BIAS, VARIANCE, SKEWNESS AND KURTOSIS OF MAXIMUM LIKELIHOOD ESTIMATORS USING MAPLE*

K.O. BOWMAN[1] and L.R. SHENTON[2]

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, P.O.Box 2008, 5700, MS-6085, Oak Ridge, TN 37831-6085, USA, bowmanko@ornl.gov

[2]Department of Statistics, University of Georgia, Athens, Georgia 30602, USA

May 11, 2005

*This submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725.

Abstract

The object of this paper is to describe the development of ideas pertaining to sample size and maximum likelihood estimators of parameters associated with a probability function or density function. About forty years ago we considered a Taylor type series for a maximum likelihood estimator $\hat{\theta}_a$ for θ_a, there being s parameters $\theta_1, \theta_2, \cdots, \theta_s$. First order bias and first order variance were included. Because of limitations in computer facilities, the skewness and kurtosis were avoided, and also because of the complicated structures involved. But toward the end of the 20th century an expression for the N^{-2} (N the sample size) term in the third central moment of $\hat{\theta}_a$ was found, and a year later a rather complicated expression for the N^{-3} term in the fourth central moment was discovered. The skewness and kurtosis expressions involved much heavier work in deriving expectations of products of log-derivatives of the probability function or density, especially when 3 or more parameters were involved. At this stage we used the Maple symbolic language to cope with the N^{-1} and N^{-2} biases, the N^{-1} and N^{-2} variances, the N^{-2} third central moment, and the N^{-3} fourth central moment. We use the $\sqrt{\beta_1} = \mu_3/\sigma^3$ to measure skewness. This ratio is location and scale free, and it takes into account the shape of the distribution involved. Since under normality $\sqrt{\beta_1} = 0$, we can set the observed value $\sqrt{\beta_1(\hat{\theta}_a)}$ for a parameter θ_a, to a small value ε and deduce a safe - sample size to achieve pseudo normality. Programs are provided in detail for the low order moments of a maximum likelihood estimator, simultaneous estimation being involved.

Key-words: Asymptotics, Lagrange Expansions, Hessian matrix, logarithmic derivative, symbolic language.

1 Introduction

Our study of small sample properties of maximum likelihood (m.l.) estimators started some forty years ago. We set up a Taylor type expansion for s parameters $\theta_1, \theta_2, \cdots, \theta_s$, and in particular $\hat{\theta}_a$, $a = 1, 2, \cdots, s$. Expectations of powers of the derivatives of the logarithm of the density were derived, up to and including the fourth power. In view of the extensive algebra involved our thoughts became directed to computer implementation, especially for the skewness and kurtosis of estimators.

The 'References' include not only subjects pertinent to the present study, but also peripheral topics such as moment methods and divergent series. In addition summaries of previous studies on estimation problems are given in Appendix B.

Since the basic structure is set out clearly in a paper we read at the Prague (1973) Symposium on Asymptotic Statistics (Shenton and Bowman, 1974), we now include an abbreviated version.

2 Asymptotic statistics (Prague Symposium)

2.1 Introduction

Suppose a population consists of categorized data, the probability of an occurrence in the jth class being $p_j(\theta)$, with n_j observations for a sample of size N. We have given expressions (Bowman and Shenton, 1965) for the N^{-1}, N^{-2} terms in the bias of $\hat{\theta}$ (the m.l. estimator), N^{-1}, N^{-2}, N^{-3} terms in $\mu_2(\hat{\theta})$, N^{-2}, N^{-3} terms in $\mu_3(\hat{\theta})$, N^{-2}, N^{-3}, N^{-4} terms in $\mu_4(\hat{\theta})$, μ_2, μ_3, μ_4 being central moments. In certain special cases higher order terms in the moments can be found by special techniques. Previously, Haldane (1953) and Haldane and Smith (1956) discussed properties of the moments of m.l. estimators in the case of one and two parameters. Actually, as has been pointed out by Cox and Snell (1968), Bartlett (1952) gave expressions for the N^{-1} biases in his paper on large - sample confidence intervals.

Here we give the first few terms in the asymptotic expansion (Lagrange) for a m.l. estimator (mentioned by name by Haldane and Smith, 1956, p.99) in the single parameter case. The N^{-2} biases and covariances in the multi-parameter case are also given with a brief outline of the derivation; these expressions have only appeared in report form previously (Bowman and Shenton, 1965). A number of miscellaneous asymptotic moments for m.l. and for comparison moment estimators, are included.

2.2 Taylor series approach

2.2.1 Single Parameter

We assume a population consists of a denumerable set of classes, there being n_j relative observations in the jth class for a sample of N with $En_j = p_j(\theta)$, where p_j depends on the single parameter θ. The log likelihood is proportional to

$$L(n,\theta) = \sum n_j \log p_j(\theta) \tag{1}$$

summed over the classes. If $\hat{\theta}$ is the m.l. estimator, then under certain regularity conditions (with $L(\theta)$ for $L(n,\theta)$)

$$L^{(1)}(\theta) + xL^{(2)}(\theta) + \frac{x^2}{2!}L^{(3)}(\theta) + \cdots = 0 \tag{2}$$

where $x = \hat{\theta} - \theta$. Now define

$$c_s = L^{(s)}(\theta)/L^{(2)}(\theta), \qquad s = 1, 2, \cdots,$$

so that $c_2 = 1$. Note that in general $E\ L^{(2)}(\theta) \neq 0$. Equation (2) may be written

$$c_1 + x\left\{1 + \frac{x}{2!}c_3 + \frac{x^2}{3!}c_4 + \cdots\right\} = 0$$

whence from Lagrange's expansion, formally, for a well-behaved function $f(\cdot)$,

$$f(x) = f(0) + \sum_1^\infty \frac{1}{s!}\frac{d^{s-1}}{dx^{s-1}}\left\{\frac{-c_1}{1+\frac{x}{2}c_3+\cdots}\right\}^s f^{(1)}(x)\,|_{x=0}\,.$$

In particular

$$\hat{\theta} = \theta + \sum_{s=1}^\infty c_1^s C_s \tag{3}$$

where

$$\begin{aligned}
&C_1 = -1, \qquad C_2 = -c_3/2,\\
&C_3 = (c_4 - 3c_3^2)/6,\\
&C_4 = -(c_5 - 10c_3c_4 + 15c_3^3)/24,\\
&C_5 = (c_6 - 10c_4^2 - 15c_3c_5 + 105c_3^2c_4 - 105c_3^4)/120,\\
&C_6 = -(c_7 - 21c_3c_6 - 35c_4c_5 + 210c_3^2c_5 + 280c_3c_4^2 - 1260c_5^3c_4 + 945c_3^5)/720
\end{aligned}$$

and so on. But $L^{(s)}(n,\theta) = \sum n_j(d^s/d\theta^s)\log p_j$, and defining the discrepancies $n_j - p_j = \varepsilon_j$, we have

$$\begin{aligned}
L^{(s)}(n,\theta) &= Lp_j\frac{d^s}{d\theta^s}\log p_j + \sum \varepsilon_j \frac{d^s}{d\theta^s}\log p_j\\
&= L_s + l_s(\varepsilon)
\end{aligned}$$

where in particular $L^{(1)} = l_1(\varepsilon)$. Hence

$$c_s = \frac{(L_s + l_s(\varepsilon))}{L_2} \sum_{r=0}^{\infty} (-l_2(\varepsilon)/L_2)^r$$

and substitution in (3) and similar equations for $(\hat{\theta} - \theta)^s$, $s = 2, 3, \cdots$, gives an expression which after taking expectations leads to the non-central moments of $\hat{\theta}$. Equation (3) can be powered digitally, and then the only further requirement is a scheme to produce expectations of low order products such as

$$\prod_{s=1}^{m} l_s^{r_s}(\varepsilon).$$

One possible approach to this problem is given by Bowman (1963).

2.2.2 Multiple parameter case

(a) N^{-2} biases and covariances

We now assume θ in (1) to be an h-component vector. The asymptotic multivariate moments in general now become very complicated in structure. In addition, it does not appear to be easy to set up an appropriate multivariate version of Lagrange which would be readily manipulatable in this case (to get some idea of the situation for multivariate Lagrange, see (i) I.J. Good, (1965, pp.499-517); (ii) L.R. Shenton and P.C. Consul (1975), (iii) P.C. Consul and L.R. Shenton, (1972, pp.13-23).

The stochastic Taylor expansion for $\hat{\theta}_a$, $(a = 1, 2, \cdots, h)$ is now

$$\hat{\theta}_a = \theta_a + \phi_1^a + \frac{1}{2!}\phi_2^a + \frac{1}{3!}\phi_3^a + \cdots$$

where

$$(i)\quad \phi_1^a = \varepsilon_r \frac{\bar{\partial}\theta_a}{\partial n_r}, \qquad \phi_2^a = \varepsilon_r \varepsilon_s \frac{\bar{\partial}^2 \theta_a}{\partial n_r \partial n_s}, \quad \text{etc.;}$$

$$(ii)\quad \varepsilon_r = n_r - p_r, \qquad E\varepsilon_r = 0;$$

$$(iii)\quad \frac{\bar{\partial}\theta_a}{\partial n_r} = \frac{\partial \hat{\theta}_a}{\partial n_r}\Big|_{n_r = p_r, \hat{\theta}_a = \theta_a}$$

and so.

(b) Notational

We define the multivariate derivative

$$\Gamma^r_{\alpha_1 \alpha_2 \cdots \alpha_m} = \frac{\partial^m}{\partial \theta_{\alpha_1} \partial \theta_{\alpha_2} \cdots \partial \theta_{\alpha_m}} \log p_r,$$

and summation being over classes,

$$(p_r\Gamma^r_{\alpha_1\alpha_2\cdots\alpha_l}\Gamma^r_{\beta_1\beta_2\cdots\beta_m}\Gamma^r_{\gamma_1\gamma_2\cdots\gamma_n}) = (p\Gamma_{\alpha_1\alpha_2\cdots\alpha_l}\Gamma_{\beta_1\beta_2\cdots\beta_m}\Gamma_{\gamma_1\gamma_2\cdots\gamma_n})$$
$$= [\alpha_1\alpha_2\cdots\alpha_l, \beta_1\beta_2\cdots\beta_m, \gamma_1\gamma_2\cdots\gamma_n].$$

For example

$$[\alpha,\beta,\gamma] = (p\Gamma_\alpha\Gamma_\beta\Gamma_\gamma) = \sum_r p_r\frac{\partial\log p_r}{\partial\theta_\alpha}\frac{\partial\log p_r}{\partial\theta_\beta}\frac{\partial\log p_r}{\partial\theta_\gamma}$$

$$[\alpha\beta,\gamma] = (p\Gamma_{\alpha\beta}\Gamma_\gamma) = \sum_r p_r\frac{\partial^2\log p_r}{\partial\theta_\alpha\partial\theta_\beta}\frac{\partial\log p_r}{\partial\theta_\gamma}$$

$$[\alpha\beta\gamma] = (p\Gamma_{\alpha\beta\gamma}) = \sum_r p_r\frac{\partial^3\log p_r}{\partial\theta_\alpha\partial\theta_\beta\partial\theta_\gamma}.$$

(c) The likelihood equations for the estimates $\hat{\theta}_a$ of θ_a

We have

$$n_r\hat{\Gamma}^r_\alpha = 0 \quad (\alpha = 1, 2, \cdots, h) \tag{4}$$

where $\hat{\Gamma}^r_\alpha = \partial\log p_r(\hat{\theta})/\partial\hat{\theta}_\alpha$. Differentiating (4) partially with respect to n_r gives

$$\hat{\Gamma}^r_\alpha + (n_r\hat{\Gamma}^r_{\alpha\beta})\frac{\partial\hat{\theta}_\beta}{\partial n_r} = 0, \tag{5}$$

from which

$$\frac{\bar{\partial}\theta_\alpha}{\partial n_r} = L^{\alpha\beta}\Gamma^r_\beta \tag{6}$$

where

$$L_{\alpha\beta} = -(p\Gamma_{\alpha\beta}) = (p\Gamma_\alpha\Gamma_\beta)$$
$$L_{\alpha\beta}L^{\beta\gamma} = \varepsilon_{\alpha\gamma} \quad \text{(Kronecker delta)}$$

so that $L^{\beta\alpha}$ refers to an element in the inverse matrix $[L_{\alpha\beta}]^{-1}$ where $[L_{\alpha\beta}]$ has h rows and columns and is assumed nonsingular.

Returning to (5) and differentiating with respect to n_s, we have

$$\hat{\Gamma}^r_{\alpha\beta}\frac{\partial\hat{\theta}_\beta}{\partial n_s} + \hat{\Gamma}^s_{\alpha\beta}\frac{\partial\hat{\theta}_\beta}{\partial n_r} + (n_r\hat{\Gamma}^r_{\alpha\beta\gamma})\frac{\partial\hat{\theta}_\beta\partial\hat{\theta}_\gamma}{\partial n_r\partial n_s} + (n_r\hat{\Gamma}^r_{\alpha\beta})\frac{\partial^2\hat{\theta}_\beta}{\partial n_r\partial n_s} = 0$$

from which

$$L_{\alpha\beta}\frac{\bar{\partial}^2\theta_\beta}{\partial n_r\partial n_s} = L^{\beta\gamma}\Gamma^r_{\alpha\beta}\Gamma^s_\gamma + L^{\beta\gamma}\Gamma^s_{\alpha\beta}\Gamma^r_\gamma + [\alpha\beta\gamma]L^{\beta\delta}L^{\gamma\epsilon}\Gamma^r_\delta\Gamma^s_\epsilon. \tag{7}$$

It will be seen since that partial differentiation is usually commutative, the right hand member of (7) should also have this property with respect to r, s. This is readily verified; in fact the last member of (7) under $r \leftrightarrow s$ is invariant, using $\delta \leftrightarrow \epsilon$, $\gamma \leftrightarrow \beta$.

Similarly, expressions may be derived for the third and fourth multivariate derivatives and the associated values at $\hat{\theta} = \theta$, $n_r = p_r$ (Bowman and Shenton, 1965).

(d) Variances

For the covariances, writing $E_2\phi(\varepsilon)$ for the coefficient of N^{-2} in $E\phi(\varepsilon)$, we have

$$E_2(\hat{\theta}_a - \theta_a)(\hat{\theta}_b - \theta_b) = (6A_{12} + 2A_{13} + 3A_{22})/12, \tag{8}$$

where

$$\begin{aligned} A_{12} &= E_2(\phi_1^a\phi_2^b + \phi_2^a\phi_1^b), \\ A_{13} &= E_2(\phi_1^a\phi_3^b + \phi_3^a\phi_1^b), \\ A_{22} &= E_2\phi_2^a\phi_2^b. \end{aligned}$$

From (6) and (7)

$$\begin{aligned} E_2\phi_1^a\phi_2^b &= E_2\varepsilon_r\varepsilon_s\varepsilon_t L^{\alpha\beta}L^{\alpha b}\left\{L^{\gamma\zeta}\Gamma_\beta^r\Gamma_\gamma^s\Gamma_\zeta^t + L^{\gamma\zeta}\Gamma_\beta^r\Gamma_{\alpha\zeta}^s\Gamma_\gamma^t + [\alpha\gamma\zeta]L^{\delta\zeta}L^{\gamma\varepsilon}\Gamma_\beta^r\Gamma_\varepsilon^s\Gamma_\delta^t\right\} \\ &= 2L^{a\beta}L^{\alpha b}L^{\gamma\varepsilon}[\alpha\varepsilon,\beta,\gamma] + 2L^{ab} + L^{a\beta}L^{b\alpha}L^{\delta\zeta}L^{\gamma\varepsilon}[\alpha\gamma\zeta][\beta,\delta,\varepsilon]. \end{aligned} \tag{9}$$

Expectations of products of linear forms such as $\varepsilon_1^{r_1}\varepsilon_2^{r_2}\cdots\varepsilon_k^{r_k}$ present no particular problem, especially for low orders (see for example, Shenton, 1959, and Bowman 1963).

Expressions similar to (9) for A_{13}, A_{22} may be set up, and substituting in (8) and simplifying finally leads to

$$cov_2(\hat{\theta}_a, \hat{\theta}_b) = \Theta_1 + \Theta_3 + \Theta_4,$$

where

$$\begin{aligned} \Theta_1 &= -L^{ab}, \\ \Theta_3 &= l^{a\beta}L^{b\alpha}L^{\gamma\delta}\{[\alpha\delta,\beta,\gamma] + [\beta\delta,\alpha,\gamma] + [\alpha\beta\gamma\delta] + 3[\alpha\delta,\beta\gamma] + 2[\alpha\beta\gamma,\delta] \\ &\quad + [\beta\gamma\delta,\alpha]/2 + [\alpha\gamma\delta,\beta]/2\}, \\ \Theta_4 &= L^{a\alpha}L^{b\beta}L^{\gamma\delta}L^{\varepsilon\zeta}\{[\alpha\gamma\zeta][\beta,\delta,\varepsilon]/2 + [\beta\gamma\zeta][\alpha,\delta,\varepsilon]/2 + [\alpha\beta\gamma][\delta\varepsilon\zeta] + 5[\alpha\gamma\varepsilon][\beta\delta\zeta]/2 \\ &\quad + [\beta\gamma\zeta][\delta\varepsilon,\alpha] + [\alpha\gamma\zeta][\delta\varepsilon,\beta] + 2[\alpha\beta\zeta][\gamma\varepsilon,\delta] + 3[\beta\gamma\varepsilon][\alpha\zeta,\delta] + 3[\alpha\gamma\varepsilon][\beta\zeta,\delta] \\ &\quad + [\gamma\delta\varepsilon][\beta\zeta,\alpha]/2 + [\gamma\delta\varepsilon][\alpha\zeta,\beta]/2 + [\alpha\varepsilon,\delta][\gamma\zeta,\beta] + [\beta\varepsilon,\delta][\gamma\zeta,\alpha] + [\alpha\varepsilon,\beta][\gamma\zeta,\delta] \\ &\quad + [\beta\varepsilon,\alpha][\gamma\zeta,\delta] + [\alpha\varepsilon,\gamma][\beta\delta\zeta]\}. \end{aligned}$$

In this expression the summation is over the Greek symbols, each of which takes all values from 1 to h, the number of parameters.

(e) Biases

The N^{-2} terms here are complicated, and after extensive algebra and simplification, we find

$$E_2\hat{\theta}_a = L^{a\alpha}L^{\beta\gamma}B_2 + L^{a\alpha}L^{\beta\delta}L^{\gamma\varepsilon}B_3 + L^{a\alpha}L^{\beta\gamma}L^{\delta\zeta}L^{\varepsilon\eta}B_4 + L^{a\alpha}L^{\beta\gamma}L^{\delta\varepsilon}L^{\zeta\eta}L^{\theta i}B_5 \quad (10)$$

where

$$\begin{aligned}
2B_2 &= -[\alpha\beta\gamma] - 2[\alpha\beta,\gamma], \\
8B_3 &= [\alpha\beta\gamma\delta\varepsilon] + 4[\beta\delta\varepsilon,\alpha\gamma] + 8[\alpha\beta\gamma,\delta\varepsilon] + 4[\alpha\beta\gamma\varepsilon,\delta] + 4[\alpha\beta\gamma,\delta,\varepsilon] + 8[\alpha\beta,\gamma\delta,\varepsilon], \\
4B_4 &= (2[\alpha\delta\varepsilon\zeta][\beta\eta,\gamma] + 2[\beta\delta\varepsilon\zeta][\alpha\gamma,\eta] + 4[\alpha\beta\delta\varepsilon][\zeta\eta,\gamma]) \\
&\quad + ([\alpha\delta\varepsilon\zeta][\beta\gamma\eta] + [\alpha\beta\delta\varepsilon][\gamma\zeta\eta] + 2[\beta\delta\varepsilon\eta][\alpha\gamma\zeta]) \\
&\quad + (2[\beta\eta,\gamma\zeta][\alpha\delta\varepsilon] + 4[\beta\varepsilon,\delta\eta][\alpha\gamma\zeta] + 4[\alpha\gamma,\zeta\eta][\beta\delta\varepsilon] + 2[\alpha\eta,\delta\varepsilon][\beta\gamma\zeta]) \\
&\quad + (4[\alpha\varepsilon\zeta,\eta][\beta\delta,\gamma] + 4[\alpha\varepsilon\zeta,\gamma][\beta\delta,\eta] + 4[\beta\delta\varepsilon,\eta][\alpha\zeta,\gamma]) \\
&\quad + (2[\alpha\beta\varepsilon,\eta][\gamma\delta,\zeta] + 4[\beta\delta\varepsilon,\eta][\alpha\gamma,\zeta] + 4[\alpha\beta\varepsilon,\zeta][\gamma\delta,\eta] + 2[\delta\varepsilon\eta,\beta][\alpha\gamma,\eta]) \\
&\quad + (4[\alpha\eta,\gamma\zeta][\delta\varepsilon,\beta] + 4[\delta\varepsilon,\zeta][\alpha\eta,\beta] + 4[\alpha\eta,\delta\varepsilon][\gamma\zeta,\beta]) \\
&\quad + (4[\varepsilon\zeta,\beta\delta][\alpha\gamma,\eta] + 2[\alpha\beta,\zeta\eta][\gamma\delta\varepsilon]) + 2[\alpha\beta\delta\varepsilon][\gamma,\zeta,\eta]/3, \\
8B_5 &= [\alpha\delta\zeta]\,\{[\beta\gamma\varepsilon][\eta\theta i] + 2[\gamma\varepsilon\theta][\beta\eta i] + 4[\beta\varepsilon\eta][\gamma\theta i] + 8[\beta\eta\theta][\gamma\varepsilon i] \\
&\quad + 2[\beta\gamma\varepsilon][\eta\theta,i] + 4[\gamma\varepsilon\theta][\beta\eta,i] + 2[\eta\theta i][\beta\varepsilon,\gamma] + 4[\gamma\eta\theta][\beta\varepsilon,i] \\
&\quad + 8[\beta\eta\theta][\gamma\varepsilon,i] + 8[\beta\eta\theta][\gamma i,\varepsilon] + 8[\beta\varepsilon\eta][\gamma\theta,i] + 8[\gamma\varepsilon\theta][\beta\eta,i] \\
&\quad + 4[\gamma\theta i][\beta\eta,\varepsilon] + 4[\eta\theta,i][\beta\varepsilon,\gamma] + 4[\beta\varepsilon,i][\eta\theta,\gamma] + 8[\eta\theta,\beta][\gamma i,\varepsilon] \\
&\quad + 8[\eta\theta,\varepsilon][\gamma i,\beta] + 4[\beta\varepsilon\theta][\gamma,\eta,i]\} + [\gamma\delta i]\,\{8[\beta\eta\theta][\alpha\varepsilon,\zeta] \\
&\quad + 4[\beta\zeta\theta][\alpha\varepsilon,\theta] + 8[\alpha\varepsilon,\theta][\beta\eta,\zeta] + 8[\alpha\varepsilon,\zeta][\beta\eta,\theta] + 8[\alpha\zeta,\beta][\varepsilon\eta,\theta]\}.
\end{aligned}$$

It should be remarked that (10) is an expression involving four summatory terms and these are written as "products" merely to abbreviate.

3 Extension to the skewness and kurtosis

3.1 New formula for skewness

We see that (Bowman and Shenton, 1965) N^{-2} terms are given for bias and covariance of m.l. estimators; another account of these is given in Shenton and Bowman (1977). At that time (1960-1988) we found the problem of skewness of m.l. estimators, in the simultaneous case too complicated to consider. But after some time a formula for $\mu_{32}(\hat{\theta}_a)$ was set up; here the estimator in question is θ_a, $\hat{\theta}_a$ being the m.l. estimator, μ_{32} being the coefficient of N^{-2} in the third central moment of $\hat{\theta}_a$, a taking the values

1 to s, s being the number of parameters. For some time, the form lay in limbo. But a little later we took a second look at the correction term involved and discovered the new formula

$$\mu_{32}(\hat{\theta}_a) = L^{a1}L^{a2}L^{a3}\left\{[1,2,3] + 3[123] + 6[12,3]\right\}$$

given in Bowman and Shenton (1998, p.2751). Here, 1,2,3 run through the values 1 to s. A year later we defined in Bowman and Shenton (1999), the formula for the kurtosis (both skewness and kurtosis being measured by sample moment ratios), a rather complicated formula to say the least.

3.2 Formula for kurtosis

The formula for $\mu_{43}(\hat{\theta}_a)$ was introduced by Bowman and Shenton (1999), namely

$$\mu_{43}(\hat{\theta}_a) = A_{40} + 2A_{32} + \frac{2}{3}A_{33} + \frac{6}{4}A_{22}. \tag{11}$$

For A_{40}, A_{32}, A_{33}, and A_{22} we have

$$\begin{aligned}
A_{40} &= L^{a1}L^{a2}L^{a3}L^{a4}[1,2,3,4] - 3(L^{aa})^2,\\
A_{32} &= 2L^{a1}L^{23}L^{a4}L^{a5}L^{a6}\{[12,3][4,5,6]+[12,4][3,5,6]+[12,5][3,4,6]+[12,6][3,4,5]\}\\
&\quad + 12(L^{aa})^2 + 6L^{a1}L^{a2}L^{a3}L^{a4}[12,3,4] + 6L^{aa}L^{a1}L^{23}L^{a4}[12,3,4]\\
&\quad + [123][4,5,6]\{L^{a1}L^{23}L^{a4}L^{a5}L^{a6} + 6L^{a1}L^{a2}L^{34}L^{a5}L^{a6} + 3L^{aa}L^{a1}L^{24}L^{35}L^{a6}\}\\
&\quad - \left\{L^{a1}L^{23}\left(2[12,3]+[123]\right)\right\}\left\{L^{a4}L^{a5}L^{a6}[4,5,6]\right\}.
\end{aligned}$$

$$A_{33} = E_3(\phi_1^a)^3\phi_3^a = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4$$

where

$$\begin{aligned}
\mathcal{A}_1 &= 18L^{aa}L^{a1}L^{24}L^{a3}[123,4] + 9L^{aa}L^{a1}L^{23}L^{a4}[123,4] + 18L^{a1}L^{a2}L^{a3}L^{a4}[123,4],\\
\mathcal{A}_2 &= -18(L^{aa})^2 + 18L^{aa}L^{a1}L^{23}L^{a4}[12,34]\\
&\quad + 36L^{aa}L^{a1}L^{23}L^{45}L^{a6}[12,5][34,6] + 36L^{a1}L^{23}L^{a4}L^{a5}L^{a6}[12,5][34,6]\\
&\quad + [345][12,6]L^{a1}\{18L^{aa}L^{23}L^{46}L^{a5} + 9L^{aa}L^{23}L^{45}L^{a6} + 18L^{23}L^{a4}L^{a5}L^{a6}\},\\
\mathcal{A}_3 &= [123][56,4]\{18L^{aa}L^{a1}L^{24}L^{35}L^{a6} + 18L^{aa}L^{a1}L^{a2}L^{35}L^{46}\\
&\quad + 24L^{aa}L^{a1}L^{26}L^{35}L^{a4} + 30L^{a1}L^{a2}L^{35}L^{a4}L^{a6}\}\\
&\quad + [123][456]L^{a1}\{18L^{aa}L^{24}L^{35}L^{a6} + 9L^{aa}L^{a2}L^{35}L^{46} + 18L^{a2}L^{35}L^{a4}L^{a6}\},\\
\mathcal{A}_4 &= [1234]L^{a1}\left\{9L^{aa}L^{23}L^{a4} + 6L^{a2}L^{a3}L^{a4}\right\}.
\end{aligned}$$

$$A_{22} = E_3(\phi_1^a)^2[(\phi_2^a)^2 - 2\phi_2^a E\phi_2^a + (E\phi_2^a)^2] = B_1 - 2B_2 + B_3$$

where

$$\begin{aligned}
B_1 &= C_1 + C_2 + C_3 \\
C_1 &= -12(L^{aa})^2 + [12,34]\{4L^{aa}L^{a1}L^{23}L^{a4} + 8L^{a1}L^{a2}L^{a3}L^{a4}\} \\
&\quad + [12,3][45,6]L^{a1}L^{a4}\Big\{4L^{aa}L^{23}L^{56} + 4L^{aa}L^{26}L^{35} + 8L^{a5}L^{23}L^{a6} + 8L^{a2}L^{35}L^{a6} \\
&\quad + 8L^{a3}L^{26}L^{a5} + 8L^{a3}L^{25}L^{a6} + 8L^{a2}L^{a3}L^{56}\Big\}, \\
C_2 &= [12,3][456]L^{a1}L^{a4}\Big\{4L^{aa}L^{23}L^{56} + 4L^{aa}L^{26}L^{35} + 8L^{23}L^{a5}L^{a6} + 8L^{a2}L^{35}L^{a6} \\
&\quad + 4L^{aa}L^{25}L^{36} + 8L^{a2}L^{a5}L^{36} + 8L^{25}L^{a3}L^{a6} + 8L^{26}L^{a3}L^{a5} + 8L^{26}L^{a3}L^{56}\Big\}. \\
C_3 &= [123][456]L^{a1}L^{a6}\Big\{L^{aa}L^{23}L^{45} + 2L^{23}L^{a4}L^{a5} \\
&\quad + 2L^{aa}L^{24}L^{35} + 8L^{24}L^{a3}L^{a5} + 2L^{a2}L^{a3}L^{45}\Big\}.
\end{aligned}$$

$$\begin{aligned}
B_2 &= 2\mu'_{11}\Big\{[12,3]L^{a1}\Big\{2L^{aa}L^{23} + 4L^{a2}L^{a3}\Big\} + [123]L^{a1}\Big\{L^{aa}L^{23} + 2L^{a2}L^{a3}\Big\}\Big\}, \\
E_1\phi_2^a &= 2\mu'_{11} = L^{a1}L^{23}\{2[12,3] + [123]\}, \\
B_3 &= \Big(L^{a1}L^{23}\{2[12,3] + [123]\}\Big)^2 L^{aa}.
\end{aligned}$$

and the measure of kurtosis

$$\beta_2(\hat{\theta}_a) = 3 + K/N$$

where

$$K = \mu_{43}(\hat{\theta}_a)/\Big\{\mu_{21}(\hat{\theta}_a)\Big\}^2 - 6\mu_{22}(\hat{\theta}_a)/\mu_{21}(\hat{\theta}_a).$$

4 Some illustrative examples

4.1 The two parameter gamma density, location known

Probability function is

$$g(x,a,\rho) = \frac{e^{-x/a}(x/a)^{\rho-1}}{a\Gamma(\rho)}. \quad (x > 0, a > 0, \rho > 0)$$

Bowman and Shenton (1982) gave formulas and numerical examples for the m.l. moments (μ'_1, μ_2, μ_3, μ_4) of $\hat{a}$, the scale estimator, and $\hat{\rho}$, the shape estimator, each for up to the terms N^{-6}. We were very fortunate in finding those results with a completely independent approach, providing a check. The m.l. moments for $\hat{a}$, and $\hat{\rho}$ are given in Table 1 and Table 2 (Bowman and Shenton, 1999).

Table 1. Comparison of two methods of evaluating μ_{21} and μ_{22} $(a = 1)$

		$\hat{a}$		$\hat{\rho}$	
ρ		μ_{21}/a^2	μ_{22}/a^2	μ_{21}/ρ^2	μ_{22}/ρ^2
5.0	m	2.0759	-1.9549	1.8759	20.472
	cm	2.0759	-1.9549	1.875912	20.47226

Table 2. Comparison of two methods of evaluating μ_{32} and μ_{43} $(a = 1)$

		$\hat{a}$		$\hat{\rho}$	
ρ		μ_{32}/a^3	μ_{43}/a^4	μ_{32}/ρ^3	μ_{43}/ρ^4
0.1	m	267.82	10259.00	4.9762×10^{-3}	90.535×10^{-4}
	cm	267.8224	10259.2093	0.0050	90.5352×10^{-4}
1.0	m	15.2599	128.3511	10.4686	280.8125
	cm	15.2599	128.3511	10.4685	280.8125
5.0	m	8.1392	35.403	14.506	440.92
	cm	8.1392	35.4035	14.5062	440.9175
50.0	m	8.0813	24.905	15.841	497.18
	cm	8.0813	24.9048	15.8410	497.1785

(m refers to the Taylor series approach given in Bowman and Shenton (1982, 1988); cm refers to the present approach using the covariance matrix).

The agreement is quite satisfactory. In passing note that moments $E(\hat{\rho} - \rho)/\rho$, $Var(\hat{\rho}/\rho)$, $\mu_3(\hat{\rho}/\rho)$, $\mu_4(\hat{\rho}/\rho)$ to order N^{-1}, $N^{-2}, \cdots, N^{-6}$ are given in Bowman and Shenton (1988, pp.63-68).

4.2 The three parameter gamma density

Probability function is

$$g(x; s, a, \rho) = \frac{e^{-y}(y/a)^{\rho-1}}{a\Gamma(\rho)}. \quad (y = x - s, x > s, a > 0, \rho > 0)$$

In this case the additional parameter is s, referring to location. Details of problems relating to the m.l. estimators $(\hat{s}, \hat{a}, \hat{\rho})$ are given in Bowman and Shenton (2002). At this time we were thinking that sample size could be related not only to the skewness (scale and location free), but also to the variances in the form μ_{22}/μ_{21}, second order term to first order term. In the paper on page 397 the ratio $R(\hat{t}) = \mu_{22}(\hat{t})/\mu_{21}(\hat{t})$, t referring to $\hat{a}/a$, $\hat{\rho}$, and $\hat{s}/a$, is set at a half, in all cases for $\rho > 4$ (moments of the m.l. estimators are unreliable unless $\rho > 4$). For $\rho = 6$, $s = 0$, $a = 1$, sample sizes are 429, 462, and 348 for $\hat{s}/a$, $\hat{\rho}$, and $\hat{a}/a$ respectively.

Table 3 is an extension of Table III in Bowman and Shenton (2002); the extension refers to $\rho = 25, 30$, and 35. If the reader now considers the moments of the estimators for $\rho = 5(5)(35)$ it will be suggested that the skewness $\sqrt{\beta_1(\hat{\rho})}$ tends to a finite limit, so that for large ρ there is not asymptotic normality.

Table 3. Asymptotic Moment Profile ($a = 1$, $s = 0$), gamma density

	ρ	μ'_{11}	μ_{21}	μ_{22}	μ_{22}/μ_{21}	$\sqrt{\beta_{11}}$	K
$\hat{a}/a$	5.0	11.3830	6.5728	2352.	357.91	5.00	226.82
	10.0	7.0058	14.0331	986.	70.25	2.43	412.29
	15.0	6.2751	21.5214	1084.	50.39	1.82	653.97
	20.0	5.9734	29.0158	1255.	43.25	1.52	891.88
	25.0	5.8090	36.5125	1451.	39.74	1.33	1133.36
	30.0	5.7055	44.0103	1654.	37.59	1.19	1373.48
	35.0	5.6344	51.5088	1862.	36.16	1.09	1613.56
$\hat{\rho}$	5.0	-39.3358	388.6581	166841.	429.27	14.38	528.12
	10.0	206.4608	4384.7326	968686.	220.92	29.85	1807.28
	15.0	659.9203	16480.7562	5404615.	327.93	38.92	3101.83
	20.0	1336.3635	41176.7676	18379787.	446.36	46.14	4375.30
	25.0	2237.1429	82972.7744	47099576.	567.65	52.35	5645.25
	30.0	3362.6347	146368.7788	100989279.	689.97	57.88	6910.79
	35.0	4712.9728	235864.7819	191706855.	812.78	62.93	8174.64
$\hat{s}/a$	5.0	31.0033	62.1550	28868.	464.45	-4.97	511.90
	10.0	-32.9230	906.1200	123957.	136.80	-18.99	636.41
	15.0	-165.4481	3650.1120	648180.	177.58	-25.52	1064.05
	20.0	-372.2318	9419.1085	2180849.	231.53	-30.50	1488.77
	25.0	-653.7693	19338.1066	5582919.	288.70	-34.72	1913.89
	30.0	-1010.2007	34532.1054	11984219.	347.05	-38.46	2336.54
	35.0	-1441.5755	56126.1046	22785548.	405.97	-41.85	2758.44

In a simulation study, not designed to fit in with the present study, (Bowman and Shenton, 1988, p.136) give the following moments of the m.l. estimators for a sample of $N = 500$ (Table 4).

Table 4 Variability of Moments Over Simulation Runs

	$\hat{s}$	$\hat{a}$	$\hat{\rho}$
Mean	0.011	1.015	6.083
	(0.044)	(1.018)	(5.996)
S.D	0.583	0.136	1.458
	(0.516)	(0.127)	(1.226)
$\sqrt{\beta_1}$	-0.853	0.098	1.43
	(0.441)	(0.175)	(0.848)
β_2	4.80	3.15	7.70
	(3.72)	(3.46)	(4.468)

(In this table, sampling is from a distribution with $s = 0$, $a = 1$, and $\rho = 6$. Parenthetic entries refer to theoretical values, derived from Table III in Bowman and Shenton (2002). The simulation values are based on 5 cycles, each cycle consisting of 4,000 replications).

The agreement is fairly satisfactory. Discrepancies in the kurtosis may be expected since in practice β_2 may easily be in the range 1 to 20.

4.3 The Weibull distribution, two parameter case

The density is

$$f(x; a, b, c) = \frac{c}{b} y^{c-1} e^{-y^c} \qquad (x > a, y = (x-a)/b, b > 0, c > 0).$$

We assume a is known and consider low order moments of the m.l. estimators $\hat{b}$ of b, $\hat{c}$ of c. As might be expected, polygamma functions such as the Psi function and its derivatives occur. We have (Bowman and Shenton, 2000);

<u>For $\hat{c}$</u>

$$\mu'_{11}(\hat{c}, b) = c[-\zeta(3) + 3\zeta(2)]/[\zeta(2)]^2 = 1.3795c,$$
$$\mu_{21}(\hat{c}, b) = 6c^2/\pi^2 = 0.6079c^2, \qquad \mu_{22}(\hat{c}, b) = \mu_{21} \times 6.3161 = 3.8398c^2,$$
$$\mu_{32}(\hat{c}, b) = 216c^3(\pi^2 - 2\zeta(3))/\pi^6 = 1.6773c^3, \qquad \sqrt{\beta_{11}(\hat{c}, b)} = 3.5386,$$
$$\mu_{42}(\hat{c}, b) = 3\mu_{21}^2, \qquad \mu_{43}(\hat{c}, b) = 18.2992c^4 - c^3, \qquad K(\hat{c}, b) = 11.6174 - 2.7033/c.$$

<u>For $\hat{b}$</u>,

$$\mu'_{11}(\hat{b}) = \frac{b}{c}\left\{\frac{-q\zeta(3) - D^2 + AD - Bq}{\zeta(2)^2} + \frac{D}{c2\zeta(2)}\right\}$$

where

$$D = \zeta(2) + [\psi(2)]^2,$$

$$
\begin{aligned}
q &= 1 + p = 1 + \psi(1), \quad (\psi(1) = -\gamma) \\
A &= 2p^2 + 6p + 9/2, \\
B &= p^3 + 5p^2 + 15p/2 + 7/2 = q(p^2 + 4p + 7/2).
\end{aligned}
$$

An alternative formula to this is

$$
\mu'_{11}(\hat{b}) = \frac{b}{c}\left\{\frac{-\zeta(3) + \zeta(3)\gamma - \zeta(2)^2 + \frac{5}{12}\pi^2 - \frac{1}{3}\pi^2\gamma}{\zeta(2)^2} + \frac{\zeta(2) + \psi(2)^2}{c2\zeta(2)}\right\}.
$$

$$
\begin{aligned}
\mu'_{11}(\hat{b}, c) &= \frac{b}{c}\left\{-0.3698 + \frac{0.5543}{c}\right\}, \quad \mu_{21}(\hat{b}, c) = \left\{1 + \frac{[\psi(2)]^2}{\psi_1(1)}\right\}\left(\frac{b}{c}\right)^2 = 1.1087\left(\frac{b}{c}\right)^2, \\
\mu_{22}(\hat{b}, c) &= \frac{b^2}{c^2}\left[0.3624 - \frac{1.9881}{c} + \frac{1.8429}{c^2}\right], \\
\mu_{32}(\hat{b}, c) &= \frac{b^3(cG_1 + G_2)}{c^4\pi^6} = \frac{b^3}{c^3}\left[-1.1686 + \frac{3.6873}{c}\right], \quad \sqrt{\beta_{11}(\hat{b}, c)} = -1.0011 + \frac{3.1587}{c}, \\
\mu_{42}(\hat{b}, c) &= 3\mu_{21}^2, \quad \mu_{43}(\hat{b}, c) = \frac{b^4}{c^4}\left[-24.3889 - \frac{34.3604}{c} + \frac{34.0696}{c^2}\right]; \\
K(\hat{b}, c) &= -21.8038 - \frac{17.1924}{c} + \frac{17.7397}{c^2}
\end{aligned}
$$

where

$$
\begin{aligned}
G_1 &= 108\pi^2 - 1296\gamma^2\zeta(3) + 1296\gamma\zeta(3) + 432\gamma^3\zeta(3) - 216\pi^2\gamma + 108\pi^2\gamma^2 \\
&\quad - 18\pi^4\gamma^2 + 36\pi^4\gamma - \pi^6 - 432\zeta(3) - 18\pi^4, \\
G_2 &= 36\pi^4 + 3\pi^6 - 72\pi^4\gamma + 36\pi^4\gamma^2 - 432\pi^2\gamma^3 + 108\pi^2\gamma^4 + 648\pi^2\gamma^2 - 432\pi^2\gamma + 108\pi^2.
\end{aligned}
$$

(ζ is the zeta function, γ Euler's constant)

To check on these results, sample sizes of 50, 100, and 200 were considered using 50,000 cycles for $n = 50$, 25,000 cycles in the cases for $c = 1, 2, 3, 4, 5$.

The comparisons with theory are given in Table 5, for $n = 200$, $a = 0$, $b = 1$.

Table 5. Comparison of theoretical and simulation moments for $\hat{b}$ and $\hat{c}$

		$\hat{b}$				$\hat{c}$			
c		μ_1	σ	$\sqrt{\beta_1}$	β_2	μ_1	σ	$\sqrt{\beta_1}$	β_2
2	T	0.99977	0.0372	0.0409	2.8702	2.0138	0.1120	0.2502	3.0516
	S	0.99990	0.0372	0.0591	2.9863	2.0156	0.1121	0.2582	3.0836
5	T	0.99974	0.0149	-0.0261	2.8773	5.0345	0.2800	0.2502	3.0554
	S	0.99979	0.0149	-0.0075	2.9886	5.0390	0.2802	0.2582	3.0836

(T=Theory, S=Simulation: For the theoretical entries parameters $\hat{b}$ and $\hat{c}$, $\mu_1'(\hat{t}) = t + \mu_{11}'(\hat{t})/N$, and similarly for $\sigma(\hat{t}) = \sqrt{\mu_{21}/N + \mu_{22}/N^2}$).

Considering the skewness and kurtosis in both cases, the distributions are nearly normal in the sense that $\sqrt{\beta_1}$ is approximately zero and β_2 is approximately 3.

NAG library random number generator was used in the simulation study. Check on the basic random number generator gave about 3-4 digit accuracy for the first 4 moments of the Weibull distribution.

The agreement is good for $\hat{c}$ moments, satisfactory altogether.

4.4 The Weibull distribution and 3 parameters estimated by m.l.

In this case the asymptotic moments are complicated and involve considerable effort in algebraic manipulation. Here are examples:

$$
\begin{aligned}
Var_1(\hat{a}) &= b^2\psi_1(1)/\Delta(c),\\
Var_1(\hat{b}) &= b^2(CG - J^2)/\Delta(c),\\
Var_1(\hat{c}) &= c^4[C - \Gamma^2(2 - 1/c)]/\Delta(c),\\
Cov_1(\hat{a}, \hat{b}) &= -b^2[G\Gamma(2 - 1/c) + J\psi(2)]/\Delta(c),\\
Cov_1(\hat{a}, \hat{c}) &= -bc^2[\psi(2)\Gamma(2 - 1/c) + J]/\Delta(c),\\
Cov_1(\hat{b}, \hat{c}) &= bc^2[\psi(2)C + \Gamma(2 - 1/c)J]/\Delta(c),
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta(c) &= c^2[\psi_1(1)C - \Gamma^2\left(2 - \frac{1}{c}\right)G - J^2 - 2\Gamma\left(2 - \frac{1}{c}\right)\psi(2)J],\\
C(c) &= \Gamma\left(1 - \frac{2}{c}\right)\frac{(c-1)^2}{c^2},\\
J(c) &= \Gamma\left(2 - \frac{1}{c}\right)\left[\frac{1}{c-1} - \psi\left(2 - \frac{1}{c}\right)\right],\\
G &= \psi_1(1) + [\psi(2)]^2.
\end{aligned}
$$

To highlight the advantages of the Maple symbolic approacha we give here further examples of the algebraic background.

$$
\Delta(c) = c^2\left\{\psi_1(1)\left[C(c) - \Gamma^2\left(2 - \frac{1}{c}\right)\right] - H^2(c)\right\}
$$

where

$$
H(c) = \Gamma\left(2 - \frac{t}{2}\right)\psi_2(2) + J(c), J(c) = \Gamma\left(2 - \frac{1}{c}\right)\left[\frac{1}{c-1} - \psi\left(2 - \frac{1}{c}\right)\right] \quad (c > 2).
$$

Defining $t = 2/c$ $(0 < t < 1)$ we have

$$H(c) = \Gamma\left(2 - \frac{t}{2}\right)\left[\psi(1) - \psi\left(1 - \frac{t}{2}\right)\right] = -\Gamma\left(2 - \frac{t}{2}\right)\sum_{s=1}^{\infty}\left(-\frac{t}{2}\right)^s \frac{\psi_s(1)}{s!}.$$

After some simplification we find

$$Var_1(\hat{c}) \sim \frac{(c-1)^2(A_0 + A_1 t + A_2 t^2 + A_3 t^3 \cdots)}{B_0 + B_1 t + B_2 t^2 + B_3 t^3 \cdots},$$

where

$$\begin{aligned}
A_0 &= A_1 = 0, \qquad A_2 = \pi^2/24, \qquad A_3 = \pi^2\gamma/24 + \zeta(3)/4,\\
A_4 &= \zeta(3)\gamma/4 + \pi^2\gamma^2/48 + 29\pi^4/5760,\\
A_5 &= 29\pi^4\gamma/5760 + 7\zeta(3)\pi^2/288 + \zeta(3)\gamma^2/8 + \pi^2\gamma^3/144 + 3\zeta(5)/16,\\
A_6 &= 3\zeta(5)\gamma/16 + 457\pi^6/967680 + 5\zeta(3)^2/96 + 7\zeta(3)\pi^2\gamma/288 + 29\pi^4\gamma^2/11520\\
&\quad + \zeta(3)\gamma^3/24 + \pi^2\gamma^4/576,\\
B_0 &= B_1 = B_2 = B_3 = 0, \qquad B_4 = 11\pi^6/34560 - \zeta(3)^2/16,\\
B_5 &= -\gamma\zeta(3)^2/16 + 11\pi^6\gamma/34560 + \pi^4\zeta(3)/960 + \pi^2\zeta(5)/48,\\
B_6 &= -\gamma^2\zeta(3)^2/32 + \pi^2\zeta(5)\gamma/48 + \pi^4\zeta(3)\gamma/960 + 193\pi^8/3628800 + \pi^2\zeta(3)^2/384\\
&\quad + 11\pi^6\gamma^2/69120 - \zeta(3)\zeta(5)/32.
\end{aligned}$$

Numerically

$$\frac{Var_1(\hat{c})}{c^4} \sim 0.488214 - \frac{0.929894}{c} + \frac{0.790780}{c^2} - \frac{0.566106}{c^3}.$$

These expressions were set up partly by using the Maple system. For large c

$$Var_1(\hat{c}) \sim \lambda(c-1)^2c^2 \qquad (c \to \infty)$$

where $\lambda = 4\zeta(2)/[11\zeta(4)\zeta(2) - 4\zeta^2(3)] = 0.47665188$. Similarly

$$Var_1(\hat{a}) \sim \lambda b^2 c^2, \qquad Var_1(\hat{b}) \sim \lambda b^2(c-1)^2. \qquad (c \to \infty)$$

This excerpt is only a part of the computer input required for the implementation of the computer program (Fortran version program). Note that various aspects of the Weibull distribution have been considered by Dubay (1965). In the next section we describe the Maple factor.

5 The advancement of symbolic computer languages

These languages appeared on the scientific horizon about four decades ago. Formac was introduced by IBM but it had a short life in favor of main frame advances. A small group of researchers, sponsored by Dr. J.L. Carman (Head of the Computer Center, University of Georgia) and led by Dr. Juris Reinfelds, connected an electric typewriter to an IBM1620; difficulties arose because climate control was neglected. Reinfelds was mainly interested in algebraic processes, and Shenton and Hutcheson suggested a statistical problem, that of converting crude moments into cumulants. Examples are given below:

CRUDE MOMENTS INTO CUMULANTS

ORDER ONE

$\mu_1' = +\kappa_1.$

ORDER TWO

$\mu_2' = +\kappa_2 + \kappa_1^2,$

$\mu_{11}' = +\kappa_{11} + \kappa_{10}\kappa_{01}.$

ORDER THREE

$\mu_3' = +\kappa_3 + 3\kappa_1^2\kappa_1 + \kappa_1^3,$

$\mu_{21}' = +\kappa_{21} + 2\kappa_{11}\kappa_{10} + \kappa_{20}\kappa_{01} + \kappa_{10}^2\kappa_{01},$

$\mu_{111}' = +\kappa_{111} + \kappa_{011}\kappa_{100} + \kappa_{110}\kappa_{001} + \kappa_{101}\kappa_{010} + \kappa_{100}\kappa_{001}\kappa_{010}.$

ORDER FOUR

$\mu_4' = +\kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4,$

$\mu_{31}' = +\kappa_{31} + 3\kappa_{21}\kappa_{10} + \kappa_{30}\kappa_{01} + 3\kappa_{20}\kappa_{11} + 3\kappa_{11}\kappa_{10}^2 + 3\kappa_{20}\kappa_{10}\kappa_{01} + \kappa_{10}^3\kappa_{01},$

$\mu_{22}' = +\kappa_{22} + 2\kappa_{12}\kappa_{10} + 2\kappa_{21}\kappa_{01} + 2\kappa_{11}^2 + \kappa_{20}\kappa_{02} + \kappa_{02}\kappa_{10}^2 + 4\kappa_{11}\kappa_{10}\kappa_{01} + \kappa_{20}\kappa_{01}^2 + \kappa_{10}^2\kappa_{01}^2,$

$\mu_{211}' = +\kappa_{211} + 2\kappa_{111}\kappa_{100} + \kappa_{201}\kappa_{010} + \kappa_{210}\kappa_{001} + +2\kappa_{110}\kappa_{101} + \kappa_{200}\kappa_{011}$

$+\kappa_{011}\kappa_{100}^2 + 2\kappa_{101}\kappa_{100}\kappa_{010} + 2\kappa_{110}\kappa_{100}\kappa_{001} + \kappa_{200}\kappa_{010}\kappa_{001} + \kappa_{100}^2\kappa_{010}\kappa_{001},$

$\mu_{1111}' = +\kappa_{1111} + \kappa_{0111}\kappa_{1000} + \kappa_{1110}\kappa_{0001} + \kappa_{1011}\kappa_{0100} + \kappa_{1101}\kappa_{0010} + \kappa_{0101}\kappa_{1010} + \kappa_{1100}\kappa_{0011} +$

$+\kappa_{1001}\kappa_{0110} + \kappa_{0110}\kappa_{1000}\kappa_{0001} + \kappa_{0011}\kappa_{1000}\kappa_{0100} + \kappa_{1010}\kappa_{0001}\kappa_{0100} + \kappa_{0101}\kappa_{1000}\kappa_{0010} +$

$+\kappa_{1100}\kappa_{0001}\kappa_{0010} + \kappa_{1001}\kappa_{0100}\kappa_{0010} + \kappa_{1000}\kappa_{0001}\kappa_{0100}\kappa_{0010}.$

More details are given in the report "Tables of Crude Moments Expressed in terms of Cumulants" by Kratky, Reinfelds, Hutcheson, and Shenton, Computer Center, University of Georgia; Computer Center Report 1972(1). On the first page we find the quotation "The Marquis gazed a Moment, and nothing did he say", William Edmondstone Aytoun (1813-1865).

Other symbolic manipulative programs were appearing in the 1970's, including Reduce, Mathematika, Maple, and Macysma. Shenton used the Mathematika package at the University of Georgia in 1980 to 1990, especially the program including the conversion of series in powers of n^{-1} into continued fractions. Accuracy was not a problem since output was in integer arithmetic.

To return to our present subject, Bowman was able to use Maple to cope with the skewness of m.l. estimators implementing the formula

$$\mu_{32}(\hat{\theta}_a) = L^{a\alpha} L^{a\beta} L^{a\gamma} \left\{ [\alpha, \beta, \gamma] + 3[\alpha\beta\gamma] + 6[\alpha\beta, \gamma] \right\}.$$

In the end, Bowman was able to reduce the problem of low order moments (μ'_{11}, μ'_{12}, μ_{21}, μ_{22}, μ_{32}, μ_{43}) of a m.l. estimator to depend entirely on the probability function involved. The great advantage of the approach is seen when we consider the Weibull density. In this case 5th order derivatives appear (especially for μ'_{12}); for examples for square bracket terms such as

$$[\theta_1\theta_2\theta_3\theta_4\theta_5] = E\left(\frac{\partial^5 \log P}{\partial\theta_1 \partial\theta_2 \partial\theta_3 \partial\theta_4 \partial\theta_5} \right),$$

and

$$[\theta_1\theta_1\theta_1\theta_1\theta_1] = E\left(\frac{\partial \log P}{\partial\theta_1} \right)^5.$$

Terms like these are used in expressions for asymptotic moments. We now give the Maple program for the six basic asymptotic moments.

6 The Maple program

6.1 New Maple program

We introduced the Maple program in Bowman and Shenton (2005) to compute asymptotic variance and skewness of the m.l. estimators. We have extended the program to compute N^{-1} and N^{-2} biases, N^{-2} variance, and N^{-3} fourth central moment. The Maple program presented here is more general than the previous versions and could be easily converted to any distribution's ('pf') with parameter number ('w'). Further, the user must decide on the value of ('lim') which depends on the range of the distribution, and for taking the expectation, use ('int') for integration of a continuous distribution and ('sum') for summation of a discrete distribution. The Maple program of the two parameter gamma distribution is presented in this section.

For example, to change from the two parameter gamma distribution to the three parameter gamma distribution, we carry out following;
(i) change 'w :=2;' to 'w :=3;'
(ii) change 'pf' to 3 parameter gamma distribution;
(iii) change 't :=[t1,t2]' to 't :=[t1,t2,t3]';
(iv) add the third parameter value, to include this change in all the 'subs' statement.

To change from the two parameter gamma distribution to the two parameter Weibull distribution, change 'pf' accordingly and supply values of the two Weibull parameter values.

For the three parameter Weibull distribution follow the example of the two parameter gamma distribution to the three parameter gamma distribution. Bowman and Shenton (2000) computed kurtosis of the three parameter Weibull distribution by writing Fortran program which consisted of 3500 or so lines and it was highly individualized. The Maple program is consisted of about 300 lines and is generalized, it is a great advancement.

Users could make further improvement by using only bias section of the program and correct a bias and run the rest of program using unbiased m.l. estimators.

6.2 Computer program

```
 #Find kurtosis, skewness, u2 and u1 of ml estimators
 #number of parameters = w
 #pf is a probability function or density (two parameter gamma density)
 #t is a vector of parameters t1,t2,...,tw
 #lim is the upper range of the distribution, constant or infinity
 #Output results are U11, U12, L (covariance matrix), U22, U32, rb1,
 #U43, and K.

 with(linalg);
 w :=2;  w2 :=w^2;  w3 :=w^3;  w4 :=w^4;  w5 :=w^5;  w6 :=w^6;
 w7 :=w^7; w8 :=w^8; w9 :=w^9; t :=[t1,t2]; lim :=infinity; Llim :=0;
 pf :=exp(-x/t1)*x^(t2-1)/(t1^t2*GAMMA(t2)); LL :=log(pf);

#Take derivatives up to 5th order of log of pf
 for i1 from 1 to w do
    D1[i1] :=diff(LL,t[i1]);
 for i2 from 1 to w do
    D2[i1,i2] :=diff(D1[i1],t[i2]);
    D11[i1,i2] :=D1[i1]*D1[i2];
 for i3 from 1 to w do
    D3[i1,i2,i3] :=diff(D2[i1,i2],t[i3]);
    D21[i1,i2,i3] :=D2[i1,i2]*D1[i3];
    D111[i1,i2,i3] :=D11[i1,i2]*D1[i3];
 for i4 from 1 to w do
    D4[i1,i2,i3,i4] := diff(D3[i1,i2,i3],t[i4]);
    D31[i1,i2,i3,i4] :=D3[i1,i2,i3]*D1[i4];
```

```
   D22[i1,i2,i3,i4] :=D2[i1,i2]*D2[i3,i4];
   D211[i1,i2,i3,i4] :=D21[i1,i2,i3]*D1[i4];
   D1111[i1,i2,i3,i4] :=D111[i1,i2,i3]*D1[i4];
 for i5 from 1 to w do
   D5[i1,i2,i3,i4,i5] := diff(D4[i1,i2,i3,i4],t[i5]);
   D41[i1,i2,i3,i4,i5] :=D4[i1,i2,i3,i4]*D1[i5];
   D32[i1,i2,i3,i4,i5] :=D3[i1,i2,i3]*D2[i4,i5];
   D311[i1,i2,i3,i4,i5] :=D31[i1,i2,i3,i4]*D1[i5];
   D221[i1,i2,i3,i4,i5] :=D22[i1,i2,i3,i4]*D1[i5]; od;od;od;od;od;

#Digits :=15; (specify the number of digits to use for accuracy)
 Digits :=15;

#Input ml estimator of parameters and take expectation
 a :=1; r :=2;   f :=subs(t1=a,t2=r,pf);
 for i1 from 1 to w do
   d1[i1] :=subs(t1=a,t2=r,D1[i1]);
 for i2 from 1 to w do
   d2[i1,i2] :=subs(t1=a,t2=r,D2[i1,i2]);
   d11[i1,i2] :=subs(t1=a,t2=r,D11[i1,i2]);
   f2[i1,i2] :=evalf(int(f*d2[i1,i2],x=Llim..lim));
   f11[i1,i2] :=evalf(int(f*d11[i1,i2],x=Llim..lim));
 for i3 from 1 to w do
   d3[i1,i2,i3] :=subs(t1=a,t2=r,D3[i1,i2,i3]);
   d21[i1,i2,i3] :=subs(t1=a,t2=r,D21[i1,i2,i3]);
   d111[i1,i2,i3] :=subs(t1=a,t2=r,D111[i1,i2,i3]);
   f3[i1,i2,i3] :=evalf(int(f*d3[i1,i2,i3],x=Llim..lim));
   f21[i1,i2,i3] :=evalf(int(f*d21[i1,i2,i3],x=Llim..lim));
   f111[i1,i2,i3] :=evalf(int(f*d111[i1,i2,i3],x=Llim..lim));
 for i4 from 1 to w do
   d4[i1,i2,i3,i4] :=subs(t1=a,t2=r,D4[i1,i2,i3,i4]);
   d31[i1,i2,i3,i4] :=subs(t1=a,t2=r,D31[i1,i2,i3,i4]);
   d22[i1,i2,i3,i4] :=subs(t1=a,t2=r,D22[i1,i2,i3,i4]);
   d211[i1,i2,i3,i4] :=subs(t1=a,t2=r,D211[i1,i2,i3,i4]);
   d1111[i1,i2,i3,i4] :=subs(t1=a,t2=r,D1111[i1,i2,i3,i4]);
   f4[i1,i2,i3,i4] :=evalf(int(f*d4[i1,i2,i3,i4],x=Llim..lim));
   f31[i1,i2,i3,i4] :=evalf(int(f*d31[i1,i2,i3,i4],x=Llim..lim));
   f22[i1,i2,i3,i4] :=evalf(int(f*d22[i1,i2,i3,i4],x=Llim..lim));
   f211[i1,i2,i3,i4] :=evalf(int(f*d211[i1,i2,i3,i4],x=Llim..lim));
   f1111[i1,i2,i3,i4] :=evalf(int(f*d1111[i1,i2,i3,i4],x=Llim..lim));
 for i5 from 1 to w do
```

```
    d5[i1,i2,i3,i4,i5] :=subs(t1=a,t2=r,D5[i1,i2,i3,i4,i5]);
    d41[i1,i2,i3,i4,i5] :=subs(t1=a,t2=r,D41[i1,i2,i3,i4,i5]);
    d32[i1,i2,i3,i4,i5] :=subs(t1=a,t2=r,D32[i1,i2,i3,i4,i5]);
    d311[i1,i2,i3,i4,i5] :=subs(t1=a,t2=r,D311[i1,i2,i3,i4,i5]);
    d221[i1,i2,i3,i4,i5] :=subs(t1=a,t2=r,D221[i1,i2,i3,i4,i5]);
    f5[i1,i2,i3,i4,i5] :=evalf(int(f*d5[i1,i2,i3,i4,i5],x=Llim..lim));
    f41[i1,i2,i3,i4,i5] :=evalf(int(f*d41[i1,i2,i3,i4,i5],x=Llim..lim));
    f32[i1,i2,i3,i4,i5] :=evalf(int(f*d32[i1,i2,i3,i4,i5],x=Llim..lim));
    f311[i1,i2,i3,i4,i5] :=evalf(int(f*d311[i1,i2,i3,i4,i5],x=Llim..lim));
    f221[i1,i2,i3,i4,i5] :=evalf(int(f*d221[i1,i2,i3,i4,i5],x=Llim..lim));
    od;od;od;od;od;

#Compute covariance matrix
  H :=Matrix(w,w,f11); L :=inverse(H);

#Computation of U22
   ii :=0;
 for i1 from 1 to w do   for i2 from 1 to w do  for i3 from 1 to w do
 for i4 from 1 to w do
   ii :=ii+1;
   AA[ii] :=f211[i1,i4,i2,i3]+f211[i2,i4,i1,i3]+f4[i1,i2,i3,i4]
        +3*f22[i1,i4,i2,i3]+2*f31[i1,i2,i3,i4]+1/2*f31[i2,i3,i4,i1]
        +1/2*f31[i1,i3,i4,i2];
 for jj from 1 to w do
   L31[jj,ii] :=L[jj,i1]*L[jj,i2]*L[i3,i4];
   L42[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i3]*L[jj,i4];
   L43[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i3]*L[jj,i4];
   L44[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i4]*L[jj,i3];
   L45[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i3,i4]*L[jj,i2];
   od;od;od;od;od;
   ii :=0;
 for i1 from 1 to w do  for i2 from 1 to w do  for i3 from 1 to w do
 for i4 from 1 to w do  for i5 from 1 to w do  for i6 from 1 to w do
   ii :=ii+1;
   CC[ii] :=f3[i1,i3,i6]*f111[i2,i4,i5]/2+f3[i2,i3,i6]*f111[i1,i4,i5]/2
          +f3[i1,i2,i3]*f3[i4,i5,i6]+5/2*f3[i1,i3,i5]*f3[i2,i4,i6]
   +f21[i4,i5,i1]*f3[i2,i3,i6]+f21[i4,i5,i2]*f3[i1,i3,i6]
   +2*f3[i1,i2,i6]*f21[i3,i5,i4]+3*f3[i2,i3,i5]*f21[i1,i6,i4]
   +3*f3[i1,i3,i5]*f21[i2,i6,i4]+f3[i3,i4,i5]*f21[i2,i6,i1]/2
   +f3[i3,i4,i5]*f21[i1,i6,i2]/2+f21[i1,i5,i4]*f21[i3,i6,i2]
   +f21[i2,i5,i4]*f21[i3,i6,i1]+f21[i1,i5,i2]*f21[i3,i6,i4]
```

```
   +f21[i2,i5,i1]*f21[i3,i6,i4]+f21[i1,i5,i3]*f21[i2,i4,i6];
 for jj from 1 to w do
   L41[jj,ii] :=L[jj,i1]*L[jj,i2]*L[i3,i4]*L[i5,i6];
   od;od;od;od;od;od;od;
 for jj from 1 to w do
  U221[jj] :=add(L31[jj,i]*AA[i],i=1..w4);
  U222[jj] :=add(L41[jj,i]*CC[i],i=1..w6);
  U22[jj] :=-L[jj,jj]+U221[jj]+U222[jj]; od;

#Computation of U32
   ii :=0;
 for i1 from 1 to w do  for i2 from 1 to w do for i3 from 1 to w do
   ii :=ii+1;
   A[ii] :=f111[i1,i2,i3]+3*f3[i1,i2,i3]+6*f21[i1,i2,i3];
   A3[ii] :=f3[i1,i2,i3];
   A21[ii] :=f21[i1,i2,i3];
   A213[ii] :=2*f21[i1,i2,i3]+f3[i1,i2,i3];
 for jj from 1 to w do
   L32[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i3];
   L33[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i3]; od;od;od;od;
 for jj from 1 to w do
   U32[jj] :=add(L32[jj,i]*A[i],i=1..w3);
   sig[jj] :=sqrt(L[jj,jj]);
   rb1[jj] :=U32[jj]/L[jj,jj]^(3/2);  od;

#Computation of U43, see equation (12)
#Computation of A40 and part of A32
   ii :=0;
 for i1 from 1 to w do  for i2 from 1 to w do
 for i3 from 1 to w do  for i4 from 1 to w do
   ii :=ii+1;
   A4[ii] :=f4[i1,i2,i3,i4];
   A31[ii] :=f31[i1,i2,i3,i4];
   A22[ii] :=f22[i1,i2,i3,i4];
   A211[ii] :=f211[i1,i2,i3,i4];
   A1111[ii] :=f1111[i1,i2,i3,i4];od;od;od;od;
 for jj from 1 to w do
   C40[jj] :=add(A1111[i]*L42[jj,i],i=1..w4)-3*L[jj,jj]^2;
   C211[jj] :=6*(add(A211[i]*L42[jj,i],i=1..w4)
             +add(A211[i]*L43[jj,i],i=1..w4))+12*L[jj,jj]^2; od;
#Computation of A32
```

```
  ii :=0;
for i1 from 1 to w do  for i2 from 1 to w do for i3 from 1 to w do
for i4 from 1 to w do for i5 from 1 to w do  for i6 from 1 to w do
  ii :=ii+1;
  A322[ii] :=(2*f21[i1,i2,i3]+f3[i1,i2,i3])*f111[i4,i5,i6];
  A321[ii] :=f21[i1,i2,i3]*f111[i4,i5,i6]+f21[i1,i2,i4]*f111[i3,i5,i6]+
             f21[i1,i2,i5]*f111[i3,i4,i6]+f21[i1,i2,i6]*f111[i3,i4,i5];
  A61[ii] :=f3[i1,i2,i3]*f111[i4,i5,i6];
  A62[ii] :=f21[i1,i2,i3]*f111[i4,i5,i6];
  A63[ii] :=f21[i1,i2,i5]*f21[i3,i4,i6];
  A64[ii] :=f21[i1,i2,i6]*f21[i3,i4,i5];
  A65[ii] :=f21[i1,i2,i3]*f21[i4,i5,i6];
  A66[ii] :=f3[i3,i4,i5]*f21[i1,i2,i6];
  A67[ii] :=f3[i1,i2,i3]*f21[i5,i6,i4];
  A68[ii] :=f3[i4,i5,i6]*f21[i1,i2,i3];
  A69[ii] :=f3[i1,i2,i3]*f3[i4,i5,i6];
  A610[ii] :=2*A62[ii]+A61[ii];
for jj from 1 to w do
  L51[jj,ii] :=L[jj,i1]*L[i2,i3]*L[jj,i4]*L[jj,i5]*L[jj,i6];
  L52[jj,ii] :=L[jj,i1]*L[jj,i2]*L[i3,i4]*L[jj,i5]*L[jj,i6];
  L53[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i4]*L[i3,i5]*L[jj,i6];
  L54[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i3]*L[i4,i5]*L[jj,i6];
  L55[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i3]*L[i4,i6]*L[jj,i5];
  L56[jj,ii] :=L[jj,jj]*L[jj,i1]*L[jj,i2]*L[i3,i5]*L[i4,i6];
  L57[jj,ii] :=L[jj,jj]*L[jj,i1]*L[i2,i6]*L[i3,i5]*L[jj,i4];
  L58[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i4]*L[i3,i5]*L[jj,i6];
  L59[jj,ii] :=L[jj,i1]*L[i2,i4]*L[jj,i3]*L[jj,i5]*L[jj,i6];
  L510[jj,ii] :=L[jj,jj]*L[jj,i1]*L[jj,i4]*L[i2,i3]*L[i5,i6];
  L511[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i4]*L[i3,i5]*L[jj,i6];
  L512[jj,ii] :=L[jj,i1]*L[jj,i3]*L[jj,i4]*L[i2,i6]*L[jj,i5];
  L513[jj,ii] :=L[jj,i1]*L[jj,i3]*L[jj,i4]*L[i2,i5]*L[jj,i6];
  L514[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i3]*L[jj,i4]*L[i5,i6];
  L515[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i3]*L[i4,i5]*L[jj,i6];
  L516[jj,ii] :=L[jj,jj]*L[jj,i1]*L[jj,i4]*L[i2,i5]*L[i3,i6];
  L517[jj,ii] :=L[jj,i1]*L[jj,i2]*L[jj,i4]*L[jj,i5]*L[i3,i6];
  L518[jj,ii] :=L[jj,i1]*L[jj,i3]*L[jj,i4]*L[i2,i5]*L[jj,i6];
  L519[jj,ii] :=L[jj,i1]*L[jj,i3]*L[jj,i4]*L[jj,i5]*L[i2,i6];
  od;od;od;od;od;od;od;
for jj from 1 to w do
   C320[jj] :=-(add(A322[i]*L51[jj,i],i=1..w6));
   C321[jj] :=2*(add(A321[i]*L51[jj,i],i=1..w6))
```

```
              +add(A61[i]*L51[jj,i],i=1..w6)
             +6*(add(A61[i]*L52[jj,i],i=1..w6))
      +3*(add(A61[i]*L53[jj,i],i=1..w6));
    C32[jj] :=C211[jj]+C320[jj]+C321[jj]; od;
#Computation of A33
  for jj from 1 to w do
   CA1[jj] :=18*(add(A31[i]*L44[jj,i],i=1..w4))
             +9*(add(A31[i]*L43[jj,i],i=1..w4))
            +18*(add(A31[i]*L42[jj,i],i=1..w4));
   CA21[jj] :=-18*L[jj,jj]^2+18*(add(A22[i]*L43[jj,i],i=1..w4));
   CA22[jj] :=36*(add(A63[i]*L54[jj,i],i=1..w6)
             +add(A63[i]*L51[jj,i],i=1..w6));
   CA23[jj] :=18*(add(A66[i]*L55[jj,i],i=1..w6))
              +9*(add(A66[i]*L54[jj,i],i=1..w6))
             +18*(add(A66[i]*L51[jj,i],i=1..w6));
   CA31[jj] :=18*(add(A67[i]*L53[jj,i],i=1..w6))
             +18*(add(A67[i]*L56[jj,i],i=1..w6))
             +24*(add(A67[i]*L57[jj,i],i=1..w6))
             +30*(add(A67[i]*L58[jj,i],i=1..w6));
   CA32[jj] :=18*(add(A69[i]*L53[jj,i],i=1..w6))
             +9*(add(A69[i]*L56[jj,i],i=1..w6))
             +18*(add(A69[i]*L58[jj,i],i=1..w6));
   CA4[jj] :=3*(add(A4[i]*L43[jj,i],i=1..w4))
             +6*(add(A4[i]*L42[jj,i],i=1..w4))
             +3*(add(A4[i]*L45[jj,i],i=1..w4))
             +3*(add(A4[i]*L44[jj,i],i=1..w4));
   C33[jj] :=CA1[jj]+CA21[jj]+CA22[jj]+CA23[jj]+CA31[jj]+CA32[jj]
            +CA4[jj]; od;
#Computation of A22
  for jj from 1 to w do for i from 1 to w4 do
   c0[jj,i] :=4*L43[jj,i]+8*L42[jj,i]; od;od;
  for jj from 1 to w do for i from 1 to w6 do
   c1[jj,i] :=4*L510[jj,i]+4*L57[jj,i]+8*L51[jj,i]+8*L58[jj,i]
                +8*L512[jj,i]+8*L514[jj,i]+8*L513[jj,i];
   c2[jj,i] :=4*L510[jj,i]+4*L57[jj,i]+8*L51[jj,i]+8*L58[jj,i]
             +4*L516[jj,i]+8*L517[jj,i]+8*L518[jj,i]+8*L519[jj,i]
                +8*L514[jj,i];
   c3[jj,i] :=L54[jj,i]+2*L51[jj,i]+2*L53[jj,i]+8*L59[jj,i]
             +2*L515[jj,i];od;od;
  for jj from 1 to w do
   C0[jj] :=add(A22[i]*c0[jj,i],i=1..w4)-12*L[jj,jj]^2;
```

```
 C1[jj] := add(A65[i]*c1[jj,i],i=1..w6);
 C2[jj] := add(A68[i]*c2[jj,i],i=1..w6);
 C3[jj] := add(A69[i]*c3[jj,i],i=1..w6); od;
 ii :=0;
for i1 from 1 to w do for i2 from 1 to w do for i3 from 1 to w do
  ii :=ii+1;
for jj from 1 to w do
  L2[jj,ii] :=L[jj,i1]*L[i2,i3]; od;od;od;od;
for jj from 1 to w do for i from 1 to w3 do
  b1[jj,i] :=2*L33[jj,i]+4*L32[jj,i];
  b2[jj,i] :=L33[jj,i]+2*L32[jj,i]; od;od;
for jj from 1 to w do
  u11[jj] :=add(L2[jj,i]*A213[i],i=1..w3);
  B2[jj] :=u11[jj]*(add(A21[i]*b1[jj,i],i=1..w3)
                 +add(A3[i]*b2[jj,i],i=1..w3));
  B3[jj] :=u11[jj]^2*L[jj,jj];
  C22[jj] :=C0[jj]+C1[jj]+C2[jj]+C3[jj]-2*B2[jj]+B3[jj];
  U43[jj] :=C40[jj]+2*C32[jj]+2/3*C33[jj]+3/2*C22[jj];
  K[jj] :=U43[jj]/L[jj,jj]^2-6*U22[jj]/L[jj,jj];od;

#Computation of U11 and U12
  ii :=0;
for i1 from 1 to w do  for i2 from 1 to w do  for i3 from 1 to w do
for i4 from 1 to w do  for i5 from 1 to w do
  ii :=ii+1;
  BB3[ii] :=f5[i1,i2,i3,i4,i5]+4*f32[i2,i4,i5,i1,i3]
           +8*f32[i1,i2,i3,i4,i5]+4*f41[i1,i2,i3,i5,i4]
   +4*f311[i1,i2,i3,i4,i5]+8*f221[i1,i2,i3,i4,i5];
for jj from 1 to w do
  L34[jj,ii] :=L[jj,i1]*L[i2,i4]*L[i3,i5]; od;od;od;od;od;od;
  ii :=0;
for i1 from 1 to w do  for i2 from 1 to w do  for i3 from 1 to w do
for i4 from 1 to w do  for i5 from 1 to w do  for i6 from 1 to w do
for i7 from 1 to w do
  ii :=ii+1;
  BB4[ii] :=(2*f4[i1,i4,i5,i6]*f21[i2,i7,i3]+2*f4[i2,i4,i5,i6]
           *f21[i1,i3,i7]+4*f4[i1,i2,i4,i5]*f21[i6,i7,i3])
           +(f4[i1,i4,i5,i6]*f3[i2,i3,i7]+2*f4[i1,i2,i4,i5]
           *f3[i3,i6,i7]+2*f4[i2,i4,i5,i7]*f3[i1,i3,i6])
           +(2*f22[i2,i7,i3,i6]*f3[i1,i4,i5]+4*f22[i4,i7,i2,i5]
           *f3[i1,i3,i6]+4*f22[i1,i3,i6,i7]*f3[i2,i4,i5]
```

```
            +2*f22[i1,i7,i4,i5]*f3[i2,i3,i6])+(4*f31[i1,i5,i6,i7]
            *f21[i2,i4,i3]+4*f31[i1,i5,i6,i3]*f21[i2,i4,i7]
            +4*f31[i2,i4,i5,i7]*f21[i1,i6,i3])+(2*f31[i1,i2,i5,i7]
            *f3[i3,i4,i6]+4*f31[i2,i4,i5,i7]*f3[i1,i3,i6]
            +4*f31[i1,i2,i5,i6]*f3[i3,i4,i7]+2*f31[i4,i5,i7,i2]
            *f3[i1,i3,i6])+(4*f22[i1,i7,i3,i6]*f21[i4,i5,i2]
            +4*f22[i4,i5,i3,i6]*f21[i1,i7,i2]+4*f22[i1,i7,i4,i5]
            *f21[i3,i6,i2])+(4*f211[i5,i6,i2,i4]*f3[i1,i3,i7]
            +2*f211[i1,i2,i6,i7]*f3[i3,i4,i5])
            +2*f4[i1,i2,i4,i5]*f111[i3,i6,i7]/3;
for jj from 1 to w do
   L46[jj,ii] :=L[jj,i1]*L[i2,i3]*L[i4,i6]*L[i5,i7];
   od;od;od;od;od;od;od;od;
 ii :=0;
for i1 from 1 to w do  for i2 from 1 to w do  for i3 from 1 to w do
for i4 from 1 to w do  for i5 from 1 to w do  for i6 from 1 to w do
for i7 from 1 to w do  for i8 from 1 to w do  for i9 from 1 to w do
  ii :=ii+1;
  BB5[ii] :=f3[i1,i4,i6]*
       (f3[i2,i3,i5]*f3[i7,i8,i9]+2*f3[i3,i5,i8]*f3[i2,i7,i9]
       +4*f3[i2,i5,i7]*f3[i3,i8,i9]+8*f3[i2,i7,i8]*f3[i3,i5,i9]
       +2*f3[i2,i3,i5]*f21[i7,i8,i9]+4*f3[i3,i5,i8]*f21[i2,i7,i9]
       +2*f3[i7,i8,i9]*f21[i2,i5,i3]+4*f3[i3,i7,i8]*f21[i2,i5,i9]
       +8*f3[i2,i7,i8]*f21[i3,i5,i9]+8*f3[i2,i7,i8]*f21[i3,i9,i5]
       +8*f3[i2,i5,i7]*f21[i3,i8,i9]+8*f3[i3,i5,i8]*f21[i2,i7,i9]
       +4*f3[i3,i8,i9]*f21[i2,i7,i5]
       +4*f21[i7,i8,i9]*f21[i2,i5,i3]+4*f21[i2,i5,i9]*f21[i7,i8,i3]
       +8*f21[i7,i8,i2]*f21[i3,i9,i5]+8*f21[i7,i8,i5]*f21[i3,i9,i2]
       +4*f3[i2,i5,i8]*f111[i3,i7,i9])+f3[i3,i4,i9]*
       (8*f3[i2,i7,i8]*f21[i1,i5,i6]+4*f3[i2,i6,i7]*f21[i1,i5,i8]
       +8*f21[i1,i5,i8]*f21[i2,i7,i6]+8*f21[i1,i5,i6]*f21[i2,i7,i8]
       +8*f21[i1,i6,i2]*f21[i5,i7,i8]);
for jj from 1 to w do
  L520[jj,ii] :=L[jj,i1]*L[i2,i3]*L[i4,i5]*L[i6,i7]*L[i8,i9];
  od;od;od;od;od;od;od;od;od;od;
for jj from 1 to w do
   U11[jj] := u11[jj]/2;
   U12[jj] := -U11[jj]+1/8*(add(BB3[i]*L34[jj,i],i=1..w5))
                      +1/4*(add(BB4[i]*L46[jj,i],i=1..w7))
                      +1/8*(add(BB5[i]*L520[jj,i],i=1..w9)); od;
```

The case of the two parameter gamma distribution has been checked with the values in the table of Bowman and Shenton (1988). The case of the three parameter gamma distribution has been checked with the values in the paper of Bowman and Shenton (2002); see §4.2. The case of the two parameter Weibull Distribution has been checked with the paper of Bowman and Shenton (2000). The general usage of Maple language we refer to Heck (2003).

6.3 An application

6.3.1 Mixture distribution

A mixture of a Poisson-Poisson distribution and a Poisson distribution is considered. The Poisson-Poisson distribution is a Lagrange distribution depending on two transformations, (i) $t = ug(t)$, and (ii) $f(t) = G(u)$, $g(\cdot)$, and $G(\cdot)$ being probability functions for a Poisson random variable. Previous studies considered binomial, negative binomial, Gram-Charlier and Pearson discrete distributions (Bowman and Shenton, 1998). Here we consider solutions associated with sister chromatid exchange lymphocyte data, the data base quite large, the objective being to determine whether smoking played a significant role. We study for example, in the case associated with female nonsmokers (FNS), male nonsmokers (MNS), female smokers (FS), and male smokers (MS), deriving asymptotic biases, variances, skewness and kurtosis for the four m.l. estimators. The probability function is

$$P(x; t_1, t_2, t_3, t_4) = t_4 t_1 (t_1 + t_2 x)^{x-1} e^{-(t_1+t_2 x)}/x! + (1 - t_4) e^{-t_3} t_3^x / x!$$

for $x = 0, 1, \cdots, 0 < t_4 < 1, 0 \le t_2 < 1, t_1 > 0, t_3 > 0$. It will be seen that when $t_2 = 0$, the first component reduces to a Poisson probability function. Central moments are:

$$\begin{aligned}
\mu_1' &= \frac{t_1}{(1-t_2)}, \\
\mu_2 &= \frac{t_1}{1-t_2} + \frac{t_1 t_2}{(1-t_2)^3}, \\
\mu_3 &= \frac{3\mu_2 t_1}{(1-t_2)^2} + \frac{t_1(1-t_2) + t_1 t_2}{(1-t_2)^3}, \\
\mu_4 &= 3\mu_2^2 + \frac{\mu_2\{15t_1^2 + 4t_2(1-t_2)\}}{(1-t_2)^4} + \frac{t_1(1-6t_2)}{(1-t_2)^5} + \frac{6t_1 t_2^2}{(1-t_2)^6}.
\end{aligned}$$

Clearly $0 \le t_2 < 1$.

6.3.2 The data sets and the analysis of data

The complete data set is given in Table 6.

Table 6

Distribution of SCE

Sample	number of subj.	SCE per cell 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Total	393	16	77	338	744	1374	1888	2285	2416	2499	2264	1631	1227	911	618	443	296
Females	179	7	24	141	312	587	809	994	1045	1122	1049	762	622	471	304	225	147
Males	214	9	53	197	432	787	1079	1291	1371	1377	1215	869	605	440	314	218	149
Non-Smokers	290	11	63	269	604	1078	1454	1762	1805	1907	1695	1175	878	582	419	287	188
Females	124	4	20	103	247	447	570	722	726	806	740	519	432	293	197	128	87
Males	166	7	43	166	357	631	884	1040	1079	1101	955	656	446	289	222	159	101
Smokers	73	4	8	42	91	198	305	349	408	427	390	351	248	243	148	117	85
Females	30	1	5	14	40	78	107	140	165	174	157	154	110	101	63	45	39
Males	43	3	3	28	51	120	198	209	243	253	233	197	138	142	85	72	46

Sample	SCE per cell (continued) 16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Total	209	145	99	49	41	33	18	11	2	4	4	3	2	0	0	2	1
Females	101	72	64	23	23	15	9	7	2	3	4	2	1	0	0	2	1
Males	108	73	35	26	18	18	9	4	0	1	0	1	1	0	0	0	0
Non-Smokers	116	73	57	29	16	12	8	6	1	1	2	1	0	0	0	1	0
Females	52	35	31	12	9	6	4	5	1	0	2	1	0	0	0	1	0
Males	64	38	26	17	7	6	4	1	0	1	0	0	0	0	0	0	0
Smokers	75	57	33	17	19	18	9	5	0	0	0	1	1	0	0	1	0
Females	32	31	6	8	10	11	5	3	0	0	0	0	1	0	0	0	0
Males	43	26	27	9	9	7	4	2	0	0	0	1	0	0	0	1	0

Taken from Bowman et al (1998).

The distributions appear to be unimodal, and frequencies reduce to zero when x is near to 32.

Byers and Shenton (1999) gave the m.l. estimator values, and corresponding standard errors in parenthesis, and they are given in Table 7.

Table 7. m.l. estimates of generalized Poisson mixture parameters

	t_4	t_1	t_2	t_3
FNS	0.9823(0.0079)	7.1748(0.0875)	0.0954(0.0141)	16.7068(1.2750)
MNS	0.9213(0.0260)	7.1005(0.0785)	0.0272(0.0197)	12.5499(0.6092)
FS	0.8997(0.0469)	7.6431(0.2094)	0.0969(0.0435)	15.3648(1.0808)
MS	0.8496(0.0621)	7.5262(0.1765)	0.0564(0.0474)	13.8857(0.8611)

In Table 8 we give our values of biases, standard errors, skewness and kurtosis.

Table 8. Poisson-Poisson and Poisson mixture distribution

		m.l.e.	Unbiased m.l.e.	Bias	σ	$\sqrt{\beta_1}$	β_2
FNS	t_1	7.1748	7.1688	0.0060	0.0859	0.0761	2.9496
$N = 6200$	t_2	0.0954	0.0974	-0.0020	0.0135	-0.2624	2.9050
	t_3	16.7068	16.7141	-0.0073	1.2010	0.1192	2.9730
	t_4	0.9823	0.9845	-0.0022	0.0075	-1.1937	5.3808
MNS	t_1	7.1005	7.1001	0.0004	0.0778	0.0555	3.0394
$N = 8300$	t_2	0.0272	0.0297	-0.0025	0.0208	-0.2940	2.8395
	t_3	12.5499	12.5294	0.0205	0.6822	0.2461	2.7918
	t_4	0.9213	0.9269	-0.0056	0.0292	-0.9198	3.7515
FS	t_1	7.6431	7.6262	0.0169	0.2101	0.1458	3.0142
$N = 1500$	t_2	0.0969	0.1048	-0.0079	0.0454	-0.3266	2.3979
	t_3	15.3648	15.2877	0.0771	1.1542	0.4732	2.9089
	t_4	0.8997	0.9102	-0.0105	0.0505	-0.8526	2.4153
MS	t_1	7.5262	7.5160	0.0102	0.1754	0.1317	3.0419
$N = 2150$	t_2	0.0564	0.0621	-0.0057	0.0453	-0.1892	2.4201
	t_3	13.8857	13.8274	0.0583	0.8242	0.4681	2.9360
	t_4	0.8496	0.8568	-0.0072	0.0589	-0.4666	1.7824

(a) **Standard Errors**:

These are given in Table 7 (parenthetic entries) and are due to Byers and Shenton (1999). Byers used the Splus program based on the m.l. estimator values of t_4, t_1, t_2, t_3, namely $\hat{t}_4$, $\hat{t}_1$, $\hat{t}_2$, $\hat{t}_3$. By and large our values (Table 8) of the standard errors (or standard deviation) agree with the Byers' values. Notice that, (i) the values of t_3 (the second component Poisson parameter) are all greater than twelve, yet the standard errors are all less than 1.3, (ii) the proportion in the mixture t_4 of the Poisson component $(1 - t_4)$ is quite small, (iii) the first component in the mixture Poisson-Poisson is basically Poisson, t_2 being small.

(b) **Bias** (Table 8):

Using first and second order terms in $E(\hat{t})$, the bias for the estimators is negligible except for the proportion parameter t_4, for which there is a correction of -0.01 for FS. Note that here the sample size is $N = 1500$, the smallest in the group of four.

(c) **Skewness and Kurtosis**:

For the Poisson-Poisson components t_1 and t_2, if $t_2 = 0$ the $P \cap P$ reduces to a Poisson probability function with parameter t_1. From the Table 7, t_1 is about 7.0 with small sigma; also $\sqrt{\beta_1}$ and β_2 are nearly normal values (0,3). Hence we may assume the distribution is approximately normal. Looking at t_2 (a discrepancy parameter from the Poisson), it is small in comparison to t_1, with small variance; asymptotic normality is quite possible. These remarks apply to the four groups, FNS, MNS, FS, and MS.

Now t_3 relates to the Poisson of the second component in the mixture; it is in the range 12-17 and its standard deviation is small in comparison. Again asymptotic normality is acceptable.

The skewness of the proportion parameter t_4 is negative in the all four groups, but for MS the kurtosis indicate a platkurtic distribution. The skewness can not be neglected and is largest in value for FNS.

Altogether, the sample sizes are large, and the first component in the mixture is close to a Poisson distribution, whereas there is a small proportion of the second component which is Poisson with large parameter.

For m.l. estimators, variance and skewness have been implemented using the Maple system. The asymptotic variance of the four parameter estimators check out against a previous study by Byers and Shenton (1999); in this paper Byers set up a Splus program to compute the standard errors of the four parameter model of SCE data.

7 Conclusions

Maple symbolic programs have been set up for the moments of a m.l. estimator $\hat{\theta}_a$, given a density (or probability function) defined by s parameters $\theta_1, \theta_2, \cdots, \theta_s$. Moments of the corresponding random variable are assumed to exist.

Moments such as $\mu_{rs}(\hat{\theta}_a)$ are considered; here r refers to the rth central moment, s to the coefficient N^{-s} in this moment. The bias formula is expanded to N^{-2}, variance to N^{-2}, μ_3 to N^{-2}, and μ_4 to N^{-3}.

Asymptotic series are set up; for example

$$E(\hat{\theta}_a) \sim \theta_a + \frac{\phi_1(\theta)}{N} + \frac{\phi_2(\theta)}{N^2} + \cdots \qquad (N \to \infty)$$

and similarly for $E(\hat{\theta}_a - \theta_a)^m$, $m = 2, 3, 4$.

Advances in symbolic languages are sketched. It is quite possible that the first use of a digital computer to solve a perturbation series in celestial mechanics occurred about 50 or so years ago. Van Dyke (1975) mentioned the case of a French astronomer who basically considered a quintuple Taylor series carried out to a term of order nine.

One of us has developed the Maple programs given here. The main results concern $\mu_{32}(\hat{\theta}_a)$, and $\mu_{43}(\hat{\theta}_a)$ for m.l. estimators. In our example of an application, the density ('pf') can involve 4 parameters, and extensions to 5 or more parameters are possible.

A four parameter discrete distribution is given as an example. Here standard errors check up from an independent approach.

The asymptotic skewness, being location free, and scale free, is an addition to our knowledge of the behavior of m.l. estimators. The basic requirement is the existence of expectation of logarithmic derivatives of the density or probability function.

Acknowledgment

The first author's research was sponsored by the Applied Mathematical Sciences Research Program, Oak Ridge National Laboratory. It is operated by UT-Battelle, LLC for the U. S. Department of Energy, under contract number DE-AC05-00OR22725.

Appendix

A The Gamma function and asymptotic series

A.1 Euler

Both the gamma distribution and the Weibull distribution have strong associations with the gamma functions. How do asymptotics appear?

The gamma integral,

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \qquad (\Re(z) > 0)$$

is due to Euler (1700-1783). It is called, according to Whittaker and Watson (1915), **Euler's Integral of the second kind**.

A.2 Binet

In the early part of the 19th century, Binet (1839) initiated work on $\ln \Gamma(z)$, giving the expression

$$\ln \Gamma(z) = \left(z - \frac{1}{2}\right) \ln z - z + \frac{1}{2} \ln(2\pi) + \int_0^\infty \left(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\right) \frac{e^{-tz}}{t} dt. \quad (\Re(z) > 0) \tag{12}$$

Some time later Binet produced the improved result

$$\ln \Gamma(z) = \left(z - \frac{1}{2}\right) \ln z - z + \frac{1}{2} \ln(2\pi) + 2 \int_0^\infty \frac{\arctan(t/z)}{e^{-2\pi t} - 1} dt. \quad (\Re(z) > 0)$$

A.3 The "Remainder" term $J(z)$

The integral in (11) may be written

$$2 \int_0^\infty \frac{(\arctan t/z) e^{-2\pi t}}{1 - e^{-2\pi t}} dt$$

and by integration by parts becomes

$$\frac{1}{\pi} \int_0^\infty \left(\ln \frac{1}{1 - e^{-2\pi t}}\right) \frac{z dt}{z^2 + t^2}. \quad (\Re(z) > 0)$$

Hence

$$\ln \Gamma(z) = I(z) + J(z),$$

where

$$I(z) = \left(z - \frac{1}{2}\right) \ln z - z + \frac{1}{2} \ln(2\pi),$$
$$J(z) = -\frac{1}{\pi} \int_0^\infty \left\{\ln(1 - e^{-2\pi t})\right\} \frac{z dt}{z^2 + t^2}. \quad (\Re > 0)$$

A.4 Series for $J(z)$

Now $J(z)$ can, at least formally, be expanded in series in descending powers of z, i.e.

$$J(z) = \frac{c_0}{z} - \frac{c_1}{z^2} + \frac{c_2}{z^3} \cdots,$$

where

$$c_0 = -\frac{1}{\pi} \int_0^\infty \ln(1 - e^{-2\pi t}) dt,$$
$$c_1 = -\frac{1}{\pi} \int_0^\infty \ln(1 - e^{-2\pi t}) t^2 dt,$$
$$c_s = -\frac{1}{\pi} \int_0^\infty \ln(1 - e^{-2\pi t}) t^{2s} dt, \quad (s = 0, 1, \cdots).$$

Expanding the logarithmic terms, we have, with $2\pi t = u$,

$$\begin{aligned}
c_s &= \frac{1}{\pi}\int_0^\infty \left(\frac{1}{2\pi}\right)\left(\ln\frac{1}{1-e^{-u}}\right)\left(\frac{u}{2\pi}\right)^{2s} du \\
&= \frac{1}{\pi(2\pi)^{2s+1}}\int_0^\infty \left(e^{-u}+\frac{e^{-2u}}{2}+\frac{e^{-3u}}{3}+\frac{e^{-4u}}{4}+\cdots\right) u^{2s} du \\
&= \frac{(2s)!}{\pi(2\pi)^{2s+1}}\left(1+\frac{1}{2^{2s+2}}+\frac{1}{3^{3s+2}}+\frac{1}{4^{2s+2}}+\cdots\right) \\
&= \frac{(2s)!}{\pi(2\pi)^{2s+1}}\zeta(2s+2),
\end{aligned}$$

in terms of the Riemann zeta function. From N.B.S. (1970, p807),

$$\zeta(2s+2) = (2\pi)^{2s+2}|B_{2s+2}|/\{2(2s+2)!\}$$

so that

$$c_s = \frac{|B_{2s+2}|}{(2s+2)(2s+1)} \qquad (s = 0, 1, \cdots)$$

in terms of Bernoulli numbers,

$$B_0 = 1,\;\; B_1 = -\frac{1}{2},\;\; B_2 = \frac{1}{6},\;\; B_3 = 0,\;\; B_4 = -\frac{1}{30},$$
$$B_5 = 0,\;\; B_6 = \frac{1}{42},\;\; B_7 = 0,\;\; B_8 = -\frac{1}{30}.$$

Hence there is the asymptotic series for $J(z)$, namely

$$\begin{aligned}
J(z) &\sim \frac{|B_2|}{1\cdot 2z} - \frac{|B_4|}{3\cdot 4\cdot z^3} + \frac{|B_6|}{5\cdot 6\cdot z^5} - \frac{|B_8|}{7\cdot 8\cdot z^7}\cdots \qquad (z\to\infty \text{ in } |z|<\pi) \\
&\sim \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5} - \frac{1}{1680z^7} + \frac{5}{5940z^9}.
\end{aligned} \tag{13}$$

A.5 Semi-convergent series

This series (13) is not divergent, but is semi-convergent, i.e. the error in using s terms is less in value than the first term omitted provided z is real and positive. For we have, formally the remainder after s term is

$$\begin{aligned}
R_s(z) &= (-1)^s\left\{\frac{c_s}{z^{2s+1}} - \frac{c_{s+1}}{z^{2s+3}} + \cdots\right\} \\
&= \frac{(-1)^{s+1}}{\pi}\int_0^\infty \ln(1-e^{-2\pi u})\left\{\frac{u^{2s}}{z^{2s+1}} - \frac{u^{2s+2}}{z^{2s+3}}\cdots\right\} \\
&= \frac{(-1)^{s+1}}{\pi z}\int_0^\infty \ln(1-e^{-2\pi u})\frac{u^{2s}}{z^{2s}}\left(1+\frac{u^2}{z^2}\right)^{-1} du \\
&= \frac{(-1)^{s+1}}{\pi z}\int_0^\infty \ln(1-e^{-2\pi u})\frac{u^{2s}}{z^{2s}}\frac{z^2}{z^2+u^2} du.
\end{aligned}$$

for z real and positive. Thus

$$|R_s(z)| < \frac{1}{\pi z}\int_0^\infty \left(\ln\frac{1}{1-e^{-2\pi u}}\right)\frac{u^{2s}}{z^{2s}}du = \frac{c_s}{z^{2s+1}},$$

i.e. the first term omitted.

A.6 G.H. Hardy and log $n!$

Hardy (1949) considered the expression

$$\log n! = \sum_1^n \log m = \left(n-\frac{1}{2}\right)\log n - n + C + \frac{B_1}{1\cdot 2n} + \frac{B_2}{3\cdot 4n^3} + \frac{B_3}{5\cdot 6n^5}\cdots$$

where

$$C = 1 - \frac{B_1}{1\cdot 2} + \frac{B_2}{3\cdot 4} + \frac{B_3}{5\cdot 6} + \frac{B_4}{7\cdot 8}\cdots.$$

Hardy remarks that the series is semi-convergent and can be used to calculate $\log n!$, and C with $z=1$. Note that he uses

$$B_1 = \frac{1}{6},\quad B_2 = \frac{1}{30},\quad B_3 = \frac{1}{42},\quad B_4 = \frac{1}{30},\quad B_5 = \frac{5}{66}.$$

The series for C may be set up from our expression in (13); since in this expression the fifth term is larger in value than the forth, the first four showing a decreasing sequence in values, we use

$$C \sim 1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} = 0.91865$$

which agrees with $\frac{1}{2}\ln(2\pi)$ to 3 significant digits.

There are surprising inconsistencies in Hardy's short note (Hardy, 1949, p.329). To use the equality symbol with no mention of an appropriate domain is rather surprising.

Returning to $\ln\Gamma(z)$, Wall's (1948) account and association with the theory of continued fractions is interesting. However there is no mention of the fact that Stieltjes discussed the continued fraction form in 1889; Wall however includes two new partial numerators in the continued fraction.

There are two errors to note on Wall's account. First in his (93.5) the sign should be negative. Second the infinite series for $J(z)$ in terms of Bernoulli numbers is not totally divergent.

B Summaries of some previous papers

COMMU.STATIST.-THEORY METH., 18940, 1511-1579 (1990)

THE APPROXIMATE DISTRIBUTION OF FOUR MOMENT STATISTICS FROM TYPE III DISTRIBUTIONS

K.O. Bowman
Oak Ridge National Laboratory
P.O.Box 2009
Oak Ridge, Tennessee 37831-8083

L.R. Shenton
University of Georgia
Athens, Georgia 30602

Key Words and Phrases: continued fractions; divergent series; moments; multivariae Taylor series; sample moments; simulation cycles.

ABSTRACT

Taylor series in the sample size are set up for the first four moments of the standard deviation, skewness, kurtosis, and coefficient of variation, the populations being χ^2 (gamma, Pearson Type III). These moments being out of reach of purely mathematical development, the study proceeds along two independent lines. For the one, simulation methods are used, an attempt being made to fix a cycle length to ensure some stability - this cycles length is pivoted on the fourth moment of the kurtosis, an expression involving sixteenth powers of the basic χ^2 - random variable. The second line of attack uses the Taylor moment series which are taken out to at most sixty terms in the total derivatives. An algorithm is used to derive the expectation of a product of powers of elements which consist of non-central sample deviates; there are four of these involved in the kurtosis, three in the skewness, and two in the standard deviation. There is an added parameter for sample size. This expectation of products of powers of sample deviates generates a set of coefficients, each coefficient multiplied by a power of n^{-1}; the larger the moment product, the greater is the span of the powers of n^{-1}. If a final moment series is desired to include all contributions up to n^{-s}, then at least $2s$ terms will be required in the Taylor expansion; moreover the series may turn out to be divergent as far as can be judged by the behavior of the terms computed. At this point, since the series are not seen to be one-signed, and since divergence is not too chaotic (as far as the triple factorial, say), rational fraction sequences are set up to dilute divergence (or accelerate apparent convergence); the approach is often successful but there are problems with small sample sizes and large skewness of the population sampled. Lastly, gross errors in relying on basic asymptotes are noted. The study brings out unusual confluences - computer oriented numerical analysis, distributional theory and approximation, and the power of rational fraction a divergency reducing tools.

J.Statist.Comput.Simul.,Vol 43,pp.217-224 (1992)
Reports available directly from the publisher:

PARAMETER ESTIMATION FOR THE BETA DISTRIBUTION

K.O. BOWMAN
Mathematical Science Section, Engineering Physics and Mathematics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6367
L.R. SHENTON
University of Georgia, Athens, Georgia 30602

(Received November 29, 1990: in final form March 18, 1992)

Moment estimators, based on the first two sample moments, for the two index parameters of the beta density (known end-points) are studied. Four moments of these estimators are set up using Computer Oriented Extended Taylor Series (COETS) to 60 terms followed by rational fraction approximations. These indicate, over a limited parameter space, that allowing for simplicity of calculation and other characteristics they are preferable to maximum likelihood estimators.

KEYWORDS: Extended Taylor series, M.l. comparisons, moment estimators, rational fraction approximation.

Mutation Research 403 (1998) 159-169

Sister chromatid exchange data and Gram-Charlier series

K.O. Bowman[a], Wesley Edding[b], Marvin A. Kastenbaum[c]. L.R. Shenton[d]

[a] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, P. O. Box 2008, Bldg 6012, Oak Ridge, TN 37831-6367, USA*
[b] *Vanderbilt University, Box 5313 Station B, Nashville, TN 37235, USA*
[c] *16933 Timberlakes Drive, S.W., Fort Myers, FL 33908, USA*
[d] *Department of Statistics, University of Georgia, Athens, GA 30602, USA*

Received 30 January 1998; revised March 1998; accepted 2 April 1998.

Abstract

Bowman et al. [K.O. Bowman, M.A. Kastenbaum, L.R. Shenton, Fitting multi-parameter distributions to SEC data, Mutation Res., 358 (1996) 15-24] showed how discrete Pearson and discrete Johnson translation-system distributions may be fitted to sister chromatid exchange (SCE) data presented by Bender et al. [M.A. Bender, R.J. Preston, R.C. Leonard, B.E. Pyatt, P.C. Gooch, On the distribution of spontaneous SCE in human peripheral blood lymphocytes, Mutation Res. 281 (1992) 227-232]. When their performances were measured by the chi-square test of goodness of fit, these distributions proved to be only moderately better alternatives to the poorly fitting Poisson, binomial, and negative binomial distributions. In this paper we extend our search for better characterizations of the SCE data by calling upon the Gram-Charlier type B approximation of the negative binomial distribution. We introduce an innovative extension of methods described in a little-known paper by Aitken and Gonin [A.C. Aiken, H.T. Gonin, On fourfold sampling with and without replacement, Proc. R.Soc.Edinburgh, 55 (1934) 114-125.], and show how a theorem by Cramér [H.Cramér, Mathematical Methods of Statistics, Princeton Univ.Press, 1946.], relating to the scale factor m_2/m_1' and its asymptotic distribution may be used to discriminate between smokers and non-smokers of the same gender.

Keywords; Chi-squared; Factorial moment; Gram-Charlier distribution; Negative binomial distribution; Distribution; Orthogonal polynomial.

COMMUN.STATIST-THEORY METH.,28(10),2497-2508 (1999)

THE ASYMPTOTIC MOMENT PROFILE AND MAXIMUM LIKELIHOOD: APPLICATIONS TO GAMMA AND GAMMA RATIO DENSITIES

K. O. Bowman

Computer Science and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008, Oak Ridge, Tennessee 37831-6367

L. R. Shenton

Department of Statistics, University of Georgia
Athens, Georgia 30602

Key words covariance matrix; gamma density; generating functions; products of random variables; Stieltjes integrals.

ABSTRACT

In previous papers (Bowman and Shenton, 1998, 1999a) we have given expressions for the asymptotic skewness and kurtosis for maximum likelihood estimators in the case of several parameters. Skewness is measured by the third standardized central moment, and kurtosis by the fourth standardized central moment. Moments of the basic structure are assumed to exist. The overarching entity is the covariance matrix (Hessian form), and elements of its inverse. These entities involve Stiltjies integrals relating to sums of products of multiple derivatives linked to the basic structure. The first paper dealt with skewness and gives a simple expression readily computerized. The second paper is devoted to the forth standardized central moment and although a certain simplification is discovered, the resulting formula is still somewhat complicated. It is surprising to find that the asymptotic kurtosis in general requires the evaluation of several hundred components. The present paper studies cases involving one, two, and three parameters and mentions strategies aimed at avoiding algebraic and numerical errors.

COMMUN.STATIST-THEORY METHOD., 28(11), 2641-2654 (1999)

THE ASYMPTOTIC KURTOSIS FOR MAXIMUM LIKELIHOOD ESTIMATORS

K. O. Bowman
Computer Science and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008, Oak Ridge, Tennessee 37831-6367

L. R. Shenton
Department of Statistics, University of Georgia
Athens, Georgia 30602

Key words: asymptotic series; expectation of random variable products; Fisher's linkage; percentage points; moment series; polarization operator; products of random variables.

ABSTRACT

In general, when moments exist, the dominant term in the fourth central moment of an estimator is three times the square of the asymptotic variance; this leads to the value three for the asymptotic kurtosis. Working on the approach given in Bowman and Shenton (1998) we now complete the basic asymptotic moment profile by giving an expression for the third order term in the fourth central moment of a maximum likelihood estimator, assuming the existence of derivatives of a density and also the existence of the covariance matrix inverse. A four moment distributional model, such as the Pearson system, or Johnson translation system, may be used to approximate percentage points of the estimators.

Far East J.Theo.Stat. 4(2)(2000), 391-422

MAXIMUM LIKELIHOOD AND THE WEIBULL DISTRIBUTION

K.O. BOWMAN and L.R. SHENTON

Computer Science and Mathematics Division
Oak Ridge National Laboratory
P.O.Box 2008
Oak Ridge, Tennessee 37831-6367, U.S.A.
e-mail: bowmanko@ornl.gov

Department of Statistics
University of Georgia
Athens, Georgia 30602

Abstract

The Weibull distribution has three parameters, location a, scale b and shape c. Maximum likelihood estimators are $\hat{a}$, $\hat{b}$. $\hat{c}$, and solutions may not always exist; for example the location estimate $\hat{a}$ must be less than the smallest member of the sample. We consider three estimation problems: (1) Estimation of one parameter when the other two are assumed to be known. (2) Estimating the scale and shape parameters when the location parameter is known. (3) Estimating the three parameters simultaneously.

Results being based on the covariance matrix and its cofactors, we give explicit expressions for the asymptotic bias, 2nd order variances, skewness to order $1/\sqrt{N}$, and asymptotic kurtosis to order $1/N$, N being the sample size. Except for the simultaneous estimation of a, b, c, the expressions for these asymptotic moments and moment ratios are simple in form involving gamma and Riemann Zeta functions. They provide a new basic supplement to our knowledge of maximum likelihood estimator moments.

A surprising discovery is the part played by the location parameter whenever it has to be estimated. For the three parameter estimation case it is already known that asymptotic covariance only exist if $c > 2$. It turns out that the asymptotic skewness only exist if $c > 3$, and the asymptotic kurtosis only exist $c > 4$. This applies to the asymptotic distribution of $\hat{a}$, $\hat{b}$, and $\hat{c}$. The source of this characteristic is the singularity appearing in the expectation of logarithmic derivatives. When less than 3 parameters are to be estimated the problem arises whenever $\hat{a}$ intrudes.

For the 3 parameter case, a new expression is developed for the asymptotic variance of $\hat{c}$. Lastly, wherever possible simulation studies are invoked for verification purposes.

ELSEVIER Computational Statistics & Data Analysis 36 (2001) 299-310

Weibull distributions when the shape parameter is defined

K.O. Bowman[a], L.R.Shenton[b]
[a] *Computer Science and Mathematics Division, Building 6012, Oak Ridge National Laboratory,*
P.O.Box 2008, MS-6367, Oak Ridge, TN 37831-6367. USA
[b] *Department of Statistics, University of Georgia, Athens, GA 30602. USA*

Received 1 April 2000; received in revised from 1 October 2000.

Abstract

The Weibull distribution, depending on parameters of location, scale, and shape, is often useful as a model for fracture data sets. If the location parameter is to be estimated then we have shown that maximum likelihood methods are not recommended. In the data set considered here the shape parameter is known to lie between 2 and 3 or so. We therefore studied the 2 parameter model for which the shape parameter is known, or has a probability structure. Simple moment estimators are used and some moments of these are studied and verified by simulation.

Keyword: Envelope distribution, moment estimator, moments of sample moments, Padé sequences, Taylor series, unbiased estimators.

J.Statist.Comput.Simul., 2002,Vol.72(5),pp.391-401

PROBLEMS WITH MAXIMUM LIKELIHOOD ESTIMATION AND THE 3 PARAMETER GAMMA DISTRIBUTION

K.O. BOWMAN[a] and L.R. SHENTON[b]

[a] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, P.O.Box 2008, MS-6367, Oak Ridge, Tennessee 37831-6367;*
[b] *Department of Statistics, University of Georgia, Ahtens, Georgia 30602*

(*Received 21 March 2001: In final form 3 October 2001*)

The three parameters involved are scale a, shape ρ, and location s. Maximum likelihood estimators are $(\hat{a}, \hat{\rho}, \hat{s})$. Using recent work on the second order variances, skewness, and kurtosis we establish the facts, that if the location parameter s is to be estimated, then the asymptotic variances only exist if $\rho > 2$, asymptotic skewness only exists if $\rho > 3$, and 2nd order variances and third order fourth central moments only exist if $\rho > 4$. The result of these limitations is that in general very large sample sizes may be needed to avoid inference problems. We also include new continued fractions for the asymptotic covariances of the maximum likelihood estimators considered.

SIAM J.Appl.Math.
Vol 23,No 2,September 1972

USE OF LAGRANGE EXPANSION FOR GENERATING DISCRETE GENERALIZED PROBABILITY DISTRIBUTIONS

P.C. CONSUL and L.R. SHENTON

Abstract. Considering $g(t)$ and $f(t)$ as two probability generating functions defined on nonnegative integers with $g(0) \neq 0$. We use Lagrange's expansion, together with the transformation $t = u \cdot g(t)$. to define families of discrete generalized probability distributions by the name of Lagrange distributions as

$$Pr[X = 0] = L(g : f : 0) = f(0),$$
$$Pr[X = x] = L(g : f : x) = \frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\{(g(t))^x \cdot f'(t)\}|_{t=0}$$

for $x = 1, 2, 3, \cdots$, where the different families are generated by assigning different values to $g(t)$ and $f(t)$. General formulas for writing down the central moments of Lagrange distributions are obtained and it is shown that they satisfy the convolution property. The double binomial family of Lagrange distributions is studied in greater detail as it gives a large number of discrete distributions, including Borel-Tanner distribution, Haight's distribution, generalized Poisson and generalized negative binomial distributions, as particular cases.

Sankhyà: The Indian Journal of Statistics
1974, Volume 36, Series B, Pt.2, pp.154-162.

MAXIMUM LIKELIHOOD ESTIMATION FOR THE PARAMETERS OF THE HERMITE DISTRIBUTION

By Y.C. PATEL, L.R. SHENTON
Universitiy of Georgia
and K.O. BOWMAN
Oak Ridge National Laboratory

Properties of the maximum likelihood estimators of the parameters (a, b) in the Hermite distribution, with probability generating function $\exp\{a(t-1)+b(t^2-1)\}$ are discussed. Numerical assessments of the first and second order coefficients in the biases and covariances are given for a limited region of the parameter space.

Rep.Stat.Appl.Res., JUSE
Vol.17, No.3, 1970

TABLES OF THE MOMENTS OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE TWO PARAMETER GAMMA DISTRIBUTION

By L.R. Shenton
Professor of Statistics, Computer Center,
University of Georgia, Athens, Georgia 30601
K.O. Bowman
Mathematical Statistician, Computer Technology Center,
Union Carbide Corporation, Nuclear Division
Oak Ridge, Tennessee, USA

The moments of the maximum likelihood estimators of the parameters ρ, a for the gamma density $f(x) = k(x/a)^{\rho-1}\exp(-x/a)$ are briefly tabulated. These include the biases $E(\hat{\rho}-\rho)/\rho$, $E(\hat{a}-a)/a$, the variances $\mathrm{Var}(\hat{\rho}/\rho)$, $\mathrm{Var}(\hat{a}/a)$, skewness and kurtosis. The range of values considered is approximately $\rho = 0.2$ to 3.0 and n (the sample size) lying between 12 and 100. Some comparisons of the moments with the usual first-order asymptotic values are made.

References

Abramowitz, M and Stegun, I.A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series 55.

Aiken, A.C. and Gonin, H.T. (1934). On fourfold sampling with and without replacement, *Proc.R.Soc.Edinburgh*, 55, 114-125.

Bender, M.A., Preston, R.J., Leonard, R.C., Pyatt, B.E. and Gooch, P.C. (1992). On the distribution of Spontaneous SCE in human peripheral blood lymphocytes, *Mutation Research*, 281, 227-232.

Bartlet, M.S. (1952). Approximate confidence intervals, II, *Biometrika*, 40, 306-317.

Binet, J.P.M. (1839). Journal de l'École Polytechnique, xvi, 123-143.

Bowman, K.O. (1963). *Moments to higher orders for maximum likelihood estimators, with an application to the negative binomial distribution*, Thesis in part fulfillment of Ph.D. requirements at Virginia Polytechnic Institute and State University.

Bowman, K.O., Eddings, W., Kastenbaum, M.A., and Shenton, L.R. (1998). Sister chromatid exchange data and Gram-Charlier series, *Mutation Research*, 403, 159-169.

Bowman, K.O., Kastenbaum, M.A., and Shenton, L.R. (1996). *Fitting Multi-Parameter Distributions to Human Chromosome Data*, Technical Report Number 96-3, Department of Statistics, University of Georgia.

Bowman, K.O. and Shenton, L.R. (1965). *Biases and Covariances of Maximum Likelihood Estimators*, K-1633, Union Carbide Corporation Nuclear Division.

Bowman, K.O. and Shenton, L.R. (1965). *Asymptotic Covariances for the Maximum Likelihood Estimators of the Parameters of a Negative Binomial Distribution*, K-1643, Union Carbide Corporation Nuclear Division.

Bowman, K.O. and Shenton, L.R. (1966). *Biases of Estimator for the Negative Binomial Distribution*, ORNL-4005, Oak Ridge National Laboratory.

Bowman, K.O. and Shenton, L.R. (1967). *Remarks on Estimation Problems for the Parameters of the Neyman Type A Distribution*, ORNL-4102, Oak Ridge National Laboratory.

Bowman, K.O. and Shenton, L.R. (1968). *Properties of Estimators for the Gamma Distribution*, CTC-1, Union Carbide Corporation Nuclear Division.

Bowman, K.O. and Shenton, L.R. (1969). Remarks on maximum likelihood estimators for the gamma distribution, *Proc. ICQC*, Tokyo, 519-522.

Bowman, K.O. and Shenton, L.R. (1970). Properties of the Maximum Likelihood Estimator for the Parameter of the Logarithmic Series Distribution, *Random Counts in Scientific Work*, 1, ed., G.P. Patil, Pennsylvania State University Press, 127-150.

Bowman, K.O. and Shenton, L.R. (1971). Small sample properties of the maximum likelihood estimator associated with Fisher's linkage problem, *Jour. Stat. Comput. and Simulation*, 1, 71-85.

Bowman, K.O. and Shenton, L.R. (1975). Comments on the distribution of the variance in sampling from the logistic distribution, *Proc. of the 40th Session of ISI*, Warsaw.

Bowman, K.O. and Shenton, L.R. (1982). Properties of estimators for the gamma distribution, *Communications in Statistics B - Simulation and Computation*, B11, 377-519.

Bowman, K.O. and Shenton, L.R. (1983). Maximum likelihood estimators for the gamma distribution revisited, *Communications in Statistics, B*, 12(6), 697-710.

Bowman, K.O. and Shenton, L.R. (1988). *Properties of Estimators for the Gamma Distribution,* Marcel Dekker, Inc.

Bowman, K.O., Shenton, L.R. and Karlof, C. (1995). Estimation problems associated with the three Parameter gamma distribution, *Commun. Statist. Theory Meth.*, 24(5), 1355-1376.

Bowman, K.O. and Shenton, L.R. (1998). Asymptotic skewness and the distribution of maximum likelihood estimators, *Commun.Statist. Theory and Method*, 27(11), 2743-2760.

Bowman, K.O. and Shenton, L.R. (1999). The asymptotic moment profile and maximum likelihood: Applications to gamma and gamma ratio densities, *Commun. Statist. Theory and Method*, 28(10), 2497-2508.

Bowman, K.O. and Shenton, L.R. (1999). The asymptotic kurtosis for maximum likelihood estimators, *Commun. Statist. Theory and Method*, 28(11), 2641-2654.

Bowman, K.O. and Shenton, L.R. (2000). Maximum likelihood and the Weibull distribution, *Far East Journal of Theoretical Statistics*, 4(2), 391-422.

Bowman, K.O. and Shenton, L.R. (2002). Problems with maximum likelihood estimators and the 3 parameter gamma distribution, *J. Statistical Computation and Simulation*, 72(5),391-401.

Bowman, K.O. and Shenton, L.R. (2005). The asymptotic variance and skewness of maximum likelihood estimators using Maple, *Journal of Statistical Computation and Simulation*, to appear.

Byers, H.R. Jr and Shenton, L.R. (1999). Sister chromatid exchange data fit with a mixture of Poisson distributions, *Mutation Research*, 427, 157-162.

Consul, P.C. and Shenton, L.R. (1972). On the multivariate generalization of the family of discrete Lagrangeian distributions, Proceedings of the Multivariate Statistical Inference Research Seminar, Halifax, 13-25.

Consul, P.C. and Shenton, L.R. (1972). Use of Lagrange expansion for generating discrete generalized probability distributions, *SIAM J. App. Math,* 23(2), 239-248.

Cox, D.R. and Snell, E.J. (1968). A general definition of residuals, *J.R.Statist.Soc.B*, 30, 248-275.

Cramèr, H. (1946). *Mathematical Methods of Statistics*, Princeton Univ.Press.

Dubay, S.D. (1965). Asymptotic properties of several estimators of Weibull parameters, *Technomerics*, 7, 423-434.

Good, I.J. (1965). Generalization to several variables of Lagranges expansion $\cdots$, *Proc.Cambridge Phylos.Soc.*,61, 499-517.

Haldane, J.B.S. (1953). The estimation of two parameters from a sample, *Sankhya*, 12, 313-320.

Haldane, J.B.S. and Smith, S.M. (1956). The sampling distribution of a maximum likelihood estimate, *Biometrika*, 43, 96-103.

Hardy, G.H. (1949). *Divergent Series*, Oxford, Clarendon Press.

Heck, A. (1997). *Introduction to Maple, Second Edition*, Springer-Verlag, New York.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1, Sec. Ed.*, John Wiley & Sons, New York.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2, Sec. Ed.*, John Wiley & Sons, New York.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1993). *Univariate Discrete Distributions, Sec. Ed* , John Wiley & Sons, New York.

Kratky, Reinfelds, Hucheson and Shenton (1972). *Tables of Crude Moments Expressed in terms of Cumulants*, Computer Center Report 1972(1), Computer Center, University of Georgia.

Patel, Y.C., Shenton, L.R. and Bowman, K.O. (1974), Maximum likelihood estimation for the parameters of the Hermite distribution, *Sankhya*, 36, 154-162.

Shenton, L.R. (1959). The distribution of moment estimators, *Biometrika*, 46, 296-305.

Shenton, L.R. and Bowman, K.O. (1963). Higher moments of a maximum-likelihood estimator, *J. Roy. Statist. Soc., Ser. B.*, 25, 305-317.

Shenton, L.R. and Bowman, K.O. (1967). Remarks on large sample estimators for some discrete distributions, *Technometrics*, 9, 587-598.

Shenton, L.R. and Bowman, K.O. with Appendix by Reinfelds, J. (1967). Sampling moments of moments and maximum likelihood estimators for discrete distributions, 36th Session of the International Statistical Institute, Sydney, Australia.

Shenton, L.R. and Bowman, K.O. (1970). Tables of the moments of the maximum likelihood estimators of the two parameter gamma distribution, *Rep. Statist. Appl. Res., JUSE*, 17(3), 25-44.

Shenton, L.R., Bowman, K.O. and Sheehan, D. (1971). Sampling moments of moments associated with univariate distributions, *J.R.Statist.Soc.B*, 33, 444-457.

Shenton, L.R. and Bowman, K.O. (1972). Further remarks on maximum likelihood estimators for the gamma distribution, *Technometrics*, 14, 725-733.

Shenton, L.R. and Bowman, K.O. (1973). Notes on the sample size to achieve normality for maximum likelihood estimators for the gamma distribution, *Monthly Weather Rev.*, 101(12), 891-892.

Shenton, L.R. and Bowman, K.O. (1974). Comments on higher order terms in moment of moments, moments of maximum likelihood estimators, regression residuals, *Proc. of the Prague Symposium on Asymptotic Statistics, September 3-6, 1973., Vol. 2*, ed., Jaroslan Jajek, Charles University, Prague, 323-344.

Shenton, L.R. and Bowman, K.O. (1977). *Maximum Likelihood Estimation in Small Samples*, Charles Griffin and Co., Ltd.

Shenton, L.R. and Consul, P.C. (1975). On bivariate Lagrange and Borel-Tanner distributions and their use in queueing theory, *Sankhya, Series A*.

Van Dyke, M. (1975). Computer extension of perturbation series in fluid mechanics, *Siam J. of Applied Math.*, 28(3), pp726-734.

Whittaker, E.T. and Watson, G.N. (1915). *A Course of Modern Analysis*, 2nd edition, completly revised (This book was the "bible" on mathematical analysis in the early part of the 20th century.) Cambridge University Press.

Wall, H.S. (1948). *Continued Fraction*, Chelsea Publishing Company, Bronx, N.Y.