Int. J. Math. & Stat. Sci.
Vol. 4, No. 2, December 1995, 135–150

ML ESTIMATION IN THE BIVARIATE 'SHORT' DISTRIBUTIONS BY THE EM ALGORITHM

K. Adamidis and S. Loukas

Department of Mathematics, University of Ioannina
451 10 Ioannina, Greece

Abstract

The maximum likelihood estimation of parameters of two bivariate 'Short' distributions, introduced by Papageorgiou (1986), is considered by applying the EM Algorithm (Dempster et al. 1977). The observed Fisher informations are derived (Louis 1982) and numerical examples based on real data are presented where the convergence of the EM algorithm is accelerated substantially by the methods of conjugate gradients (Jamshidian and Jennrich 1993) and Louis (1982).

Keywords: Accident data, Bivariate 'Short' distributions, EM algorithm.
1991 A.M.S. Subject Classification: Primary 62F10 , Secondary 62F15 .

1. Introduction

The analysis of accident data has stimulated a considerable amount of interest among researchers during the present century. The major innovations in accident theory have been made along with hypotheses on the occurrence of the event of an accident. These were based on the concepts of pure chance, contagion, risk, proneness and 'spells' (time periods in which all accidents must occur). The formulations of such ideas have led to the development of many well known probability models and fruitful statistical theory. A comprehensive review is provided by Kemp (1970). Cresswell and Froggatt (1963) proposed a three parameter discrete distribution as a 'spells' model for accident data, which they called the 'Short' distribution. Later Irwin (1964) pointed out that it can be given a proneness interpretation. The definition of the 'Short' distribution, its statistical properties and estimation of parameters by the methods of moments and maximum likelihood can be found in Kemp (1967) and Kerr (1969).

Papageorgiou (1986) introduced two bivariate extensions of the 'Short' distribution with univariate 'Short' marginals. Kocherlakota and Kocherlakota (1992) refer to these distributions as bivariate 'Short' type I and II (henceforth type I and II). In their notation the distributions are defined by assuming pairs of discrete random variables $Y_k = (Y_{1k}, Y_{2k})$, $k = 1, 2$, with the structure

$$(1.1) \qquad (Y_{jk} = X_{jk} + X_{3k}; j = 1, 2) \quad k = 1, 2,$$

where $X_{jk}, j = 1, 2, 3$ are mutually independent Poisson and Neyman Type A variables. Specifically, let X_{11}, X_{21} follow Neyman Type A distributions with parameters $(\lambda_j, \theta_j), j = 1, 2$ and X_{31} follow a Poisson distribution with parameter λ. The distribution of Y_1 is type I with parameter vector $\phi_1 = (\theta_1, \theta_2, \lambda, \lambda_1, \lambda_2)$

and probability function

$$(1.2)\qquad P(y_1;\phi_1) = e^{-\lambda} \sum_{x_{31}=0}^{\min(y_{11},y_{21})} \frac{\lambda^{x_{31}}}{x_{31}!} \times \prod_{j=1}^{2} \left(\frac{e^{-\lambda_j}\theta_j^{y_{j1}-x_{31}}}{(y_{j1}-x_{31})!} \sum_{m_j=0}^{\infty} \frac{(\lambda_j e^{-\theta_j})^{m_j}}{m_j!} m_j^{y_{j1}-x_{31}} \right).$$

Similarly, by taking X_{12}, X_{22} to be Poisson variables with parameters λ_1, λ_2 respectively and X_{32} to be a Neyman Type A variable with parameters (λ, θ) then the distribution of Y_2 is type II with parameter vector $\phi_2 = (\lambda_1, \lambda_2, \theta, \lambda)$ and probability function

$$(1.3)\qquad P(y_2;\phi_2) = e^{-\lambda} \sum_{x_{32}=0}^{\min(y_{12},y_{22})} \frac{\theta^{x_{32}}}{x_{32}!} \times \sum_{m=0}^{\infty} \frac{(\lambda e^{-\theta})^{m}}{m!} m^{x_{32}} \prod_{j=1}^{2} \frac{e^{-\lambda_j}\lambda_j^{y_{j2}-x_{32}}}{(y_{j2}-x_{32})!}.$$

The statistical properties of the distributions and estimation of their parameters by the method of moments have been studied by Papageorgiou (1986) where it is pointed out that the efficiency of the method of moments is consistently low for both distributions. Therefore, maximum likelihood estimators are strongly desired. A review and generalizations of Papageorgiou's results are provided by Kocherlakota and Kocherlakota (1992).

It is the purpose of this paper to consider the maximum likelihood estimation problems for the type I and II distributions. The likelihoods corresponding to (1.2) and (1.3) are rather unpleasant functions to maximize by computing their first or second derivatives, let alone that it seems unlikely to attain maximizers in closed forms. In this paper, our interest is in approaching the maximization problems by means of the statistical reasoning underlying the theme of the EM algorithm (Dempster *et al.* 1977). The observed Fisher informations of the EM estimates are obtained (Louis 1982, Section 3) and examples based on real data are included. In each of the examples the convergence of the EM algorithm is accelerated by the methods of Louis (1982, Section 5) and Jamshidian and Jennrich (1993).

2. Estimation using the EM Algorithm

We assume that data $Y_k^{obs} = (y_{ijk}; i = 1, \ldots, n,\ j = 1, 2)$, $k = 1, 2$ are observed and are known to be i.i.d from the type I and II distributions with probability functions (1.2), (1.3) respectively. We are concerned with the use of the EM algorithm (Dempster *et al.* 1977) to obtain maximum likelihood estimates of parameters of the distributions.

To implement the EM algorithm we view the actual data as incomplete data from the distributions of $X_k = (X_{jk}; j = 1, \ldots, 6-k)$, $k = 1, 2$ with parameter vectors ϕ_k as in (1.2), (1.3) and probability functions

$$P(x_k; \phi_k) = \prod_{j=1}^{6-k} e^{-\mu_{jk}} \frac{\mu_{jk}^{x_{jk}}}{x_{jk}!} \qquad k = 1, 2, \tag{2.1}$$

where μ_{jk} are the jth elements of $\mu_1 = (x_{41}\theta_1, x_{51}\theta_2, \lambda, \lambda_1, \lambda_2)$, $\mu_2 = (\lambda_1, \lambda_2, x_{42}\theta, \lambda)$. Thus, for $k = 1, 2$ we postulate complete-data specifications based on $Y_k^c = (x_{ijk}; i = 1, \ldots, n, j = 1, \ldots, 6-k)$ for the observed data Y_k^{obs} with the $Y's$ having the structure (1.1). The marginal distributions of X_{j1} are Neyman Type A with parameters (λ_j, θ_j) for $j = 1, 2$ and Poisson with parameters $\lambda, \lambda_1, \lambda_2$ for $j = 3, 4, 5$ respectively. Also the variables are pair wise independent except $(X_{j1}, X_{(j+3)1}), j = 1, 2$. Similarly, the marginal distributions of X_{j2} are Poisson with parameters $\lambda_1, \lambda_2, \lambda$ for $j = 1, 2, 4$ respectively, Neyman Type A with parameters (λ, θ) for $j = 3$ and the variables are pair wise independent except (X_{32}, X_{42}). The distributions defined by (2.1) were constructed to have these specific marginals. Since the Neyman Type A is obtained by compounding a Poisson distribution by a Poisson distribution, the product of their probability functions with those of a further of two independent Poisson distributions forms the probability function of a four variate distribution with Neyman Tupe A and Poisson marginals (see also Marshall and Olkin 1988 and references therein).

The E-step of the method estimates the unobserved complete-data log likelihoods by their conditional expectations, given the observed data and current estimates of $\phi_k, \phi_k^{(t)}, k = 1, 2$. The latter are linear in the following sufficient statistics for ϕ_k

$$s_k = (s_{jk} = \sum_{i=1}^{n} x_{ijk}; j = 1, \ldots, 6-k) \quad k = 1, 2.$$

Consequently, when $Y_k, k = 1, 2$ are observed the EM algorithm eliminates the unobserved X_k by finding the conditional expectations of the sufficient statistics with respect to the distributions of $(X_k|y_k; \phi_k)$. The relevant computations can be reduced by using the structure (1.1). Since $X_{jk} = Y_{jk} - X_{3k}$ for $j = 1, 2$ the conditional expectations of X_{1k}, X_{2k} given y_k can be derived using those of X_{3k} given y_k. In fact the central moments of $(X_k|y_k; \phi_k)$ can be calculated using the central moments of $(X_{jk}|y_k; \phi_k, j = 3, \ldots, 6-k)$. Using (1.1) we can easily show that for $k = 1, 2$

$$\begin{aligned} (2.2) \qquad & E\left\{\prod_{j=1}^{6-k}[X_{jk} - E(X_{jk}|Y_k)]^{r_j}|Y_k\right\} \\ & = (-1)^{r_1+r_2} E\{[X_{3k} - E(X_{3k}|Y_k)]^{r_1+r_2+r_3} \\ & \times \prod_{j=4}^{6-k}[X_{jk} - E(X_{jk}|Y_k)]^{r_j}|Y_k\}. \end{aligned}$$

When $k = 1, X_{41}$ is independent of X_{51} and X_{31} is independent of $X_{j1}, j = 1, 2, 4, 5$ so that

$$\begin{aligned} &P(X_{31} = x_{31}, X_{41} = x_{41}, X_{51} = x_{51}, Y_1 = y_1) \\ &\quad = P(X_{11} = y_{11} - x_{31}, X_{21} = y_{21} - x_{31} | x_{41}, x_{51}) \\ &\quad \times \prod_{j=3}^{5} P(X_{j1} = x_{j1}). \end{aligned}$$

Futhermore, X_{11} is independent of X_{51} and X_{41} is independent of X_{21}. It follows that

$$\begin{aligned} &P(X_{11} = y_{11} - x_{31}, X_{21} = y_{21} - x_{31} | x_{41}, x_{51}) \\ &\quad = \prod_{j=1}^{2} P(X_{j1} = y_{j1} - x_{31} | x_{(j+3)1}). \end{aligned}$$

Therefore

$$\begin{aligned} &P(X_{31} = x_{31}, X_{41} = x_{41}, X_{51} = x_{51} | y_1) \\ &\quad = \prod_{j=1}^{2} P(X_{j1} = y_{j1} - x_{31} | x_{(j+3)1}) \\ &\quad \times \prod_{j=3}^{5} P(X_{j1} = x_{j1}) / P(y_1), \end{aligned} \tag{2.3}$$

where the random variables $X_{j1}|x_{(j+3)1}$ and X_{j1} follow Poisson distributions with parameters $x_{(j+3)1}\theta_j$ for $j = 1, 2$ and $\lambda, \lambda_1, \lambda_2$ for $j = 3, 4, 5$ respectively, and Y_1 follows the type I distribution with probability function given by (1.2). The analogous result for the case $k = 2$ is derived by noting that X_{12} is independent of X_{22} and both are independent of X_{32}, X_{42}. We find that

$$\begin{aligned} &P(X_{32} = x_{32}, X_{42} = x_{42} | y_2) \\ &\quad = \prod_{j=1}^{2} P(X_{j2} = y_{j2} - x_{32}) P(X_{32} = x_{32} | x_{42}) \\ &\quad \times P(X_{42} = x_{42}) / P(y_2). \end{aligned} \tag{2.4}$$

with $X_{j2}, X_{32}|x_{42}$ following Poisson distributions with parameters $\lambda_1, \lambda_2, \lambda, j = 1,2,4$ and $x_{42}\theta$ respectively and Y_2 being a type II variable with probability function (1.3). The conditional moments of the two distributions are given by

$$(2.5) \qquad E\left(\prod_{j=3}^{6-k} x_{jk}^{t_j}|y_k;\phi_k\right) = \sum_{x_{3k}=0}^{\min(y_{1k},y_{2k})} \sum_{x_{4k}=0}^{\infty} \cdots \sum_{x_{(6-k)k}=0}^{\infty} \prod_{j=3}^{6-k} x_{jk}^{t_j} P(x_{3k},\ldots,x_{(6-k)k}|y_k;\phi_k),$$

for appropriate values of $t_j \in N \cup \{0\}, j = 3,\ldots,6-k, k = 1,2$ and the probability functions are obtained from (2.3), (2.4). The required conditional expectations are computed for each experimental unit and the E-steps are completed by accumulating over all units to obtain

$$(2.6) \quad E(S_{jk}|Y_k^{obs};\phi_k^{(t)}) = \sum_{i=1}^{n} y_{ijk} - \sum_{i=1}^{n} E(X_{i3k}|y_{i1k},y_{i2k};\phi_k^{(t)}) \quad j = 1,2,$$
$$= \sum_{i=1}^{n} E(X_{ijk}|y_{i1k},y_{i2k};\phi_k^{(t)}) \quad j = 3,\ldots,6-k$$

for $k = 1,2$.

The M-steps of the EM algorithms determine updated values of ϕ_k, $\phi_k^{(t+1)}$ say, by simulating the maximum likelihood estimations which would have been carried out if the data were complete. This is done by maximizing the expected log likelihoods of the complete data given the observed data over ϕ_k. Hence we require the maximizers for the log likelihoods corresponding to (2.1) with the sufficient statistics replaced by their conditional expectations. The maximizations of the complete-data log likelihoods over $\phi_k, k = 1,2$ are easily accomplished to give $\hat{\phi}_{jk} = s_{jk}/s_{(5-2k+j)k}$ for $j = 2k-1,\ldots,k+1$ and $\hat{\phi}_{jk} = s_{jk}/n$

for $j = 5-2k, 6-2k, 6-k$, whereupon the M-steps dictate the new estimates

$$\begin{aligned}\phi_{jk}^{(t+1)} &= E(S_{jk}|Y_k^{obs};\phi_k^{(t)})/E(S_{(5-2k+j)k}|Y_k^{obs};\phi_k^{(t)}) \\ &\quad j = 2k-1,\ldots,k+1, \\ &= \tfrac{1}{n}E(S_{jk}|Y_k^{obs};\phi_k^{(t)}) \quad j = 5-2k, 6-2k, 6-k,\end{aligned}$$

for $k = 1, 2$.

The transitions from $\phi_k^{(t)}$ to $\phi_k^{(t+1)}$ through the E and M steps given above define the EM algorithms for the estimation problems under consideration. Repeated EM steps maximize the observed data log likelihoods over ϕ_k and at convergence, where $\phi_k^{(t)} = \phi_k^{(t+1)} = \hat{\phi}_k$, the estimates satisfy the maximum likelihood equations

$$\text{(2.7)} \qquad \begin{aligned}\hat{\phi}_{jk} &= (\bar{y}_{jk} - \hat{\phi}_{3k}\hat{\phi}_{4k}^{k-1})/\hat{\phi}_{(j+3)k}^{2-k} \quad j = 1, 2, \\ &= \frac{\sum_{i=1}^{n} E(X_{ijk}|y_{i1k}, y_{i2k};\hat{\phi}_k)}{n\left[\sum_{i=1}^{n} E(X_{i42}|y_{i12}, y_{i22};\hat{\phi}_2)/n\right]^{(4-j)(k-1)}} \\ &\quad j = 3,\ldots,6-k,\end{aligned}$$

where the expectations are given using (2.5) with $\phi_k = \hat{\phi}_k, k = 1, 2$.

3. The observed information matrices

Using the result of Louis (1982, Section 3) (see also Meilijson 1989), the observed Fisher informations of the EM estimates can be derived from the conditional expectations and variances of the complete-data observed information matrices and score functions respectively, given the observed data. These involve calculating the conditional variances of $X_{pk}, p = 1,\ldots,6-k$ and covariances of the pairs $(X_{pk}, X_{qk}), p, q = 1, \ldots, 6-k, p \neq q$, given $y_k, k = 1, 2$. However, using (2.2) with appropriate values of $(r_j; j = 1, \ldots, 6-k)$ we find that

$$\begin{aligned}
\mathrm{Var}(X_{pk}|Y_{1k}, Y_{2k}) &= \mathrm{Var}(X_{3k}|Y_{1k}, Y_{2k}) \quad p = 1, 2, \\
\mathrm{Cov}(X_{pk}, X_{qk}|Y_{1k}, Y_{2k}) &= (-1)^q \mathrm{Var}(X_{3k}|Y_{1k}, Y_{2k}) \\
&\quad p = 1, \quad q = 2; \\
&\quad p = 1, 2, \qquad q = 3, \\
\mathrm{Cov}(X_{pk}, X_{qk}|Y_{1k}, Y_{2k}) &= -\mathrm{Cov}(X_{3k}, X_{qk}|Y_{1k}, Y_{2k}) \\
&\quad p = 1, 2, \qquad q = 4, \ldots, 6-k,
\end{aligned}$$

with $k = 1, 2$.

Using the latter results the elements in the upper triangular part, of the $(6-k)$th order observed information matrices $\mathrm{I}(\phi_k|Y_k^{obs}), k = 1, 2$ of the observed data, are found to be

$$\begin{aligned}
(3.1) \qquad I_{pp} &= \frac{1}{\phi_{pk}^2} \sum_{i=1}^{n} \{ y_{ipk} - E(X_{i3k}) - \mathrm{Var}(X_{i3k}) \\
&\quad - (2-k)[\phi_{pk}^2 \mathrm{Var}(X_{i(p+3)k}) \\
&\quad + 2\phi_{pk} \mathrm{Cov}(X_{i(p+3)k}, X_{i3k})]\}, \quad p = 1, 2, \\
&= \frac{1}{\phi_{pk}^2} \sum_{i=1}^{n} \{ E(X_{ipk}) - \mathrm{Var}(X_{ipk}) \\
&\quad - (k-1)(4-p)[\theta^2 \mathrm{Var}(X_{i42}) \\
&\quad - 2\theta \mathrm{Cov}(X_{i32}, X_{i42})]\}, \quad p = 3, \ldots, 6-k,
\end{aligned}$$

$$\begin{aligned}
(3.2) \quad I_{pq} &= \frac{-1}{\phi_{pk}\phi_{qk}} \sum_{i=1}^{n} \{ \mathrm{Var}(X_{i3k}) + (2-k)[\theta_1 \mathrm{Cov}(X_{i41}, X_{i31}) \\
&\quad + \theta_1\theta_2 \mathrm{Cov}(X_{i41}, X_{i51}) + \theta_2 \mathrm{Cov}(X_{i51}, X_{i31})]\}, \\
&\quad p = 1, q = 2, \\
&= \frac{1}{\phi_{pk}\phi_{qk}} \sum_{i=1}^{n} [\mathrm{Cov}(X_{iqk}, X_{i3k}) + (2-k)\phi_{pk}
\end{aligned}$$

$$
\begin{aligned}
&\times \operatorname{Cov}(X_{i(p+3)k}, X_{iqk}) - (k-1)(4-q)\\
&\times \theta \operatorname{Cov}(X_{i32}, X_{i42})], \quad p = 1,2,\ q = 3, \ldots, 6-k,\\
= &\frac{-1}{\phi_{pk}\phi_{qk}} \sum_{i=1}^{n} [\operatorname{Cov}(X_{ipk}, X_{iqk}) + (1-k)\lambda \operatorname{Var}(X_{i42})],\\
&p = 3, \ldots, 5-k,\ q = p+1, \ldots, 6-k,
\end{aligned}
$$

where the expectations are conditional on $(y_{i1k}, y_{i2k}; \phi_k)$, $k = 1, 2$ and are obtained using (2.5).

The inverses of the matrices defined by (3.1), (3.2) are required in every iteration of Louis (1982, Section 5) method which may be employed to accelerate the convergence of the EM algorithm. As pointed out in Meilijson (1989), Louis's method is a slight modification of the classical Newton-Raphson iterative scheme. The analytical work required is the same for both methods provided that Fisher's (1925) and Louis's identities are used for the computations of the observed-data scores and observed informations, so that tedious differentiations, as they are in our problems, are avoided. On the other hand, the conjugate gradient acceleration method of Jamshidian and Jennrich (1993) makes no reference to observed information matrices nor to their inverses. Every cycle consists of an EM step treated as generalized gradient and used as search direction. The accelerated estimate is effected by first modifying the direction of the EM step and then optimizing its length by a simple linear search. The computation of the new direction involves calculations of first derivatives of the observed-data log likelihood; for the problems considered here, these can be obtained by simple modifications of the EM equations. Specifically, in virtue of Fisher's (1925) identity the gradients of the observed-data log likelihoods $g(\phi_k)$, $k = 1, 2$ are found to be

$$\begin{aligned} g(\phi_{jk}) &= \frac{1}{\phi_{jk}} E(S_{jk}|Y_k^{obs}; \phi_k) - E(S_{(5-2k+j)k}|Y_k^{obs}; \phi_k) \\ &\quad j = 2k-1, \ldots, k+1, \\ &= \frac{1}{\phi_{jk}} E(S_{jk}|Y_k^{obs}; \phi_k) - n \\ &\quad j = 5-2k, 6-2k, 6-k, \end{aligned}$$

where the expectations are given by (2.6) for $k = 1, 2$.

If the conjugate gradient acceleration method (or the normal EM scheme) is prefered the observed Fisher information matrices given above need to be evaluated only at the last interation. In this case the maximum likelihood equations given by (2.7) can be used in (3.1), (3.2) to yield simpler expressions. The resulting matrices are inverted to give the asymptotic variance-covariance matrix of the maximum likelihood estimates. Of course, an alternative to be derivation of analytic forms of the observed informations matrices, is use of a numerical differentiation method; see for example Meilijson (1989), Meng and Rubin (1991).

4. Examples

Three examples serve as illustrations of the foregoing theory. In the first and third examples the data are observations on the number of accidents sustained by groups of 708 and 79 bus drivers respectively in two consecutive 2-year time periods; in the second example the data are the distribution of the number of plants of two species in each of 100 systematically laid out contiguous quadrats obtained in the course of a study of a secondary rain forest in Trinidad. The data can be found in Kocherlakota and Kocherlakota (1992, pp. 294, 295, 243) respectively and are not reproduced here.

In each of the examples the maximum likelihood estimates of parameters of the relevant distributions and matrix of their asymptotic variances and covariances based on the sampled data are obtained. The fit of the associated distribution is examined by the usual χ^2-test using the maximum likelihood estimates and row-by-row grouping to ensure that the expected number in any group does not fall below unity. The grouping schemes are similar to those in Kocherlakota and Korcherlakota (1992).

Initial values for the parameters were those provided by their moment estimates. The EM algorithm in its usual scheme was painfully slow in all cases considered. The convergence was accelerated substantially by the methods of conjugate gradients and Louis, starting after the first iteration. Following the suggestion of Jamshidian and Jennrich (1993), the line search in the former method was performed by secant iterations and in all methods convergence was assumed when the absolute differences between successive estimates were less than 10^{-5}. The total number of iterations to convergence for each method is reported below. Furthermore, following the suggestion of a referee we carried out direct Newton-Raphson on the log likelihoods from the observed data to compare its speed to Louis's method. We found that for our second example both methods converged in the same number of iterations while for the first and third examples the Newton Raphson method converged three iterations later.

Example 1: type I distribution, $n = 708$.

The normal EM procedure needed 2312 iterations to converge whereas the conjugate gradients and Louis's methods converged in 43 and 7 iterations respectively. The maximum likelihood estimates attained and their asymptotic variances and covariances are given below

$$\hat{\theta}_1 = .275, \hat{\theta}_2 = .239, \hat{\lambda} = .267, \hat{\lambda}_1 = 2.668, \hat{\lambda}_2 = 4.294,$$

$$
I(\hat{\phi}_1|Y_1^{obs})^{-1} = \begin{pmatrix} .0081 & .0003 & .0008 & -.0798 & -.0087 \\ & .0064 & .0005 & -.0045 & -.1153 \\ & & .0019 & -.0136 & -.0150 \\ & & & .8267 & .1387 \\ & & & & 2.145 \end{pmatrix}
$$

These resulted in a χ^2 value of 27.74 on 28 degrees of freedom (compared with 33.40 on 29 degrees of freedom reported in Kocherlakota and Kocherlakota using moment estimates).

Example 2: type I distribution, $n = 100$.

The normal EM procedure needed 564 iterations to converge whereas the conjugate gradients and Louis's methods converged in 36 and 5 iterations respectively. The maximum likelihood estimates attained and their asymptotic variances and covariances are given below

$$
\hat{\theta}_1 = .698, \hat{\theta}_2 = .309, \hat{\lambda} = .259, \hat{\lambda}_1 = .976, \hat{\lambda}_2 = 1.233,
$$

$$
I(\hat{\phi}_1|Y_1^{obs})^{-1} = \begin{pmatrix} .0755 & .0039 & .0026 & -.0993 & -.0240 \\ & .0683 & .0001 & -.0056 & -.2630 \\ & & .0052 & -.0075 & -.0091 \\ & & & .1593 & .0467 \\ & & & & 1.092 \end{pmatrix}
$$

These resulted in a χ^2 value of 9.63 on 8 degrees of freedom (compared with 12.33 on 8 degrees of freedom obtained by Gillings 1974, who fitted a bivariate Neyman Type A distribution using maximum likelihood estimates; see also Kocherlakota and Kocherlakota).

Example 3: type II distribution, $n = 79$.

The normal EM procedure needed 1172 iterations to converge whereas the conjugate gradients and Louis's methods converged in 19 and 5 iterations respectively. The maximum likelihood estimates attained and their asymptotic variances and covariances are given below

$$\hat{\lambda}_1 = 1.090, \hat{\lambda}_2 = 1.331, \hat{\theta}_1 = .358, \hat{\lambda} = 1.624,$$

$$I(\hat{\phi}_2|Y_2^{obs})^{-1} = \begin{pmatrix} .0425 & .0287 & .0343 & -.2359 \\ & .0456 & .0343 & -.2359 \\ & & .1568 & -.7953 \\ & & & 4.293 \end{pmatrix}$$

These resulted in a χ^2 value of 14.69 on 17 degrees of freedom (compared with 20.57 on 18 degrees of freedom reported in Kocherlakota and Kocherlakota).

Acknowledgements

We would like to thank the referees for their comments and Professor T. Papaioannou for helpful discussions.

References

[1] Cresswell, W.L. ad Froggatt, P. (1963), *The causation of bus driver accidents: an epidemiological study*, Oxford University Press, London.

[2] Dempster, A.P., Laird, N. M. and Rubin, D. B. (1977), *Maximum likelihood from incomplete data via the EM algorithm (with discussion)*, J. Roy. Statist. Soc. B, **39**, 1-38.

[3] Fisher, R. A. (1925), *Theory of statistical estimation*, Proc. Camb. Philos. Soc., **22**, 700-725.

[4] Gillings, D.B. (1974), *Some further results for bivariate generalizations of the Neyman Type A distribution*, Biometrics, **30**, 619-628.

[5] Irwin, J. O. (1964), *The personal factor in accidents*, J. Roy. Statist. Soc. A, **127**, 438-451.

[6] Jamshidian, M. and Jennrich, R. I. (1993), *Conjugate gradient acceleration of the EM algorithm*, J. Amer. Statist. Assoc., **88**, 221-228.

[7] Kemp, C. D. (1967), *On a contagious distribution suggested for accident data*, Biometrics, **23**, 241-255.

[8] Kemp, C. D. (1970), *'Accident proneness' and discrete distribution theory*, Random counts in scientific work 2 (G. P. Patil Ed.), The Pennsylvania State University Press, Pennsylvania, 41-65.

[9] Kerr, J. D. (1969), *The estimation of the 'Short' distribution*, Biometrics, **25**, 417-420.

[10] Kocherlakota, S. and Kocherlakota, K. (1992), *Bivariate discrete distributions*, Marcel Dekker Inc., New York.

[11] Louis, T.A. (1982), *Finding the observed information when using the EM algorithm*, J. Roy. Statist. Soc. B, **44**, 226-233.

[12] Marshall, A. W. and Olkin, I. (1988), *Families of multivariate distributions*, J. Amer. Statist. Assoc., **83**, 834-841.

[13] Meilijson, I. (1989), *A fast improvement to the EM algorithm on its own terms*, J. Roy. Statist. Soc. B, **51**, 127-138.

[14] Meng, X. and Rubin, D. B. (1991), *Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm*, J. Amer. Statist. Assoc., **86**, 899-909.

[15] Papageorgiou, H. (1986), *Bivariate 'Short' distributions*, Commun. Statist. Theory Meth., **15**, 893-906.

Received by S. Kocherlakota *on* June, 1994. Revised March 1995 .

Refereed by Ramesh C. Gupta, Dept. of Mathematics, Univ. of Maine, 5752 Neville Hall, Orono, Maine, U.S.A. 04469-5752 and Mortaza Jamshidian, School of Mathematics, Isfahan University of Technology, Isfahan, Iran. .

Accepted on June 05, 1995