Far East J.Theo.Stat., 20(1), (2006), 73-95.

BINOMIAL AND POISSON MIXTURES, MAXIMUM LIKELIHOOD, AND MAPLE CODE*

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

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, P.O.Box 2008, 4500N, MS-6191, Oak Ridge, TN 37831-6191, USA, bowmanko@ornl.gov

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

May 10, 2006

Abstract

The bias, variance, and skewness of maximum likelihood estimators are considered for binomial and Poisson mixture distributions. The moments considered are asymptotic, and they are assessed using the Maple code. Question of existence of solutions and Karl Pearson's study are mentioned, along with the problems of valid sample space. Large samples to reduce variances are not unusual; this also applies to the size of the asymptotice skewness.

Keywords and phrases: Asymptotic skewness, asymptotic variance, kurtosis, negative binomial distribution, probability functions, ratio statistics for skewness.

*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 nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

1 Introduction

Our recent papers (Bowman and Shenton, 2005a, 2005b, 2006) have considered basic properties of maximum likelihood estimators with respect to low order moments. Given a probability function involving s parameters we derive values for the asymptotic variances derived from the Hessian matrix, and also the skewness in the form of the moment ratio $\mu_3/\mu_2^{3/2}$ (third central moment/$s.d.^3$).

In the present case we concentrate on Poisson (or binomial) mixtures consisting of s components and $s-1$ proportions. Thus

$$P(X=x)=\sum_{\lambda=1}^{s}\pi_\lambda P_\lambda(x,\theta_\lambda)=P(x)=P, \tag{1}$$

where $x=0,1,\cdots$, $\theta_\lambda>0$, $0<\pi_\lambda<1$, $\lambda=1,2,\cdots,s$. In particular we shall consider $3pP$ (3 parameter Poisson), $5pP$, and $7pP$ using the Maple code (see Bowman and Shenton, 2005b) for computational numerical value, and symbolic for algebraic results. Several examples for the discrete case are given in Bowman and Shenton (2006).

2 Maximum likelihood equations for Poisson mixture

The Poisson mixture probability function for a random variable X is given in (1) with individual component being

$$P_\lambda(x,\theta_\lambda)=e^{-\theta_\lambda}\theta_\lambda^x/x!. \quad (x=0,1,\cdots,\theta_\lambda>0)$$

Maximum likelihood estimators are $\hat\theta_1$, $\hat\theta_2$, $\cdots$, $\hat\theta_s$; $\hat\pi_1$, $\hat\pi_2$, $\cdots$, $\hat\pi_{s-1}$, these allowing for the constraint $\sum_{i=1}^{s}\pi_i=1$. Note $0<\pi_i<1$.

For a random sample of N from the probability function in (1) n_x is the number of occurrences of the variable x, the range of x being 0 to n. From the likelihood function

$$\frac{\partial L}{\partial\pi_1}=\sum_{x=0}^{n}n_x\frac{(\hat P_1-\hat P_s)}{\hat P}=0 \tag{2}$$

where

$$L=\ln\prod_{x=0}^{n}P(x)^{n_x}.$$

Since finite sums are involved

$$\sum_{x=0}^{n}\frac{n_x}{N}\frac{\hat P_{s-1}}{\hat P}=\sum_{x=0}^{n}\frac{n_x}{N}\frac{\hat P_s}{\hat P}$$

$$\sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_{s-2}}{\hat{P}} = \sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_s}{\hat{P}}$$

$$\vdots \quad \vdots \quad \vdots$$

$$\sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_1}{\hat{P}} = \sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_s}{\hat{P}}$$

so that multiplying each equation by the appropriate π and adding we have

$$\sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_\lambda}{\hat{P}} = 1. \quad (\lambda = 1, 2, \cdots, s-1)$$

For an iterative solution to the maximum likelihood equations, we then have

$$_{new}\pi_\lambda = \pi_\lambda \sum_{x=0}^{n} \frac{n_x}{N} \frac{\tilde{P}_\lambda}{\tilde{P}}. \quad (\lambda = 1, 2, \cdots, s-1)$$

Initial values of $\hat{\theta}_1$, $\hat{\theta}_2$, $\cdots$, $\hat{\theta}_s$; $\hat{\pi}_1$, $\hat{\pi}_2$, $\cdots$, $\hat{\pi}_{s-1}$, are given. To complete the cycle, we have for example,

$$\frac{\partial L}{\partial \theta_1} = \frac{(x-\theta_1)}{\theta_1} \frac{\pi_1 n_x \tilde{P}_1}{\tilde{P}}$$

so that

$$\hat{\theta}_1 = \sum_{x=0}^{n} \frac{n_x}{N} \frac{x\tilde{P}_1}{\tilde{P}}$$

and similarly for $\hat{\theta}_2, \cdots, \hat{\theta}_s$. Symbols like $\tilde{P}_1$ and $\tilde{P}$ refer to the up-dated parameters in the completion of each new approximation.

Example 1: 2 component 3 parameter Poisson mixture, θ_1, θ_2, π_1. For initial estimators use moment estimators. The example considers data due to Bender et al (1992) about sister chromatid exchange (SCE) frequencies, 19,650 people being involved. There were 10,700 males and 8,950 females. The data consists of from 0 to 32 SCE occurrences, so that a discrete frequency model is appropriate. Models such as one parameter Poisson, one parameter binomial, two parameter negative binomial and a mixture of Poisson and Poisson Poisson have been tried, the latter reducing the χ^2 measure. We fit Poisson mixture models of two and three components to the SCE data consisting of 30% smokers, 70% nonsmokers. The moment estimators are

$$\theta_1^* = 7.21, \quad \theta_2^* = 12.36, \quad \pi_1^* = 0.86$$

for 2 component Poisson mixtures (see Bowman and Shenton, 2002).

For maximum likelihood estimators, use moment estimators for the initial values for the cycle

$$_{new}\hat{\pi}_1 = \pi_1 \sum_{x=1}^{n} \frac{n_x}{N} \frac{\tilde{P}_1}{\tilde{P}}, \quad _{new}\hat{\theta}_1 = \sum_{x=1}^{n} \frac{n_x}{N} \frac{x\tilde{P}_1}{\tilde{P}}, \quad _{new}\hat{\theta}_2 = \sum_{x=1}^{n} \frac{n_x}{N} \frac{x\tilde{P}_2}{\tilde{P}}$$

to complete the cycles with desired accuracy.

Similarly for the 5 parameter Poisson mixture,

$$_{new}\hat{\pi}_1 = \pi_1 \sum_{x=1}^{n} \frac{n_x}{N} \frac{\tilde{P}_1}{\tilde{P}}, \quad _{new}\hat{\pi}_2 = \pi_2 \sum_{x=1}^{n} \frac{n_x}{N} \frac{\tilde{P}_2}{\tilde{P}},$$

$$_{new}\hat{\theta}_1 = \sum_{x=1}^{n} \frac{n_x}{N} \frac{x\tilde{P}_1}{\tilde{P}}, \quad _{new}\hat{\theta}_2 = \sum_{x=1}^{n} \frac{n_x}{N} \frac{x\tilde{P}_2}{\tilde{P}}, \quad _{new}\hat{\theta}_3 = \sum_{x=1}^{n} \frac{n_x}{N} \frac{x\tilde{P}_3}{\tilde{P}}$$

to complete the cycles with desired accuracy.

Tabulations of the maximum likelihood moments and Fisher's efficiencies are given in Tables 1 and 2.

Table 1. Poisson mixture parameter estimates

	2 component			3 component		
	moment	m.l.e	mle bias	moment	m.l.e	mle bias
θ_1	7.2062	7.2694	-0.0020	6.0571	6.0511	-0.3535
θ_2	12.3640	12.5778	0.0003	7.9717	7.9001	0.3860
θ_3				13.6276	13.4669	0.1505
π_1	0.8610	0.8769	0.0008	0.2388	0.2245	0.1911
π_2	0.1390	0.1231		0.6890	0.6968	-0.1875
π_3				0.0722	0.0787	0.0032

Table 2. Poisson mixture parameter standard deviation and efficiency

	2 component			3 component		
	moment	m.l.e	efficiency	moment	m.l.e	efficiency
$\sigma(\theta_1)$	0.0575	0.0502	0.76	1.4942	1.1514	0.59
$\sigma(\theta_2)$	0.2820	0.2722	0.93	0.9822	0.6463	0.43
$\sigma(\theta_3)$				0.9330	0.5926	0.40
$\sigma(\pi_1)$	0.0154	0.0127	0.68	0.4717	0.3412	0.52
$\sigma(\pi_2)$	0.0154	0.0127	0.68	0.4401	0.3233	0.54
$\sigma(\pi_3)$				0.0366	0.0228	0.38

Comment: Since the efficiencies are in the interval (0,1) the example lends support to the validity of the Maple code.

3 Variances and skewness

3.1 $2pP$, π_1 known

Details of a few cases are given in Table 3. (excerpt from Bowman and Shenton, 2003a).

Table 3. Moments of $\hat{\theta}_1$ and $\hat{\theta}_2$
2 component Poisson mixture π_1 known

π	θ_1	μ_{11}	μ_{21}	$\sqrt{\beta_{11}}$	θ_2	μ_{11}	μ_{21}	$\sqrt{\beta_{11}}$
0.1	0.5	214.19	161.90	96.03	1.0	-23.79	3.05	-50.30
	1.0	490.84	394.39	142.54	1.5	-54.53	6.41	-93.82
	5.0	1017.71	1648.55	143.62	6.0	-113.07	26.41	-96.74
	10.0	3400.99	5778.52	261.31	11.0	-377.88	82.33	-210.51
0.2	0.5	58.22	39.53	50.12	1.0	-14.55	3.57	-28.26
	1.0	135.16	97.71	75.74	1.5	-33.79	7.67	-53.36
	5.0	289.51	424.43	77.16	6.0	-72.37	32.47	-56.60
	10.0	962.94	1469.98	143.11	11.0	-240.73	102.50	-121.22
0.5	0.5	12.40	5.75	23.40	1.0	-12.40	6.75	-17.83
	1.0	30.55	14.82	38.67	1.5	-30.55	15.82	-34.79
	5.0	71.74	70.90	41.61	6.0	-71.74	72.91	-39.78
	10.0	241.81	240.98	83.16	11.0	-241.81	242.98	-82.09
0.8	0.5	8.12	1.88	20.71	1.0	-32.50	25.11	-26.96
	1.0	22.87	5.12	42.31	1.5	-91.50	69.41	-54.38
	5.0	60.38	26.96	50.75	6.0	-241.57	361.31	-66.51
	10.0	217.13	92.37	114.24	11.0	-868.53	1332.87	-133.60
0.9	0.5	9.94	1.33	26.31	1.0	-89.50	72.94	-47.49
	1.0	30.99	3.75	63.13	1.5	-278.99	228.62	-97.08
	5.0	87.16	20.61	80.80	6.0	-784.51	1279.31	-121.03
	10.0	327.33	71.61	192.81	11.0	-2946.01	5010.52	-240.50

Comments:

(i) The asymptotic variance (μ_{21} is the coefficient of the N^{-1} term in the variance, N the sample size) increases as θ_1 increases.

(ii) For $\theta_1 = 10$, $\theta_2 = 11$, the variance of $\hat{\theta}_1$ is a 4 digit integer.

(iii) The asymptotic skewness term ($\sqrt{\beta_{11}}$ is the coefficient $1/\sqrt{N}$ in the skewness) is large in value in most cases.

(iv) "Safe" sample size (setting $\sqrt{\beta_{11}}$ to be 1/2) is large, ranging from about 3000 to over 2 million. These are given in Table 2, Bowman and Shenton (2003a).

3.2 $5pP$ case

3.2.1 Some examples

Component values, θ_1, θ_2, and θ_3. Proportions, π_1, π_2, and $1-\pi_1-\pi_2=\pi_3$.

Details are in the Table 4 (excerpt from Table 2, Bowman and Shenton, 2003b).

Table 4. Variances of estimators in $5pP$ case

θ_1	θ_2	θ_3	π_1	π_2	$Var_1(\hat{\theta_1})$	$Var_1(\hat{\theta_2})$	$Var_1(\hat{\theta_3})$	$Var_1(\hat{\pi}_1)$	$Var_1(\hat{\pi}_2)$
1.0	1.5	2.5	1/3	1/3	97800	595789	12551	328185	217113
			0.1	0.5	1223631	304824	10287	373094	244840
1.0	2.0	2.5	1/3	1/3	8117	867186	211822	10606	470467
			0.1	0.5	100332	454024	176029	12078	563404
0.5	0.8	1.5	1/3	1/3	256831	1258381	16610	2270547	1747291
			0.1	0.5	3395444	668659	13861	2707559	2081280
0.5	1.0	1.2	1/3	1/3	75958	13291081	3775517	410677	57042156
			0.1	0.5	1050209	7446537	3315073	513599	72101446

Comments: In general for this table, variances are large, unusually so, and a relation between parameter values and variance is anything but clear.

3.2.2 The efficiency of moment estimators for a $5pP$ case

Table 5 gives the bias (μ'_{11}), $\sigma(\sqrt{\mu_{21}/N})$, and skewness ($\sqrt{\beta_1}$ approximation) for a $5pP$ case and sample size $N = 10,000$. From Fisher's work on estimation efficiency, the efficiency of the method of moment should be less than unity; having set up the Maple maximum likelihood estimators, the efficiency values are around 1/3, a satisfactory value, supporting the validity of the Maple code. Further details are given in Bowman and Shenton (2002, Appendix).

Table 5. 5-parameter Poisson Mixture ($5pP$) and the method of Moments

	θ_1	θ_2	θ_3	π_1	π_2	N
Case 1	2.0	3.0	6.0	1/2	1/3	10,000
Bias	-0.8502	0.9861	0.2610	-0.8613	0.9024	
σ	0.5710	1.6234	0.3811	0.7206	0.6718	
$\sqrt{\beta_1}$	-8.1678	3.6798	2.7995	-6.7471	7.4244	
Efficiency	0.2075	0.1961	0.3093	0.1987	0.1990	

Due to the scale the graphs appear to be continuous.

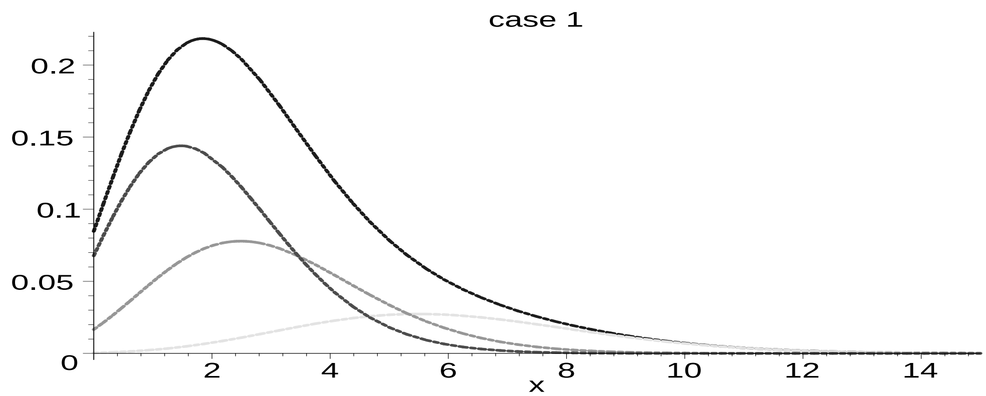

Figure 1: 3 component Poisson mixture distribution $\theta_1 = 2$, $\theta_2 = 3$, $\theta_3 = 6$, $\pi_1 = 1/2$, $\pi_2 = 1/3$

3.2.3 A five parameter case ($5pP$) mentioned by Everitt and Hand

Everitt and Hand(1981, p99) studied maximum likelihood estimators for the case $\theta_1 = 0.5$, $\theta_2 = 3.0$, $\theta_3 = 6.0$; $\pi_1 = 0.3$, $\pi_2 = 0.3$. The computation of bias, σ, and skewness by Maple code are stated in Table 6, sample size $N = 500$.

Table 6. Maximum Likelihood moments in a $5pP$ case

	θ_1	θ_2	θ_3	π_1	π_2	N
Case 2	0.5	3.0	6.0	0.3	0.3	500
mle	0.143	2.921	7.165	0.158	0.604	
Bias	-0.0730	-0.0538	0.2463	-0.0327	0.0876	
σ	0.1991	1.2464	0.6048	0.0795	0.1155	
$\sqrt{\beta_1}$	-0.9274	-0.0873	1.6380	-1.5418	1.5987	

Comments: The sample of $N = 500$ seems large enough to reduce estimator moments to moderate values, the skewness values for $\hat{\theta}_3$. $\hat{\pi}_1$ $\hat{\pi}_2$ are negative.

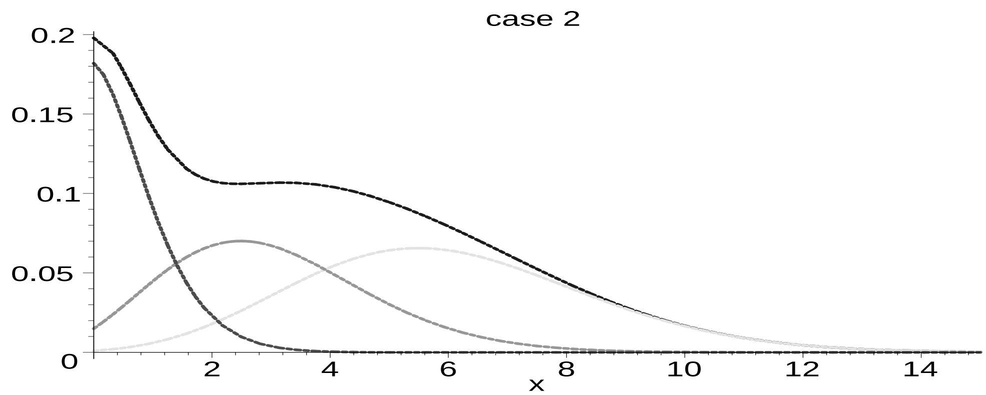

Figure 2: 3 component Poisson mixture distribution, $\theta_1 = 1/2$, $\theta_2 = 3$, $\theta_3 = 6$, $\pi_1 = \pi_2 = 0.3$

3.3 $4pP$ case

3.3.1 Two component 4 parameter case, Poisson and Negative binomial mixture

Probability function:

$$P(X = x) = \pi P_1 + (1 - \pi)P_2$$

where $0 < \pi_1 < 1$, $x = 0, 1, \cdots$, and

$$P_1 = \frac{e^{-\theta}\theta^x}{x!},$$
$$P_2 = \frac{1}{(p+1)^k}\left(\frac{p}{p+1}\right)^x \frac{\Gamma(k+x)}{\Gamma(k)x!}. \quad (k > 0)$$

For maximum likelihood estimators $\hat{\pi}$, $\hat{\theta}$, $\hat{p}$, $\hat{k}$, a possible iterated solution, if a solution exists, is;

$$
{}_{new}\hat{\pi} = \pi \sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_1}{\hat{P}}, \qquad {}_{new}\hat{\theta} = \sum_{x=0}^{n} \frac{n_x}{N} x \frac{\hat{P}_1}{\hat{P}},
$$

$$
{}_{new}\hat{p} = \frac{1}{k} \sum_{x=0}^{n} \frac{n_x}{N} x \frac{\hat{P}_2}{\hat{P}}
$$

and

$$
\psi({}_{new}\hat{k}) = \sum_{x=0}^{n} \frac{n_x}{N} \psi(k+x) \frac{\hat{P}_2}{\hat{P}} - \ln(1+p)
$$

where N is the sample size, n refers to the highest value of x in the sample, and $\psi(\cdot)$ is the Psi function.

For the low order maximum likelihood moments see Table 7.

Table 7 Maximum likelihood moment for 2 component 4 parameter case.

	θ	k	p	π	N
	1.0	1.0	1.0	0.5	1
μ_{11}	-73.90	322.84	42.92	-40.96	
μ_{21}	31.53	417.81	288.95	36.40	
σ	5.62	20.44	17.00	6.03	
$\sqrt{\beta_1}$	-57.64	68.81	32.15	-55.27	
	1.0	1.0	1.0	0.5	500
bias	-0.1478	0.6457	0.0858	-0.0819	
σ	0.2511	0.9141	0.7602	0.2698	
$\sqrt{\beta_1}$	-2.58	3.08	1.44	-2.47	
	1.0	5.0	1.0	0.5	1
μ_{11}	-0.3320	607.8813	-6.5019	-4.3383	
μ_{21}	11.5823	3576.6033	91.6750	4.4025	
σ	3.4033	59.8047	9.5747	2.0982	
$\sqrt{\beta_1}$	-6.3625	42.2589	14.4670	-23.6284	
	1.0	5.0	1.0	0.5	500
bias	-0.0007	1.2158	-0.0130	-0.0087	
σ	0.1522	2.6745	0.4282	0.0938	
$\sqrt{\beta_1}$	-0.28	1.89	0.65	-1.06	

3.3.2 4 component all proportions known

Everitt and Hand's 7 parameter example, but modified as all proportions known. Variances and skewness are notably smaller than $7pP$. Comparison to $7pP$ will be given in sequel.

Table 8. Moments of Maximum Likelihood estimators of 4-parameter Poisson Mixture $\pi_1 = \pi_2 = \pi_3 = 0.2$ known and sample size $N = 100$

	θ_1	θ_2	θ_3	θ_4
	0.5	1.5	3.0	5.0
Bias	0.2166	0.1576	-0.1039	-0.2711
σ	0.3575	0.8907	1.1124	0.7104
$\sqrt{\beta_1}$	2.9580	1.1755	-0.7498	-1.1013

Comment: The standard deviations are small and the 4 skewness less than 3 in value.

3.4 $7pP$ case

Table 9. Moments of maximum likelihood estimators of 7-parameter Poisson mixture with sample size $N = 1,000,000$

	θ_1	θ_2	θ_3	θ_4	π_1	π_2	π_3
Case 1	0.5	1.5	3.0	5.0	0.2	0.2	0.2
Bias	-0.1542	-0.3680	0.5774	0.0460	-0.1138	0.1392	0.0086
σ	0.1322	0.9151	0.8837	0.0663	0.0961	0.0717	0.1218
$\sqrt{\beta_1}$	-6.3131	-2.1612	3.6039	3.2741	-6.5680	7.2545	-0.0705
Case 2	0.5	1.5	3.0	5.0	0.25	0.25	0.25
Bias	-0.0789	-0.1728	0.2727	0.0423	-0.0718	0.0895	0.0005
σ	0.0972	0.6489	0.6027	0.0872	0.0873	0.0600	0.1072
$\sqrt{\beta_1}$	-4.3442	-1.3976	2.4919	2.2720	-4.5165	5.0458	-0.2898
Case 3	0.5	1.5	3.0	5.0	0.1	0.1	0.1
Bias	-0.8391	-2.2499	3.3459	0.0796	-0.3154	0.3702	0.0530
σ	0.2966	2.1487	2.1689	0.0482	0.1094	0.0899	0.1443
$\sqrt{\beta_1}$	-15.5559	-5.7492	8.5447	7.8658	-16.1553	16.9948	0.9169
Case 4	0.5	1.5	3.0	5.0	0.7	0.1	0.1
Bias	-0.0317	-0.5801	0.6978	0.1079	-0.0973	0.1136	0.0023
σ	0.0245	1.0808	0.9754	0.1392	0.0600	0.0383	0.0705
$\sqrt{\beta_1}$	-6.7772	-2.8361	3.9633	3.6422	-8.9720	8.1712	-0.3437
Case 5	1.0	2.0	4.5	6.0	0.3	0.3	0.1
Bias	-0.0638	-0.1594	0.0984	0.1718	-0.0592	0.0328	0.1371
σ	0.0942	0.3635	1.4494	0.1658	0.0893	0.0515	0.0541
$\sqrt{\beta_1}$	-3.1659	-2.0234	0.2350	5.5975	-3.1819	2.9646	11.9292

Case 1 is the first example of Everitt and Hand's $7pP$. Cases 2-4 are variation of Case 1, and Case 5 is their second example of $7pP$.

Table 10. Case of 7-parameter Poisson Mixture from Everitt and Hand
Sample size $N = 200$

	θ_1	θ_2	θ_3	θ_4	π_1	π_2	π_3
	0.50	1.50	3.00	5.00	0.20	0.20	0.20
mle	0.75	1.34	3.37	5.08	0.15	0.19	0.20
Bias	-770	-1840	2887	230	-569	696	43
σ	9	65	62	5	7	5	9
$\sqrt{\beta_1}$	-446	-153	255	232	-464	513	-5

The sample size $N = 200$ was used by Everitt and Hand. As it is shown in the table $N = 200$ is not large enough to make any inference on the estimators.

4 Covariance and correlation

Correlation of the case of $4pP$ is

$$R = \begin{bmatrix} 1.00 & -0.69 & 0.51 & -0.26 \\ -0.69 & 1.00 & -0.67 & 0.43 \\ 0.51 & -0.67 & 1.00 & -0.55 \\ -0.26 & 0.43 & -0.55 & 1.00 \end{bmatrix}$$

and the value of Hessian matrix is

$$H = 0.8157e - 6.$$

Correlation of the Case 1 of $5pP$ is

$$R = \begin{bmatrix} 1.00 & 0.97 & 0.73 & 0.99 & -1.00 \\ 0.97 & 1.00 & 0.84 & 0.99 & -0.98 \\ 0.73 & 0.84 & 1.00 & 0.78 & -0.75 \\ 0.99 & 0.99 & 0.78 & 1.00 & -1.00 \\ -1.00 & -0.98 & -0.75 & -1.00 & 1.00 \end{bmatrix}$$

and the value of Hessian matrix is

$$H = 0.5523e - 9.$$

Correlation of the Case 2 of $5pP$ is

$$R = \begin{bmatrix} 1.00 & 0.77 & 0.49 & 0.89 & 0.21 \\ 0.77 & 1.00 & 0.82 & 0.87 & 0.68 \\ 0.49 & 0.82 & 1.00 & 0.59 & 0.87 \\ 0.89 & 0.87 & 0.59 & 1.00 & 0.30 \\ 0.21 & 0.68 & 0.87 & 0.30 & 1.00 \end{bmatrix}$$

and the value of Hessian matrix is

$$H = 0.1402e - 4.$$

Correlation of the Case 1 of $7pP$ is

$$R = \begin{bmatrix} 1.00 & 0.98 & 0.89 & 0.72 & 1.00 & 0.73 & -0.96 \\ 0.98 & 1.00 & 0.96 & 0.82 & 0.99 & 0.86 & -1.00 \\ 0.89 & 0.96 & 1.00 & 0.93 & 0.91 & 0.96 & -0.97 \\ 0.72 & 0.82 & 0.93 & 1.00 & 0.75 & 0.94 & -0.81 \\ 1.00 & 0.99 & 0.91 & 0.75 & 1.00 & 0.78 & -0.98 \\ 0.73 & 0.86 & 0.96 & 0.94 & 0.78 & 1.00 & -0.88 \\ -0.96 & -1.00 & -0.97 & -0.81 & -0.98 & -0.88 & 1.00 \end{bmatrix}$$

and the value of Hessian matrix is

$$H = 0.2239e - 15.$$

Correlation of Case 2-4 are similar to the Case 1, and also the values of H's.

In general correlations are very high in all cases of Poisson mixture. The magnitude of the value of H decrease drastically with the increases in the number of parameter imply that the variances increase as the number of parameter increases.

5 Binomial distribution mixture

5.1 Formulas

Here the binomial random variable X has probability function

$$P(X = x) = P(x) = \sum_{i=1}^{s} \pi_i \binom{n}{x} p_i^x (1 - p_i)^{n-x} = \sum_{i=1}^{s} \pi_i P_i$$

where $n = 2, 3, \cdots$; $x = 0, 1, \cdots, n$; $0 < p_i < 1$, $\sum_{i=1}^{s} \pi_i = 1$.

For maximum likelihood estimators $\hat{\pi}_i$, $\hat{p}_i$,

$$\frac{\partial L}{\partial \pi_i} = \frac{\hat{P}_i - \hat{P}_s}{\hat{P}},$$

and in sample of N leading to

$${}_{new}\hat{\pi}_i = \hat{\pi}_i \sum_{x=0}^{n} \frac{n_x}{N} \frac{\hat{P}_i}{\hat{P}},$$

the iterative cycle being completed by

$${}_{new}\hat{p}_i = \sum_{x=0}^{n} x \frac{n_x}{N} \frac{\hat{P}_i}{\hat{P}},$$

where n is the highest values of x in the sample of N, and n_x is the frequency of x.

5.2 The Hessian

For simplicity we take $s = 4$. Then the Hessian determinant is of order 7x7. By elementary determinant operations we find after simplification

$$|H_7^*| = c^2 \left| \begin{array}{cccc|ccc} y_1^2 & y_1y_2 & y_1y_3 & y_1y_4 & y_1z_1 & y_1z_2 & y_1z_3 \\ y_2y_1 & y_2^2 & y_2y_3 & y_2y_4 & y_2z_1 & y_2z_2 & y_2z_3 \\ y_3y_1 & y_3y_2 & y_3^2 & y_3y_4 & y_3z_1 & y_3z_2 & y_3z_3 \\ y_4y_1 & y_4y_2 & y_4y_3 & y_4^2 & y_4z_1 & y_4z_2 & y_4z_3 \\ \hline z_1y_1 & z_1y_2 & z_1y_3 & z_1y_4 & z_1^2 & z_1z_2 & z_1z_3 \\ z_2y_1 & z_2y_2 & z_2y_3 & z_2y_4 & z_2z_1 & z_2^2 & z_2z_3 \\ z_3y_1 & z_3y_2 & z_3y_3 & z_3y_4 & z_3z_1 & z_3z_2 & z_3^2 \end{array} \right|$$

where $y_i = (x - np_i)P_i/P$, $z_i = (P_i - P_4)/P$, $i = 1, 2, 3, 4$, and each element in the final determinant consisting of $n + 1$ terms, a finite series, with $1/P$ in the denominator of each term. Note that since $0 < P < 1$ in the parameter domain, $1/P$ is bounded. Not also the $\min P$ over this domain could be defined. In H_7 the partitioning with respect to component 4x4 and proportions 3x3 are indicated, we have

$$c^2 = \prod_{i=1}^{4} \left(\frac{\pi_i}{p_i(1 - p_i)} \right)^2,$$

and

$$(i) \quad y_r y_s = \sum_{x=0}^{n} \frac{(x - np_r)(x - np_s)P_r P_s}{P} \quad (r, s = 1, 2, 3, 4)$$

$$(ii) \quad y_r z_s = \sum_{x=0}^{n} \frac{(x - np_r)P(P_r - P_4)}{P} \quad (r = 1, 2, 3, 4, s = 1, 2, 3)$$

$$(iii) \quad z_r z_s = \sum_{x=0}^{n} \frac{(P_r - P_4)(P_s - P_4)}{P} \quad (r, s = 1, 2, 3)$$

each of the expressions involving $n + 1$ terms with denominators $P(x)$, and $0 < P(x) < 1$.

In H_7, the Hessian, the partitions relate to components (4x4) and proportions (3x3), and are delineated.

Comparisons of the probability functions for Poisson and binomial mixtures are shown in Figures 3 and 4.

5.3 Some examples of low order maximum likelihood estimators from Maple

These are given in Table 11 and binomial distribution with n=10, 30, 50, and 100.

Table 11 Low order moment for binomial mixture

	p_1	p_2	p_3	p_4	π_1	π_2	π_3
	0.1	0.2	0.6	0.9	0.2	0.2	0.2
				$n = 10$			
μ_{11}	-176.53	247.29	40.94	4.08	-358.35	391.00	-15.30
σ	3.4128	6.2219	1.8292	0.3154	17.0090	16.1045	0.8589
$\sqrt{\beta_{11}}$	-271.94	236.40	68.21	15.74	-107.73	119.78	-51.11
				$n = 30$			
μ_{11}	-0.9600	1.6350	0.1817	0.0281	-0.3048	0.4478	-0.0541
σ	0.3521	0.4653	0.2399	0.0952	1.4842	1.4624	0.4197
$\sqrt{\beta_{11}}$	-6.5922	15.2739	0.0689	-1.3593	0.8644	0.3844	1.5094
				$n = 50$			
μ_{11}	-0.2039	0.3253	0.0194	0.0040	-0.0448	-0.0593	-0.0030
σ	0.1854	0.2351	0.1616	0.0688	0.7899	0.7877	0.4031
$\sqrt{\beta_{11}}$	-0.3084	3.8629	0.0748	-0.8399	1.1161	0.9821	1.5023
				$n = 100$			
μ_{11}	-0.0290	0.0515	0.0003	0.0002	0.0023	-0.0022	0.0003
σ	0.0922	0.1188	0.1097	0.0475	0.4840	0.4840	0.4001
$\sqrt{\beta_{11}}$	1.6062	-0.0446	-0.0659	-0.4394	1.4276	1.4449	1.5000

Comments: The skewness are large in values for the components in particular when the binomial index $n = 10$ and $n = 30$. However there is a steady improvement as n increases to $n = 100$. The corresponding correlation determinants are:

$$R_{10} = \left[\begin{array}{cccc|ccc} 1.00 & 0.96 & 0.65 & 0.34 & 0.99 & -0.99 & -0.50 \\ 0.96 & 1.00 & 0.76 & 0.42 & 0.99 & -0.98 & -0.53 \\ 0.65 & 0.76 & 1.00 & 0.71 & 0.70 & -0.67 & -0.19 \\ 0.34 & 0.42 & 0.71 & 1.00 & 0.37 & -0.36 & 0.23 \\ \hline 0.99 & 0.99 & 0.70 & 0.37 & 1.00 & -1.00 & -0.52 \\ -0.99 & -0.98 & -0.67 & -0.36 & -1.00 & 1.00 & 0.51 \\ -0.50 & -0.53 & -0.19 & 0.23 & -0.52 & 0.51 & 1.00 \end{array}\right]$$

$$R_{30} = \left[\begin{array}{cccc|ccc} 1.00 & 0.76 & 0.09 & 0.01 & 0.87 & -0.87 & -0.04 \\ 0.76 & 1.00 & 0.16 & 0.02 & 0.87 & -0.86 & -0.06 \\ 0.09 & 0.16 & 1.00 & 0.17 & 0.11 & -0.10 & 0.08 \\ 0.01 & 0.02 & 0.17 & 1.00 & 0.02 & -0.01 & 0.11 \\ \hline 0.87 & 0.87 & 0.11 & 0.02 & 1.00 & -0.94 & -0.11 \\ -0.87 & -0.86 & -0.10 & -0.01 & -0.94 & 1.00 & -0.03 \\ -0.04 & -0.06 & 0.08 & 0.11 & -0.11 & -0.03 & 1.00 \end{array}\right]$$

$$R_{50} = \left[\begin{array}{cccc|ccc} 1.00 & 0.61 & 0.01 & 0.00 & 0.71 & -0.71 & -0.00 \\ 0.61 & 1.00 & 0.03 & 0.00 & 0.70 & -0.70 & -0.01 \\ 0.01 & 0.03 & 1.00 & 0.05 & 0.02 & -0.01 & 0.02 \\ 0.00 & 0.00 & 0.05 & 1.00 & 0.00 & -0.00 & 0.02 \\ \hline 0.71 & 0.70 & 0.02 & 0.00 & 1.00 & -0.81 & -0.13 \\ -0.71 & -0.70 & -0.01 & -0.00 & -0.81 & 1.00 & -0.12 \\ -0.00 & -0.01 & 0.02 & 0.02 & -0.13 & -0.12 & 1.00 \end{array}\right]$$

$$R_{100} = \left[\begin{array}{cccc|ccc} 1.00 & 0.39 & 0.00 & 0.00 & 0.37 & -0.37 & 0.00 \\ 0.39 & 1.00 & 0.00 & 0.00 & 0.36 & -0.36 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ \hline 0.37 & 0.36 & 0.00 & 0.00 & 1.00 & -0.49 & -0.21 \\ -0.37 & -0.36 & 0.00 & 0.00 & -0.49 & 1.00 & -0.21 \\ 0.00 & 0.00 & 0.00 & 0.00 & -0.21 & -0.21 & 1.00 \end{array}\right]$$

Comments: For $n = 100$, the correlations are all nearly zero.

6 Conclusion

Some ideas of the present literature on mixture distributions may be gained from the recent paper by Karlis and Xekalaki (2005). Theoretical properties are given, with special attention to Poisson mixtures. Another recent account of Poisson and binomial mixtures is given by Everitt and Hand (1981). Little seems to be knwon about variance of maximum likelihood estimators and their skewness. In our paper we use Maple code to set up

Bias:	N^{-1} and N^{-2} terms
Variance:	N^{-1} and N^{-2} terms $(\mu_{21}/N + \mu_{22}/N)$
Skewness:	the $1/\sqrt{N}$ term
Kurtosis:	The N^{-3} term in the fourth central moment.

A large value of the coefficient of variation ($V = 100\sigma/\text{mean}$) may suggest that the model is at fault, or that there may not be a solution. The same can be said about a large value of the asymptotic skewness.

There is also the question of sample space and validity for the estimators, $\hat{\theta}_1$, $\hat{\theta}_2$, $\cdots$; $\hat{\pi}_1$, $\hat{\pi}_2$, $\cdots$. Bowman and Shenton (1973) solved this problem for a moment approach to the mixture of two normal densities, using the three sample cumulants κ_3, κ_4, κ_5. It turns out that the acceptable space is quite small. In this connection see Karl Pearson (1894).

References

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 Res.*, 281, 227-232.

Bowman, K.O. and Shenton, L.R. (1973). Space of solutions for a normal mixture, *Biometrika*, 60, 629-636.

Bowman, K. O. and Shenton, L. R. (2002), Poisson Mixture of k Components, *Far East Journal of Theoretical Statistics*, 7(2), 93-110.

Bowman, K. O. and Shenton, L. R. (2003a), Sample size assessment and the two component Poisson mixture, proportion known, *Far East Journal of Theoretical Statistics*, 9(2), 97-115.

Bowman, K. O. and Shenton, L. R. (2003b), Canonical dichotomized alternant determinants and Poisson mixture, *Far East Journal of Theoretical Statistics*, 10(2), 87-109.

Bowman, K.O. and Shenton, L.R. (2005a), The asymptotic variance and skewness of maximum likelihood estimators using Maple, *Journal of Statistical Computation and Simulation*, 75(12), 975-986.

Bowman, K.O. and Shenton, L.R. (2005b), Bias, variance, skewness and kurtosis of maximum likelihood estimators using Maple, *Far East Journal of Theoretical Statistics*, 17(2), 137-195.

Bowman, K.O. and Shenton, L.R. (2006), Moments of maximum likelihood estimators in the discrete case, *Far East Journal of Theoretical Statistics*, 18(1), 61-91.

Everitt, B.S. and Hand, D.J. (1981), *Finite Mixture Distribution*, London: Chapman and Hall.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson Distributions, *International Statistical Review*, 73(1), 35-58.

Pearson, K (1894). Contribution to the mathematical theory of evolution. *Phil. Trans. A*, 185, 71-110.

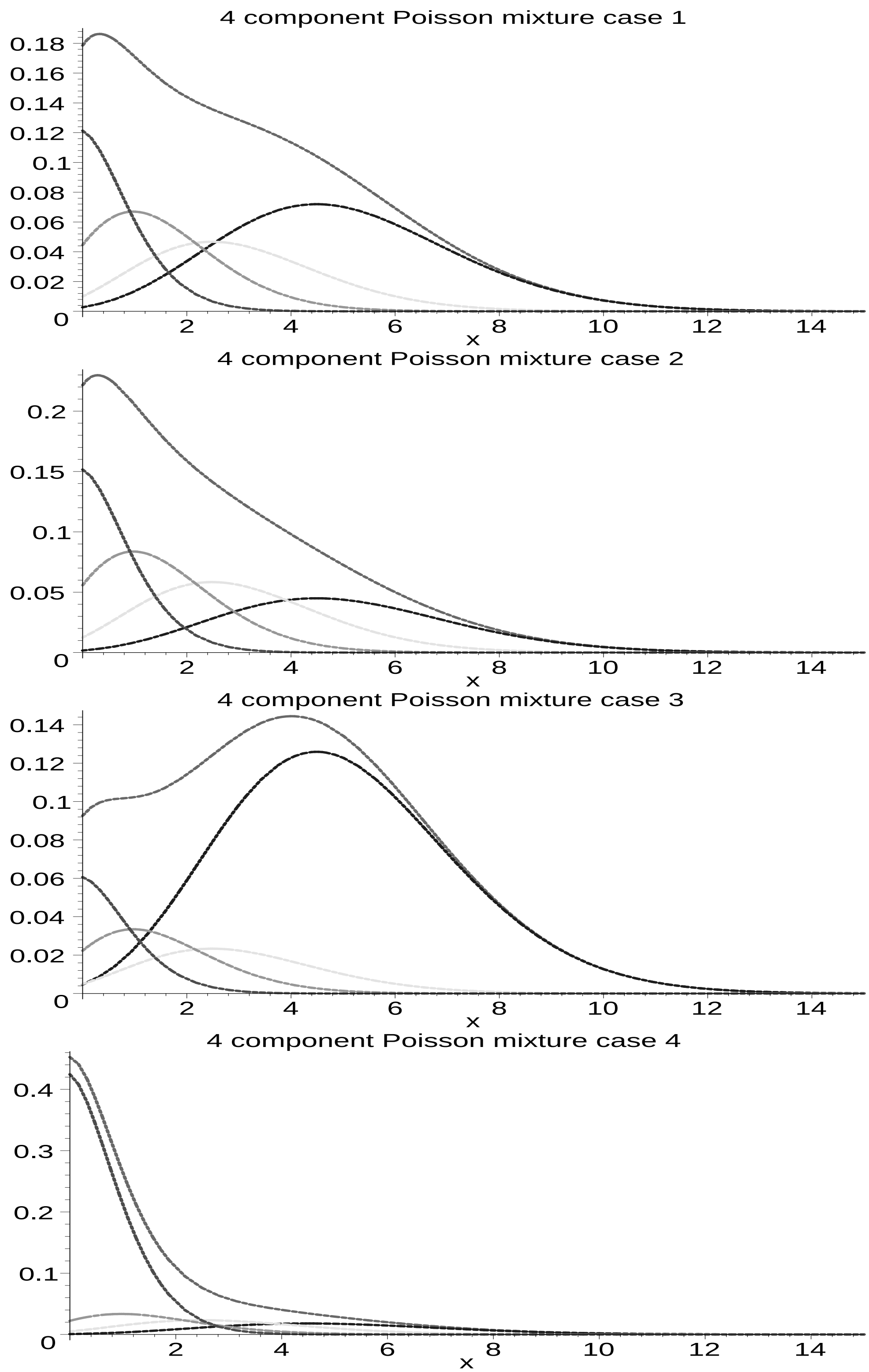

Figure 3: 4 component Poisson mixture distribution

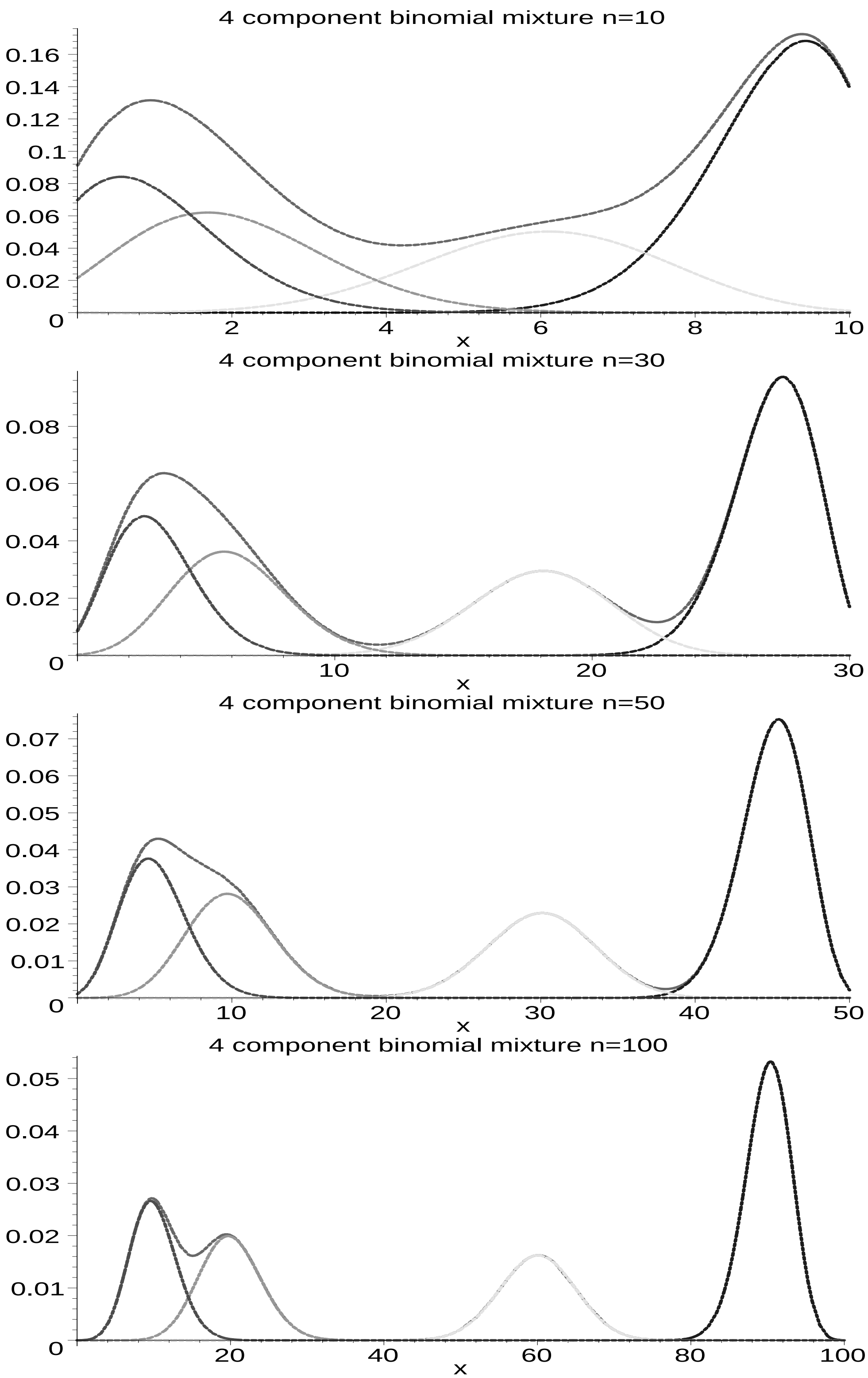

Figure 4: 4 component binomial mixture distribution