framework. In practice, the resulting estimates of a nonparametric function is used as the alternatives for testing the adequacy of the parametric covariate effects. Brumback *et al.* (1999) showed that a nonparametric function estimated via penalized splines or smoothing splines has a mixed effects representation. An appealing feature of using the mixed effects representation is that one can cast the hypothesis test of parametric against nonparametric covariate effects as a variance component test. Zhang and Lin (2003) developed the variance component score test to construct a goodness-of-fit test of polynomial regression in semi parametric additive mixed models (SAMMs), a special case of GAMMs. Due to the special structure of the smoothing matrix, the distribution of statistic score is approximated by a scaled chi-squared distribution. Simulation studies showed that the score test is conservative and not very powerful for binary response.

For checking the adequacy of parametric covariate effects, Huang and Zhang (2008) have presented an overview on score test applied in the context of SAMMs. Their simulations indicate that the score test shows less performance for binary data. In this paper, we propose to use the technique of Monte Carlo (MC) tests in order to improve the test score, for small size sample. Indeed, we adapte MC test to solve the problem of control the power of score test. The MC approach allows us to introduce a new test that differs in two respects from the tests existing in the literature. First, the test is exact in the sense that the probability of rejecting the null hypothesis when it is true is always less than or equal to the nominal level of the test. Secondly, this approach allows to obtain exact randomized test using very small numbers of MC replications of the original test statistic under the null hypothesis. Finally, MC test is a reliable and easy instrument for testing polynomial degree of non parametric function. So, the aim of this paper is to solve the problem of distortion of power of score test. By conducting simulations studies, we show that the MC technique can improve the power of test.

The paper is organized as follows. Section 1 introduces the SAMMs model. In Section 2, we describe the polynomial tests in SAMMs and we describe how exact MC tests can be implemented. In Section 3, we present the results from a small simulation study to compare the performance of the asymptotic score test and the MC test. The paper is concluded in Section 4 with some discussion.

2. THE MODEL SPECIFICATION

In this section, we briefly present SAMMs for clustered data, and this estimation procedure. These models are special cases of GAMMs considered by Lin and Zhang (1999). Let the data consist of a response variable y_{ij} for the j^{th} observation $(j = 1, .., n)$ of the i^{th} cluster $(i = 1, .., N)$, a scalar covariate x_{ij}, and a scalar covariate s_{ij} associated with fixed effects, and a scalar covariate z_{ij} associated with random effects. Conditional on a $(q, 1)$ vector of random effects b_i, the y_{ij} are assumed to be independent with conditional means $E(y_{ij}|b_i) = \mu^b_{ij}$ and conditional variances $\operatorname{var}(y_{ij}|b_i) = \phi\, \varpi^{-1}_{ij} v(\mu^b_{ij})$, where ϕ is a dispersion parameter, ϖ_{ij} is a known prior weight, and v is a variance function. The SAMM assumes that the conditional mean μ^b_{ij} takes this form:

$$g(\mu^b_{ij}) \;=\; s_{ij}\,\gamma + f(x_{ij}) + z_{ij}\, b_i \;, \tag{2.1}$$

where γ is a fixed effect, $f(x)$ is an arbitrary smooth function and g is a known link function. b_i is a random effect associated with covariates z_{ij}. It is further assumed that the random effects b_i are independent and have a normal distribution $N(0, \sigma^2_b)$.

We propose to transforme the model (2.1) to fully parametric model where the unknown smoothing function $f(x_{ij})$ may be expressed as a linear combination of proper basis functions. We consider the truncated power basis usually used in this context, as by Ruppert *et al.* (2003) or Ngo and Wand (2004). A penalized linear spline model for (2.1) is

$$g(\mu^b_{ij}) \;=\; s_{ij}\,\gamma + \sum_{h=1}^{H} \delta_h\, x^h_{ij} + \sum_{k=1}^{K} a_k (x_{ij} - \kappa_k)_+ + z_{ij}\, b_i \;, \tag{2.2}$$

where $\kappa_1, ..., \kappa_K$ is a set of distinct knots in the range of x_{ij} and $x_+ = \max(0; x)$. The number of knots K is fixed and large enough (in this case $K = 40$) to ensure the exibility of the curve. The knots are chosen as quantiles of x with probabilities $1/(K+1), ..., K/(K+1)$. We use truncated lines as the basis for regression since their simple mathematical form is very useful in formulating complicated models. More complex basis such as B-splines and radial basis functions (with better numerical properties) could also be used.

Let Y, X, B and b denote the matrix obtained from stacking up the N subject-specific vectors of the same symbol. Also, let $Z = \mathrm{diag}(Z_1, ..., Z_N)$ and $a = (a_1, ..., a_K)'$. Zhang and Lin (2003) suggested that a can be treated as random effects following $N(0, \tau I)$, so the model (2.2) is considered as a linear combination of the fixed effects β and the random effects a and b. Under this mixed-model representation of the smoothing spline estimator of f, the SAMM (2.1) can be written as the following GLMM:

$$g(\mu^b) = X\beta + Ba + Zb\ , \tag{2.3}$$

where β is the fixed effect associated with covariates matrix X. The vector a is Normal $(0, \tau I)$, the independent random effect b is Normal $(0, \sigma_b^2)$. This GLMM representation takes the same form as that Lin and Zhang (1999) used for natural cubic spline estimators. For detailed justification of the estimation procedure, see Lin and Zhang (1999).

3. THE POLYNOMIAL TEST

3.1. Asymptotic score test

Zhang and Lin (2003) considered the problem of testing the nonparametric function $f(x)$ in model (2.1) being a h-order polynomial. They first estimated $f(x)$ by a h-order smoothing spline and expressed f with a mixed effects representation. Then, they tested if $f(x)$ is h order. Testing $f(x)$ in SAMM (2.1) being a h-order polynomial is equivalent to testing $H_0\colon \tau = 0$ in the induced GLMM in (2.3). Under the induced GLMM in (2.3), Zhang and Lin (2003) showed that the score U_τ for testing $H_0\colon \tau = 0$ takes the following form:

$$\begin{aligned} U_\tau(\hat\psi) &= \left.\frac{\partial l_M(\tau, \psi; y)}{\partial \tau}\right|_{\tau=0} \\ &= \frac{1}{2}\Big\{(y - X\beta)' V^{-1} B B' V^{-1} (y - X\beta) - \mathrm{tr}(PBB')\Big\}\Big|_{\beta=\hat\beta, \hat\psi}\ , \end{aligned} \tag{3.1}$$

where $\psi = (\sigma_b^2, \phi)$ is the nuisance parameter vector, and $l_M(\tau, \psi)$ is the marginal log likelihood function of τ and ψ (by integrating out random effects b and fixed effects β). $\hat\beta$ is the maximum likelihood estimator (MLE) of β and $\hat\psi$ is the restricted maximum likelihood estimator (REML) of ψ, and $Y = X\beta + Zb + \Delta(y - \mu^b)$ is the working vector from the following null GLMM

$$g(\mu^b) = X\beta + Zb \ , \tag{3.2}$$

where $\Delta = \operatorname{diag}\{g'(\mu_{ij}^b)\}$, $W = \operatorname{diag}\{\omega_{ij}\}$ is the working weight and $\omega_{ij} = \{\phi\, \varpi_{ij}^{-1} v(\mu_{ij})\, [g'(\mu_{ij})]^2\}^{-1}$, $b \sim N(0, \sigma_b^2)$, $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'$ and $V = W^{-1} + ZDZ'$. W is working weight matrix evaluated at the conditional expectation μ^b and taken under the null hypothesis $\tau = 0$. One can use the existing software such as the R packages (glmmPQL) to obtain the estimates $\hat{\beta}$ and $\hat{\psi}$ by fitting the model (3.2).

Zhang and Lin (2003) showed that the null distribution of U_τ can be approximated by a scaled chi-squared distribution. A major problem in the score test context comes from the fact that applicable procedure rely heavily on asymptotic approximations whose accuracy can be quiet poor. This is evident from the study simulation reported in Zhang and Lin (2003). In any case, it is widely acknowledged that score asymptotic test is unreliable in finite sample, in the sense that the test was a little conservative and not very powerful. We reemphasize this fact and propose to use the technique of Monte Carlo test [Dwass (1957), Barnard (1963), Dufour and Khalaf (2002)] in order to obtain provably exact procedures.

3.2. Monte Carlo test

In this paper, we describe the MC test methodology for testing the polynomial degree of $f(x)$. In effect, it is possible to apply the test of MC because the statistic of score U_τ under the null distribution is a continuous pivotal function (its distribution does not depend on unknown parameter). Let U_0 denote the observed test statistic of score calculated on the basis of data observed. Then the critical region of a test with level α can be expressed as $G(U_0) \le \alpha$ such as $G(U_0) = P(U \ge U_0/H_0)$ is the critical function for a right tailed test. $G(U_0)$ is unknown and it will be estimated by generating under null assumption M independent replications or exchangeable statistics $U_1, .., U_M$ [see Dwass (1957) and Dufour *et al.* (1998)]. For the application of the technique of the tests of MC, we define

$$\hat{G}_M(U_0) = \frac{1}{M}\sum_{i=1}^{M} I_{[0,\infty)}\big(U_i - U_0\big)\ , \qquad I_A(z) = \begin{cases} 1, & \text{if } z \in A\,, \\ 0, & \text{if } z \notin A\,. \end{cases} \tag{3.3}$$

In other words, $M\hat{G}_M(U_0)$ is the number of simulated statistics which are greater or equal to U_0, $\hat{R}_M(U_0) = M - M\hat{G}_N(U_0) + 1$ gives the rank of U_0 among the variables $U_0, U_1, .., U_M$. The estimated critical function is then given by this formula:

$$\hat{p}_M(U_0) \;=\; \frac{M\hat{G}_M(U_0) + 1}{M+1} \;. \tag{3.4}$$

Thus the critical region of level α associated with a test MC is expressed by $\hat{p}_M(U_0) \leq \alpha$ such as $\hat{p}_M(U_0)$ represents the empirical probability that the value more superior than U_0 is realized if the null hypothesis is true. Hence $\hat{p}_M(U_0)$ may be viewed as a MC-value. Note that the MC decision rule may also be expressed in terms of $\hat{R}_M(U_0)$. Indeed the critical region $\frac{M\hat{G}_M(kU_0)+1}{M+1} < \alpha$ is equivalent to $\hat{R}_M(U_0) \geq (M+1)(1-\alpha)+1$.

In other words, the MC test is significant at a 5% level if the rank of U_0 in the series $U_0, U_1, .., U_M$ is at least equal to 96. If the null distribution of U_0 is nuisance-parameter-free and $\alpha(M+1)$ is an integer, then the critical region is probably exact, in the sense

$$P\Big[\hat{p}_M(U_0) \leq \alpha\Big] = \alpha \tag{3.5}$$

or alternatively

$$P\Big[\hat{R}_M(U_0) \geq (M+1)(1-\alpha)+1\Big] \;=\; \alpha \;. \tag{3.6}$$

The proof of the equation (3.5) and (3.6) is based on the theorem concerning the distribution of the ranks associated with a finite dimensional array of exchangeable statistics; see Dufour (2006) for more informations.

The determination of the Monte Carlo p-Value for the polynomial degree test to the model (2.1), is described as follows:

- Fit the model on original data set $Y^{(0)}$ and calculate the ML estimates $\hat{\beta}$, $\hat{\psi} = (\hat{\sigma}_b^2, \hat{\sigma}_\varepsilon^2)$ and $\hat{\tau}$.
- Obtain the score statistic based on $\hat{\psi}$ and denote it $U_\tau^{(0)}$.
- Treat $\hat{\psi}$ as fixed and fitted from the null model $g(\mu) = X\beta + Zb$ (under the null hypothesis $H_0\colon \tau = 0$ and $\psi = \hat{\psi}$), repeat the following steps for $m = 1, .., M$.
  - Draw the vector $\tilde{b}^{(m)}$ as i.i.d. $N(0, \hat{\sigma}_b^2)$ and the vector $\tilde{\varepsilon}^{(m)}$ as i.i.d. $N(0, \hat{\sigma}_\varepsilon^2)$.
  - Obtain the simulated independent variable $\tilde{Y}^{(m)} = X\beta + Z\tilde{b}^{(m)} + \tilde{\varepsilon}^{(m)}$ where $\tilde{Y} = X\beta + Ba + Zb + \Delta(y - \mu^b)$ is working vector, such as $\Delta = \mathrm{diag}\{g'(\mu_{ij}^b)\}$.
  - Regress $\tilde{Y}^{(m)}$ on X and B (fit the model $g(\mu) = X\beta + Ba + Zb$ on simulated data set).

  – Derive the score statistic test $U_1, ..., U_M$ associated with the regression of $\tilde{Y}^{(m)}$ on X and B.

- Obtain the rank $\hat{R}_M(U_0)$ in the series $U_0, U_1, ..., U_M$.
- Reject the null $H_0\colon \tau = 0$ if $\hat{R}_M(U_0) \geq (M+1)(1-\alpha)+1$.

Furthermore, a MC p-value may be obtained as $\hat{p}_M(U_0) = \frac{M+1-\hat{R}_M(U_0)}{M+1}$. We choose M so that $\alpha(M+1)$ is an integer (for example, for $\alpha = 0.05$; we can take $M = 19; 39; 99...$).

MC test can be interpreted as a parametric bootstrap method applied to statistics whose null distribution does not depend on nuisance parameters. However the central additional observation is that the randomization allows one to exactly control the size of the test for a given (possibly small) number of MC simulations. For further discussion of Monte Carlo tests (including its relation with the bootstrap), Kiviet and Duffour (1997) and Dufour *et al.* (1998).

4. SIMULATION EXPERIMENTS

In order to assess the performance of two test procedures discussed above, we conduct a small simulation study. The performance of the polynomial test are evaluated and compared for clustered data with different types of responses and different magnitudes of correlation. For illustration purpose, we consider testing the linearity of covariate effects under the partially linear model framework, i.e. whether $f(x)$ is a linear function of x in model (2.1). Following the penalized spline, we formulate the asymptotic score test as variance component test based on the GLMM representation (2.3) as discussed above. In addition, for testing the same null hypothesis, we also formulate the Monte Carlo test which is exact in the sense that the probability of rejection the null hypothesis when is true is always less than or equal to the nominal size of the test. In our case, we are testing whether $f(x)$ is a 1-degree polynomial of x. Conditional on the cluster-specific random intercept $b_i \sim N(0, \sigma_b)$ with $\sigma_b = 0.5$ and $\sigma_b = 1$, independent Gaussian and Binary responses y_{ij} (for $i = 1, ..., N$ and $j = 1, .., n$) were generated respectively under the model

$$g(\mu_{ij}) \;=\; \gamma_0 + s_{ij}\,\gamma_1 + f(x_{ij}) + b_i \;, \tag{4.1}$$

where $g(\mu) = \mu$ for Gaussian response, and $g(\mu) = \mathrm{logit}(\mu)$ for Binary responses. The scale parameter ϕ was estimated for Gaussian responses and was set to be one for Binary responses. Where s_{ij} is generated from Normal law $N(0, 0.1)$, x_{ij} is generated from Uniform law ($U[0,1]$), the true values of γ_0 and γ_1 are taken to be $\gamma_0 = 1$ and $\gamma_1 = 2$; two sample sizes are used ($N = 2,\; n = 5$) and ($N = 4,\; n = 5$);

five different functions of $f(x)$ are considered:

$$f_c(x) = (0.25\,c)\,x \cdot \exp(2 - 2x) - x + 0.5\,, \qquad \text{for} \quad c = (0, 1, 2, 3, 4)\ .$$

Note that when $c = 0$, $f_c(x)$ is a linear function of x and $f_c(x)$ deviates further from linearity with increasing c. The functions $f_c(x)$ are plotted in Figure 1.

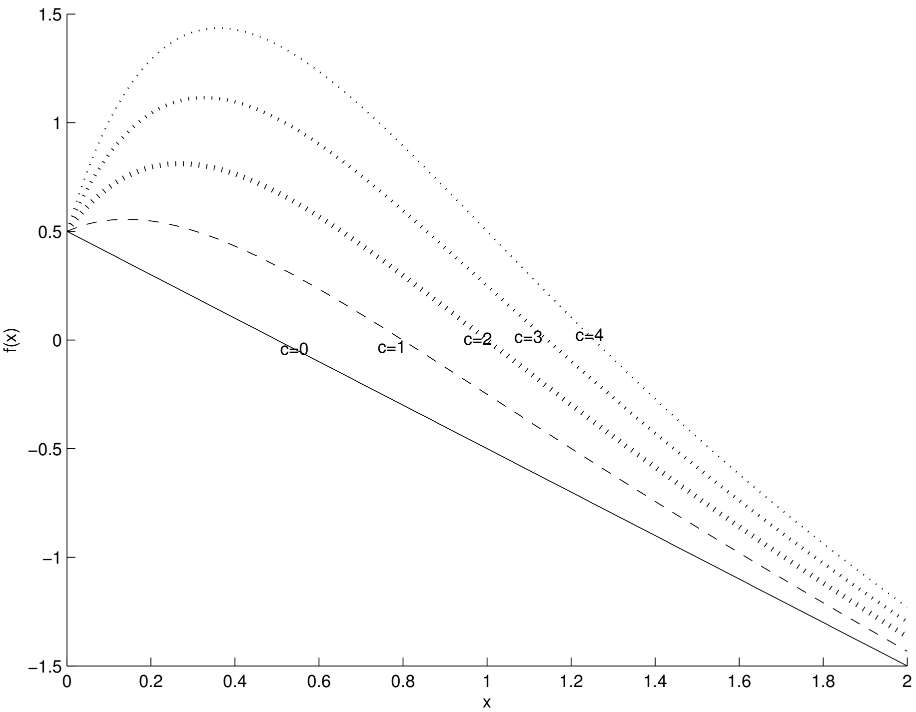


Figure 1: Functions $f_c(x)$ for $c = (0, 1, 2, 3, 4)$ used in the simulation studies for the polynomial test.

We apply the Asymptotic score (*Asy.Sco*) and the Monte Carlo (*MC.Sco*) testing procedures to each simulated data set. The simulation results are based on 1000 Monte Carlo simulation runs. For testing the null hypothesis that $f(x)$ is a linear function of x, the size and the power of each testing procedures are calculated by setting $c = 0$ and $c \neq 0$ respectively. We used a penalized spline to estimate $f(x)$, the number of knots for the penalized spline is set to be 40. The number of trials for the MC test was set to 19. The number of overall replications was 1000. All experiments were performed with language R (version 7.2.1). The simulation results are presented in the table 1 and 2, which report rejection percentages from 1000 replications at the nominal level 5% under the null hypothesis.

Table 1: Empirical sizes and powers of linearity test for two types of data where $N = 2$ and $n = 5$.

Variance random effect	Data Type	Test	Size	Powers			
			$c = 0$	$c = 1$	$c = 2$	$c = 3$	$c = 4$
$\sigma_b = 0.05$	Gaussian	Asy	0.055	0.178	0.624	0.934	0.995
		MC	0.051	0.211	0.645	1.000	1.000
	Binary	Asy	0.033	0.073	0.167	0.260	0.511
		MC	0.054	0.291	0.711	0.887	1.000
$\sigma_b = 1$	Gaussian	Asy	0.048	0.202	0591	0.942	0.995
		MC	0.054	0.251	0.671	1.000	1.000
	Binary	Asy	0.040	0.068	0.120	0.271	0.442
		MC	0.061	0.125	0.741	0.910	1.000

Table 2: Empirical sizes and powers of linearity test for two types of data where $N = 4$ and $n = 5$.

Variance random effect	Data Type	Test	Size	Powers			
			$c = 0$	$c = 1$	$c = 2$	$c = 3$	$c = 4$
$\sigma_b = 0.05$	Gaussian	Asy	0.055	0.199	0.627	0.914	0.990
		MC	0.050	0.223	0.775	1.000	1.000
	Binary	Asy	0.047	0.095	0.211	0.310	0.621
		MC	0.052	0.325	0.812	0.970	1.000
$\sigma_b = 1$	Gaussian	Asy	0.045	0.207	0.603	0.922	0.995
		MC	0.050	0.304	0.789	1.000	1.000
	Binary	Asy	0.042	0.077	0.211	0.314	0.511
		MC	0.050	0.301	0.805	0.960	1.000

The results showed that the asymptotic score test for Binary responses was a conservative and not very powerful. The increased sample size brings the empirical sizes of the two tests closer to the nominal levels, whereas the variance component seems to have not much influence on them. These tests show decreased power where the variance component increases. Regarding the empirical size, our simulation results show that the linearity test with Monte Carlo is very performant for Gaussian responses for different magnitudes of the variance component. The empirical sizes were very close to the nominal size and the powers of the test were high, and were not significantly affected by the magnitude of the variance component. Indeed, the table 1 and 2 show that MC test achieves a perfect size control for Binary responses for different magnitudes of the variance component. As expected, the increased sample size improves the overall power.

In fact the simulation results show clearly that the technique of MC test correct size distortion due to poor large sample approximations. In general, our simulation indicates that the MC test is more powerful than the asymptotic score test. For simplicity, only the linearity test is considered in the current simulation; however in practice, one might be interested in testing higher-order polynomial covariate effects (i.e. $h > 1$), which can be easily carried out by using a different values of h.

5. DISCUSSION

We have reviewed in this paper a test procedure for testing whether the nonparametric function is some fixed-degree polynomial. The key idea is based on the mixed-effect representation of the natural spline estimator of the nonparametric function. Zhang and Lin (2003) developed score test and approximated its distribution by a scaled Chi-square distribution. For Binary data, the simulation studies show that the performance of the test is less satisfactory. This is mainly due to the less satisfactory performance of the Laplace approximation for the score statistic and the implicit Gaussian fourth-moment assumption when estimating the variance of the score statistic. We have hence proposed the simulation based procedures to derived exact p-value for polynomial test for SAMM. We have exploited the fact the score test is pivotal under the null hypothesis which allows one to apply the technique of MC tests.

The feasibility of our approach was illustrated trough a simulation experiment. The results show that asymptotic score test is unreliable for binary response in contrast MC test achieve perfect size control and have a good power. It is important to emphasis that MC procedure require less calculation with modern computer facilities. Zhang and Davidian (2004) have proposed a conditional estimation procedure built on likelihood inference for generalized additive mixed models. It is interesting for future research to extend our Monte Carlo test considering the conditional estimation procedure.

However, The score test is sensitive to outliers. Recently, Qin and Zhu (2008) focus on robust estimation of mean parameters of partial linear mixed model. They proposed to approximate the nonparametric function f by a regression spline and to estimate both the linear parameter and the spline coefficients by a M-estimator. It is interesting for future research to extend our Monte Carlo test considering the robust score test.

ACKNOWLEDGMENTS

The authors acknowledge helpful suggestions by the editor and the referees of this paper.

REFERENCES

[1] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.

[2] Barnard, G.A. (1963). Comment on 'The Spectral Analysis of Point Processes' by M.S. Bartlett, *Journal of the Royal Statistical Society, Series*, **25**, 294.

[3] Brumback, Babette A.; Ruppert, David and Wand, M.P. (1999). Variable selection and function estimation in additive nonparametric regression using a data-based prior: Comment, *Journal of the American Statistical Association*, **94**, 794–7.

[4] Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics, *Journal of Econometrics*, **133**(2), 443–477.

[5] Dufour, J.-M. and Khalaf, L. (2002). Simulation based finite and large sample tests in multivariate regressions, *Journal of Econometrics*, **111**, 303–322.

[6] Dufour, J.-M.; Farhat, A.; Gardol, L. and Khalaf, L. (1998). Simulation-based finite sample normality tests in linear regressions, *The Econometrics Journal*, **1**, 154–173.

[7] Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses, *Annals of Mathematical Statistics*, **28**, 181–187.

[8] Huang, M. and Zhang, D. (2008). Testing polynomial covariate effects in linear and generalized linear mixed models, *Statistics Surveys*, **2**, 154–169.

[9] Qin, G.Y. and Zhu, Z.Y. (2008). Robust estimation in partial linear mixed model for longitudianl data, *Acta Mathematica Scientia*, **28**, 334–348.

[10] Kiviet, J.F. and Duffour, J.-M. (1997). Exact tests in single equation autoregressive distributed lag models, *Journal of Econometrics,* **80**, 325–353.

[11] Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data, *Biometrics*, **38**, 963–974.

[12] Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models using smoothing splines, *Journal of the Royal Statistical Society*, **61**, 381–400.

[13] Ngo, L. and Wand, M.P. (2004). Smoothing with mixed model software, *Journal of Statistical Software*, **9**, 1–54.

[14] Ruppert, D.; Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press.

[15] Zhang, D. and Davidian, M. (2004). Likelihood and conditional likelihood inference for generalized additive mixed models for clustered data, *Journal of Multivariate Analysis*, **91**, 90–106.

[16] Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models, *Biostatistics*, **4**, 57–74.

[17] Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects; A Gibbs sampling approach, *Journal of the American Statistical Association* **86**, 413, 79-86.

REVSTAT – Statistical Journal
Volume 10, Number 2, June 2012, 181–210

ON AN EXTREME VALUE VERSION OF THE BIRNBAUM–SAUNDERS DISTRIBUTION

Authors: Marta Ferreira
– Departamento de Matemática, Universidade do Minho, Braga, Portugal
msferreira@math.uminho.pt

M. Ivette Gomes
– D.E.I.O. (F.C.U.L.) and C.E.A.U.L., Universidade de Lisboa, Lisboa, Portugal
ivette.gomes@fc.ul.pt

Víctor Leiva
– Departamento de Estadística, Universidad de Valparaíso, Valparaíso, Chile
victor.leiva@uv.cl

Received: September 2011 Revised: November 2011 Accepted: November 2011

Abstract:

- The Birnbaum–Saunders model is a life distribution originated from a problem of material fatigue that has been largely studied and applied in recent decades. A random variable following the Birnbaum–Saunders distribution can be stochastically represented by another random variable used as basis. Then, the Birnbaum–Saunders model can be generalized by switching the distribution of the basis variable using diverse arguments allowing to construct more general classes of models. Extreme value distributions are useful to determinate the probability of events that are larger or smaller than others previously observed. In this paper, we propose, characterize, implement and apply an extreme value version of the Birnbaum–Saunders distribution.

Key-Words:

- *domain of attraction; extreme data; likelihood method; R computer language.*

AMS Subject Classification:

- 60E05, 62G32, 62N02, 62N99.

1. INTRODUCTION

Extreme value (EV) models are appropriate to establish the probability of events that are larger or smaller than others previously observed. As an example where these models can be used, suppose that a sea-wall projection requires a coastal defence from all sea levels for the next 100 years. These EV models are a precious tool that enables this type of extrapolations. Actually, the EV theory is widely used by many researchers in applied sciences when high values of certain phenomena are modeled. For instance, ocean wave, thermodynamics of earthquakes, wind energy, risk assessment on financial markets, and medical phenomena can be mentioned. Some books on EV theory are Leadbetter *et al.* (1983), Galambos (1987), Embrechts *et al.* (1997), Beirlant *et al.* (2004), and de Haan and Ferreira (2006). For a more practical view on this topic, see Coles (2001), and for more recent references, see Ferreira and Canto e Castro (2008), Gomes *et al.* (2008a,b), Beirlant *et al.* (2012), and Scarrot and MacDonald (2012), among others.

Life distributions are usually positively skewed, unimodal, two-parameter models and with non-negative support; see Marshall and Olkin (2007) and Saunders (2007). A life distribution that has received a considerable attention in recent decades is the Birnbaum–Saunders (BS) model. This model was originated from a problem of material fatigue and has been largely applied to reliability and fatigue studies; see Birnbaum and Saunders (1969). The BS distribution relates the total time until the failure to some type of cumulative damage normally distributed. This attention for the BS distribution is due to its many attractive properties and its relationship with the normal distribution.

Extensive work has been done on the BS model with regard to its properties, inference and applications. A comprehensive treatment on this model until mid 90's can be found in Johnson *et al.* (1995, pp. 651–662). For more detail about new applications of the BS model, see Leiva *et al.* (2009a). For applications in fields beyond engineering allowing business, environmental and medical data to be analyzed by using this model, see Leiva *et al.* (2007, 2008b, 2009b, 2010a,b, 2011), Podlaski (2008), Barros *et al.* (2008), Bhatti (2010), Ahmed *et al.* (2010), and Vilca *et al.* (2010). Thus, at the present, the BS model can be widely used as a statistical distribution rather than restricted to a life distribution.

Because a random variable (r.v.) following the BS distribution can be represented by another basis r.v., generalizations of this distribution can be obtained switching the distribution of the basis variable by diverse arguments allowing to construct more general classes of models. Several generalizations of the BS distribution have been recently proposed by a number of authors, including Díaz-García and Leiva (2005), Vilca and Leiva (2006), Sanhueza *et al.* (2008), Gómez *et al.* (2009) and Guiraud *et al.* (2009), which allow us to obtain a major degree of flexibility for this distribution. Usual and generalized versions of the BS distribution are implemented in the R software (`http://www.R-project.org`) by packages called `bs` and `gbs`, which can be downloaded from `http://CRAN.R-project.org`; see Leiva *et al.* (2006) and Barros *et al.* (2009). These packages contain functions for computing probabilities, estimating parameters, generating random numbers and carrying out goodness-of-fit and hazard analysis. Leiva *et al.* (2008a) studied three generators of random numbers from the BS and generalized BS (GBS) distributions.

The main aim of this work is to obtain an EV version of the BS distribution relevant not only by itself as a model, but also for a parametric statistical analysis of extreme or rare events. The paper is organized as follows. In Section 2, we provide a preliminary notion of different aspects related to BS and EV distributions. In Section 3, we characterize extreme value Birnbaum–Saunders (EVBS) distributions. In Section 4, we focus on extremal domains of attraction of a general class of BS models that we call BS type (BST) distributions. In Section 5, we carry out a hazard analysis of EVBS distributions mainly based on the hazard rate (h.r.). In Section 6, we discuss about the estimation procedure based on the maximum likelihood (ML) method and model checking. In Section 7, we conduct out the numerical application of this work, which includes an exploratory data analysis (EDA) and a parametric statistical analysis based on the EVBS distribution. Finally, in Section 8, we sketch some concluding remarks.

2. A PRELIMINARY NOTION

In this section, we provide preliminary aspects about BS, BST and EV distributions.

2.1. BS and BST distributions

An r.v. T with usual BS distribution is characterized by its shape and scale parameters $\alpha > 0$ and $\beta > 0$, respectively. This is denoted by $T \sim \mathrm{BS}(\alpha, \beta)$, where β is also the median of the distribution. BS and standard normal r.v.'s, denoted respectively by T and Z for now, are related by

$$(2.1) \qquad T = \beta\Big(\alpha Z/2 + \sqrt{\{\alpha Z/2\}^2 + 1}\Big)^2 \quad \text{and} \quad Z = \Big(\sqrt{T/\beta} - \sqrt{\beta/T}\Big)/\alpha\ .$$

Let $T \sim \mathrm{BS}(\alpha, \beta)$. Then, the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of T are respectively given by

$$(2.2) \qquad f_T(t) = \phi\big(a(t)\big)\, a'(t) \quad \text{and} \quad F_T(t) = \Phi\big(a(t)\big)\ , \qquad t > 0\ ,$$

where ϕ and Φ are the standard normal p.d.f. and c.d.f., respectively,

$$(2.3)\quad a(t) \equiv a_t = \left(\sqrt{t/\beta} - \sqrt{\beta/t}\right)/\alpha \quad \text{and} \quad a'(t) \equiv A_t = t^{-3/2}\left(t+\beta\right)/\left(2\alpha\sqrt{\beta}\right),$$

with $a'(t) = da(t)/dt$ being the derivative of $a(t)$ with respect to t. The quantile function (q.f.) of T is expressed as

$$(2.4)\qquad t(q) \equiv t_q = F_T^{-1}(q) = \beta\left(\alpha\,\xi_q/2 + \sqrt{\{\alpha\,\xi_q/2\}^2 + 1}\right)^2, \qquad 0 < q < 1,$$

where $F_T^{-1}(t) := \inf\{x\colon F(x) \geq t\}$ is the generalized inverse function of the c.d.f. of T and ξ_q is the q^{th} quantile of the r.v. $Z \sim \mathrm{N}(0,1)$. Note from (2.4) that, as mentioned, the median of T is $t_{0.5} = \beta$.

Important properties of $T \sim \mathrm{BS}(\alpha,\beta)$ are: (i) $c\,T \sim \mathrm{BS}(\alpha, c\,\beta)$, $c > 0$; (ii) $1/T \sim \mathrm{BS}(\alpha, 1/\beta)$; and (iii) $V = (T/\beta + \beta/T - 2)/\alpha^2 \sim \chi^2(1)$, i.e., V follows the χ^2 distribution with one degree of freedom (d.f.).

The assumption given in (2.1) can be relaxed supposing that Z follows any other distribution with p.d.f. f_Z. Thus, we obtain the general class of BST distributions earlier mentioned, which is denoted by $T \sim \mathrm{BST}(\alpha,\beta; f_Z)$ for an associated r.v. T and whose p.d.f. is given by

$$(2.5)\qquad f_T(t) = f_Z\big(a(t)\big)\,a'(t), \qquad t > 0.$$

In particular, if Z follows a standard symmetric distribution in the real number set, denoted by $Z \sim \mathrm{S}(f_Z)$, we then find the GBS distribution, i.e., $T \sim \mathrm{GBS}(\alpha,\beta; g)$, where g is the kernel of the p.d.f. of Z given by $f_Z(z) = c\,g(z^2)$, with $z \in \mathbb{R}$ and c being the normalization constant, i.e., the positive value such that $\int_{-\infty}^{+\infty} g(z^2)\,dz = 1/c$; see Díaz-García and Leiva (2005). Then, if $f_Z(z) = \phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$, for $z \in \mathbb{R}$, the standard normal p.d.f., we obviously recover the usual BS distribution, i.e., an r.v. $T \sim \mathrm{BST}(\alpha,\beta;\phi) \equiv \mathrm{BS}(\alpha,\beta)$; see Birnbaum and Saunders (1969). For the GBS case, $V = (T/\beta + \beta/T - 2)/\alpha^2 \sim \mathrm{G}\chi^2(1; f_Z)$, i.e., V follows the generalized χ^2 class of distributions with one d.f., which has the $\chi^2(1)$ distribution as a special case if f_Z is the standard normal density; see Sanhueza *et al.* (2008).

2.2. EV distributions and extremal domains of attraction

The central limiting result in EV theory states the following. Consider an independent identically distributed sequence of r.v.'s $\{X_n,\, n \geq 1\}$, with marginal c.d.f. F. Hence, if there are constants $a_n > 0$ and $b_n \in \mathbb{R}$, and a non-degenerate c.d.f. G such that, as $n \to \infty$,

$$(2.6)\qquad \mathbb{P}\Big(\max\{X_1, ..., X_n\} \leq a_n\,x + b_n\Big) \;\to\; G(x),$$

then G must be the c.d.f. of a generalized extreme value (GEV) r.v., depending on a parameter $\gamma \in \mathbb{R}$. The notation $X \sim \text{GEV}(\gamma)$ is used in this case and the corresponding c.d.f. is given by

$$(2.7) \qquad G(x) \equiv G_\gamma(x) = \begin{cases} \exp\bigl(-\{1+\gamma x\}^{-1/\gamma}\bigr)\,; & 1+\gamma x > 0, \quad \gamma \in \mathbb{R}\backslash\{0\}\ , \\ \exp\bigl(-\exp(-x)\bigr)\,; & x \in \mathbb{R}, \quad \gamma = 0\ , \end{cases}$$

with $G_0(x)$ obtained from $G_\gamma(x)$, for $\gamma \in \mathbb{R}\backslash\{0\}$, as $\gamma \to 0$. As a consequence, we say that F belongs to the max-domain of attraction of G_γ, in short $F \in \mathcal{D}_{\mathrm{M}}(G_\gamma)$. The parameter γ, known as the EV index, is a shape parameter that determines the right-tail behavior of F, being so a crucial parameter in EV theory. Specifically, if $\gamma < 0$, we have the Weibull max-domain of attraction, i.e., light right-tails, with a finite right endpoint. In addition, $\gamma = 0$ corresponds to the Gumbel max-domain of attraction (exponential right-tails). And if $\gamma > 0$, we have the Fréchet max-domain of attraction corresponding to heavy right-tails (polynomial tail decay), with an infinite right endpoint.

The GEV distribution with c.d.f. given in (2.7) is also known as the von Mises–Jenkinson representation. This is a general form from which we derive the three above mentioned distribution types, i.e.,

$$G_\gamma(x) = \begin{cases} \Psi_{-1/\gamma}(-1-\gamma x)\,; & \gamma < 0\ , \\ \Lambda(x)\,; & \gamma = 0\ , \\ \Phi_{1/\gamma}(1+\gamma x)\,; & \gamma > 0\ , \end{cases}$$

where, for $\varrho > 0$, $\Psi_\varrho(x) = \exp\bigl(-\{-x\}^\varrho\bigr)$ with $x < 0$ (Weibull distribution for maxima), $\Lambda(x) = \exp\bigl(-\exp(-x)\bigr)$ with $x \in \mathbb{R}$ (Gumbel distribution for maxima), and $\Phi_\varrho(x) = \exp(-x^{-\varrho})$ with $x > 0$ (Fréchet distribution for maxima). The Gumbel distribution for maxima and the Fréchet distribution for maxima are the commonly known Gumbel and Fréchet distributions, respectively. Location ($\mu \in \mathbb{R}$) and scale ($\sigma > 0$) parameters can be introduced in the GEV distribution by considering $G_\gamma(\{x-\mu\}/\sigma)$, denoted by $X \sim \text{GEV}(\mu, \sigma, \gamma)$.

All results developed for maxima can easily be reformulated for minima because $\min\{X_1, ..., X_n\} = -\max\{-X_1, ..., -X_n\}$. Actually, if we are interested in the lower tail, we can rewrite a result similar to the one given in (2.6) for minima, with a limiting c.d.f. $G(x) \equiv G^*_\gamma(x)$, which is now denoted as $X \sim \text{GEV}^*(\gamma)$, such that $G^*_\gamma(x) = 1 - G_\gamma(-x)$, i.e.,

$$(2.8) \qquad G^*_\gamma(x) = \begin{cases} 1-\exp\bigl(-\{1-\gamma x\}^{-1/\gamma}\bigr)\,; & 1-\gamma x > 0, \quad \gamma \in \mathbb{R}\backslash\{0\}\ , \\ 1-\exp\bigl(-\exp(x)\bigr)\,; & x \in \mathbb{R}, \quad \gamma = 0\ . \end{cases}$$

As a consequence, we say that F belongs to the min-domain of attraction of G^*_γ, in short $F \in \mathcal{D}_{\mathrm{m}}(G^*_\gamma)$. Analogously to the GEV distribution, the GEV* case (minima) is a general form from which we derive the following three possible

EV limiting cases:

$$G^*_\gamma(x) = \begin{cases} \Psi^*_{-1/\gamma}(1-\gamma x); & \gamma < 0\,, \\ \Lambda^*(x); & \gamma = 0\,, \\ \Phi^*_{1/\gamma}(-1+\gamma x); & \gamma > 0\,, \end{cases}$$

where, for $\varrho > 0$, $\Phi^*_\varrho(x) = 1 - \exp\bigl(-\{-x\}^{-\varrho}\bigr)$ with $x < 0$ (Fréchet distribution for minima), $\Lambda^*(x) = 1 - \exp\bigl(-\exp(x)\bigr)$ with $x \in \mathbb{R}$ (Gumbel distribution for minima), and $\Psi^*_\varrho(x) = 1 - \exp(-x^\varrho)$ with $x > 0$ (Weibull distribution for minima, commonly known as the Weibull distribution).

3. EXTREME VALUE BS DISTRIBUTIONS

In this section, we propose and characterize the EVBS model based on limiting EV models for maxima, as well as for minima, denoted as EVBS* distributions. In addition, a shape analysis for the EVBS and EVBS* distributions is provided. Specifically, consider that

$$Z \sim \mathrm{GEV}(\gamma) \equiv \mathrm{GEV}(0,1,\gamma)\,, \tag{3.1}$$

i.e., Z has c.d.f. as given in (2.7). Then,

$$T = \beta\Bigl(\alpha Z/2 + \sqrt{\alpha^2 Z^2/4+1}\Bigr)^2 \sim \mathrm{EVBS}(\alpha,\beta,\gamma)\,.$$

Directly from the GEV p.d.f., $g_\gamma(t) = dG_\gamma(t)/dt$, associated with the GEV c.d.f. $G_\gamma(t)$ given in (2.7), and considering $F_T(t) = G_\gamma(a_t)$, $t_q = F_T^{-1}(q)$ and $f_T(t) = A_t\, g_\gamma(a_t)$, with a_t and A_t as given in (2.3), the EVBS r.v. T can be defined in the following ways:

I. The p.d.f. of T is given by

$$f_T(t) = \begin{cases} A_t(1+\gamma a_t)^{-1-1/\gamma}\exp\bigl(-\{1+\gamma a_t\}^{-1/\gamma}\bigr); & \gamma \neq 0\,, \\ A_t \exp\bigl(-\exp(-a_t) - a_t\bigr); & \gamma = 0\,, \end{cases} \tag{3.2}$$

where $t > (\alpha^2\beta + 2\,\beta\gamma^2)/(2\,\gamma^2) - \sqrt{(\alpha^4\beta^2 + 4\,\alpha^2\beta^2\gamma^2)/\gamma^4}\big/2$ if $\gamma > 0$; $t>0$ if $\gamma=0$; and $0<t<(\alpha^2\beta+2\,\beta\gamma^2)/(2\,\gamma^2)+\sqrt{(\alpha^4\beta^2+4\,\alpha^2\beta^2\gamma^2)/\gamma^4}\big/2$ if $\gamma<0$.

II. The c.d.f. of T is expressed as

$$F_T(t) = \begin{cases} \exp\bigl(-\{1+\gamma a_t\}^{-1/\gamma}\bigr); & \gamma \neq 0\,, \\ \exp\bigl(-\exp(-a_t)\bigr); & \gamma = 0\,. \end{cases} \tag{3.3}$$

III. The q.f. of T is as given in (2.4) by replacing ξ_q with z_q, the q^{th} quantile of the c.d.f. $G_\gamma(x)$, as expressed in (2.7), i.e., $z_q = \bigl(\{-\log(q)\}^{-\gamma} - 1\bigr)\big/\gamma$ if $\gamma \neq 0$, and $z_q = -\log\bigl(-\log(q)\bigr)$ if $\gamma = 0$.

Analogously, if we consider in (3.1) the GEV distribution for minima given in (2.8), we use the notation $T^* \sim \mathrm{EVBS}^*(\alpha,\beta,\gamma)$ for an associated r.v. T^*, and, as before, noting that $F_{T^*}(t) = G^*_\gamma(a_t) = 1 - G_\gamma(-a_t)$ and that $f_{T^*}(t) = A_t\, g^*_\gamma(a_t) = A_t\, g_\gamma(-a_t)$, the EVBS* r.v. T^* can be defined in the following ways:

I′. The p.d.f. of T^* is given by

$$f_{T^*}(t) = \begin{cases} A_t(1-\gamma a_t)^{-1-1/\gamma} \exp\bigl(-\{1-\gamma a_t\}^{-1/\gamma}\bigr)\,; & \gamma \neq 0\,, \\ A_t \exp\bigl(-\exp(a_t) + a_t\bigr)\,; & \gamma = 0\,, \end{cases} \tag{3.4}$$

where $t > (\alpha^2\beta + 2\,\beta\gamma^2)/(2\,\gamma^2) - \sqrt{(\alpha^4\beta^2 + 4\,\alpha^2\beta^2\gamma^2)/\gamma^4}/2$ if $\gamma < 0$; $t > 0$ if $\gamma = 0$; and $0 < t < (\alpha^2\beta + 2\beta\gamma^2)/(2\gamma^2) + \sqrt{(\alpha^4\beta^2 + 4\alpha^2\beta^2\gamma^2)/\gamma^4}/2$ if $\gamma > 0$.

II′. The c.d.f. of T^* is defined as

$$F_{T^*}(t) = \begin{cases} 1 - \exp\bigl(-\{1-\gamma a_t\}^{-1/\gamma}\bigr)\,; & \gamma \neq 0\,, \\ 1 - \exp\bigl(-\exp(a_t)\bigr)\,; & \gamma = 0\,. \end{cases} \tag{3.5}$$

III′. The q.f. of T^* is also as given in (2.4), but by replacing ξ_q with $z^*_q = z^*_q(\gamma)$, the q^{th} quantile of the c.d.f. $G^*_\gamma(x)$, as expressed in (2.8), i.e., with $z_q(\gamma)$ being the q^{th} quantile of the c.d.f. $G_\gamma(x)$, as given in (2.7), $z^*_q = -z_{1-q}(\gamma) = \bigl(1 - \{-\log(1-q)\}^{-\gamma}\bigr)/\gamma$ if $\gamma \neq 0$, and $z^*_q = \log\bigl(-\log(1-q)\bigr)$ if $\gamma = 0$.

Next, as a direct application of the change of variable method, some properties of the EVBS and EVBS* distributions are provided.

Proposition 3.1. *Let $T \sim \mathrm{EVBS}(\alpha,\beta,\gamma)$ and $T^* \sim \mathrm{EVBS}^*(\alpha,\beta,\gamma)$. Then,*

(i) *$c\,T \sim \mathrm{EVBS}(\alpha, c\,\beta, \gamma)$ and $c\,T^* \sim \mathrm{EVBS}^*(\alpha, c\,\beta, \gamma)$, with $c > 0$;*

(ii) *$1/T \sim \mathrm{EVBS}^*(\alpha, 1/\beta, \gamma)$ and $1/T^* \sim \mathrm{EVBS}(\alpha, 1/\beta, \gamma)$.*

Figure 1 (first and second panels) displays shapes for the EVBS and EVBS* densities for different values of their parameters. In all of these graphs, we consider $\beta = 1$, without loss of generality, because β is a scale parameter, such as stated in Proposition 3.1(i). In these plots, we further use the notation $\mathrm{EVBS}(\alpha,\gamma) \equiv \mathrm{EVBS}(\alpha,1,\gamma)$. For the EVBS densities presented in Figure 1 (first panel), we see how the shape parameter α modifies the shape of these densities. In the case of the parameter γ, we detect changes in the kurtosis, as expected. Similar aspects are observed when we consider the EVBS* densities presented in Figure 1 (second panel).

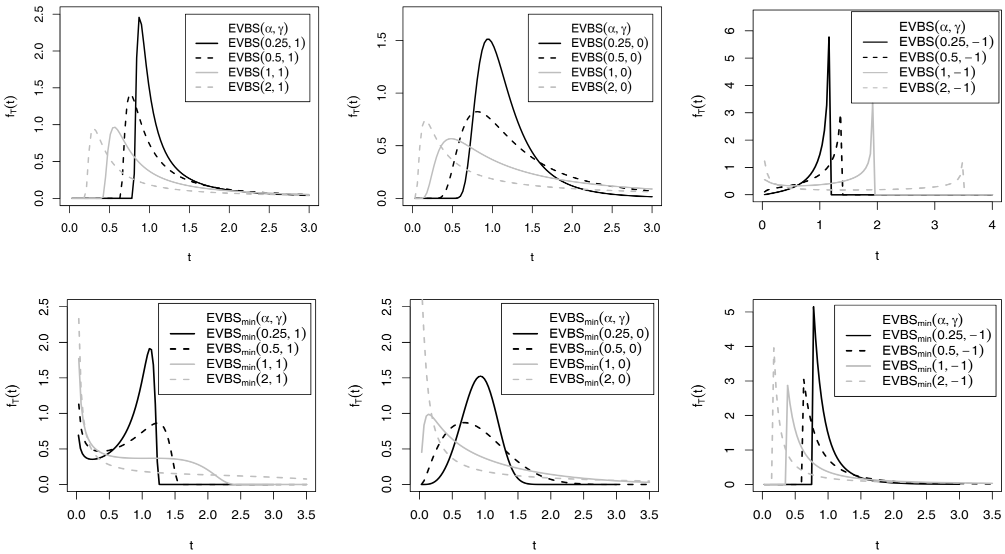


Figure 1: p.d.f. plots of the EVBS (1st panel) and EVBS* (2nd panel) distributions for $\beta = 1$ and the indicated values of (α, γ), where $\mathrm{EVBS}_{\min} \equiv \mathrm{EVBS}^*$.

4. BST EXTREMAL DOMAINS OF ATTRACTION

In this section, we obtain the extremal domains of attraction for BST distributions.

We analyze the extremal domain of attraction of the c.d.f. of an r.v.

$$T \,=\, \beta\Big(\alpha\, Z/2 + \sqrt{\alpha^2 Z^2/4+1}\,\Big)^2, \qquad \alpha>0, \quad \beta>0\;, \tag{4.1}$$

not necessarily following an EVBS distribution, whenever the c.d.f. of the r.v. Z, compulsory given by

$$Z \,=\, \Big(\sqrt{T/\beta} - \sqrt{\beta/T}\Big)/\alpha\;,$$

belongs to some extremal domain of attraction either for maxima or for minima.

4.1. Max-domains of attraction

We start with the Fréchet case and use the following necessary and sufficient condition for $F \in \mathcal{D}_{\mathrm{M}}(G_\gamma)$ with $\gamma > 0$, derived in Gnedenko (1943) (see also de Haan and Ferreira, 2006, Theorem 1.2.1-1):

$$F \in \mathcal{D}_{\mathrm{M}}(G_\gamma), \quad \gamma>0 \quad\Longleftrightarrow\quad \lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} \,=\, x^{-1/\gamma}\,, \tag{4.2}$$

for all $x > 0$, and the right endpoint of F, namely $x^F := \inf\{x\colon F(x) \geq 1\}$, is necessarily infinite.

Theorem 4.1. *Let the c.d.f. of the r.v. Z be in the Fréchet max-domain of attraction, necessarily with a positive EV index, i.e., $\gamma_Z > 0$. Then, the c.d.f. of the r.v. T given in* (4.1) *is also in the Fréchet max-domain of attraction, i.e., $F_T \in \mathcal{D}_{\mathrm{M}}(G_{\gamma_T})$, with $\gamma_T = 2\,\gamma_Z$.*

Proof: By hypothesis, $F_Z \in \mathcal{D}_{\mathrm{M}}(G_{\gamma_Z})$ for $\gamma_Z > 0$. Thus, F_Z satisfies (4.2) for γ_Z. Then, we have that, as $t \to \infty$,

$$\frac{1 - F_T(tx)}{1 - F_T(t)} = \frac{1 - F_Z(a_{tx})}{1 - F_Z(a_t)} \approx \frac{1 - F_Z\big(\{tx/\beta\}^{1/2}/\alpha\big)}{1 - F_Z\big(\{t/\beta\}^{1/2}/\alpha\big)} \approx x^{-1/(2\gamma_Z)}\,,$$

with the notation $u_t \approx v_t$ being valid if and only if $u_t/v_t \to 1$, as $t \to \infty$. □

For light right-tails, i.e., for the Weibull max-domain of attraction, we can prove a result similar to that of Theorem 4.1, if we use the following necessary and sufficient condition for $F \in \mathcal{D}_{\mathrm{M}}(G_\gamma)$ with $\gamma < 0$ (also derived in Gnedenko, 1943):

$$\text{(4.3)} \qquad F \in \mathcal{D}_{\mathrm{M}}(G_\gamma), \quad \gamma < 0 \quad \Longleftrightarrow \quad \lim_{t\to\infty} \frac{1 - F\big(x^F - 1/\{tx\}\big)}{1 - F(x^F - 1/t)} = x^{1/\gamma}\,,$$

for all $x > 0$, and the right endpoint of F, namely x^F, is finite.

Theorem 4.2. *Let the c.d.f. of the r.v. Z be in the Weibull max-domain of attraction, necessarily with a negative EV index, i.e., $\gamma_Z < 0$. Then, the c.d.f. of the r.v. T given in* (4.1) *is also in the Weibull max-domain of attraction and $\gamma_T = \gamma_Z$.*

Proof: We have

$$\lim_{t\to\infty} \frac{1 - F_T\big(t^F - 1/\{tx\}\big)}{1 - F_T(t^F - 1/t)} = \lim_{t\to\infty} \frac{1 - F_Z\big(a_{t^F - 1/\{tx\}}\big)}{1 - F_Z\big(a_{t^F - 1/t}\big)}\,,$$

with t^F being the right endpoint of F_T. But we can assume, without loss of generality, that z^F, the right endpoint of F_Z, is null, i.e., $z^F = 0$. Hence, $t^F = \beta$ and, as $t \to \infty$,

$$\frac{1 - F_Z\big(a_{t^F - 1/\{tx\}}\big)}{1 - F_Z\big(a_{t^F - 1/t}\big)} \approx \frac{1 - F_Z\big(-\{\alpha\beta\,tx\}^{-1}\big)}{1 - F_Z\big(-\{\alpha\beta\,t\}^{-1}\big)} \approx x^{1/\gamma_Z}\,. \qquad □$$

We next work on the slight more restrictive class of twice-differentiable c.d.f.'s $F \in \mathcal{D}_{\mathrm{M}}(G_\gamma)$, the so-called twice-differentiable domain of attraction of G_γ, denoted by $\widetilde{\mathcal{D}}_{\mathrm{M}}(G_\gamma)$. A possible characterization of the twice-differentiable domain of attraction of G_γ is due to Pickands (1986). Let us then assume that there exists F'', $f = F'$, and consider the function

$$k(x) \;=\; -f(x)\big/\big\{F(x)\log\big(F(x)\big)\big\} \;=\; \big\{-\log\big(-\log F(x)\big)\big\}' .$$

Hence, with $\gamma(x) = \{1/k(x)\}'$, we have

$$F \in \widetilde{\mathcal{D}}_{\mathrm{M}}(G_\gamma) \quad\Longleftrightarrow\quad \lim_{x\uparrow x^F} \gamma(x) = \gamma \;. \tag{4.4}$$

Consequently, if $x^F = +\infty$, $\lim_{x\to\infty} x\,k(x) = 1/\gamma$, and if $x^F < +\infty$, $\lim_{x\uparrow x^F} (x^F - x)\,k(x) = -1/\gamma$, i.e., $\lim_{x\to\infty} k(x) = 0$, if $\gamma > 0$, and $\lim_{x\uparrow x^F} k(x) = +\infty$, if $\gamma < 0$. If $\gamma = 0$, we can have $k(x) \to 0$, $k(x) \to +\infty$, or $k(x) \to c$, for $0 < c < +\infty$. Observe also that, after some simple calculations, we can write

$$\gamma(x) \;=\; F''(x)\,F(x)\log\big(F(x)\big)\big/f^2(x) - \log\big(F(x)\big) - 1 \;. \tag{4.5}$$

Theorem 4.3. *Let $T \sim \mathrm{BST}(\alpha,\beta;f_Z)$, with Z in the subset of the max-domain of attraction of G_{γ_Z} constituted by twice-differentiable c.d.f.'s, the so-called twice-differentiable max-domain of attraction of G_{γ_Z}, and assume that*

$$c \;=\; \lim_{t\uparrow t^F} F_Z(a_t)\log\big(F_Z(a_t)\big)\,A'_t\big/\big\{A_t^2\,f_Z(a_t)\big\}$$

is finite, where t^F is the right endpoint of the r.v. T and a_t and A_t are as given in (2.3), with $A'_t = dA_t/dt$. Then, $F_T \in \mathcal{D}_{\mathrm{M}}(G_{\gamma_T})$, with $\gamma_T = \gamma_Z + c$.

Proof: By hypothesis, the necessary and sufficient condition (4.4) holds for Z, with F, γ and $\gamma(x)$ replaced by F_Z, γ_Z and $\gamma_Z(x)$, respectively. Now, just observe that, by applying (4.5), and then (2.3)–(2.5) and $A'_t = -\big(\sqrt{t/\beta} + 3\sqrt{\beta/t}\big)\big/(4\alpha t^2)$, we have

$$\begin{aligned}\gamma_T(t) \;&=\; \frac{F''_T(t)\,F_T(t)\log\big(F_T(t)\big)}{f_T^2(t)} - \log\big(F_T(t)\big) - 1\\ &=\; \frac{F_Z(a_t)\log\big(F_Z(a_t)\big)\,A'_t}{A_t^2\,f_Z(a_t)} + \frac{F''_Z(a_t)\,F_Z(a_t)\log\big(F_Z(a_t)\big)}{f_Z^2(a_t)} - \log\big(F_Z(a_t)\big) - 1\,.\end{aligned} \tag{4.6}$$

On the basis of the limit in Theorem 4.3, the first term in the second line of (4.6) approaches c as $t \uparrow t^F$. Because the following term approaches γ_Z, the result follows. □

Corollary 4.1. *Under the conditions of Theorem 4.3, we have $c = \gamma_Z$ if $\gamma_Z > 0$ and $c = 0$ if $\gamma_Z < 0$.*

Example 4.1. We now provide a few illustrations of Corollary 4.1:

(**i**) If Z has Fréchet or Pareto distributions (in the Fréchet max-domain of attraction, i.e., $\gamma_Z > 0$), then the limit in Theorem 4.3 is $c = \gamma_Z$ and so $\gamma_T = 2\,\gamma_Z$. Indeed, as stated in Theorem 4.1, this result holds more generally in $\mathcal{D}_\mathrm{M}(G_{\gamma_Z})$, with $\gamma_Z > 0$.

(**ii**) If Z has Weibull or uniform distributions (in the Weibull max-domain of attraction, i.e., $\gamma_Z < 0$), then $c = 0$ and $\gamma_T = \gamma_Z$. In fact, as stated in Theorem 4.2, this result holds more generally in $\mathcal{D}_\mathrm{M}(G_{\gamma_Z})$, with $\gamma_Z < 0$.

Remark 4.1. We further conjecture that, in Corollary 4.1, we can often replace $\gamma_Z < 0$ by $\gamma_Z \le 0$. This is supported by the examples of an r.v. Z either exponential or gamma, or Gumbel or normal, all in $\mathcal{D}_\mathrm{M}(G_0)$, i.e., with $\gamma_Z = 0$. Then $c = 0$ and $\gamma_T = \gamma_Z = 0$. Also, if Z has an exponential-type (ET) distribution, with a finite right endpoint, i.e., $F_Z(x) = K\exp\bigl(-c/\{z^F - x\}\bigr)$, for $x < z^F$, $c > 0$, and $K > 0$ (again in the Gumbel max-domain of attraction), then also $c = 0$ and $\gamma_T = \gamma_Z = 0$.

Because in the twice-differentiable domain of attraction of G_γ the von Mises condition is necessary and sufficient to have $\lim_{x\uparrow x^F}\gamma(x) = \gamma$, with $\gamma(x) = \{1/k(x)\}'$ (see Pickands, 1986, Theorem 5.2), we can also state that

$$F \in \widetilde{\mathcal{D}}_\mathrm{M}(G_\gamma) \iff \lim_{x\uparrow x^F}\bigl\{1 - F(x)\bigr\}\,F''(x)/\bigl\{F'(x)\bigr\}^2 = -\gamma - 1\,. \tag{4.7}$$

Therefore, we can still write the following result.

Theorem 4.4. *Under the conditions and notations of Theorem 4.3, let us assume that*

$$c^* = \lim_{t\uparrow t^F}\bigl\{1 - F_Z(a_t)\bigr\}\,A'_t/\bigl\{A_t^2\,F'_Z(a_t)\bigr\} < \infty\,.$$

Then, $F_T \in \mathcal{D}_\mathrm{M}(G_{\gamma_T})$, with $\gamma_T = \gamma_Z - c^*$.

Proof: Just observe that

$$\begin{aligned}\frac{\bigl\{1 - F_T(t)\bigr\}\,F''_T(t)}{\bigl\{F'_T(t)\bigr\}^2} &= \frac{\bigl\{1 - F_Z(a_t)\bigr\}\,\bigl\{A'_t F'_Z(a_t) + A_t^2 F''_Z(a_t)\bigr\}}{A_t^2\,\bigl\{F'_Z(a_t)\bigr\}^2} \\ &= \frac{\bigl\{1 - F_Z(a_t)\bigr\}\,A'_t}{A_t^2\,F'_Z(a_t)} + \frac{\bigl\{1 - F_Z(a_t)\bigr\}\,F''_Z(a_t)}{\bigl\{F'_Z(a_t)\bigr\}^2}\,.\end{aligned} \tag{4.8}$$

By hypothesis, as $t \uparrow t^F$, the last term in (4.8) converges to $-\gamma_Z - 1$, and the result follows. □

Corollary 4.2. *Under the conditions and notations of Theorem 4.4, if we further assume that Z has an infinite right endpoint, then $F_T \in \mathcal{D}_\mathrm{M}(G_{\gamma_T})$, with $\gamma_T = \gamma_Z$, provided there exists a finite limit for $\{1 - F_Z(x)\}/F'_Z(x)$, as $x \to \infty$.*

4.2. Min-domains of attraction

We now analyze the domains of attraction for minima. To emphasize the possible difference between the right and left EV indices, we denote this last one as γ^*.

We reformulate conditions (4.2) and (4.3) for minima obtaining respectively

$$F \in \mathcal{D}_{\rm m}(G^*_{\gamma^*}),\ \ \gamma^* > 0 \iff \lim_{t\to-\infty} \frac{F(tx)}{F(t)} = x^{-1/\gamma^*}, \ \ \forall\, x > 0 \ , \tag{4.9}$$

and

$$F \in \mathcal{D}_{\rm m}(G^*_{\gamma^*}),\ \ \gamma^* < 0 \iff \lim_{t\to-\infty} \frac{F\big(x_F - 1/\{tx\}\big)}{F\big(x_F - 1/t\big)} = x^{1/\gamma^*}, \ \ \forall\, x > 0 \ , \tag{4.10}$$

where the left endpoint $x_F := \inf\{x\colon F(x) > 0\}$ is finite; see, e.g., Galambos (1987, Theorem 2.1.5). Observe that a BST r.v. T cannot be in the Fréchet min-domain of attraction because its left endpoint is not $-\infty$; see, e.g., Galambos (1987, Theorem 2.1.4).

In the sequel, the notations $\text{Weibull}_{\min}$, $\text{Frechet}_{\min}$ and $\text{Gumbel}_{\min}$ are used for denoting Weibull, Fréchet and Gumbel distributions for minima, respectively, with parameter γ^*, and z_F and t_F denoting the left endpoints of Z and T, respectively.

Theorem 4.5. *Let the c.d.f. of the r.v. Z be in the Weibull min-domain of attraction, necessarily with a negative EV index, i.e., $\gamma^*_Z < 0$. Then, the c.d.f. of the r.v. T given in* (4.1) *is in the Weibull min-domain of attraction and $\gamma^*_T = \gamma^*_Z$.*

Proof: Assume, without loss of generality, that $z_F = 0$, with z_F being the left endpoint of F_Z (i.e., $t_F = \beta$, with t_F being the left endpoint of F_T). Then,

$$\lim_{t\to-\infty} \frac{F_T\big(t_F - 1/\{tx\}\big)}{F_T\big(t_F - 1/t\big)} = \lim_{t\to-\infty} \frac{F_Z\big(a_{t_F - 1/\{tx\}}\big)}{F_Z\big(a_{t_F - 1/t}\big)} = \lim_{t\to-\infty} \frac{F_Z\big(-\{\alpha\beta\, t x\}^{-1}\big)}{F_Z\big(-\{\alpha\beta\, t\}^{-1}\big)}$$

and the result follows from the fact that F_Z satisfies (4.10) for γ^*_Z. □

Theorem 4.6. *Let the c.d.f. of the r.v. Z be in the Fréchet min-domain of attraction, necessarily with a positive EV index, i.e., $\gamma^*_Z > 0$. Then, the c.d.f. of the r.v. T given in* (4.1) *is in the Weibull min-domain of attraction, and $\gamma^*_T = -2\,\gamma^*_Z$.*

Proof: Consider, without loss of generality, $t_F = 0$. Then, $z_F = -\infty$ and

$$\begin{aligned}\lim_{t\to-\infty} \frac{F_T\big(t_F - 1/\{tx\}\big)}{F_T\big(t_F - 1/t\big)} &= \lim_{t\to-\infty} \frac{F_Z\big(a_{-1/\{tx\}}\big)}{F_Z\big(a_{-1/t}\big)} \\ &= \lim_{t\to-\infty} \frac{F_Z\big(-\{-\beta\, t x\}^{1/2}/\alpha\big)}{F_Z\big(-\{-\beta t\}^{1/2}/\alpha\big)} = x^{-1/(2\gamma^*_Z)} \ ,\end{aligned}$$

where the last step is due to the fact that F_Z satisfies (4.9) for γ^*_Z. □

Next, we again work on the slight more restrictive class of twice-differentiable c.d.f.'s, such as in Subsection 4.1. Analogously to the domain of attraction for maxima, the von Mises condition in (4.7), reformulated for minima, enables us to state that

$$F \in \widetilde{\mathcal{D}}_{\mathrm{m}}(G^*_{\gamma^*}) \iff \lim_{x \downarrow x_F} F(x)\, F''(x) / \big(F'(x)\big)^2 = \gamma^* + 1 \,, \tag{4.11}$$

where x_F is the left endpoint of F and $\widetilde{\mathcal{D}}_{\mathrm{m}}(G^*_{\gamma^*})$ denotes the twice-differentiable domain of attraction of $G^*_{\gamma^*}$.

Theorem 4.7. *Let $T \sim \mathrm{BST}(\alpha, \beta; f_Z)$, with Z in the subset of the min-domain of attraction of $G^*_{\gamma^*_Z}$ constituted by the twice-differentiable c.d.f.'s, the so-called twice-differentiable min-domain of attraction of $G^*_{\gamma^*_Z}$, and assume that*

$$d \;=\; \lim_{t \downarrow t_F} F_Z(a_t)\, A'_t \big/ \big(A_t^2\, F'_Z(a_t)\big)$$

is finite, where t_F is the left endpoint of T and a_t and A_t are as given in (2.3)*, with $A'_t = dA_t/dt$. Then, $F_T \in \mathcal{D}^*_{\mathrm{m}}(G^*_{\gamma^*_T})$, with $\gamma^*_T = \gamma^*_Z + d$.*

Proof: The result is easy to prove because

$$\frac{F_T(t)\, F''_T(t)}{f_T(t)^2} \;=\; \frac{F_Z(a_t)\,\big\{A'_t f_Z(a_t) + A_t^2 F''_Z(a_t)\big\}}{A_t^2 f_Z(a_t)^2} \;=\; \frac{F_Z(a_t)\, A'_t}{A_t^2 f_Z(a_t)} + \frac{F_Z(a_t)\, F''_Z(a_t)}{f_Z(a_t)^2} \,. \qquad \square$$

Corollary 4.3. *Under the conditions of Theorem 4.7, we have $d = 0$ if $\gamma^*_Z < 0$ and $d = -3\,\gamma^*_Z$ if $\gamma^*_Z > 0$.*

Example 4.2. We next provide a few illustrations of Corollary 4.3:

- **(i)** If Z has a Weibull distribution for minima (in the Weibull min-domain of attraction, i.e., $\gamma^*_Z < 0$), or an exponential, Pareto or uniform distribution (also in the Weibull min-domain of attraction, with $\gamma^*_Z = -1$), or even a Gamma(p, q) distribution (in the Weibull min-domain of attraction, with $\gamma^*_Z = -1/p$), then $d = 0$ and $\gamma^*_T = \gamma^*_Z$. Indeed, as stated in Theorem 4.5, this result holds more generally in $\mathcal{D}_{\mathrm{m}}(G^*_{\gamma^*_Z})$, with $\gamma^*_Z < 0$.
- **(ii)** If Z has a Fréchet distribution for minima (in the Fréchet min-domain of attraction, i.e., $\gamma^*_Z > 0$), then $d = -3\,\gamma^*_Z$ and $\gamma^*_T = -2\,\gamma^*_Z$. In fact, as stated in Theorem 4.6, this result holds more generally in $\mathcal{D}_{\mathrm{m}}(G^*_{\gamma^*_Z})$, with $\gamma^*_Z > 0$.
- **(iii)** If Z has an ET distribution as in Remark 4.1 (in the Fréchet min-domain of attraction, with $\gamma^*_Z = 1$), then $d = -3\,\gamma^*_Z = -3$ and $\gamma^*_T = -2$, i.e., T belongs to the Weibull min-domain of attraction.

Remark 4.2. Similarly to what we mentioned in Remark 4.1, we further conjecture that, in Corollary 4.3, we can often replace $\gamma^*_Z < 0$ by $\gamma^*_Z \leq 0$. This is supported by the fact that if Z has a Gumbel distribution for minima, or any of the limiting distributions for maxima (Fréchet, Gumbel, Weibull), or a normal distribution (all in the Gumbel min-domain of attraction, i.e. $\gamma^*_Z = 0$), then the limit in Theorem 4.7 is $d = 0$ and $\gamma^*_T = \gamma^*_Z = 0$.

5. HAZARD ANALYSIS

We may define a hazard as a dangerous event that could conduct to an emergency or disaster. The origin of this event may be due to a situation that could have an adverse effect. Thus, a hazard is a potential and not an actual possibility, i.e., it can be statistically evaluated. A hazard analysis is the assessment of a risk that is present in a particular environment. Therefore, hazard assessment allows us to evaluate potential risk by the estimated frequency or intensity of the r.v. of interest. In this section, we study the EVBS h.r. and its change point.

5.1. Hazard rate

Statistically, a hazard analysis can be carried out by the h.r. function. Apart from hazard rate, this function is also known as chance function, failure rate, force of mortality, intensity function, or risk rate, among other names. In actuarial science, for example, the h.r. is the annualized probability that a person at a specified age will die in the next instant, expressed as a death rate per year. For more details about the concept of h.r., see Marshall and Olkin (2007, pp. 10–13).

A nice property of the h.r. is that it allows us to characterize the behavior of statistical distributions. For example, the h.r. may have several different shapes such as increasing (IHR), constant (exponential distribution), decreasing (DHR), bathtube (BT), inverse bathtube (IBT or upside-down) approaching to a non-null constant and IBT approaching to zero. An incorrect specification of the h.r. could have serious consequences in the analysis; see, e.g., Bhatti (2010) for a study about this issue.

The h.r. of an r.v. T is given in general by $h_T(t) = f_T(t)/R_T(t)$, for $t > 0$, and $0 < F_T(t) < 1$, where $R_T(t) = 1 - F_T(t)$, for $t > 0$, is the reliability function (r.f.), and f_T and F_T are the p.d.f. and c.d.f. of the r.v. T. The change point of $h_T(t)$, denoted by t_c, is defined as the moment where the h.r. attains either a maximum or a minimum value and it is the solution of the equation $f(t_c) = -f'(t_c)/h(t_c)$, whenever F is twice-differentiable, and such a solution exists.

5.2. TTT curve

The h.r. of an r.v. T can be characterized by its corresponding total time on test (TTT) function given by

$$H_T^{-1}(u) = \int_0^{F_T^{-1}(u)} \big(1 - F_T(y)\big)\, dy$$

or by its scaled version given by $W_T(u) = H_T^{-1}(u)/H_T^{-1}(1)$, for $0 \le u \le 1$, where once again F_T^{-1} is the generalized inverse function of the c.d.f. of T. Now, W_T can be empirically approximated, allowing to construct the empirical scaled TTT curve by plotting the consecutive points $\big[k/n, W_n(k/n)\big]$, where $W_n(k/n) = \big\{\sum_{i=1}^{k} T_{(i)} + (n-k)\, T_{(k)}\big\}/\sum_{i=1}^{n} T_{(i)}$, for $k = 1, ..., n$, with $T_{(i)}$ being the corresponding i^{th} ascending order statistic, for $1 \le i \le n$.

From Figure 2 (left), we observe different theoretical shapes for the scaled TTT curve. Thus, a TTT plot expressed by a curve that is concave (or convex) corresponds to the IHR (or DHR) class. A concave (or convex) and then convex (or concave) shape for the TTT curve corresponds to a IBT (or BT) h.r. A TTT plot represented by a straight line is an indication that the exponential distribution must be used. Thus, a graphical plot of the empirical scaled TTT curve could provide to us the type of distribution that the data have. See also in Figure 2 the theoretical scaled TTT curves for EVBS (center) and EVBS* (right) models. In these plots, we again use the notation $\text{EVBS}(\alpha, \gamma) \equiv \text{EVBS}(\alpha, 1, \gamma)$.

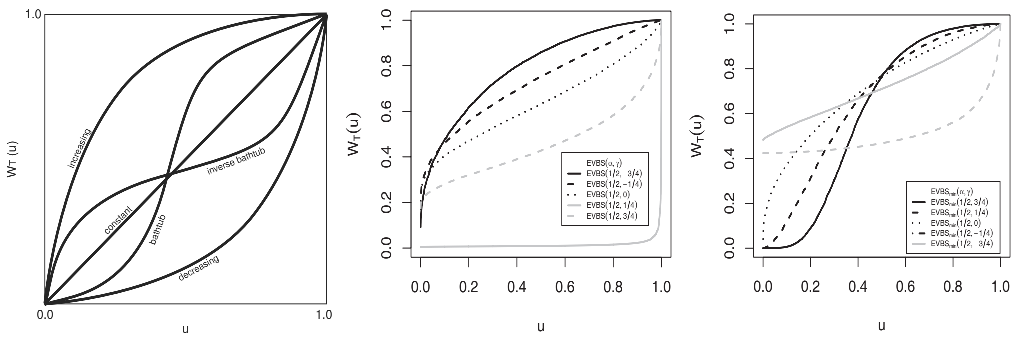


Figure 2: Theoretical scaled TTT curves for a general distribution with the indicated h.r. shape (left) and for the $\text{EVBS}(0.5, \gamma)$ (center) and $\text{EVBS}^*(0.5, \gamma)$ (right) distributions for the indicated values of γ.

5.3. EVBS hazard rate

The normal distribution is in the IHR class. The gamma and Weibull distributions can be either in IHR or DHR classes (of course, the case of the exponential distribution with constant h.r. is considered by these two models). However, the lognormal (LN) distribution has a non-monotonic h.r., because it is initially increasing until its change point and then it decreases to zero, i.e., the LN model is in the IBT h.r. class. The BS h.r. behaves similarly to the LN h.r., i.e., it is initially increasing until its change point and then decreasing not to zero, but to a positive constant. Thus, although BS, gamma, LN and Weibull distributions have densities with similar shapes, their h.r.'s are totally different.

Let $T \sim \text{EVBS}(\alpha, \beta, \gamma)$. Then, directly from the definition of the p.d.f., $f_T(t)$, and the c.d.f., $F_T(t)$, of the r.v. $T \sim \text{EVBS}(\alpha, \beta, \gamma)$, given in (3.2) and (3.3), respectively, we have that:

A. The r.f. of T is expressed as $R_T(t) = 1 - F_T(t)$, with $F_T(t)$ given in (3.3).

B. Again with a_t and A_t as given in (2.3), the h.r. of T is defined as

$$h_T(t) = \frac{f_T(t)}{R_T(t)} = \begin{cases} A_t (1+\gamma a_t)^{-1-1/\gamma} \big/ \big(\exp\big(\{1+\gamma a_t\}^{-1/\gamma}\big) - 1\big)\,; & \gamma \neq 0\,, \\ A_t \exp(-a_t) \big/ \exp\big(\exp(-a_t)\big) - 1\,; & \gamma = 0\,, \end{cases}$$

where $t > (\alpha^2\beta + 2\beta\gamma^2)/(2\gamma^2) - \sqrt{(\alpha^4\beta^2 + 4\alpha^2\beta^2\gamma^2)/\gamma^4}/2$ if $\gamma > 0$; $t > 0$ if $\gamma = 0$; and $0 < t < (\alpha^2\beta + 2\beta\gamma^2)/(2\gamma^2) + \sqrt{(\alpha^4\beta^2 + 4\alpha^2\beta^2\gamma^2)/\gamma^4}/2$ if $\gamma < 0$.

C. With the notation $b_{t_c} = 1 + \gamma a_{t_c}$, the change point t_c of the h.r. of T is obtained as the solution of the equations:

$$\begin{cases} \Big(\Big\{A'_{t_c} - A^2_{t_c}(1+\gamma)\, b_{t_c}^{-1}\Big\} \Big\{\exp\big(b_{t_c}^{-1/\gamma}\big) - 1\Big\} + \\ \qquad + A^2_{t_c}\{1+\gamma a_{t_c}\}^{-1-1/\gamma} \exp\big(b_{t_c}^{-1/\gamma}\big)\Big)\, b_{t_c}^{-1-1/\gamma} = 0\,; & \gamma \neq 0\,, \\ A^2_{t_c}\Big(1 + \big\{\exp(-a_{t_c}) - 1\big\} \exp\big(\exp(-a_{t_c})\big)\Big) + \\ \qquad + A'_{t_c}\Big(\exp\big(\exp(-a_{t_c})\big) - 1\Big) = 0\,; & \gamma = 0\,. \end{cases}$$

Let $T^* \sim \text{EVBS}^*(\alpha, \beta, \gamma)$. Then, now, directly from the definition of the p.d.f., $f_{T^*}(t)$, and the c.d.f., $F_{T^*}(t)$, of the r.v. $T^* \sim \text{EVBS}^*(\alpha, \beta, \gamma)$, given in (3.4) and (3.5), respectively, we have that:

D. The r.f. of T^* is expressed as $R_{T^*}(t) = 1 - F_{T^*}(t)$, with $F_{T^*}(t)$ as given in (3.5).

E. Again with a_t and A_t as given in (2.3), the h.r. of T^* is defined as

$$h_{T^*}(t) \;=\; \frac{f_{T^*}(t)}{R_{T^*}(t)} \;=\; \begin{cases} A_t(1-\gamma\, a_t)^{-1-1/\gamma}; & \gamma \neq 0\,, \\ A_t \exp(a_t)\,; & \gamma = 0\,, \end{cases}$$

where $t > (\alpha^2\beta + 2\,\beta\gamma^2)/(2\,\gamma^2) - \sqrt{(\alpha^4\beta^2 + 4\,\alpha^2\beta^2\gamma^2)/\gamma^4}\big/2$ if $\gamma < 0$; $t > 0$ if $\gamma = 0$; and $0 < t < (\alpha^2\beta + 2\beta\gamma^2)/(2\gamma^2) + \sqrt{(\alpha^4\beta^2 + 4\alpha^2\beta^2\gamma^2)/\gamma^4}\big/2$ if $\gamma > 0$.

F. The change point t_c of the h.r. of T^* is obtained as the solution of the equations:

$$\begin{cases} (1-\gamma\, a_{t_c})^{-1-1/\gamma}\left(A'_{t_c} + \{1+\gamma\}\, A^2_{t_c}\{1-\gamma\, a_{t_c}\}^{-1}\right) \,=\, 0; & \gamma \neq 0\,, \\ A^2_{t_c} + A'_{t_c} = 0; & \gamma = 0\,. \end{cases}$$

As can be seen in Figure 3, the EVBS and EVBS* h.r.'s present several different shapes going through all the h.r. shape classes mentioned above. In these plots, once again, we use the notation $\text{EVBS}(\alpha,\gamma) \equiv \text{EVBS}(\alpha, 1, \gamma)$. The h.r. of T can also approach ∞, zero or a positive constant, as $t \to \infty$. These are strong points in favor of our models, as they become interesting for modeling purposes.

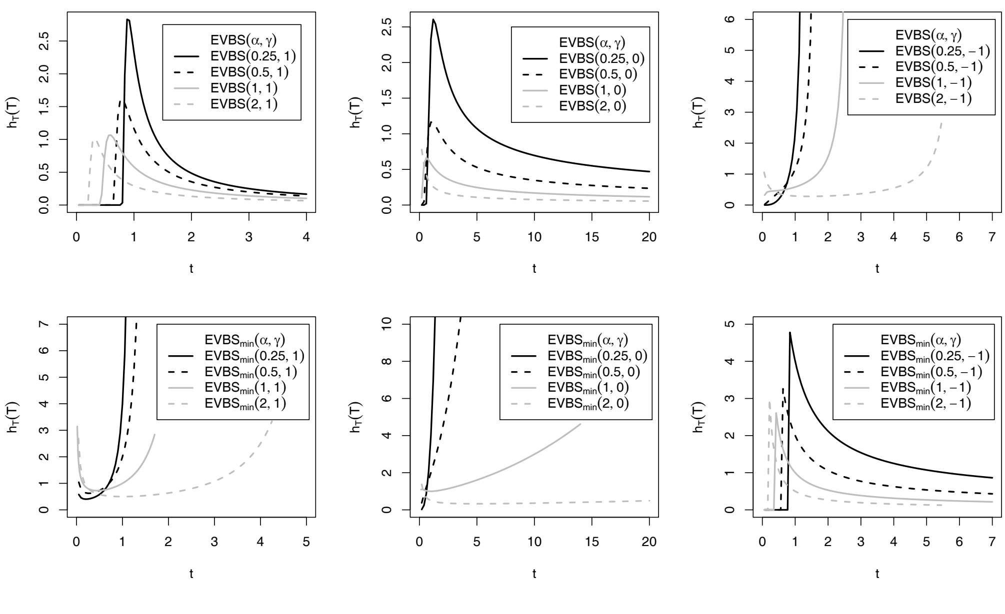


Figure 3: h.r. plots of the EVBS (1st panel) and EVBS* (2nd panel) distributions for $\beta = 1$, and the indicated values of (α, γ), where $\text{EVBS}_{\min} \equiv \text{EVBS}^*$.

6. ESTIMATION AND MODEL CHECKING

In this section, we present some results related to estimation aspects and model checking for EVBS distributions.

6.1. ML estimation

As is well-known, ML estimates are obtained from the solution of the system $\dot{\ell}(\boldsymbol{\theta}) \equiv \mathbf{0}$, where $\dot{\ell}(\boldsymbol{\theta})$ denotes the score vector of first derivatives of the logarithm of the likelihood function for $\boldsymbol{\theta}$, namely $\ell(\boldsymbol{\theta})$. In our case, if we consider the EVBS model (the procedure is similar for the EVBS* model), this function is given by $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, where, for $i = 1, ..., n$,

$$\ell_i(\boldsymbol{\theta}) = \begin{cases} \log(A_{t_i}) - \left(1 + \frac{1}{\gamma}\right)\log(1 + \gamma\, a_{t_i}) - (1 + \gamma\, a_{t_i})^{-1/\gamma}; & \gamma \neq 0\,, \\ \log(A_{t_i}) - \exp(-a_{t_i}) - a_{t_i}\,; & \gamma = 0\,, \end{cases}$$

with $\boldsymbol{\theta} = [\alpha, \beta, \gamma]^{\top}$. The score vector $\dot{\ell}(\boldsymbol{\theta}) = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = [\dot{\ell}_{\theta_1}]$, with $\theta_1 = \alpha$, β, or γ, is given by

$$\begin{bmatrix} -\dfrac{n}{\alpha} + \displaystyle\sum_{i=1}^{n}\left(\frac{\{1+\gamma\}\, a_{t_i}}{\alpha\{1+\gamma\, a_{t_i}\}} - \frac{1}{\alpha}\, a_{t_i}\{1+\gamma\, a_{t_i}\}^{-1-1/\gamma}\right) \\ \displaystyle\sum_{i=1}^{n}\left(-\frac{\alpha}{2\,\beta}\frac{a_{t_i}}{\sqrt{\frac{t_i}{\beta}}+\sqrt{\frac{\beta}{t_i}}} + \frac{\{1+\gamma\}}{2\,\alpha\beta\{1+\gamma\, a_{t_i}\}}\left\{\sqrt{\frac{t_i}{\beta}}+\sqrt{\frac{\beta}{t_i}}\right\} - \frac{\sqrt{\frac{t_i}{\beta}}+\sqrt{\frac{\beta}{t_i}}}{2\,\alpha\beta}\{1+\gamma\, a_{t_i}\}^{-1-1/\gamma}\right) \\ \displaystyle\sum_{i=1}^{n}\left(-\frac{\{1+\gamma\, a_{t_i}\}^{-1/\gamma}}{\gamma^2}\left\{\log(1+\gamma\, a_{t_i}) - \frac{\gamma\, a_{t_i}}{1+\gamma\, a_{t_i}}\right\} - \frac{\{\gamma+1\}\, a_{t_i}}{\gamma\{1+\gamma\, a_{t_i}\}} + \frac{\log(1+\gamma\, a_{t_i})}{\gamma^2}\right) \end{bmatrix},$$

whenever $\gamma \neq 0$ and, if $\gamma = 0$, is given by

$$\begin{bmatrix} -\dfrac{n}{\alpha} - \displaystyle\sum_{i=1}^{n}\left(\frac{1}{\alpha}\exp(-a_{t_i})\, a_{t_i} - \frac{1}{\alpha}\, a_{t_i}\right) \\ \displaystyle\sum_{i=1}^{n}\left(-\frac{1}{2\,\alpha\beta}\exp(-a_{t_i})\left\{\sqrt{\frac{t_i}{\beta}}+\sqrt{\frac{\beta}{t_i}}\right\} + \frac{1}{2\,\alpha\beta}\left\{\sqrt{\frac{t_i}{\beta}}+\sqrt{\frac{\beta}{t_i}}\right\} - \frac{\alpha}{2\beta}\frac{a_{t_i}}{\sqrt{t_i/\beta}+\sqrt{\beta/t_i}}\right) \end{bmatrix}.$$

In this case, the system of likelihood equations $\dot{\ell}(\boldsymbol{\theta}) \equiv \mathbf{0}$ does not produce an explicit solution so that a numerical procedure is necessary. To this end, initial values for the parameters α, β and γ can be obtained using the methods to be described in Subsection 6.2. In addition, these likelihood equations seem to be often unstable.

We propose to use the following approach for solving this problem of instability. The approach consists of obtaining the optimum value for the parameter γ assuming it to be known, for example, following a similar algorithm to that proposed by Rinne (2009, pp. 426–433) and called by him as non-failing (NF); see also Barros *et al.* (2009) and Leiva *et al.* (2011). In these works, they fixed values for their parameter, in our case γ, within a set of several possible values for this parameter, and they then estimate the structural parameters, in our case, α and β. Finally, we consider the fixed γ that maximizes the likelihood function. Specifically, this approach is based on a partition of the real number set into a suitable amount of sub-intervals. Fixing γ in each of these intervals, we estimate α and β by using the ML method and then we look for the value of γ that maximizes the likelihood function. In this case, the NF algorithm is given by:

NF1 For a fixed value of γ:

- **NF1.1** Estimate the parameters α and β of the EVBS model using the estimates of α and β from the procedure to be described in Subsection 6.2 as starting values.
- **NF1.2** Compute the associated likelihood function.

NF2 Choose the value of γ that maximizes the likelihood function and then consider the obtained ML estimates of α and β as result.

6.2. Starting estimation

Firstly, to find initial values for the numerical optimization procedure needed for the ML estimation of the EVBS distribution parameters described in Subsection 6.1, we introduce a graphical method analogous to the probability plots; see Leiva *et al.* (2008a). This method is useful for goodness-of-fit and can also be used as an estimation method or, at least, to find initial values for an iterative procedure. The method consists of transforming the data forming pairs of values that should follow a linear relationship if these data would come from the EVBS distribution. Then, by using a simple linear regression method, the slope and intercept of this linear relationship are estimated. The line is used for goodness-of-fit such as a quantile versus quantile (QQ) plot. Specifically, if we consider the EVBS c.d.f. as given in (2.2), but with Φ replaced by F_Z, we have $t = \beta + \alpha\sqrt{\beta}\sqrt{t}\,F_Z^{-1}(F_T(t))$, where F_Z^{-1} is the generalized inverse c.d.f. of the generator EV distribution and F_T is the EVBS c.d.f. However, it is difficult to derive a linear function over t in the above expression, which is fundamental for probability plots. We consider $p = \sqrt{t}\,F_Z^{-1}(F_T(t))$ obtaining the linear function $y \approx a + b\,x$, where $x = p$, $y = t$, the intercept is $a = \beta$, and the slope is $b = \alpha\sqrt{\beta}$. Now, suppose we have n ordered observations, say $t_{(1)} \leq ... \leq t_{(n)}$. Because we can estimate $F_T(t_{(i)})$ by $q_i = (i - 0.3)/(n + 0.4)$, for $i = 1, ..., n$, the graphical plot of $t_{(i)}$ versus $\bar{p}_i$, where $\bar{p}_i = \sqrt{t_{(i)}}\, F_Z^{-1}(q_i)$ is approximately a straight line whenever the data come from some EVBS distribution. Goodness-of-fit can be visually and analytically studied using the coefficient of determination of the fit of regression between $t_{(i)}$ and $\bar{p}_i$. Therefore, the parameters α and β of this distribution can be estimated by using the least square method obtaining $\bar{\beta} = \bar{a}$ and $\bar{\alpha} = \bar{b}/\sqrt{\bar{a}}$.

Secondly, to find an initial value for the parameter γ, we can use a landmark from the EV theory. This landmark is the result about the limiting generalized Pareto (GP) behavior of the scaled excesses; see, e.g., Balkema and de Haan (1974) and Pickands (1975). This enables the development of the so-called ML EV index estimators, which we can take as an initial value for γ to be used for the numerical optimization procedure needed for the ML estimation of the EVBS distribution parameters. We refer the peaks over threshold methodology of estimation (see Smith, 1987) as well as the methodology used by Drees *et al.* (2004), named peaks over random threshold in Araújo Santos *et al.* (2006).

6.3. Model checking

Once the EVBS distribution parameters are estimated, a natural question that arises is checking how good is the fit of the model to the data. We can use the invariance property of the ML estimators for fitting the p.d.f. and c.d.f. of the EVBS model. Also, to compare the EVBS distributions to other distributions, we can use model selection procedure based on loss of information, such as Akaike (AIC), Schwarz's Bayesian (BIC) and Hannan–Quinn (HQIC) information criteria. These criteria allow us to compare models for the same data set and are given by $\mathrm{AIC} = -2\,\ell(\widehat{\boldsymbol{\theta}}) + 2d$, $\mathrm{BIC} = -2\,\ell(\widehat{\boldsymbol{\theta}}) + d\log(n)$, and $\mathrm{HQIC} = -2\,\ell(\widehat{\boldsymbol{\theta}}) + 2\,d\log(\log(n))$, where, as mentioned, $\ell(\widehat{\boldsymbol{\theta}})$ is the log-likelihood function for the parameter $\boldsymbol{\theta}$ associated with the model evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, n is the sample size, and d is the dimension of the parameter space.

AIC, BIC and HQIC are based on a penalization of the likelihood function as the model becomes more complex, i.e., with more parameters. Thus, a model whose information criterion has a smaller value is better. This is an important point, because the EVBS distribution has more parameters than the usual BS distribution. Because models with more parameters always provide a better fit, AIB, BIC and HQIC allow us to compare models with different numbers of parameters due to the penalization incorporated in such criteria. This methodology is very general and can be applied even to non-nested models, i.e., those models that are not particular cases of a more general model; see Vilca *et al.* (2011) and references therein.

Generally, differences between two values of the information criteria are not very noticeable. In that case, the Bayes factor (BF) can be used to highlight such differences, if they exist. To define the BF, assume the data D belong to one of two hypothetical models, namely M_1 and M_2, according to probabilities $\mathbb{P}(\mathrm{D}|\mathrm{M}_1)$ and $\mathbb{P}(\mathrm{D}|\mathrm{M}_2)$, respectively. Given probabilities $\mathbb{P}(\mathrm{M}_1)$ and $\mathbb{P}(\mathrm{M}_2) = 1 - \mathbb{P}(\mathrm{M}_1)$, the data produce conditional probabilities $\mathbb{P}(\mathrm{M}_1|\mathrm{D})$ and $\mathbb{P}(\mathrm{M}_2|\mathrm{D}) = 1 - \mathbb{P}(\mathrm{M}_1|\mathrm{D})$, respectively. Then, the BF that allow us to compare M_1 (model considered as correct) to M_2 (model to be contrasted with M_1) is given by

$$B_{12} = \frac{\mathbb{P}(\mathrm{D}|\mathrm{M}_1)}{\mathbb{P}(\mathrm{D}|\mathrm{M}_2)} \ . \tag{6.1}$$

Based on (6.1), we can use the approximation

$$2\log(B_{12}) \approx 2\big[\ell(\widehat{\boldsymbol{\theta}}_1) - \ell(\widehat{\boldsymbol{\theta}}_2)\big] - \big[d_1 - d_2\big]\log(n) \ , \tag{6.2}$$

where $\ell(\widehat{\boldsymbol{\theta}}_k)$ is the log-likelihood function for the parameter $\boldsymbol{\theta}_k$ under the model M_k evaluated at $\boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k$, d_k is the dimension of $\boldsymbol{\theta}_k$, for $k = 1, 2$, and n is the sample size. Notice that the approximation in (6.2) is computed subtracting the BIC value from the model M_2, given by $\mathrm{BIC}_2 = -2\,\ell(\boldsymbol{\theta}_2) + d_2\log(n)$, to the BIC value of the model M_1, given by $\mathrm{BIC}_1 = -2\,\ell(\boldsymbol{\theta}_1) + d_1\log(n)$. In addition, notice that if model M_2 is a particular case of M_1, then the procedure corresponds to applying the likelihood ratio (LR) test. In this case, $2\log(B_{12}) \approx \chi^2_{12} - \mathrm{df}_{12}\log(n)$, where χ^2_{12} is the LR test statistic for testing M_1 versus M_2 and $\mathrm{df}_{12} = d_1 - d_2$ are the d.f.'s associated with the LR test, so that one can obtain the corresponding p-value from $2\log(B_{12}) \dot\sim \chi^2(d_1 - d_2)$, with $d_1 > d_2$.

In general, the BF is informative because it presents ranges of values in which the degree of superiority of one model with respect to another can be quantified. An interesting interpretation of the BF is displayed in Table 1; see Vilca *et al.* (2011) and references therein.

Table 1: Interpretation of $2\log(B_{12})$ associated with the BF.

$2\log(B_{12})$	Evidence in favor of M_1
< 0	Negative (M_2 is accepted)
$[0, 2)$	Weak
$[2, 6)$	Positive
$[6, 10)$	Strong
≥ 10	Very strong

7. APPLICATION

In this section, to illustrate some of the results obtained in this study, we fit the EVBS* model (for minimum) to a real data set corresponding to air pollutant concentrations. We assume that the data are uncorrelated and independent and, therefore, a diurnal or cyclic trend analysis is not necessary. This assumption has been supported by some authors for different reasons; see, e.g., Vilca *et al.* (2010) and references therein. For example, environmental data are sometimes reported as average or total values and so spatiotemporal dependence is missing. In this analysis, we first discuss an implementation in `R` code of the EVBS model. Next, the data set upon analysis is introduced. Then, an EDA is produced. Finally, estimation and EVBS model checking are carried out.

7.1. Implementation in R code

Several `R` packages for analyzing data from different distributions are available from CRAN (for example, the `bs` and `gbs` packages). An `R` package named `evbs` to analyze data from EVBS models is being developed by the authors, whose "in progress" version is available upon request. This package contains diverse indicators and methodologies useful for EVBS distributions. In addition, the `evbs` package incorporates the scaled TTT curve as a descriptive tool to identify the possible shape of the h.r.

7.2. The data set

The data correspond to daily ozone concentrations that were collected in New York during May–September, 1973. These data were taken from Nadarajah (2008) and have been provided by the New York State Department of Conservation. This set of daily ozone level measurements (in ppb = ppm $\times$1000), that we call from now simply `ozone`, are: 41, 36, 12, 18, 28, 23, 19, 8, 7, 16, 11, 14, 18, 14, 34, 6, 30, 1, 11, 4, 32, 23, 45, 115, 37, 29, 71, 39, 23, 21, 37, 20, 12, 13, 49, 32, 64, 40, 77, 97, 97, 85, 10, 27, 7, 48, 35, 61, 79, 63, 16, 108, 20, 52, 82, 50, 64, 59, 39, 9, 16, 78, 35, 66, 122, 89, 110, 44, 65, 22, 59, 23, 31, 44, 21, 9, 45, 168, 73, 76, 118, 84, 85, 96, 78, 91, 47, 32, 20, 23, 21, 24, 44, 21, 28, 9, 13, 46, 18, 13, 24, 16, 23, 36, 7, 14, 30, 14, 18, 20, 11, 135, 80, 28, 73, 13.

7.3. Exploratory data analysis

Firstly, an analysis of autocorrelation indicates that there is not such autocorrelation so that the dependence over time can be discarded. Thus, the use of a methodology based on univariate random samples is adequate for ozone. Secondly, Table 2 presents a descriptive summary of these data, which includes standard deviation (SD) and coefficients of variation (CV), skewness (CS) and kurtosis (CK). Figure 4 (left) displays their histogram. This table and histogram indicate a positively skewed distribution. Thirdly, the TTT plot shown in Figure 4 (right) indicates that ozone seems to have a h.r. that is coherent with that of the EVBS distributions. However, maybe the more relevant aspect of the EDA of these ozone levels is noted when we analyze the original boxplot and the adjusted boxplot for asymmetric distributions. Interestingly, the original boxplot displayed in Figure 4 (first plot on the center figure) shows some atypical observations lying on the right-tail of the distribution of ozone, but this boxplot was constructed for symmetric data. When we produce the adjusted boxplot for asymmetric distributions using ozone, there are not atypical observations on the right-tail. Nevertheless, this type of observations appear on the left-tail of the distribution of the data; see Figure 4 (second plot on the center figure). For more details about this adjusted boxplot for asymmetric data, see Hubert and Vandervieren (2008), an `R` package called `robustbase` and its function `adjbox`. Therefore, the EDA provides to us diverse evidences for supporting the use of the EVBS model to describe ozone.

Table 2: Descriptive statistics for ozone (in ppb = ppm × 1000).

Median	Mean	SD	CV	CS	CK	Range	Min.	Max.	n
31.50	42.13	32.99	78.30%	1.21	3.11	167.00	1.00	168.00	116

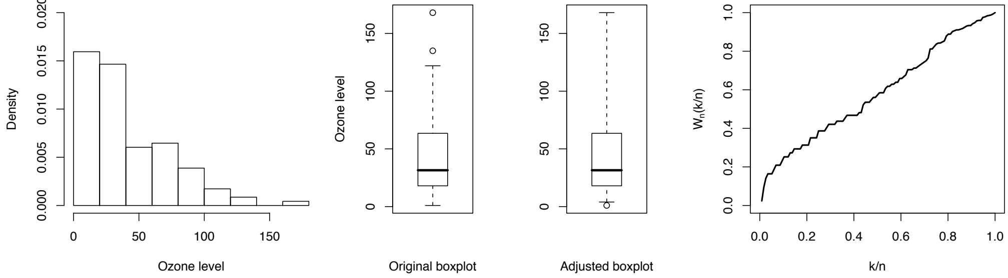


Figure 4: Histogram (left), indicated boxplots (center) and TTT plot (right) for ozone.

The EVBS* distribution should accommodate the observations concentrated on the left-tail well. Then, we think the EVBS* distribution based on the $\text{Gumbel}_{\min}$ model should be an appropriate model for describing ozone. Because this model belongs to the Gumbel min-domain of attraction (see Subsection 4.2), we carry out a semi-parametric EV test to analyze whether ozone belongs to this domain or not. Specifically, we want to test $H_0\colon F \in \mathcal{D}_{\mathrm{m}}(G^*_\gamma)$, with $\gamma \geq 0$. For details about this test, see Dietrich *et al.* (2002). In Figure 5, we see the sample path of the test statistic as a function of the k largest order statistics and the critical value (horizontal line) above which we reject the null hypothesis. We do not reject H_0 for $1 \leq k \leq 60$, which is a credible result in EV theory to keep this hypothesis. Observe that we cannot have $\gamma > 0$ because the left endpoint $-\infty$ does not make sense for these data (the daily ozone measures must be greater or equal to 0).

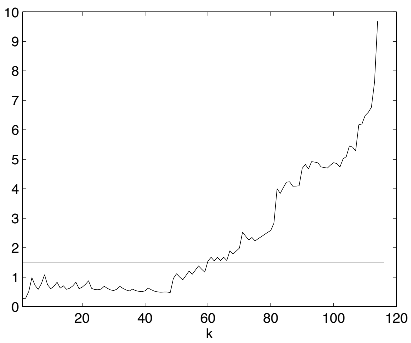


Figure 5: Sample path of the extreme value condition test applied to the ozone data (horizontal line: critical value above which we reject $F \in \mathcal{D}_{\mathrm{m}}(G^*_\gamma)$, with $\gamma \geq 0$).

Next, we fit the EVBS* model based on the $\text{Gumbel}_{\min}$ distribution to ozone. GEV and GP models are also considered as comparative models. In addition, a skew-normal BS (SNBS) model is also fitted, because according to Vilca *et al.* (2011), a SNBS distribution has heavier tails than the usual BS distribution. This characteristic can also be obtained by, for instance, a BS model based on the Student-t distribution. Moreover, when the distribution of the data is concentrated on the left-tail, a SNBS distribution should be a better alternative than the usual BS distribution or than a BS model based on any symmetric distribution, such as the Student-t model. In fact, if ozone comes from a SNBS model and we fit the usual BS distribution, we overestimate the lower percentiles. However, the EVBS distributions introduced here are also good alternatives for modeling data following a distribution with heavier tails than the usual BS distribution.

7.4. Estimation and checking model

To find the ML estimates of the EVBS distribution parameters, we use the procedure described in Subsections 6.1 and 6.2. Thus, based on ozone, we obtain the ML estimates along with the values of AIC, BIC, HQIC and BF used for model selection; see Table 3.

Table 3: ML estimates, information criteria and Bayes factors in the indicated models for ozone.

Distribution	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$-\ell$	AIC	BIC	HQIC	$2\log(B_{12})$
$\text{EVBS}^*(\alpha,\beta,\gamma)$	0.80	45.68	0.00	541.31	1088.62	1088.81	1084.51	—
$\text{GEV}(\mu,\sigma,\gamma)$	24.00	18.34	0.36	543.78	1093.55	1093.75	1089.44	4.93
$\text{GP}(\sigma,\gamma)$	—	52.49	−0.25	546.19	1096.37	1096.50	1093.63	5.00
$\text{SNBS}(\alpha,\beta,\lambda)$	1.27	14.84	1.07	545.61	1097.21	1097.40	1093.10	8.59
$\text{BS}(\alpha,\beta)$	0.98	28.02	—	549.10	1102.19	1102.32	1099.45	10.82

From Table 3, we note that the EVBS* model based on the Gumbel$_{\min}$ distribution has lower values of AIC, BIC and HQIC with respect to the BS, EV, GP and SNBS models for ozone data. This is a first indication of the superiority of the proposed model. Then, we use the BF to establish the magnitude of the differences between the values of the BIC of the proposed model and of its competitors. Thus, according to Table 1 and the BF's (approximated by the BIC's) given in Table 3, we detect for the EVBS* model (i) a very strong evidence in its favor with respect to the BS model, (ii) a strong evidence with respect to the SNBS model and (iii) a positive evidence with respect to the GEV and GP models. This is a second more power indication of the superiority of the proposed model.

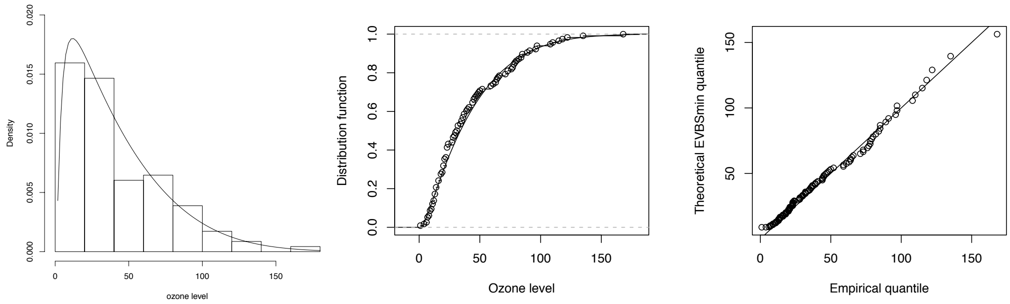


Figure 6: Histogram with estimated EVBS* (Gumbel$_{\min}$) p.d.f. (left), empirical and estimated EVBS* theoretical c.d.f. plots (center) and QQ plot (right) for ozone.

Now, from Figure 6 (center), we see the excellent coherence between the empirical and EVBS theoretical c.d.f.'s. for `ozone`. Moreover, a QQ plot for the EVBS distribution shown in Figure 6 (right) confirms such a coherence between the EVBS* model and the data. In fact, the histogram and the estimated EVBS* p.d.f. based on the $\text{Gumbel}_{\text{min}}$ distribution provided in Figure 6 (left) also shows an excellent fit of the EVBS* model to these ozone data. Therefore, we conclude that the EVBS* distribution provides a much better fit than the other considered models for the ozone data analyzed in this study.

8. CONCLUDING REMARKS

This article has dealt with an extreme value version of the Birnbaum–Saunders distribution. Specifically, we have found the density of the extreme value Birnbaum–Saunders distribution and discussed its shape. We have obtained the cumulative distribution and quantile functions of this distribution as well as highlighted some of their properties. Extremal domains of attraction for Birnbaum–Saunders type distributions have been studied. A characterization of the hazard rate of extreme value Birnbaum–Saunders distributions has been also carried out. We have developed an `R` package with the obtained results and used part of it for analyzing a real data set of ozone concentrations. This analysis has allowed us to show the adequacy of these new statistical distributions.

ACKNOWLEDGMENTS

Marta Ferreira was partially supported by the Research Centre of Mathematics of the University of Minho through the FCT Pluriannual Funding Program and by PTDC/FEDER grants from Portugal. M. Ivette Gomes was partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia) by PEst-OE/MAT/UI0006/2011, FCT/OE, POCI 2010 and PTDC/FEDER grants from Portugal. Víctor Leiva was supported by FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico) by 1120879 grant and project DIPUV 43/2010, both from Chile.

REFERENCES

[1] Ahmed, S.E.; Castro-Kuriss, C.; Flores, E.; Leiva, V. and Sanhueza, A. (2010). A truncated version of the Birnbaum–Saunders distribution with an application in financial risk, *Pak. J. Statist.*, **26**, 293–311.

[2] Araújo Santos, P.; Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation, *Revstat*, **4**, 227–247.

[3] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.*, **2**, 792–804.

[4] Barros, M.; Paula, G.A. and Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics, *Lifetime Data Anal.*, **14**, 316–332.

[5] Barros, M.; Paula, G.A. and Leiva, V. (2009). An R implementation for generalized Birnbaum–Saunders distributions, *Comp. Statist. Data Anal.*, **53**, 1511–1528.

[6] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Application*, John Wiley and Sons, New York.

[7] Beirlant, J.; Caeiro, F. and Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes, *Revstat*, **10**, 1–31.

[8] Bhatti, C.R. (2010). The Birnbaum–Saunders autoregressive conditional duration model, *Math. Comp. Simul.*, **80**, 2062–2078.

[9] Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions, *J. Appl. Probab.*, **6**, 319–327.

[10] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

[11] Díaz-García, J.A. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions, *J. Statist. Plan. Infer.*, **128**, 445–457 (Erratum: *J. Statist. Plan. Infer.*, **137**, 1512–1513).

[12] Dietrich, D.; de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions, *Extremes*, **5**, 71–85.

[13] Drees, H.; Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Ann. Appl. Probab.*, **14**, 1179–1201.

[14] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[15] Ferreira, M. and Canto e Castro, L. (2008). Tail and dependence behavior of levels that persist for a fixed period of time, *Extremes*, **11**, 113–133.

[16] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, Krieger, FL.

[17] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**, 423–453.

[18] Gomes, M.I.; Canto e Castro, L.; Fraga Alves, M.I. and Pestana, D. (2008a). Statistics of extremes for iid data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions, *Extremes*, **11**, 3–34.

[19] Gomes, M.I.; de Haan, L. and Henriques-Rodrigues, L. (2008b). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses, *J. R. Statist. Soc. B*, **70**, 31–52.

[20] Gómez, H.W.; Olivares, J. and Bolfarine, H. (2009). An extension of the generalized Birnbaum–Saunders distribution, *Statist. Probab. Letters*, **79**, 331–338.

[21] Guiraud, P.; Leiva, V. and Fierro, R. (2009). A non-central version of the Birnbaum–Saunders distribution for reliability analysis, *IEEE Trans. Rel.*, **58**, 152–160.

[22] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[23] Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions, *Comp. Statist. Data Anal.*, **52**, 5186–5201.

[24] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions* – Vol. 2, John Wiley and Sons, New York.

[25] Leadbetter, M.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

[26] Leiva, V.; Hernández, H. and Riquelme, M. (2006). A new package for the Birnbaum–Saunders distribution, *R Journal*, **6**, 35–40.

[27] Leiva, V.; Barros, M.; Paula, G.A. and Galea, M. (2007). Influence diagnostics in log-Birnbaum–Saunders regression models with censored data, *Comp. Statist. Data Anal.*, **51**, 5694–5707.

[28] Leiva, V.; Sanhueza, A. and Sen, P.K. (2008a). Random number generators for the generalized Birnbaum–Saunders distribution, *J. Statist. Comp. Simul.*, **78**, 1105–1118.

[29] Leiva, V.; Barros, M.; Paula, G. and Sanhueza, D. (2008b). Generalized Birnbaum–Saunders distributions applied to air pollutant concentration, *Environmetrics*, **19**, 235–249.

[30] Leiva, V.; Sanhueza, A. and Saunders, S.C. (2009a). New developments and applications on life distributions under cumulative damage, Technical Report No. 2009-04. `http://www.cimfav.cl/reports.html#2009`

[31] Leiva, V.; Sanhueza, A. and Angulo, J.M. (2009b). A length-biased version of the Birnbaum–Saunders distribution with application in water quality, *Stoch. Environ. Res. Risk Assess.*, **23**, 299–307.

[32] Leiva, V.; Sanhueza, A. and Kotz, S. (2010a). A unified mixture model based on the inverse Gaussian distribution, *Pak. J. Statist.*, **26**, 445–460.

[33] Leiva, V.; Vilca, F.; Balakrishnan, N. and Sanhueza, A. (2010b). A skewed sinh-normal distribution and its properties and application to air pollution, *Comm. Statist. Theor. Meth.*, **39**, 426–443.

[34] Leiva, V.; Athayde, E.; Azevedo, C. and Marchant, C. (2011). Modeling wind energy flux by a Birnbaum–Saunders distribution with unknown shift parameter, *J. Appl. Statist.*, **38**, 2819–2838.

[35] Marshall, A.W. and Olkin, I. (2007). *Life Distributions*, Springer, New York.

[36] Nadarajah, S. (2008). A truncated inverted beta distribution with application to air pollution data, *Stoch. Environ. Res. Risk Assess.*, **22**, 285–289.

[37] Pickands, J. (1986). The continuous and differentiable domains of attraction of the extreme-value distribution, *Ann. Probab.*, **14**, 996–1004.

[38] Podlaski, R. (2008). Characterization of diameter distribution data in near-natural forests using the Birnbaum–Saunders distribution, *Can. J. Forest Res.*, **18**, 518–526.

[39] Rinne, H. (2009). *The Weibull Distribution*, Chapman and Hall/CRS press, Boca Raton, FL.

[40] Sanhueza, A.; Leiva, V. and Balakrishnan, N. (2008). The generalized Birnbaum–Saunders distribution and its theory, methodology and application, *Comm. Statist. Theor. Meth.*, **37**, 645–670.

[41] Saunders, S.C. (2007). *Reliability, Life Testing and Prediction of Services Lives*, Springer, New York.

[42] Scarrot, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *Revstat*, **10**, 33–60.

[43] Smith, R.L. (1987). Estimating tails of probability distributions, *Ann. Statist.*, **15**, 1174–1207.

[44] Vilca, F. and Leiva, V. (2006). A new fatigue life model based on the family of skew elliptic distributions, *Comm. Statist. Theor. Meth.*, **35**, 229–244.

[45] Vilca, F.; Sanhueza, A.; Leiva, V. and Christakos, G. (2010). An extended Birnbaum–Saunders model and its application in the study of environmental quality in Santiago, Chile, *Stoch. Environ. Res. Risk Assess.*, **24**, 771–782.

[46] Vilca, F.; Santana, L.; Leiva, V. and Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum–Saunders distributions, *Comp. Statist. Data Anal.*, **55**, 1665–1678.

REVSTAT – Statistical Journal
Volume 10, Number 2, June 2012, 211–227

COMPARISON OF CONFIDENCE INTERVALS FOR THE POISSON MEAN: SOME NEW ASPECTS

Authors: V.V. Patil
– Department of Statistics, Tuljaram Chaturchand College, Baramati, Maharashtra, India
vaishu_kop@indiatimes.com

H.V. Kulkarni
– Department of Statistics, Shivaji University, Kolhapur, Maharashtra, India
kulkarnih1@rediffmail.com

Received: February 2010 Revised: May 2011 Accepted: January 2012

Abstract:

- We perform a comparative study among nineteen methods of interval estimation of the Poisson mean, in the intervals (0,2), [2,4] and (4,50], using as criteria coverage, expected length of confidence intervals, balance of noncoverage probabilities, E(P-bias) and E(P-confidence). The study leads to recommendations regarding the use of particular methods depending on the demands of a particular statistical investigation and prior judgment regarding the parameter value if any.

Key-Words:

- *confidence intervals (CI); Poisson mean; comparison.*

AMS Subject Classification:

- 62-07, 62F25.

1. INTRODUCTION

Construction of CIs in discrete distributions is a widely addressed problem. The standard method of obtaining a $100 \times (1-\alpha)\%$ CI for the Poisson mean μ is based on inverting an equal tailed test for the null hypothesis $H_0: \mu = \mu_0$. This is an "exact" CI, in the sense that it is constructed using the exact distribution.

Exact CIs are very conservative and too wide. A large number of alternate methods for obtaining CIs for μ based on approximations for the Poisson distribution are suggested in the literature to overcome these drawbacks. Desirable properties of those approximate CIs are:

- for $(1-\alpha)$ confidence interval the infimum over μ of the coverage probability should be equal to $(1-\alpha)$;
- confidence interval can not be shortened without the infimum of the coverage falling below $(1-\alpha)$.

We attempt to perform an exhaustive review of the existing methods for obtaining confidence intervals for the Poisson parameter and present an extensive comparison among these methods based on the following criterion:

1) Expected length of confidence intervals (E(LOC)),

2) Percent coverages (Coverage),

3) E(P-bias) and E(P-confidence),

4) Balance of right and left noncoverage probabilities.

Section 2 enumerates several methods for interval estimation of μ, giving appropriate references. Section 3 describes criteria used for comparison, Section 4 reports details of the comparative study and Section 5 presents concluding remarks.

2. A REVIEW OF THE EXISTING METHODS

Table 1 presented next reports 19 CIs for the Poisson mean. In the table, "Schwertman and Martinez" is abbreviated as SM, "Freeman and Tukey" by FT, "Wilson and Hilferty" by WH, "continuity correction" by CC, "Second Normal" by SN and "Likelihood Ratio" by LR. Furthermore, $\alpha_1 = \alpha/2$, $\alpha_2 = 1-\alpha/2$, and $x_c = x + c$ for any number c.

Table 1: Confidence limits for the nineteen methods.

Name and reference	Lower Limit	Upper Limit
1: Garwood (GW) (1936)	$(\chi^2_{(2x,\alpha_1)})/2$	$(\chi^2_{(2x_1,\alpha_2)})/2$
2: WH (WH) (1931)	$x\big(1-1/9x+Z_{\alpha_1}/3\sqrt{x}\big)^3$	$x_1\big(1-1/9x_1+Z_{\alpha_2}/3\sqrt{x_1}\big)^3$
3: Wald (W) SM (1994)	$x+Z_{\alpha_1}\sqrt{x}$	$x+Z_{\alpha_2}\sqrt{x}$
4: SN (SN) SM (1994)	$x+Z^2_{\alpha_1}/2+Z_{\alpha_1}\sqrt{x+Z^2_{\alpha_1}/4}$	$x+Z^2_{\alpha_1}/2+Z_{\alpha_2}\sqrt{x+Z^2_{\alpha_2}/4}$
5: Wald CC (FNCC) SM (1994)	$x_{-0.5}+Z_{\alpha_1}\sqrt{x_{-0.5}}$	$x_{0.5}+Z_{\alpha_2}\sqrt{x_{0.5}}$
6: SN CC (SNCC) SM (1994)	$x_{-0.5}+Z^2_{\alpha_1}/2+Z_{\alpha_1}$ $\big(x_{-0.5}+Z^2_{\alpha_2}/4\big)^{.5}$	$x_{0.5}+Z^2_{\alpha_2}/2+Z_{\alpha_2}$ $\big(x_{0.5}+Z^2_{\alpha_1}/4\big)^{.5}$
7: Molenaar (MOL) (1970)	$x_{-0.5}+\big(2Z^2_{\alpha_1}+1\big)/6+Z_{\alpha_1}$ $\big(x_{-0.5}+\big(Z^2_{\alpha_1}+2\big)/18\big)^{.5}$	$x_{0.5}+\big(2Z^2_{\alpha_2}+1\big)/6+Z_{\alpha_2}$ $\big(x_{0.5}+\big(Z^2_{\alpha_2}+2\big)/18\big)^{.5}$
8: Bartlett (BART) (1936)	$\big(\sqrt{x}+Z_{\alpha_1}/2\big)^2$	$\big(\sqrt{x}+Z_{\alpha_2}/2\big)^2$
9: Vandenbroucke (SR) (1982)	$\big(\sqrt{x_c}+Z_{\alpha_1}/2\big)^2$	$\big(\sqrt{x_c}+Z_{\alpha_2}/2\big)^2$
10: Anscombe (ANS) (1948)	$\big(\sqrt{x+3/8}+Z_{\alpha_1}/2\big)^2-3/8$	$\big(\sqrt{x+3/8}+Z_{\alpha_2}/2\big)^2-3/8$
11: FT (FT) (1950)	$0.25\Big(\big(\sqrt{x}+\sqrt{x_1}+Z_{\alpha_1}\big)^2-1\Big)$	$0.25\Big(\big(\sqrt{x}+\sqrt{x_1}+Z_{\alpha_2}\big)^2-1\Big)$
12: Hald (H) (1952)	$\big(\sqrt{x_{-.5}}+Z_{\alpha_1}/2\big)^2+.5$	$\big(\sqrt{x_{-.5}}+Z_{\alpha_2}/2\big)^2+.5$
13: Begaud (BB) (2005)	$\big(\sqrt{x_{.02}}+Z_{\alpha_1}/2\big)^2$	$\big(\sqrt{x_{.96}}+Z_{\alpha_2}/2\big)^2$
14: Modified Wald (MW) Barker (2002)	For $x=0$; 0 For $x>0$; Wald limit	For $x=0$; $-\log(\alpha_1)$ For $x>0$; Wald limit
15: Modified Bartlett (MB) Barker (2002)	For $x=0$; 0 For $x>0$; Bartlett limit	For $x=0$; $-\log(\alpha_1)$ For $x>0$; Bartlett limit
16: LR (LR) Brown *et al.* (2003)	No closed form	No closed form
17: Jeffreys (JFR) Brown *et al.* (2003)	$G\big(\alpha_1,x_{0.5},1/r\big)$	$G\big(\alpha_2,x_{0.5},1/r\big)$
18: Mid-P Lancaster (1961)	No closed form	No closed form
19: Approximate Bootstrap Confidence (ABC) Swift (2009)	$x+\dfrac{Z_0+Z_{\alpha_1}}{\big(1-a(Z_0+Z_{\alpha_1})\big)^2}\sqrt{x}$ where $a=Z_0=1/(6\sqrt{x})$	$x+\dfrac{Z_0+Z_{\alpha_2}}{\big(1-a(Z_0+Z_{\alpha_2})\big)^2}\sqrt{x}$

3. CRITERIA FOR COMPARISON

The criteria considered for the comparison among the above mentioned CIs are E(LOC) of CIs, coverage probability, ratio of the left to right noncoverage probabilities, E(P-confidence) and E(P-bias).

Here we explain the details of the three criterion for comparison mentioned in Section 1. Without loss of generality a sample of size $n = 1$ is considered. The comparisons are carried out over $\mu \in (0, 50]$.

The expected value of a function $g(x)$ is computed as $\sum_{x=0}^{\infty} g(x)\, p_\mu(x)$ where $p_\mu(x) = e^{-\mu}\mu^x/x!$. The infinite sums in the computation of these quantities were approximated by appropriate finite ones up to 0.001 margin of error.

The coverage probability $C(\mu)$, noncoverage probability on the left $L(\mu)$, noncoverage probability on the right $R(\mu)$, and corresponding expected length E(LOC) of a CI $\big(l(x), u(x)\big)$ are respectively computed by taking $g(x) = I\big(l(x) \le \mu \le u(x)\big)$, $I\big(\mu > u(x)\big)$, $I\big(\mu < l(x)\big)$ and $\big(u(x) - l(x)\big)$, where $I(\cdot)$ is the indicator function of the bracketed event.

3.1. Computation of E(P-confidence) and E(P-bias)

Let CI(x) be the CI obtained for the observation x having nominal level $(1-\alpha)\,100\%$. The P-bias and P-confidence are defined in terms of the standard equal tailed P-value function $P(\mu, x) = \min\big(2\,P_\mu(X \le x), 2\,P_\mu(X \ge x), 1\big)$. The P-confidence of the CI that measures how strongly the observation x rejects parameter values outside CI is defined as $C_p\big(\mathrm{CI}(x), x\big) = \big(1 - \sup_{\mu \notin \mathrm{CI}(x)} P(\mu, x)\big) \times 100\%$.

The P-bias of a CI which quantifies the largeness of P-values for values of μ outside the CI in comparison with those inside the CI is given by $b\big(\mathrm{CI}(x), x\big) = \max\big(0,\ \sup_{\mu \notin \mathrm{CI}(x)} P(\mu, x) - \inf_{\mu \in \mathrm{CI}(x)} P(\mu, x)\big) \times 100\%$. For the Poisson distribution $P(\mu, x)$ is continuous and a monotone function in μ in opposite directions to the left and right of the interval for each value of x. Hence the supremums and infimums occur at the upper or lower end points of the CIs. Consequently the formulae of P-bias and P-confidence are reduced to

$$C_p\big(\mathrm{CI}(x), x\big) = \Big(1 - \max\Big\{2\,P\big(X \ge x;\ \mu = l(x)\big), 2\,P\big(X \le x;\ \mu = u(x)\big)\Big\}\Big) \times 100\%\,,$$

$$b\big(\mathrm{CI}(x), x\big) = \max\Big\{0, \Big\{2\,P\big(X \ge x;\ \mu = l(x)\big) - 2\,P\big(X \le x;\ \mu = u(x)\big)\Big\}\Big\} \times 100\%\,.$$

Their expected values are computed as described above.

It was observed that when the actual value of μ is a fraction, the CI with their endpoints rounded to the nearest integer (for lower limit, rounding to an integer less than the limit and reverse for the upper limit) improved coverage probabilities to a very large extent at the cost of increasing E(LOC) at most by one unit. This is clearly visible from Figure 1 which displays the Box plot of coverages for the rounded and unrounded CIs obtained using Wald method. Similar pattern was observed for other methods.

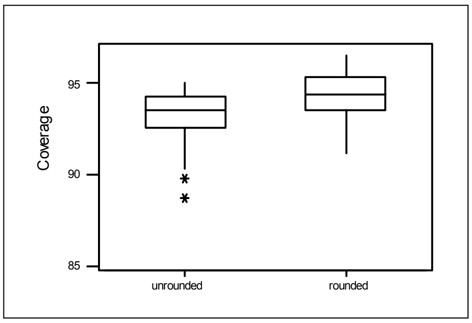


Figure 1: Impact of rounding on coverage of Wald CI.

Consequently the E(LOC) and percent coverages reported here correspond to these rounded intervals and the comparison carried out among the methods in the sequel is based on rounded intervals.

4. COMPARISON AMONG THE METHODS

4.1. Comparison based on coverages and E(LOC)

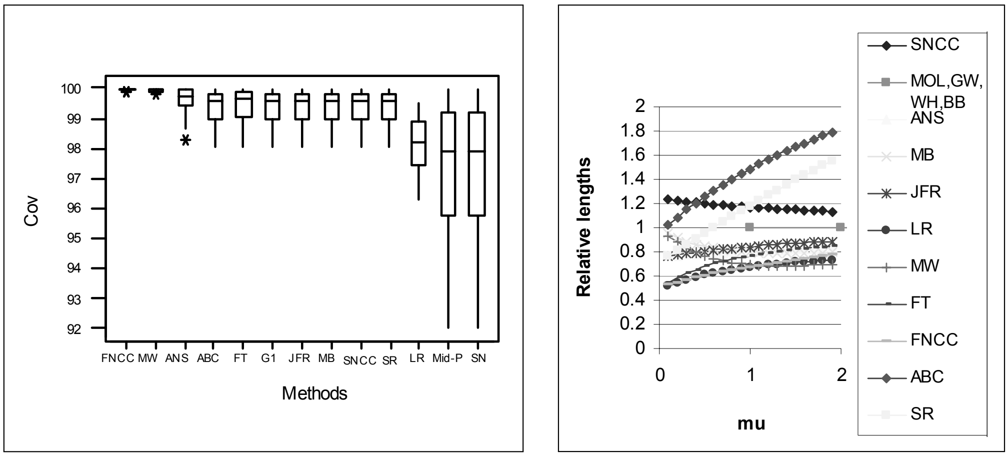
On careful examination revealed that different methods perform differently in certain subsets of the parameter space.

Consequently the performance of each method was studied separately on the three regions, namely (0,2), [2,4] and (4,50] in the parameter space. Panels (a) and (b) of Figures 2A to 4A display respectively the boxplots of coverages and graphs of relative E(LOC) of conservative methods (i.e. ratio of E(LOC) for the concerned method to the same for Garwood exact CI) for different regions defined above. Figures 2B to 4B display similar plots for nonconservative methods.

The observations from these graphs are tabulated in Table 2. The methods displayed in bold face have shortest length among the concerned group. Here $G1 = \{\mathrm{GW, MOL, WH, BB}\}$ and $G2 = \{\mathrm{BART, W, H}\}$.

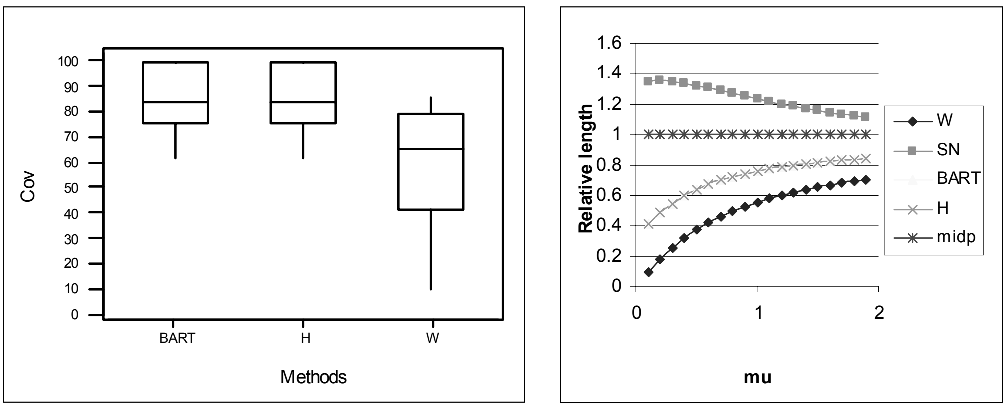
Table 2: Coverage performance of the nineteen methods.

Type	$\mu \in$ (0,2)	$\mu \in$ [2,4]	$\mu \in$ (4,50]
Conservative	**FNCC**, **LR**, ANS, $G1$, FT, JFR, MB, MW, SN SNCC, ABC, SR, Mid-P	**MB**, **ANS**, SN, $G1$ ABC, SR, JFR SNCC, Mid-P	$\boldsymbol{G1}$, SNCC, ABC, LR H, BART, MW, ANS FT, MB, SN, Mid-P JFR, FNCC, W, SR
Non-Conservative	$G2$	$G2$, FNCC, FT, LR MW	—



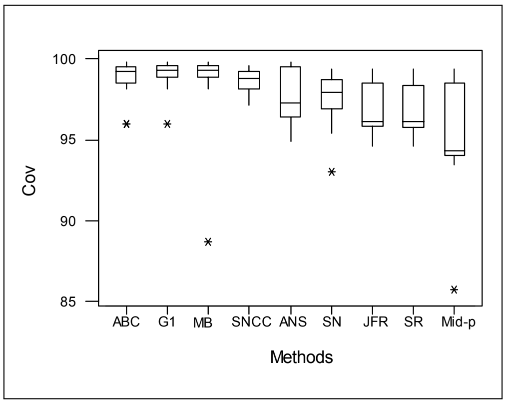
(a) Boxplot of coverages for conservative methods. **(b)** Relative lengths of conservative methods.

Figure 2A: Coverages and relative E(LOC) for conservative methods for parametric space (0,2), where $G1 = \{\text{GW}, \text{MOL}, \text{WH}, \text{BB}\}$.

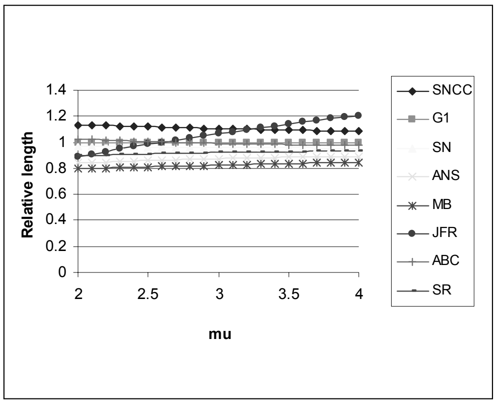


(a) Boxplot of coverages for nonconservative methods. **(b)** Relative lengths of nonconservative methods.

Figure 2B: Coverages and relative E(LOC) for nonconservative methods for parametric space (0,2).

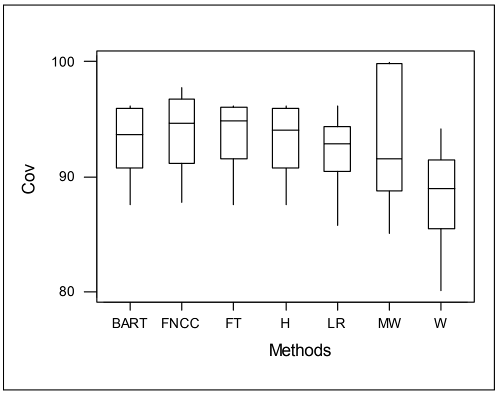


(a) Boxplot of coverages for conservative methods.

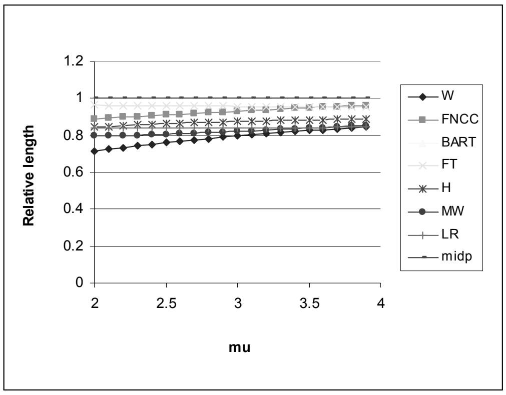


(b) Relative lengths of conservative methods.

Figure 3A: Coverages and relative E(LOC) for conservative methods for parametric space [2,4], where $G1 = \{\mathrm{GW}, \mathrm{MOL}, \mathrm{WH}, \mathrm{BB}\}$.

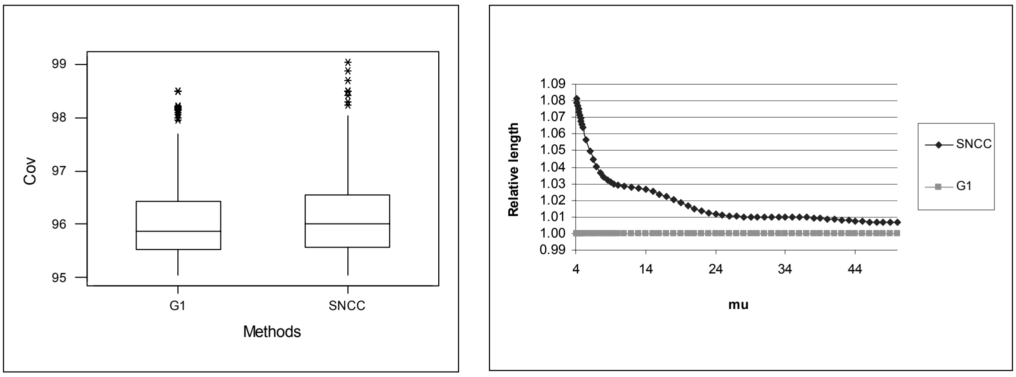


(a) Boxplot of coverages for nonconservative methods.



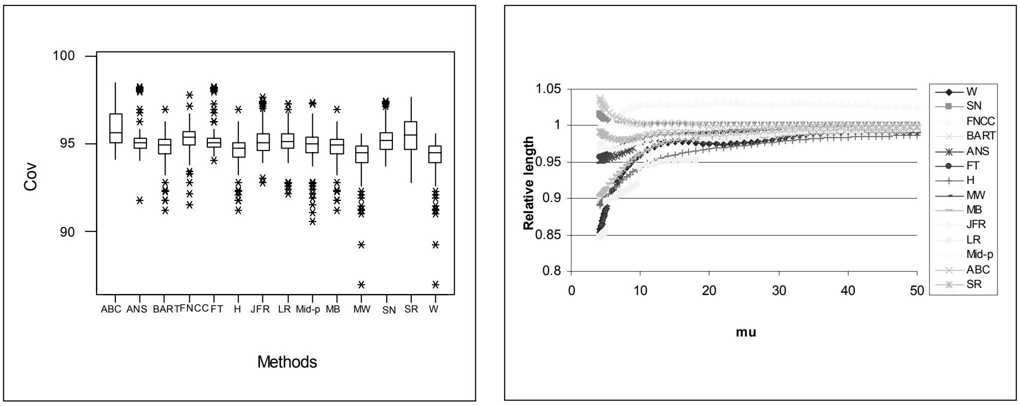
(b) Relative lengths of nonconservative methods.

Figure 3B: Coverages and relative E(LOC) for nonconservative methods for parametric space [2,4].



(a) Boxplot of coverages for conservative methods. **(b)** Relative lengths of conservative methods.

Figure 4A: Coverages and relative E(LOC) for conservative methods for parametric space (4,50], where $G1 = \{\mathrm{GW}, \mathrm{MOL}, \mathrm{WH}, \mathrm{BB}\}$.



(a) Boxplot of coverages for nonconservative methods. **(b)** Relative lengths of nonconservative methods.

Figure 4B: Coverages and relative E(LOC) for nonconservative methods for parametric space (4,50].

4.2. Comparison with respect to balance of noncoverage probabilities

For a two sided CI procedure it is desirable to have the right and left noncoverage probabilities to be fairly balanced. We plot the ratio of the left to right noncoverage probabilities as a function of Poisson mean for the nineteen methods in Figure 5A and 5B for regions (2,4) and (4,50). For balanced noncoverage, ratio should oscillate in the close neighborhood of 1. For region (0,2) all methods are well below 1, with the exception of Wald method.

A careful observation of figures leads to the following region wise performance of methods with respect to right-to-left noncoverage balance reported in Table 3.

Table 3: Performance based on right-to-left noncoverage balance.

Performance	$\mu \in (2,4)$	$\mu \in (4,50)$
Fairly balanced around 1	—	$G1$, ABC, LR, JFR, SR
Uniformly below 1	SNCC, SN, $G1$, ABC, MB	SN, SNCC, Mid-P
Uniformly above 1	LR, JFR, SR, Mid-P FT, ANS, FNCC, MW, $G2$	FT, MB, ANS, FNCC, MW, $G2$

4.3. Comparison based on E(P-bias) and E(P-confidence)

For comparison of methods on the basis of E(P-bias) and E(P-confidence), we consider three regions of sample space (0,2), (2,4) and (4,50). Three panels (a) to (c) of Figures 6 and 7 represent boxplots of E(P-confidence) and E(P-bias) for these three regions. Recommendations on the basis of E(P-bias) and E(P-confidence) for two regions tabulated in Table 4.

Table 4: Recommendations on the basis of E(P-bias) and E(P-confidence).

Performance	$\mu \in (0,2)$	$\mu \in (2,4)$	$\mu \in (4,50)$
Smallest E(P-bias) Largest E(P-confidence)	FNCC, MW, W SNCC, SN, ABC, $G1$	FNCC, SNCC SNCC, SN, $G1$	SNCC, Mid-P, SN SNCC, $G1$, SN

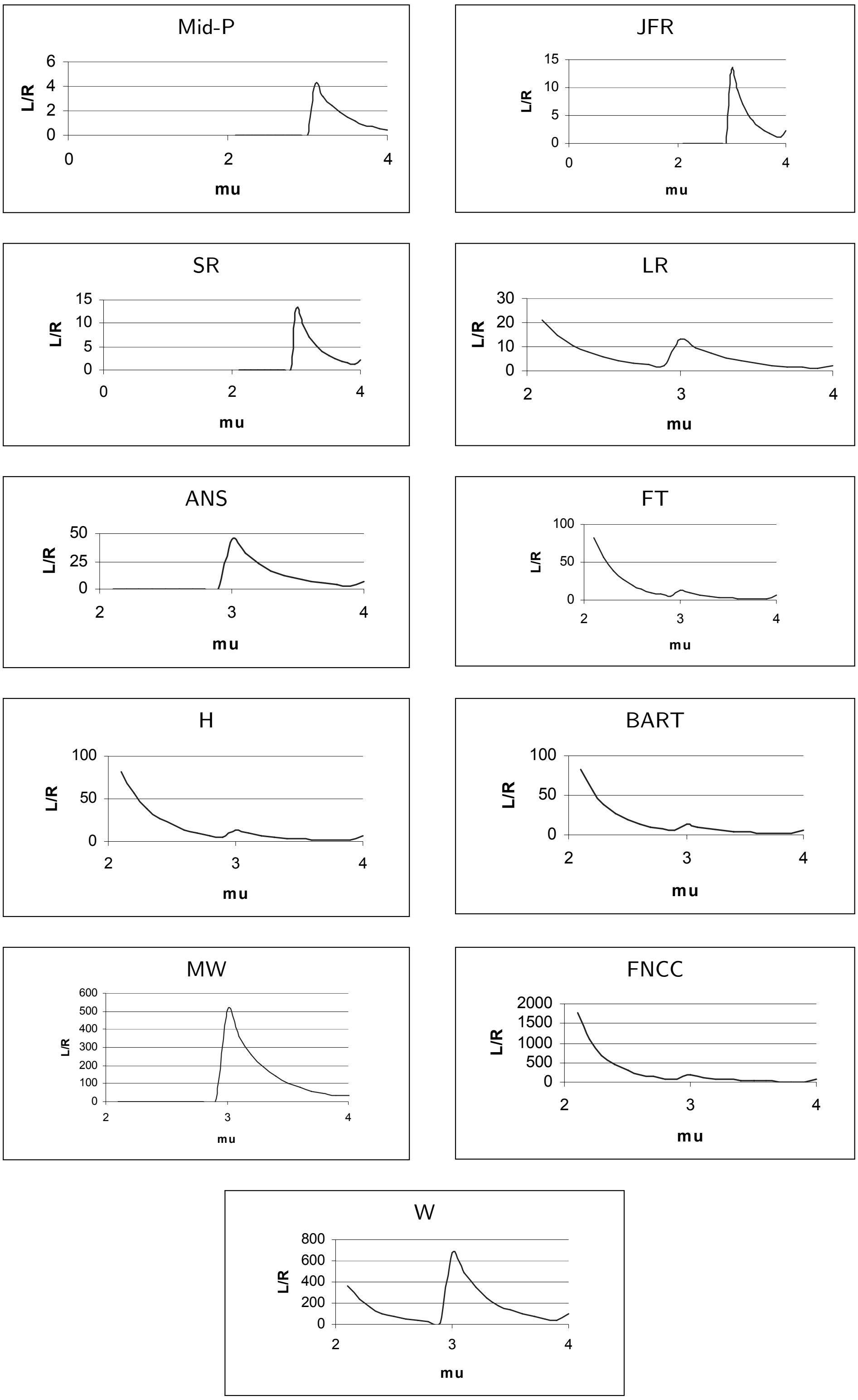
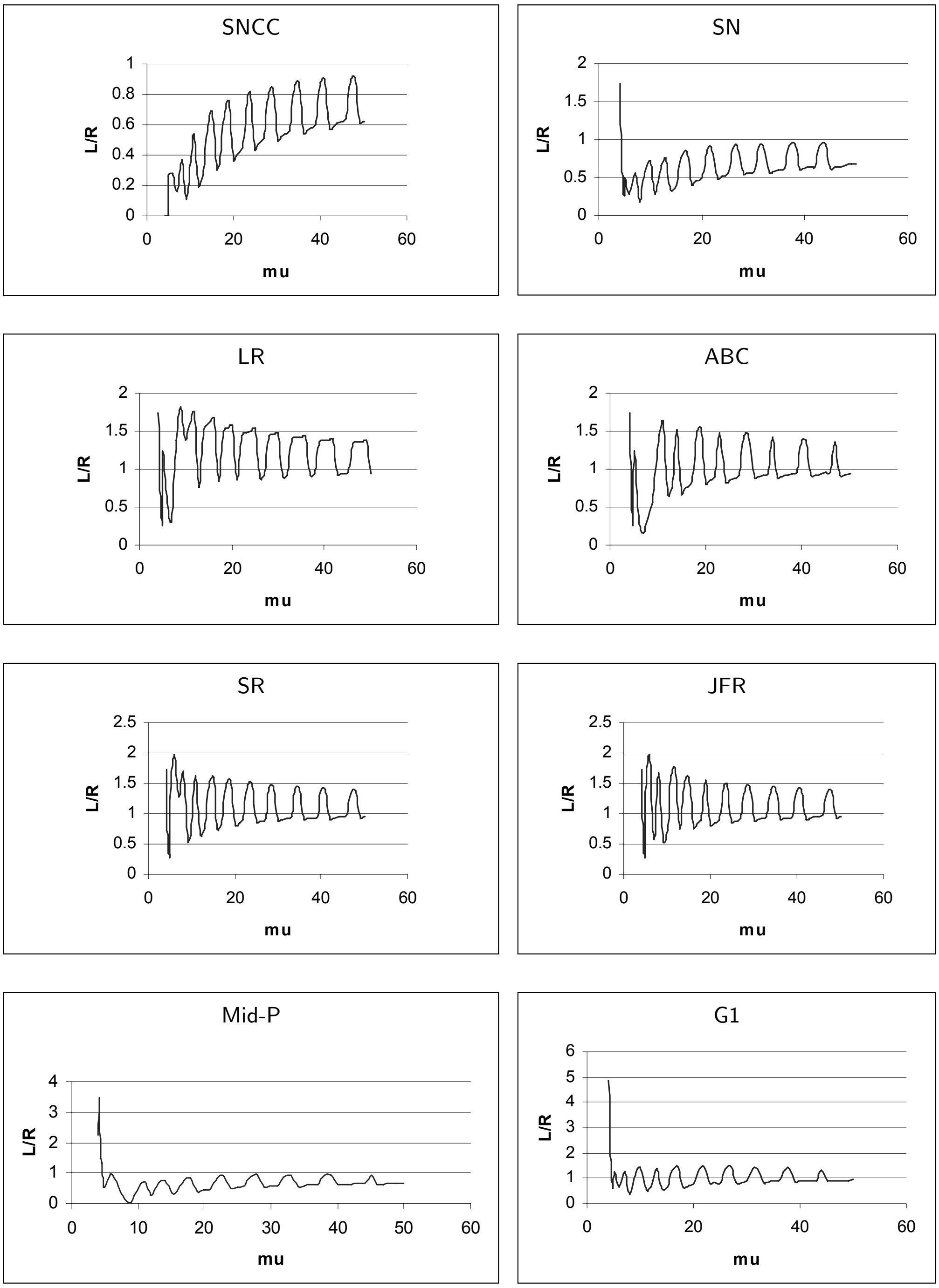


Figure 5A: Graph of ratio of noncoverage probabilities for parametric space (2,4]. The ratio of noncoverage probabilities for methods SNCC, SN, $G1$, ABC and MB are zero for parametric space (2,4].


SNCC
L/R
mu
SN
L/R
mu
LR
L/R
mu
ABC
L/R
mu
SR
L/R
mu
JFR
L/R
mu
Mid-P
L/R
mu
G1
L/R
mu

(continues)

(continued)

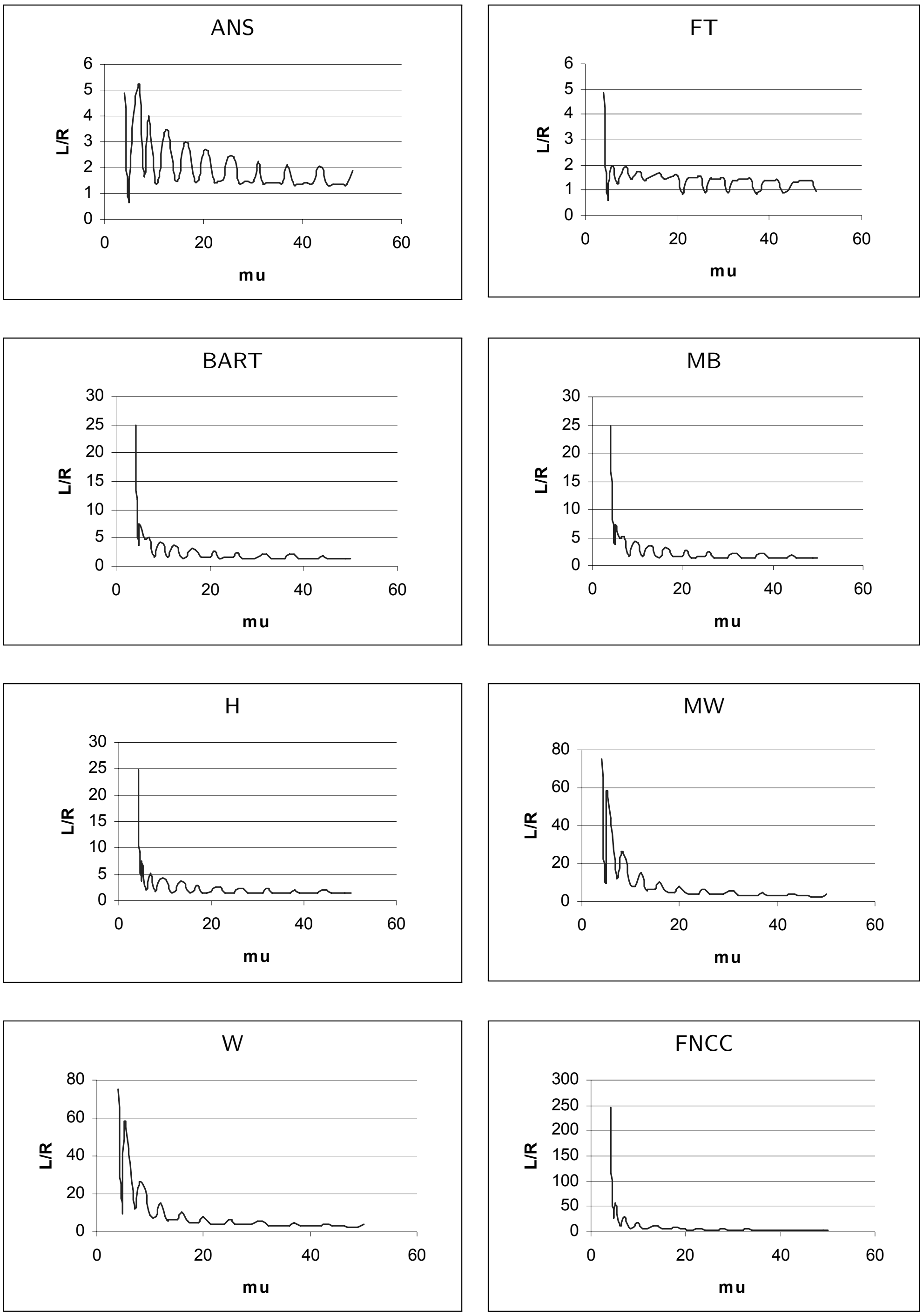


Figure 5B: Graph of ratio of non coverage probabilities for parametric space (4,50].

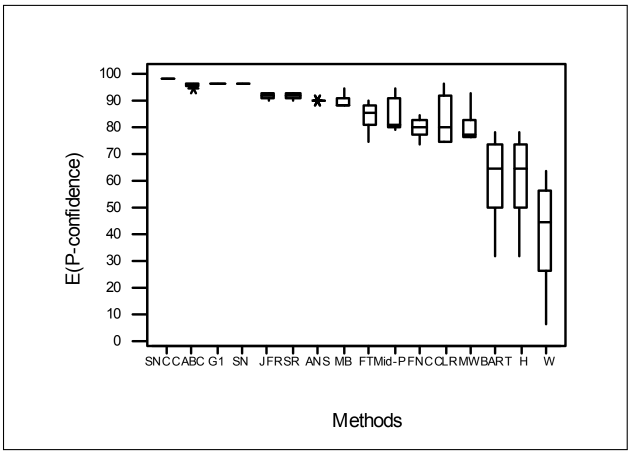


Figure 6(a): Boxplot of E(P-confidence) for parametric space (0,2].

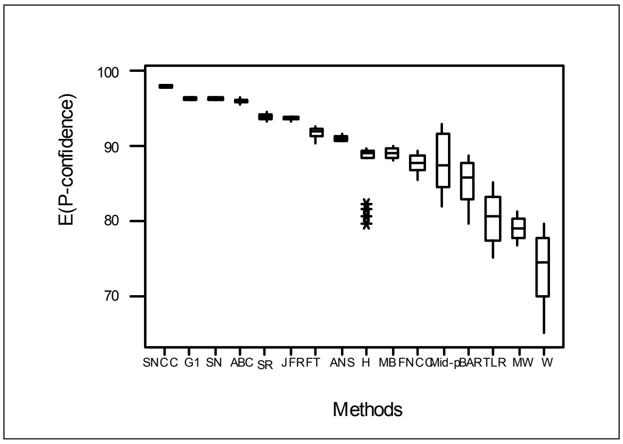


Figure 6(b): Boxplot of E(P-confidence) for parametric space (2,4].

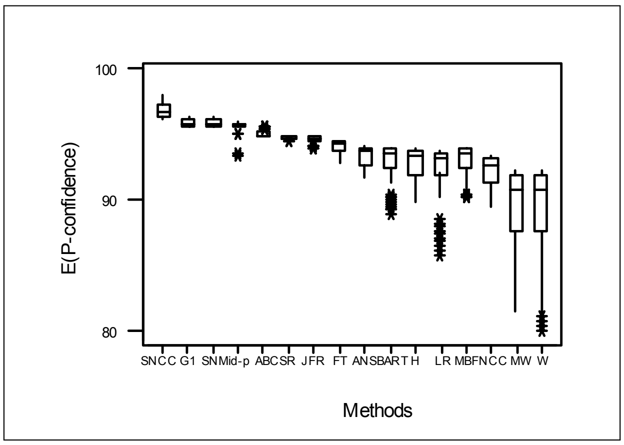


Figure 6(c): Boxplot of E(P-confidence) for parametric space (4,50).

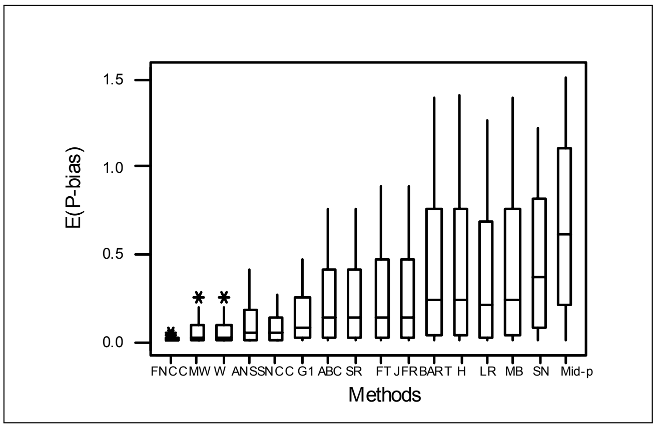


Figure 7(a): Boxplot of E(P-bias) for parametric space (0,2].

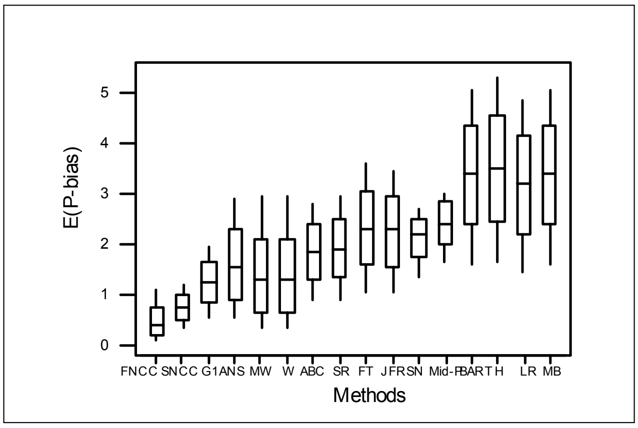


Figure 7(b): Boxplot of E(P-bias) for parametric space (2,4].

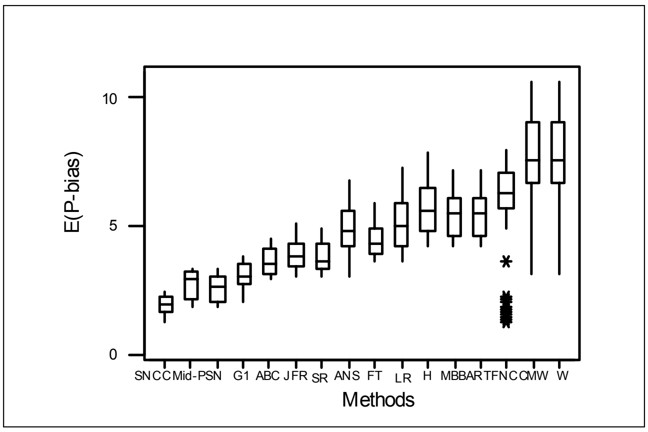


Figure 7(c): Boxplot of E(P-bias) for parametric space (4,50).

5. CONCLUDING REMARKS

Rounding of end points of CI considerably improves the coverages of CI. Our remarks are based on rounded intervals. A best choice for CI depends on the objectives of the underlying investigations and a broad prior knowledge about the underlying parameter if any.

Finally, our investigation suggests the following recommendations:

1) In the analysis of rare events where μ is expected to be very small *in between 0 to 2*, we recommend MW and FNCC method on the basis of highest coverage probabilities with shortest expected length and smallest expected P-bias and reasonable expected P-confidence. In this region LR is also recommendable on the basis of all the criteria except E(P-bias).

2) For the situations where the parameter is expected to be large *more than 4*, methods involved in $G1$ are the best choice. In fact the performance of methods in $G1$ is uniformly satisfactory (if not best) on the entire parameter space with respect to all the criteria, so in the absence of any knowledge regarding the underlying parameter, we recommend these methods for use.

3) We strongly recommend to avoid using W, BART, and H methods in all kinds of applications, since these are uniformly nonconservative for all parameter values, have large E(P-bias) and smallest E(P-confidence) and highly imbalanced noncoverage on the right and left side.

These recommendations are useful guidelines for consulting professionals, in data analysis, software development, and can be an interesting addition to the discussion of case studies in Applied Statistic courses.

ACKNOWLEDGMENTS

We are very much grateful to anonymous referees and the Associated Editor for many suggestions that greatly improved the manuscript. H.V. Kulkarni was supported by the grants received by Government of India, Department of Science and Technology, India, under the project Reference No. SR/S4/MS: 306/05.

REFERENCES

[1] Anscombe, F.J. (1948). The transformation of Poisson, Binomial and Negative Binomial data, *Biometrika*, **35**, 246–254.

[2] Barker, L. (2002). A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is ≤ 5, *The American Statistician*, **56**(2), 86–89.

[3] Bartlett, M.S. (1936). The square root transformation in the analysis of variance, *Journal of Royal Statistical Society Supplement*, **3**, 68–78.

[4] Begaud, B.; Karin, M.; Abdelilah, A.; Pascale, T.; Nicholas, M. and Yola, M. (2005). An easy to use method to approximate Poisson confidence limits, *European Journal of Epidemiology*, **20**(3), 213–216.

[5] Brown, Cai and Dasgupta (2003). Interval estimation in exponential families, *Statistica Sinica*, **13**, 19–49.

[6] Freeman, M.F. and Tukey, J.W. (1950). Transformations related to the angular and square root, *Annals of Mathematical Statistics*, **21**, 606–611.

[7] Garwood, F. (1936). Fiducial limits for the Poisson distribution, *Biometrika*, **28**, 437–442.

[8] Hald, A. (1952). *Statistical Theory with Engineering Applications*, John Wiley and Sons, New York.

[9] Lancaster, H.O. (1961). Significance tests in discrete distributions, *Journal of American Statistical Association*, **56**, 223–234.

[10] Molenaar, W. (1970). *Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions*, Mathematical Center Tracts 31, Mathematisch Centrum, Amsterdam.

[11] Schwertman, N.C. and Martinez, R.A. (1994). Approximate Poisson confidence limits, *Communication in Statistics — Theory and Methods*, **23**(5), 1507–1529.

[12] Swift, M.B. (2009). Comparison of confidence intervals for a Poisson Mean-Further considerations, *Communication in Statistics — Theory and Methods*, **38**, 748–759.

[13] Vandenbroucke, J.P. (1982). A shortcut method for calculating the 95 percent confidence interval of the standardized mortality ratio, (Letter), *American Journal of Epidemiology*, **115**, 303–304.

[14] Wilson, E.B. and Hilferty, M.M. (1931). The distribution of chi-square, *Proceedings of the National Academy of Sciences*, **17**, 684–688.

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 $i \le 3$. 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

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

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

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 $T^{a_1,\dots,a_r}_{\mathbf{x}_1,\dots,\mathbf{x}_r}$ (for example, ${a_1 \atop \mathbf{x}_1}$ and ${a_2 \atop \mathbf{x}_2}$) will not alter its value. So, $T^{a,\dots,a,b,\dots,b}_{\mathbf{x}_1,\dots,\mathbf{x}_r,\mathbf{y}_1,\dots,\mathbf{y}_s}$ is also the ordinary functional derivative of $S(F_a) = T^{b,\dots,b}_{\mathbf{y}_1,\dots,\mathbf{y}_r}$ at $(\mathbf{x}_1,\dots,\mathbf{x}_s)$.

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

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

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

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

and

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

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

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

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 $i \leq 3$) 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

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

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

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

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

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

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

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

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]:

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

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 $\left\{n^{-i}\, T_i(\widehat{F}),\ 0 \le i \le q\right\}$ 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)

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

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

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

where

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

and

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

So,

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

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

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

Set

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

Then

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

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

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

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

and

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

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

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

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

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),

$$\begin{aligned}
C_1 &= \sum_a \lambda_a\, g_{a,a}\, \mu_2[a]/2 \ , \\
C_2 &= \sum_a \lambda_a^2\, g_{a,a,a}\, \mu_3[a]/6 + \sum_{a,b} \lambda_a \lambda_b\, g_{a,a,b,b}\, \mu_2[a]\, \mu_2[b]/8 \ , \\
C_3 &= \sum \lambda_a^3\, g_{a,a,a,a} \big\{\mu_4[a] - 3\,\mu_2[a]^2\big\}/24 \\
&\quad + \sum \lambda_a \lambda_b^2\, g_{a,a,b,b,b}\, \mu_2[a]\, \mu_3[b]/12 + \sum \lambda_a \lambda_b \lambda_c\, g_{a,a,b,b,c,c}\, \mu[a]\, \mu_2[b]\, \mu_2[c]/48 \ ,
\end{aligned}$$

$$\begin{aligned}
T_1(F) &= -C_1 \ , \\
T_2(F) &= \sum \lambda_a^2\, g_{a,a,a}\, \mu_3[a]/3 + \sum \lambda_a \lambda_b\, g_{a,a,b,b}\, \mu_1[a]\, \mu_2[b]/8 - \sum \lambda_a^2\, g_{a,a}\, \mu_2[a]/2 \ , \\
T_3(F) &= -\sum \lambda_a^3\, g_{a,a}\, \mu_2[a]/2 + \sum \lambda_a^3\, g_{a,a,a}\, \mu_3[a] \\
&\quad - \sum \lambda_a^3\, g_{a,a,a,a} \big\{\mu_4[a]/4 + \mu_2[a]^2/2\big\} \\
&\quad + \sum \lambda_a^2 \lambda_b\, g_{a,a,b,b}\, \mu_2[a]\, \mu_2[b]/4 - \sum \lambda_a \lambda_b^2\, g_{a,a,b,b,b}\, \mu_2[a]\, \mu_3[b]/6 \\
&\quad - \sum \lambda_a \lambda_b \lambda_c\, g_{a,a,b,b,c,c}\, \mu_2[a]\, \mu_2[b]\, \mu_2[c]/48 \ .
\end{aligned}$$

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),

$$\begin{aligned}
C_1 &= -A_2 \ ,\\
C_2 &= A_3 - 6\,A_2B_2 \ ,\\
C_3 &= -A_4 + 3\,A_{0,2,2} + 6\,A_2B_3 + 9\,A_3B_2 - 15\,A_2B_2^2 \ ,\\
T_1(F) &= A_2 \ ,\\
T_2(F) &= 2\,A_3 - 3\,A_2B_2 + A_{1,2,1},\\
T_3(F) &= A_{2,2,1} - 9\,A_{1,3,1} - 3\,A_3 + 6\,A_4 - 12\,A_{0,2,2} - 3\,A_{1,2,1}B_2 - 3\,A_2B_{1,2,1}\\
&\quad - 8\,A_2B_3 - 12\,A_3B_2 + 15\,A_2B_2^2 \ .
\end{aligned}$$

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

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

For example,

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

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

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

For example,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

By (A.8),

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

where

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

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)

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

and

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

where

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

so that

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

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),

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

Also $\left[\begin{smallmatrix}i\\r\end{smallmatrix}\right]$ 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,

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

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

$$
\begin{aligned}
&(5.18) \quad C_1 = \mu_4 - 3\,\mu_2^2\;, \quad T_1(F) = -\mu_4 + 3\,\mu_2^2\;, \\
&(5.19) \quad C_2 = -2\,\mu_4 + 5\,\mu_2^2\;, \quad T_2(F) = -5\,\mu_4 + 10\,\mu_2^2\;, \quad S_2(F) = -4\,\mu_4 + 7\mu_2^2\;.
\end{aligned}
$$

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, $\{a_i = a_{i,2,2}(F)\}$ 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

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

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,

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

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 $\{A_i,\, i \le p/2\}$:

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

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 $\{\text{s.d.}\}/\text{bias}\,\{\sigma(\widehat{F})\}=\lambda_1+O(n^{-1})$, 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 $\{S_{n,2}(\widehat{F})\}/\text{bias}\,\{\text{s.d.}\}$ $=\lambda_2\,n^{-1}+O(n^{-2})$, 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

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

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

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

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 $a = b = 1$,

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

So, by (A.24)

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

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

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

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 $g(\mathbf{S})=S_1^2(S_2S_3)^{-1}$, so $g_1=2\,T(F)\,S_1^{-1}$, $g_2=-T(F)\,S_2^{-1}$, $g_3=-T(F)\,S_3^{-1}$, $g_{i,i}=2\,T(F)\,S_i^{-2}$, $g_{1,2}=-2\,T(F)\,(S_1S_2)^{-1}$, $g_{1,3}=-2\,T(F)\,(S_1S_3)^{-1}$, and $g_{2,3}=T(F)\,(S_2S_3)^{-1}$. 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],

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

where

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

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

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

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

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

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

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

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

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

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

where

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

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

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

and so

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

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

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

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

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

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

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 $a=b=1$

$$\begin{aligned} K_1^{\alpha,\beta}(F) &= T^{\alpha,\beta}(a,a) = g_i^\alpha\, g_j^\beta\, \mu[i,j]\ , \\ T^{\alpha,\beta}(ab,ab) &= g^\alpha_{i,j}\, g^\beta_{k,l}\, \mu[i,k]\,\mu[j,l]\ , \\ T^{\alpha,\beta}(a^2,a) &= g^\alpha_{i,j}\, g^\beta_k\, \mu[i,j,k]\ , \\ T^{\alpha,\beta}(a^2b,b) &= g^\alpha_{i,j,k}\, g^\beta_l\, \mu[i,j]\,\mu[k,l]\ , \end{aligned}$$

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 $a = b = 1$

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

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

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

For example, $\operatorname{var}\{\widehat{\mu}^{-1}\}$ or (if Proposition 4.2 needs to be applied), $\operatorname{var}\{\widehat{\mu}^{-1} I(|\widehat{\mu}|>l)\}$, 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

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

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

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

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

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

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

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

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

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

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

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

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

where $\sum^{m(\mathbf{n})}$ sums over all $m(\mathbf{n}) = r!/\prod_{i=1}^{r}(i!^{n_i}\, n_i!)$ partitions $(\pi_1, \dots, \pi_k)$ of $1, \dots, r$ 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, $\{B_{r,k}(\mathbf{S})_{i_1,\dots,i_k}\}$:

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

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

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

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 $(\mathbf{x}_1, \dots, \mathbf{x}_r)$ by $\binom{a_1,\dots,a_r}{\mathbf{x}_1,\dots,\mathbf{x}_r}$ wherever it occurs. So, in the notation of (3.1), (A.3)–(A.5) imply

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

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

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

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

where

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

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

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

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

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

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

$$
\begin{aligned}
B^{i} &= S_i(a^2b^2c^2)\ ,\\
B^{i,j} &= B_1^{i,j} + B_2^{i,j} + B_3^{i,j}\ ,\\
B_1^{i,j} &= 2\sum^3 S_{i,j}(a, ab^2c^2)\ ,\\
B_2^{i,j} &= \sum^3 S_i(a^2)\, S_j(b^2c^2) \ +\ 4\sum^3 S_{i,j}(ab, abc^2)\ ,\\
B_3^{i,j} &= 2\sum^3 S_{i,j}(a^2b, bc^2) \ +\ 4\, S_{i,j}(abc, abc)\ ,\\
B^{i,j,k} &= B_1^{i,j,k} + B_2^{i,j,k} + B_3^{i,j,k}\ ,\\
B_1^{i,j,k} &= \sum^3 S_{i,j}(a,a)\, S_k(b^2c^2) \ +\ 4\sum^3 S_{i,j,k}(a,b,abc^2)\ ,\\
B_2^{i,j,k} &= 2\sum^6 S_{i,k}(a, ac^2)\, S_j(b^2) \ +\ 4\sum^6 S_{i,j,k}(a, ab, bc^2)\\
&\quad + 8\sum^3 S_{i,j,k}(a, bc, abc)\ ,\\
B_3^{i,j,k} &= S_i(a^2)\, S_j(b^2)\, S_k(c^2) + 2\sum^3 S_i(a^2)\, S_{j,k}(bc,bc) + 8\, S_{i,j,k}(ab, bc, ca)\ ,\\
B^{i,j,k,l} &= B_1^{i,j,k,l} + B_2^{i,j,k,l}\ ,\\
B_1^{i,j,k,l} &= 2\sum^6 S_{i,j}(a^2b, b)\, S_{k,l}(c,c) + 8\, S_{i,j,k,l}(abc, a, b, c)\ ,\\
B_2^{i,j,k,l} &= \sum^3 \Big\{ S_{i,j}(a,a)\, S_k(b^2)\, S_l(c^2) + 2\, S_{i,j}(a,a)\, S_{k,l}(bc,bc)\\
&\quad + 4\, S_{i,j,k}(a,b,ab)\, S_l(c^2) + 8\, S_{i,j,k,l}(a,b,ac,bc)\Big\},\\
B^{i_1,\dots,i_5} &= \sum^3 \Big\{ S_{i_1}(a^2)\, S_{i_2,i_3}(b,b)\, S_{i_4,i_5}(c,c) + S_{i_1,i_2,i_3}(ab,a,b)\, S_{i_4,i_5}(c,c)\Big\},\\
B^{i_1,\dots,i_6} &= S_{i_1,i_2}(a,a)\, S_{i_3,i_4}(b,b)\, S_{i_5,i_6}(c,c)\ ,
\end{aligned}
$$

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\big(g_{1,1-}\,S_{1,1-} + 2\,g_{1,2-}\,S_{1,2-} + g_{2,2-}\,S_{2,2-}\big) \times&\\ \times\,\big(g_{-1,1-}\,S_{-1,1-} + 2\,g_{-1,2-}\,S_{-1,2-} + g_{-2,2-}\,S_{-2,2-}\big) \times&\\ \times\,\big(g_{-1,1}\,S_{-1,1} + 2\,g_{-1,2}\,S_{-1,2} + g_{-2,2}\,S_{-2,2}\big)&\end{aligned}$$

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

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

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

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

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

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

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

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,

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

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

$$V_F\binom{c}{z_2} = 0$$

unless $c=b$ and

$$U_F\binom{b,c}{y_1,z_1} = 0$$

unless $b=c=a$. Also

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

and

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

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 $T_F\binom{\cdot,\cdot,\cdot}{\cdot,\cdot,\cdot}$ are interchangeable we may replace $\sum^2_{x_1,x_2}$ by 2. Since

(B.1) $$\int T_F\binom{a,a,a}{x_1,x_2,y_2}\,dF_a(x_i) = 0$$

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,

$$\int^2 d_{\mathbf{x}}\,U_F\binom{a,a}{y_1,z_1}\,T_F\binom{a,a,a}{x_1,x_2,y_2} = -2\,T_F\binom{a,a,a}{y_1,z_1,y_2}.$$

So,

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

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 $V_F\binom{a}{z}$ 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,

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

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

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

we obtain

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

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 $T(a,a) = \int T_F(\substack{a \\ x})^2\, dF_a(x)$. 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:

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

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

$$\mu_{4\cdot 1} = h_1^4 - \mu_4 - 4\,h_1\mu_3 ,$$
$$\mu_{4\cdot 1,2} = 12\,h_1 h_2\mu_2 - 4(h_1^3-\mu_3)h_2 - 4\,h_1(h_2^3-\mu_3) ,$$
$$\mu_{4\cdot 1,2,3} = 12(h_1^2-\mu_2)h_2 h_3 + 12\,h_1(h_2^2-\mu_2)h_3 + 12\,h_1 h_2(h_3^2-\mu_2) ,$$
$$\mu_{4\cdot 1,2,3,4} = -72\,h_1 h_2 h_3 h_4 ,$$

$$\mu_{5\cdot 1} = h_1^5 - \mu_5 - 5\,h_1\mu_4 ,$$
$$\mu_{5\cdot 1,2} = 20\,h_1 h_2\mu_3 - 5(h_1^4-\mu_4)h_2 - 5\,h_1(h_2^4-\mu_4) ,$$
$$\mu_{5\cdot 1,2,3} = -60\,h_1 h_2 h_3\mu_2 + 20(h_1^3-\mu_3)h_2 h_3 + 20\,h_1(h_2^3-\mu_3)h_3 + 20\,h_1 h_2(h_3^3-\mu_3) ,$$
$$\mu_{5\cdot 1,2,3,4} = -60(h_1^2-\mu_2)h_2 h_3 h_4 - 60\,h_1(h_2^2-\mu_2)h_3 h_4 - 60\,h_1 h_2(h_3^2-\mu_2)h_4 - 60\,h_1 h_2 h_3(h_4^2-\mu_2) ,$$
$$\mu_{5\cdot 1,2,3,4,5} = 480\,h_1 h_2 h_3 h_4 h_5 ,$$

$$\mu_{6\cdot 1} = h_1^6 - \mu_6 - 6\,h_1\mu_5 ,$$
$$\mu_{6\cdot 1,2} = 30\,h_1 h_2\mu_4 - 6(h_1^5-\mu_5)h_2 - 6\,h_1(h_2^5-\mu_5) ,$$
$$\mu_{6\cdot 1,2,3} = -120\,h_1 h_2 h_3\mu_3 + 30(h_1^4-\mu_4)h_2 h_3 + 30\,h_1(h_2^4-\mu_4)h_3 + 30\,h_1 h_2(h_3^4-\mu_4) ,$$
$$\mu_{6\cdot 1,2,3,4}/120 = 3\,h_1 h_2 h_3 h_4\mu_2 - (h_1^3-\mu_3)h_2 h_3 h_4 - h_1(h_2^3-\mu_3)h_3 h_4 - h_1 h_2(h_3^3-\mu_3)h_4 - h_1 h_2 h_3(h_4^3-\mu_3) ,$$
$$\mu_{6\cdot 1,2,3,4,5}/360 = (h_1^2-\mu_2)h_2 h_3 h_4 h_5 + h_1(h_2^2-\mu_2)h_3 h_4 h_5 + h_1 h_2(h_3^2-\mu_2)h_4 h_5 + h_1 h_2 h_3(h_4^2-\mu_2)h_5 + h_1 h_2 h_3 h_4(h_5^2-\mu_2) .$$

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.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, I.P., in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal's website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.