Journal of Statistical Computation and Simulation
Vol. 75, No. 12, December 2005, 975–986

Taylor & Francis
Taylor & Francis Group

The asymptotic variance and skewness of maximum likelihood estimators using Maple†

K. O. BOWMAN*‡ and L. R. SHENTON§

‡Computational Sciences and Engineering Division, Oak Ridge National Laboratory, P.O. Box 2008, 5770, MS-6085, Oak Ridge, TN 37831-6085, USA
§Department of Statistics, University of Georgia, Athens, GA 30602, USA

(Revised 26 March 2004; in final form 4 October 2004)

In 1998, Bowman and Shenton introduced an asymptotic formula for the third central moment of a maximum likelihood estimator $\hat{\theta}_\alpha$ of the parameter θ_α, $a = 1, 2, \ldots, s$. From this moment, the asymptotic skewness can be set up using the standard deviation. Clearly, the skewness, measured in this way is location free, and scale free, so that shape is accounted for. The computer program is implemented by insertion of the values of expectations of products of logarithmic derivatives, a tiresome task. But now using Maple, the only input consists of the values of the parameters and the form of the density or probability function. Cases of up to four parameters have been implemented. However, in this paper we present two- and three-parameter cases in detail. Future improvements in handling Maple may lead to the implementation of the general case. Bowman and Shenton [Bowman, K.O. and Shenton, L.R., 1999, The asymptotic kurtosis for maximum likelihood estimators. *Communications in Statistics, Theory and Methods*, **28**(11), 2641–2654.] also developed an asymptotic formula for the kurtosis, which is not used here. This study was initiated in our monograph [Shenton, L.R. and Bowman, K.O., 1977, *Maximum Likelihood Estimation in Small Samples* (Charles Griffin and Co., Ltd.)].

Keywords: Binomial mixtures; Covariance matrix; Hessian matrix; Hybrid mixtures; Logarithmic derivatives; Poisson mixtures

1. Introduction

Given a density or probability function $f(x; \underline{\theta})$, depending on s parameters $\theta_1, \theta_2, \ldots, \theta_s$, we have given [1] a formula for the asymptotic skewness for the maximum likelihood estimator $\hat{\theta}_i$, of θ_i, $i = 1, 2, \ldots, s$. Briefly, the denominator is based on the maximum likelihood covariance matrix which itself is the inverse of the Hessian matrix.

*Corresponding author. Email: bowmanko@ornl.gov

†This submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00949650412331321142

The Hessian matrix leading to the maximum likelihood covariances is

$$\Delta_s = \begin{bmatrix} [\theta_1,\theta_1] & [\theta_1,\theta_2] & \cdots & [\theta_1,\theta_s] \\ [\theta_2,\theta_1] & [\theta_2,\theta_2] & \cdots & [\theta_2,\theta_s] \\ \vdots & \vdots & \vdots & \vdots \\ [\theta_s,\theta_1] & [\theta_s,\theta_2] & \cdots & [\theta_s,\theta_s] \end{bmatrix}, \tag{1}$$

where in expectations

$$[\theta_i,\theta_j] = E\left\{\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_i}\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_j}\right\} \quad (i, j = 1, 2, \ldots, s),$$

the derivatives being assumed to exist. For the case when $f(x;\underline{\theta}\,)$ is a continuous probability density, this becomes

$$[\theta_i,\theta_j] = \int_{-\infty}^{\infty}\left\{\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_i}\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_j}\right\} f(x;\underline{\theta}\,)\,\mathrm{d}x \quad (i, j = 1, 2, \ldots, s),$$

existence being assumed, *i.e.*, convergence of the integral. For the discrete case, the support being $x = 0, 1, \ldots,$

$$[\theta_i,\theta_j] = \sum_{x=0}^{\infty}\left\{\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_i}\frac{\partial \ln f(x;\underline{\theta}\,)}{\partial\theta_j}\right\} f(x;\underline{\theta}\,),$$

the summation being assumed to be convergent. Usually a stopping rule, based on the value of the last term or terms, is involved. Asymptotic covariances are set up from the inverse of the matrix (1); thus

$$\mathrm{Cov}_1(\hat{\theta}_j,\hat{\theta}_h) = \frac{\Delta_{jh}}{|\Delta_s|},$$

where $\Delta_{j,h}$ is the minor of the element in the jth row and hth column, $|\Delta_s|$ indicating the determinant of the matrix Δ_s. For example, when $s = 3$,

$$\Delta_3 = \begin{bmatrix} [\theta_1,\theta_1] & [\theta_1,\theta_2] & [\theta_1,\theta_3] \\ [\theta_2,\theta_1] & [\theta_2,\theta_2] & [\theta_2,\theta_3] \\ [\theta_3,\theta_1] & [\theta_3,\theta_2] & [\theta_3,\theta_3] \end{bmatrix},$$

so

$$\mathrm{Var}_1(\hat{\theta}_1) = \frac{\begin{bmatrix} [\theta_2,\theta_2] & [\theta_2,\theta_3] \\ [\theta_3,\theta_2] & [\theta_3,\theta_3] \end{bmatrix}}{|\Delta_3|},$$

$$\mathrm{Var}_1(\hat{\theta}_2) = \frac{\begin{bmatrix} [\theta_1,\theta_1] & [\theta_1,\theta_3] \\ [\theta_3,\theta_1] & [\theta_3,\theta_3] \end{bmatrix}}{|\Delta_3|}.$$

The subscript 1 in $\mathrm{Var}_1(\cdot)$ indicates the coefficient of the N^{-1} term, N being sample size.

The skewness formula uses a summatory notation for the N^{-2} coefficient in the third central moment of $\hat{\theta}_\alpha$, a being $1, 2, \ldots, s$, and $\alpha, \beta, \gamma = 1, 2, \ldots, s$, namely

$$\mu_{32}(\hat{\theta}_a) = L^{a\alpha}L^{a\beta}L^{a\gamma}([\alpha,\beta,\gamma] + 3[\alpha\beta\gamma] + 6[\alpha\beta,\gamma]), \tag{2}$$

where

- $f_x = f(x; \underline{\theta})$ is the probability function or density and $\underline{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_s)$.
- $L_{\alpha\beta} = E\{(\partial \ln f_x/\partial\theta_\alpha) \cdot (\partial \ln f_x/\partial\theta_\beta)\} = \sum(1/f_x)(\partial f_x/\partial\theta_\alpha)(\partial f_x/\partial\theta_\beta)$. $L^{\beta\alpha}$ is an element in the covariance matrix, the N^{-1} factor omitted.
- $[\alpha, \beta, \gamma] = E\{(\partial \ln f_x/\partial\theta_\alpha)(\partial \ln f_x/\partial\theta_\beta)(\partial \ln f_x/\partial\theta_\gamma)\}$, $[\alpha\beta\gamma] = E\{\partial^3 \ln f_x/(\partial\theta_\alpha\partial\theta_\beta \partial\theta_\gamma)\}$, $[\alpha\beta, \gamma] = E[\{\partial^2 \ln f_x/(\partial\theta_\alpha\partial\theta_\beta)\}(\partial \ln f_x/\partial\theta_\gamma)]$. It is assumed that a set of logarithmic derivatives exist. The above expressions will be referred to as 'square bracket' terms.
- A non-singular covariance matrix is assumed. It will involve $s(s+1)/2$ covariances.

For the asymptotic skewness, we have

$$\sqrt{\beta_1(\hat{\theta}_a)} \sim \frac{\mu_{32}(\hat{\theta}_a)/(L^{aa})^{3/2}}{\sqrt{N}} = \frac{\sqrt{\beta_{11}(\hat{\theta}_a)}}{\sqrt{N}}, \quad (N \longrightarrow \infty, a = 1, 2, \ldots, s) \tag{3}$$

The implications of the summation formula in equation (2) for the asymptotic third central moment of an estimator $\hat{\theta}_a$ are summarized in table 1 for the case of two parameters θ_1 and θ_2. θ_1 and θ_2 are represented by 1, and 2, respectively. Permutations are valid in columns 3 and 5, of table 1, but not completely in $[\alpha\beta, \gamma]$, since $[\alpha\gamma, \beta]$ may not equal $[\alpha\beta, \gamma]$, but $[\alpha\beta, \gamma] = [\beta\alpha, \gamma]$.

Previously, we have evaluated numerous terms in equation (2) and derived the asymptotic skewness and lower order moments for the three-parameter gamma density [2], and the three-parameter Weibull density [3], the densities being

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

$$\text{Weibull: } f(x; a, b, c) = \frac{c}{b} y^{c-1} \mathrm{e}^{-y^c} \quad \left(y = \frac{x-a}{b}, \ \ x > 0, \ \ b > 0\right).$$

Using a Maple program, it is now possible to avoid evaluating the derivatives in equation (2) and merely specifying the density or probability function. The density or probability function must have derivatives with respect to the parameters which exist in a certain domain.

Here, we describe applications to the two-component Poisson mixture distribution and the two-component hybrid mixture (Poisson and binomial) distribution. It is gratifying to note that the abbreviated Maple approach checks out for the gamma density [two cases, (1) s is known, (2) s is unknown], and the Weibull density.

The asymptotic skewness $\sqrt{\beta_1(\hat{\theta}_a)}$ is chosen as a guide to sample size because skewness measured by moments $(\mu_3/\mu_2^{3/2})$ is location free and scale free; bias and variance do not in

Table 1. Formula (2), θ_a, $a = 1$ or 2.

α	β	γ	$L^{a\alpha}L^{a\beta}L^{a\gamma}$	$[\alpha\beta\gamma]$	$[\alpha\beta, \gamma]$	$[\alpha, \beta, \gamma]$
1	1	1	$(L^{a1})^3$	[111]	[11, 1]	[1, 1, 1]
1	1	2	$(L^{a1})^2L^{a2}$	[112]	[11, 2]	[1, 1, 2]
1	2	1	$(L^{a1})^2L^{a2}$	[121]	[12, 1]	[1, 2, 1]
2	1	1	$(L^{a1})^2L^{a2}$	[211]	[12, 1]	[2, 1, 1]
1	2	2	$L^{a1}(L^{a2})^2$	[122]	[12, 2]	[1, 2, 2]
2	1	2	$L^{a1}(L^{a2})^2$	[212]	[21, 2]	[2, 1, 2]
2	2	1	$L^{a1}(L^{a2})^2$	[221]	[22, 1]	[2, 2, 1]
2	2	2	$(L^{a2})^3$	[222]	[22, 2]	[2, 2, 2]

generally represent the shape of a distribution. If we set $\sqrt{\beta_1(\hat{\theta}_a)}$ to be a quantity ϵ, then sample size is

$$N^*(\hat{\theta}_a, \epsilon) = \frac{\left\{\sqrt{\beta_{11}(\hat{\theta}_a)}\right\}^2}{\epsilon^2},$$

and we could take $1 \le \epsilon \le 2$, or more demanding $\epsilon = 0.1$.

An application of the Maple system to sets of data on sister chromatid exchanges on blood lymphocytes is briefly considered.

We first of all take the case of the two-component Poisson mixture distribution.

2. The two component Poisson mixture distribution

Probability function:

$$P(x; \underline{\theta}, \underline{\pi}) = \pi_1 \frac{e^{-\theta_1}\theta_1^x}{x!} + (1-\pi_1)\frac{e^{-\theta_2}\theta_2^x}{x!} \quad (0 < \pi_1 < 1,\ \ \theta_1 > 0,\ \ \theta_2 > 0,\ \ x = 0, 1, \ldots).$$

Some examples of variance and skewness are given in table 2.

Comments to table 2:

- In cases 1 and 2, note the improvement in the moments as the difference $|\theta_1 - \theta_2|$ increases from 1 to 5. Note also the next case in which the difference narrows again to unity with a consequent increase in moment values. A similar phenomena is evident in cases 8 and 9.
- Case 10 shows the decrease in most values when the difference $|\theta_1 - \theta_2| = 10$.
- Our tabulations are all for $\theta_2 > \theta_1$. To produce the moment for $\theta_1 > \theta_2$, we must use the transformations $\theta_1 \leftrightarrow \theta_2$ and $\pi_1 \leftrightarrow \pi_2$. Thus for the case $\theta_1 = 2, \theta_2 = 1, \pi_1 = 0.2$, we look up in our table for $\theta_1 = 1, \theta_2 = 2, \pi_1 = 0.8$. Thus, this explains the sign change.
- When $\pi_1 = 0.2$, there is emphasis on the second component for which the proportion is 0.8. This suggests the observed fact that $\sqrt{\beta_{11}(\hat{\pi}_1)}$ is positive. When $\pi_1 = 0.8$, it follows that negative values of $\sqrt{\beta_{11}(\hat{\pi}_1)}$ are to be expected. Note that we are considering $E[\hat{\pi}_1 - E(\hat{\pi}_1)]^3$.

Table 2. Two-component Poisson mixture distribution.

				Variance			$\sqrt{\beta_1}$		
Case	θ_1	θ_2	π_1	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\pi}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\pi}_1$
1	1.0	2.0	0.2	822.97	77.30	154.68	−23.19	147.56	193.02
2	1.0	6.0	0.2	15.37	11.22	0.34	5.60	0.97	3.42
3	5.0	6.0	0.2	29,604.27	2,067.12	4,935.37	−224.15	975.35	1,155.20
4	1.0	2.0	0.5	116.81	148.51	125.01	−46.44	58.00	1.32
5	3.0	4.0	0.5	1,262.71	1,340.23	1,269.85	−183.69	196.81	3.21
6	5.0	6.0	0.5	4,588.84	4,713.37	4,579.49	−369.60	382.24	3.51
7	1.0	2.0	0.8	34.75	538.54	82.20	−95.61	36.73	−134.70
8	3.0	4.0	0.8	415.37	6,039.41	987.91	−403.51	111.53	−498.42
9	3.0	9.0	0.8	6.90	112.78	0.40	0.25	0.38	−4.44
10	5.0	6.0	0.8	1,576.69	23,325.07	3,819.66	−841.21	215.89	−1,004.06
11	5.0	15.0	0.8	8.38	123.70	0.22	0.81	−0.27	−1.99

Table 3. Skewness for the hybrid and corresponding Poisson–Poisson distribution.

π_1		$\sqrt{\beta_{11}(\hat{\theta})}$	$\sqrt{\beta_{11}(\hat{p})}$	$\sqrt{\beta_{11}(\hat{\pi}_1)}$
0.2	Hybrid	28.84	−89.37	132.71
	P.P. ($\theta_1 = 2, \theta_2 = 1$)	36.73	−95.61	−134.70
0.5	Hybrid	49.57	−50.92	8.10
	P.P.	58.00	−46.44	1.32
0.8	Hybrid	130.40	−46.05	−161.04
	P.P.	147.56	−23.19	193.02

3. Asymptotic skewness for the two-component hybrid case, Poisson and binomial mixture distribution

Probability function:

$$P(x; \theta, p, \pi_1|n) = \pi_1 \frac{\mathrm{e}^{-\theta}\theta^x}{x!} + (1 - \pi_1) \binom{n}{x} p^x (1 - p)^{n-x},$$
$$(\theta > 0, \quad 0 \le p \le 1, \quad 0 < \pi_1 < 1, \quad n = 1, 2, \ldots).$$

We chose a case for which the binomial component is approximately a Poisson distribution, *i.e.*, $n = 50$, $p = 0.2, \theta = 2$. For three values of the proportion parameter, π_1 comparisons with the corresponding two-component Poisson–Poisson mixture distribution are given in table 3.

Comments to table 3:

- The parameters are chosen so as to produce a close comparison between hybrid and Poisson–Poisson mixture distribution. Thus, n is fairly large, p is fairly small.
- The agreement between the two assessments is quite satisfactory but there is a sign change for $\sqrt{\beta_{11}(\hat{\pi}_1)}$.

The asymptotic variances of the hybrid Poisson and binomial mixture distribution are of less interest, but have been studied in Bowman and Shenton [4].

In our paper [4], evidence was provided that the low-order asymptotic moments of maximum likelihood estimators for the two-component hybrid case would exist even when the components were assigned close values; for example, $\theta = 1.0$, $p = 0.1, n = 10$, for which when $\pi_1 = 0.5$, for example, we have $\sqrt{\beta_{11}(\hat{\theta})} = 2568$, $\sqrt{\beta_{11}(\hat{p})} = 1233$, and $\sqrt{\beta_{11}(\hat{\pi}_1)} = 48.8$.

4. An application to the four parameter case

Byers and Shenton [5] considered a data set concerning sister chromatid exchange and the possible effect of smoking on individuals involved. There were four categories, female non-smokers, male non-smokers, female smokers, and male smokers. A model used to fit the

Table 4. Standard errors and skewness of male non-smokers, $n = 8300$.

Parameters	$\hat{\lambda}_1$	$\hat{\theta}$	$\hat{\lambda}_2$	$\hat{\pi}_1$
ml Estimators	7.1005	0.0272	12.5499	0.9213
Asymptotic standard error	0.0778	0.0208	0.6822	0.0292
Asymptotic skewness	0.0004	0.8327	1.7450	0.7144

discrete data was a mixture of Poisson–Poisson and Poisson distributions, with probability function

$$P(x; \lambda_1, \theta, \lambda_2, \pi_1) = \frac{\pi_1 \lambda_1 (\lambda_1 + \theta_x)^{x-1} \mathrm{e}^{-(\lambda_1+\theta_x)}}{x!} + \frac{(1-\pi_1)\mathrm{e}^{-\lambda_2}\lambda_2^x}{x!}$$

for $x = 0, 1, \ldots, 0 < \pi_1 < 1, 0 \leq \theta < 1, \lambda_1 > 0, \lambda_2 > 0$. The first component is the Poisson–Poisson distribution, and this reduces to a Poisson distribution when $\theta = 0$.

In a preliminary study of asymptotic skewness using the Maple system, we give the following moments for maximum likelihood estimators for the sub-group male non-smokers, for which the sample size $n = 8300$, and there being four parameters (table 4).

Comments to table 4:

- The standard errors given here were computed by Maple, and in substance agree with those given by Byers and Shenton [5] using the Splus program. To interpret standard errors, keep in mind the associated mean. Thus, $\sigma(\hat{\lambda}_1)$ relates to the mean of $\hat{\lambda}_1$, which is 7.1005; similarly for $\sigma(\hat{\lambda}_2)$. But means of $\hat{\theta}$ and $\hat{\pi}_1$ are small in value.
- The asymptotic skewness for $\hat{\lambda}_1$ is small so that asymptotic normality might be suggested; a small skewness may imply a small kurtosis because of inherent correlation. Even with a sample size of 8300, a similar conclusion is not obvious for the $\hat{\theta}$, $\hat{\lambda}_2$ and $\hat{\pi}_1$ distributions.
- The four data sets are given in Bowman *et al.* [6], table 1.
- The maximum likelihood solutions, model fits, and other aspects including Splus assessments are given in Byers and Shenton [5].

5. The sample size and almost normality for maximum likelihood estimator

5.1 *Sample size*

In 1922, Fisher [7] introduced the notion of efficient estimator; this considered the ratio '(Variance using maximum likelihood estimator)/(Variance of other consistent estimator)'. A consistent estimator of a parameter τ, say $t(x_1, x_2, \ldots, x_n)$, is such that

$$E\{t(x_1, x_2, \ldots, x_n)\} = \tau + \tau_1/n + \cdots, \quad (n \longrightarrow \infty).$$

Fisher claimed in his 1922 paper that most moment methods were inefficient, information thereby lost. Fisher over-looked the question of sample size and admitted his error in the 1950s [8].

We are not making the same mistake with respect to our asymptotic skewness formula for maximum likelihood processes. $\sqrt{\beta_{11}(\hat{\theta})}$ is the coefficient of $N^{-1/2}$ in the real skewness. It is a statistic based on sample moments. Our suggestion is that in a sustained study of 'bell shaped' data sets, the skewness should be calculated in pilot studies to guide the researcher with the choice of sample size.

Table 5. Sample size for pseudo-normality, Poisson–Poisson mixture of two components, three parameters.

θ_1	θ_2	π_1	$N^*(\hat{\theta}_1, 0.1)$	$N^*(\hat{\theta}_2, 0.1)$	$N^*(\hat{\pi}_1, 0.1)$
1.0	2.0	0.2	5.38E+04	2.18E+06	3.73E+06
		0.5	2.16E+05	3.36E+05	1.74E+02
		0.8	9.14E+05	1.35E+05	1.81E+06
1.0	6.0	0.2	3.14E+03	9.41E+01	1.17E+03

Under regularity conditions, maximum likelihood estimators are asymptotically normal [9]. One step towards this state is to make the skewness small in values; our interpretation of small here is $|\sqrt{\beta}_1| = 1/10$. The sample sizes are shown in tables 5 and 6.

Comment to table 5: Several million are required when $\theta_2 - \theta_1 = 1$ but the number is significantly reduced for $\pi_1 = 0.2$ when $\theta_2 - \theta_1 = 5$.

Comment to table 6: Samples as large as several millions are indicated. Note the entries can be rescaled to apply to other scales of ϵ; for example, if $\epsilon = 1$, the estimates are divided by 100.

Note, however, that Geary [10] showed that there is a minimum property of the determinant of the covariance matrix which is asymptotic and holds for maximum likelihood methods.

5.2 *Further examples*

Aside from cases mentioned in this paper and references, we could think of Poisson-negative binomial, binomial-negative binomial, $\chi_1^2 - \chi_2^2$ (χ^2 with different degrees of freedom), lognormal (three parameters) distributions and others. The three volumes on distributions by Johnson *et al.* [11–13] provide many more examples.

5.3 *Programs*

5.3.1 Introduction of Maple. In this section, two Maple programs, two- and three-parameter cases are listed. Users must supply the distribution as a (pf). The parameters are t1 and t2 for the two-parameter case, t1, t2, and t3 for the three-parameter case.

Variables D1, D2, and D3 are first, second, and third derivatives with respect to parameters.

After the third derivatives are derived, user must supply the maximum likelihood estimator values for the parameters and compute expectation of 'squared bracket' terms. For the continuous distribution, integration is used, and for discrete distribution summation is used over the range of x.

Table 6. Sample sizes for pseudo-normality, two-component hybrid (Poisson binomial).

θ_1	p	n	π_1	$N^*(\hat{\theta}_1, 0.1)$	$N^*(\hat{\theta}_2, 0.1)$	$N^*(\hat{\pi}_1, 0.1)$
2.0	0.02	50	0.2	8.32E+04	7.99E+05	1.76E+06
			0.5	2.46E+05	2.59E+05	6.56E+03
			0.8	1.70E+06	2.12E+05	2.59E+06

The matrix (L) is the covariance matrix, for example L_{11} is the N^{-1} term of the variance of the $\hat{\theta}_1$.

U3t1 and U3t2 are N^{-2} terms of the μ_3 of $\hat{\theta}_1, \hat{\theta}_2$, and so on.

5.3.2 Program 1. This is the program for the maximum likelihood estimators of the two-parameter gamma distribution (continuous density), $\hat{a}$ and $\hat{\rho}$. Suppose the estimators were calculated by a standard statistical program package. This program will compute the N^{-1} terms of variances and $N^{-1/2}$ terms of skewness.

```
#Find skewness of ml estimators of 2 parameter gamma dist
#Reference paper is Bowman and Shenton [1].

with(linalg);

# pf = Gamma probability function
pf := exp(−x/t1)∗xˆ(t2-1)/(t1ˆt2∗GAMMA(t2));
LL := log(pf);
# W = 3 × (2 × 2 × 2), need 1st to 3rd order derivatives,
#   2 × 2 × 2 represents
#  all possible combination of derivatives of 2 parameter
#   case.
W := array (1..3,1..8,
  [[1,1,1,1,2,2,2,2],[1,1,2,2,1,1,2,2],[1,2,1,2,1,2,1,2]]);

D1 := simplify(vector(2,[diff(LL,t1),diff(LL,t2)]));
for i from 1 to 2 do
    D2[2∗i-1] := diff(D1[i],t1);    D2[2∗i] := diff(D1[i],t2);
    D11[2∗i-1] := D1[i]∗D1[1];      D11[2∗i] := D1[i]∗D1[2]; od;
for i from 1 to 4 do
    D3[2∗i-1] := simplify(diff(D2[i],t1));
    D3[2∗i] := simplify(diff(D2[i],t2)); od;
# Having completed all the symbolic operations.

# Substitute the parameter values for t1 and t2.
a := 1; r := 2; f := subs(t1=a,t2=r,pf);
for i from 1 to 2 do
    d1[i] := subs(t1=a,t2=r,D1[i]); od;

# Substitute the parameter values for t1 and t2; and take
#  expectation by
# integration (continuos distribution).
for i from 1 to 4 do
   d2[i] := subs(t1=a,t2=r,D2[i]); f11[i] := subs(t1=a,
   t2=r,D11[i]);
   ef11[i] := evalf(int(f∗f11[i],x=0..infinity));
   f2[i] := evalf(int(f∗d2[i],x=0..infinity)); od;
for i from 1 to 8 do
   d3[i] := subs(t1=a,t2=r,D3[i]);
   f3[i] := evalf(int(f∗d3[i],x=0..infinity)); od;
for i from 1 to 4 do
```

```
    f21[2*i-1] := evalf(int(f*d2[i]*d1[1],x=0..infinity));
    f21[2*i] := evalf(int(f*d2[i]*d1[2],x=0..infinity));
    f111[2*i-1] := evalf(int(f*f11[i]*d1[1],x=0..infinity));
    f111[2*i] := evalf(int(f*f11[i]*d1[2],x=0..infinity)); od;

# Compose Hesian matrix
H :=matrix(2,2,[[ef11[1],ef11[2]],[ef11[3],ef11[4]]]);

# Find covariance matrix by inverting H
L := inverse(H);

# Compute the value of equation (2).
for i from 1 to 8 do
    A[i] := f111[i]+3*f3[i]+6*f21[i];
    B[i] := evalf(L[1,W[1,i]]*L[1,W[2,i]]*L[1,W[3,i]]);
    C[i] := evalf(L[2,W[1,i]]*L[2,W[2,i]]*L[2,W[3,i]]); od;
    AA := vector(8,[A[1],A[2],A[3],A[4],A[5],A[6],A[7],A[8]]);
    BB := vector(8,[B[1],B[2],B[3],B[4],B[5],B[6],B[7],B[8]]);
    CC := vector(8,[C[1],C[2],C[3],C[4],C[5],C[6],C[7],C[8]]);
    BBT := transpose(BB);  CCT := transpose(CC);

    U3t1 := evalf(multiply(BBT,AA)); U3t2 := evalf(multiply
    (CCT,AA));

# Compute standardized skewness values.
rb1a := U3t1/L[1,1]^(3/2);  rb1rho := U3t2/L[2,2]^(3/2);
```

5.3.3 Program 2. This is the program for the maximum likelihood estimators of the three-parameter Poisson mixture distribution (discrete distribution), $\hat{\theta}_1$, $\hat{\theta}_2$, and π_1. Suppose that the estimators were calculated by standard statistic program package as Program 1. As the number of parameters increase, the requirement of a large sample size is expected. Note the D3 is computed differently from Program 1. There are 27 different combinations of derivatives; however, only 17 are a distinct set of combinations of three parameters. Therefore, it will save a lot of time and space to compute only those distinctive sets. It will be more so when the program is extended to four parameters. There will be 64 combinations of derivatives and it may be important to save time and space of computing.

```
#Find skewness of ml estimators of 3 parameter mixture
# distribution
#Reference paper is Bowman and Shenton [2].
#parameters of mixture of a Poisson-Poisson is t1, t2 and t3
with(linalg);
# pf = 3 parameter mixture distribution (discrete
#  distribution).
pf := t3*exp(-t1)*t1^x/x!+(1-t3)*exp(-t2)*t2^x/x!;
LL := log(pf);
# W =3 × (3 × 3 × 3), need 1st to 3rd order derivatives
#   3 × 3 × 3 represents
# all possible combination of derivatives of 3 parameter
#  case.
```

```
W :=array(1..3,1..27,
   [[1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3],
    [1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3],
    [1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3]]);
D1 := vector(3,[simplify(diff(LL,t1)),simplify(diff(LL,t2)),
simplify(diff(LL,t3))]);
for i from 1 to 3 do
   D2[3*i-2] := simplify(diff(D1[i],t1));
   D2[3*i-1] := simplify(diff(D1[i],t2));
   D2[3*i] := simplify(diff(D1[i],t3));
   d11[3*i-2] := D1[i]*D1[1];
   d11[3*i-1] := D1[i]*D1[2];
   d11[3*i] := D1[i]*D1[3]; od;
i := 1;   D3[3*i-2] := simplify(diff(D2[i],t1));
          D3[3*i-1] := simplify(diff(D2[i],t2));
          D3[3*i] := simplify(diff(D2[i],t3));
i := 2;   D3[3*i-2] := D3[2];
          D3[3*i-1] := simplify(diff(D2[i],t2));
          D3[3*i] :=simplify(diff(D2[i],t3));
i := 3;   D3[3*i-2] := D3[3];D3[3*i-1] := D3[6];
          D3[3*i] := simplify(diff(D2[i],t3));
i := 4;   D3[3*i-2] := D3[2];D3[3*i-1] := D3[5];D3[3*i] := D3[6];
i := 5;   D3[3*i-2] := D3[5];
          D3[3*i-1] := simplify(diff(D2[i],t2));
          D3[3*i] := simplify(diff(D2[i],t3));
i := 6;   D3[3*i-2] := D3[6];D3[3*i-1] := D3[15];
          D3[3*i] := simplify(diff(D2[i],t3));
i := 7;   D3[3*i-2] := D3[3];D3[3*i-1] := D3[6];D3[3*i] := D3[9];
i := 8;   D3[3*i-2] := D3[6];D3[3*i-1] := D3[15];D3[3*i] := D3[18];
i := 9;   D3[3*i-2] := D3[9];D3[3*i-1] := D3[18];
          D3[3*i] := simplify(diff(D2[i],t3));
# Having completed all the symbolic operations.

# Substitute the parameter values for t1, t2, and t3.
a := 5; r := 6; s := 0.5; f := subs(t1=a,t2=r,t3=s, pf);
  lim := r*10;

for i from 1 to 3 do
   d1[i] := subs(t1=a,t2=r,t3=s,D1[i]); od;
# Substitute the parameter values for t1, t2, and t3; and
#  take expection by
# sum over the region from 0 to limit (lim). (discrete
#  distribution).
for i from 1 to 9 do
   d2[i] := subs(t1=a,t2=r,t3=s,D2[i]);
   f11[i] := subs(t1=a,t2=r,t3=s,d11[i]);
   ef11[i] := evalf(sum(f*f11[i],m=0..lim));
   f2[i] := evalf(sum(f*d2[i],m=0..lim)); od;
for i from 1 to 27 do
   d3[i] := subs(t1=a,t2=r,t3=s,D3[i]);
   f3[i] := evalf(sum(f*d3[i],m=0..lim)); od;
```

```
for i from 1 to 9 do
   f21[3*i-2] := evalf(sum(f*d2[i]*d1[1],m=0..lim));
   f21[3*i-1] := evalf(sum(f*d2[i]*d1[2],m=0..lim));
   f21[3*i] := evalf(sum(f*d2[i]*d1[3],m=0..lim));
   f111[3*i-2] := evalf(sum(f*f11[i]*d1[1],m=0..lim));
   f111[3*i-1] := evalf(sum(f*f11[i]*d1[2],m=0..lim));
   f111[3*i] := evalf(sum(f*f11[i]*d1[3],m=0..lim)); od;

# Compose Hesian Matrix.
H := matrix(3,3,[[ef11[1],ef11[2],ef11[3]],
[ef11[4],ef11[5],ef11[6]],
[ef11[7],ef11[8],ef11[9]]]);

# Find covariance matrix by inverting H.
L := inverse(H);

# Compute the value of equation (2).
for i from 1 to 27 do
   A[i] := f111[i]+3*f3[i]+6*f21[i];
   B[i] := L[1,W[1,i]]*L[1,W[2,i]]*L[1,W[3,i]];
   C[i] := L[2,W[1,i]]*L[2,W[2,i]]*L[2,W[3,i]];
   E[i] := L[3,W[1,i]]*L[3,W[2,i]]*L[3,W[3,i]]; od;
AA := vector(27,[A[1],A[2],A[3],A[4],A[5],A[6],A[7],A[8],A[9],
       A[10],A[11],A[12],A[13],A[14],A[15],A[16],A[17],A[18],
A[19],A[20],A[21],A[22],A[23],A[24],A[25],A[26],A[27]]);
BB := vector(27,[B[1],B[2],B[3],B[4],B[5],B[6],B[7],B[8],B[9],
B[10],B[11],B[12],B[13],B[14],B[15],B[16],B[17],B[18],
B[19],B[20],B[21],B[22],B[23],B[24],B[25],B[26],B[27]]);
CC := vector(27,[C[1],C[2],C[3],C[4],C[5],C[6],C[7],C[8],C[9],
       C[10],C[11],C[12],C[13],C[14],C[15],C[16],C[17],C[18],
C[19],C[20],C[21],C[22],C[23],C[24],C[25],C[26],C[27]]);
EE := vector(27,[E[1],E[2],E[3],E[4],E[5],E[6],E[7],E[8],E[9],
       E[10],E[11],E[12],E[13],E[14],E[15],E[16],E[17],E[18],
E[19],E[20],E[21],E[22],E[23],E[24],E[25],E[26],E[27]]);
BBT := transpose(BB);
CCT := transpose(CC);
EET := transpose(EE);

 U3t1 := evalf(multiply(BBT,AA)); U3t2 := evalf(multiply(CCT,AA));
 U3t3 := evalf(multiply(EET,AA));

# Compute standardized skewness values for t1, t2, and t3.
rb1t1 := U3t1/L[1,1]^(3/2); rb1t2 := U3t2/L[2,2]^(3/2);
rb1t3 := U3t3/L[3,3]^(3/2);
```

For details of programming, see Heck [14].

6. Conclusion

We have given a simple formula for the third standardized central moment of a maximum likelihood estimator where s parameters are involved; this moment is the asymptotic skewness. In general, when modeling a ‘bell shaped’ data base, skewness is more important than kurtosis, and for the normal distribution, the skewness is zero. Therefore, reducing the asymptotic skewness in value to a small quantity enables one to set up a simple size which would provide pseudo-normality.

For mixture distributions, closeness of the components obviously leads to greater variances, which indeed might approach infinity. The Maple version of the asymptotic skewness leads to a measure of sample size; it does not completely solve the sample size problem but it does supply additional information. It may be mentioned that we have also given a complicated formula for the asymptotic kurtosis in the form

$$\beta_2 \sim 3 + \frac{K(\hat{\theta})}{n},$$

n the sample size.

Acknowledgement

The research of first author 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.

References

[1] Bowman, K.O. and Shenton, L.R., 1998, Asymptotic skewness and the distribution of maximum likelihood estimators. *Communications in Statistics, Theory and Methods*, **27**(11), 2743–2760.
[2] Bowman, K.O. and Shenton, L.R., 2002, Problems with maximum likelihood estimation and the 3 parameter gamma distribution. *Journal of Statistical Computation and Simulation*, **72**(5), 391–401.
[3] Bowman, K.O. and Shenton, L.R., 2000, Maximum likelihood and the Weibull distribution. *Far East Journal of Theoretical Statistics*, **4**(2), 391–422.
[4] Bowman, K.O. and Shenton, L.R., 2004, Mixtures, hybrid mixtures, canonical forms, and Matusits's distance. *Far East Journal of Theoretical Statistics*, **12**(1), 69–88.
[5] Byers, R.H., Jr. and Shenton, L.R., 1999, Sister chromatid exchange data fit with a mixture of Poisson distribution. *Mutational Research*, **427**, 157–162.
[6] 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.
[7] Fisher, R.A., 1922, On the mathematical foundations of theoretical statistics. *Philosophical Transactions A*, **222**, 309–368.
[8] Fisher, R.A., 1950, *Contributions to Mathematical Statistics* (John Wiley & Sons) (Authors' comments: On the mathematical foundations of theoretical statistics. *Philosophical Transaction A*, **222**, 309–368, 1992.
[9] Fisher, R.A., 1934, Two new properties of mathematical likelihood. *Proceedings of Royal Society, Series A*, **144**, 285–307.
[10] Geary, R.C., 1942, The estimation of many parameters, *Journal of Royal Statistical Society*, **LV**, 213–217.
[11] Johnson, N.L., Kotz, S. and Kemp, A.W., 1993, *Univariate Discrete Distributions* (2nd edn) (New York: John Wiley & Sons).
[12] Johnson, N.L., Kotz, S. and Balakrishnan, N., 1994, *Continuous Univariate Distributions* (2nd edn), Vol. 1 (New York: John Wiley & Sons).
[13] Johnson, N.L., Kotz, S. and Balakrishnan, N., 1995, *Continuous Univariate Distributions* (2nd edn), Vol. 2 (New York: John Wiley & Sons).
[14] Heck, A., 1997, *Introduction to Maple* (2nd edn) (New York: Springer-Verlag).
[15] Bowman, K.O. and Shenton, L.R., 1999, The asymptotic kurtosis for maximum likelihood estimators. *Communications in Statistics, Theory and Methods*, **28**(11), 2641–2654.
[16] Shenton, L.R. and Bowman, K.O., 1977, *Maximum Likelihood Estimation in Small Samples* (Charles Griffin and Co., Ltd.).