Ann. Inst. Statist. Math.
Vol. 55, No. 4, 817–830 (2003)
©2003 The Institute of Statistical Mathematics

ML ESTIMATION FOR MULTIVARIATE SHOCK MODELS VIA AN EM ALGORITHM

DIMITRIS KARLIS

Department of Statistics, Athens University of Economics and Business, 76, Patission Str., 10434 Athens, Greece, e-mail: karlis@hermes.aueb.gr

(Received January 29, 2001; revised August 13, 2002)

Abstract. Multivariate extensions of univariate distributions, though useful, have not been applied in practice mainly due to shortage of inferential procedures caused by numerical complexity. The multivariate Marshall-Olkin distribution is a multivariate extension of the exponential distribution. Its representation as a multivariate shock model makes it appealing for such applications. Unfortunately, ML estimation is not easy and special numerical techniques are needed. In this paper an EM type algorithm based on the multivariate reduction technique is described. The behavior of the algorithm is examined and a numerical example is provided.

Key words and phrases: Multivariate reduction, Marshall-Olkin distribution, maximum likelihood estimation.

1. Introduction

Multivariate extensions of univariate distributions, though useful, have not been applied in practice mainly due to shortage of inferential procedures caused by numerical complexity. Moreover, generalization of univariate models is not straightforward in the sense that certain desirable properties may hold for more than one multivariate model.

It is common to consider as multivariate extensions of univariate distributions, multivariate distributions with marginals in the given family. Statistical inference is not easy even for a model where the univariate case is trivial. This leads to restricted usage of multivariate models, even though they can capture more information from the data.

A variety of multivariate extensions of the exponential distribution have been considered at the past. These include the distributions of Gumbel (1960), Freund (1961) and its extensions discussed in Heinrich and Jensen (1995), the Marshall-Olkin (1967) multivariate exponential distributions, the distributions proposed in Dowton (1970) as well as the distribution of Block and Basu (1974).

Among them the multivariate Marshall-Olkin (MMO) distribution plays an important role since it is the only multivariate distribution with exponential marginals that fulfils a multivariate lack of memory property. The MMO distribution is not absolutely continuous. Unfortunately, statistical inference for the MMO is not an easy task due to the complicated nature of its density function.

In this article, Maximum Likelihood (ML) estimation of the parameters of the MMO distribution is considered. This is implemented via an EM algorithm that makes use of the multivariate reduction technique that leads to the MMO distribution. The multivariate reduction technique is described in Section 2. For ease of exposition, the bivariate

Marshall-Olkin (BMO) distribution is examined analytically, in Sections 3–5. By treating the simplest case in detail, the derivation of the algorithm can be more clearly presented and understood, while circumventing the considerable notational problems of the general case. Section 3 briefly reviews the BMO distribution, while Section 4 deals with existing estimation methods. In Section 5 a full derivation of the EM algorithm for the BMO is provided. Section 6 presents the case when a censored version of the BMO distribution arises. Section 7 contains a generalization of the algorithm to the general multivariate case, while in Section 8 a detailed application of the algorithm in simulated data from a trivariate distribution can be found. Cases with restricted parameters are also discussed. Finally, concluding remarks can be found in Section 9.

2. Multivariate reduction technique

Multivariate reduction is an appealing technique, used in many contexts, for constructing multivariate distributions. The idea is to create a number of dependent random variables Y_i, $i = 1, \dots, m$ from a number of independent random variables X_i, $i = 1, \dots, k$.

Let $X_i \sim F(x_i; \lambda_i), i = 0, 1, \dots, m$. A case of special interest arises if a mapping g exists such that

$$Y_i = g(X_0, X_i) \sim F(y_i; \lambda_0 + \lambda_i), \qquad i = 1, \dots, m$$

and $(Y_1, \dots, Y_m)$ has a multivariate extension of the $F(x; \lambda)$ distribution. In this case $\{F(x; \lambda) : \lambda > 0\}$ is said to form an additive family under the mapping g. The main point of such a construction is that the marginal distributions belong to the family $F(x; \lambda)$ and thus the multivariate distribution can be regarded as a multivariate version of the univariate distribution.

Mardia (1970) showed that multivariate reduction construction of multivariate distributions leads to positive correlation between Y_i and Y_j, $i \neq j = 1, \dots, m$. Some common distributions allow for a mapping g for constructing multivariate analogues. So, the function $g(X_0, X_i) = X_0 + X_i$ forms an additive family for the Poisson and the normal distributions and thus multivariate distributions can be constructed in this way. In fact, the derivation of the multivariate Poisson distribution via multivariate reduction is well known (see, e.g. Johnson *et al.* (1997)). In a similar fashion one can create multivariate Gamma distributions (see, e.g. Cherian (1941)), while the mapping $g(X_0, X_i) = \min(X_0, X_i)$ is additive for the exponential distribution.

Relaxing the assumption that all X_i's belong to the same family, a wealth of multivariate distributions with specific, marginal distributions can be constructed. For example, if X_i, $i = 1, 2$ follow Poisson distributions, while X_0 follows a binomial distribution, the bivariate Charlier Series distribution is obtained, which has marginal distributions that are the convolutions of a Poisson and a binomial random variables (see, e.g. Papageorgiou and Loukas (1995)). Moreover, Stein and Juritz (1987) discussed bivariate mixed Poisson distributions constructed via trivariate reduction, based on mixing distributions that are reproductive under addition. The MMO (Marshall and Olkin (1967)) distribution arises by a multivariate reduction with mapping $g(X_0, X_i) = \min(X_0, X_i)$, and multivariate analogues of the Weibull, the geometric and the Pareto distributions can be derived. Generalized trivariate reduction schemes are described in Zheng and Matis (1993), Heinrich and Jensen (1995) and Lai (1995).

3. The bivariate Marshall-Olkin distribution

Consider the simple exponential distribution with density $f(x) = \theta \exp(-\theta x)$, $x, \theta > 0$, denoted as $Expo(\theta)$. Marshall and Olkin (1967) proposed a bivariate extension of the exponential distribution. The derivation of this distribution is based on the following scheme. Suppose that X_0, X_1, X_2 are independent random variables and $X_i \sim Expo(\theta_i)$. Then the random variables $Y_1 = \min(X_0, X_1)$ and $Y_2 = \min(X_0, X_2)$ jointly follow the bivariate Marshall-Olkin (BMO) distribution and their joint density function is given by

$$f(y_1, y_2) = \begin{cases} \theta_1(\theta_2 + \theta_0)S(y_1, y_2) & \text{if} \quad y_2 > y_1 > 0 \\ \theta_2(\theta_1 + \theta_0)S(y_1, y_2) & \text{if} \quad y_1 > y_2 > 0 \\ \theta_0 S(y_1, y_2) & \text{if} \quad y_1 = y_2 > 0 \end{cases} \tag{3.1}$$

where $S(y_1, y_2) = P(Y_1 > y_1, Y_2 > y_2) = \exp(-\theta_1 y_1 - \theta_2 y_2 - \theta_0 \max(y_1, y_2))$ is the joint survival function. The marginal distributions are $Expo(\theta_1\theta_0/(\theta_1 + \theta_0))$ and $Expo(\theta_2\theta_0/(\theta_2 + \theta_0))$ respectively and $\theta_0 = 0$ corresponds to the case where Y_1 and Y_2 are independent. The correlation coefficient is θ_0/θ, where $\theta = \theta_0 + \theta_1 + \theta_2$. Essentially θ_0 represents the increase in the failure rates of Y_1 and Y_2.

The BMO distribution is not absolutely continuous with respect to the Lebesgue measure in $\Re^2$ and has a singular part on the diagonal $x = y \geq 0$. Note also that $P(X = Y) \geq 0$ and this singular component has a density function with respect to the Lebesgue measure in $\Re^1$. A natural interpretation of the BMO distribution is based on shock failures where there exists positive probability of simultaneous failure of exponential type for both components (see, e.g. Shamseldin and Press (1984)).

The BMO distribution is the only bivariate distribution with exponential marginal distributions satisfying the bivariate lack of memory property (Basu (1995), p. 327). Moreover, Brindley and Thompson (1972) showed that the BMO distribution and its multivariate extensions form the boundary between increasing and decreasing multivariate failure rate distributions.

4. Estimation of the parameters

Estimation procedures for the parameters of the BMO distribution have been considered by many authors. Bemis *et al.* (1972) described moment estimation for the parameters of the BMO distribution by equating the marginal means and the correlation coefficient with their sample counterparts.

Other types of estimators, the INT estimator that arise from an one step ahead iteration towards the ML estimates, have been considered by Proschan and Sullo (1976). Since, as shown in the sequel, the ML estimates cannot be obtained in closed form expressions, iterative numerical techniques have to be used. They showed that the asymptotic efficiency of the INT estimators coincides with that of the ML estimates and they are superior to moment estimates.

Let n_1, n_2, n_3 denote the number of observations for which $y_2 > y_1$, $y_2 < y_1$ and $y_1 = y_2$, respectively, and $n = n_1 + n_2 + n_3$. The log-likelihood function for a given sample of observations can be written as

$$\begin{aligned} L = & -\theta_1 \sum_{i=1}^{n} y_{1i} - \theta_2 \sum_{i=1}^{n} y_{2i} - \theta_0 \sum_{i=1}^{n} \max(y_{1i}, y_{2i}) \\ & + n_1 \ln[\theta_1(\theta_2 + \theta_0)] + n_2 \ln[\theta_2(\theta_1 + \theta_0)] + n_3 \ln(\theta_0) \end{aligned}$$

which is valid for all possible values of θ_i and n_i, if we adopt the convention that $0 \ln 0 = 0$, $c \ln 0 = -\infty (c > 0)$, $\exp(-\infty) = 0$.

The resulting system of equations cannot be solved in closed form expressions and numerical techniques are required (see, Bemis *et al.* (1972), Klein (1995), p. 334). If all $n_i > 0$ the above system has a unique solution. If $n_3 = 0$ and either $n_1 = 0$ or $n_2 = 0$ the MLE exists but it is not unique, while if $n_3 > 0$ and either $n_1 = 0$ or $n_2 = 0$ the MLE does not exist. Note that $P(n_i = 0) = [1 - \theta_i/\theta]^n$ and thus for moderate sample sizes with $\theta_i > 0$ one expects to obtain observations of all the types. The ML estimates are in general biased. Bemis *et al.* (1972) proposed using $n\theta_i/(n+1)$ to reduce the bias. Proschan and Sullo (1976) discussed ML estimation for multivariate Marshall-Olkin distributions. Similar results were obtained. Various testing hypotheses are described in Bhattacharyya and Johnson (1973), Hanagal and Kale (1992*a*, 1992*b*) and Hanagal (1992). The ML estimates are consistent, asymptotically efficient and have asymptotic trivariate normal distribution.

5. An EM algorithm for ML estimation

It is clear that the trivariate reduction technique used for the derivation of the BMO distribution, allows for a missing data interpretation, suitable for the application of the EM algorithm. A detailed description of the general EM algorithm can be found in the book of McLachlan and Krishnan (1997).

According to the multivariate reduction derivation, the missing data consist of the non-observable random variables $X_i = (X_{1i}, X_{2i}, X_{0i})$ while the observed data are the values $Y_i = (Y_{1i}, Y_{2i})$.

The EM algorithm proceeds by calculating the conditional expectation of X_i given Y_i and the current values of the parameters, while the M-step just calculates the ML estimates for a sample from exponential distributions, using the expectations of the E-step.

In the next section we fully describe the derivation of the required conditional expectations.

5.1 *The conditional expectations*

For notational convenience we drop the subscript i denoting the observation. Let $s = \min(y_1, y_2)$, $t = \max(y_1, y_2)$ and $\Theta = (\theta_0, \theta_1, \theta_2)$ denote the vector of parameters. Also, let $f_j(x)$ and $F_j(x)$, $j = 0, 1, 2$ denote the density and the survival function respectively of the random variable X_j.

Suppose further that $y_1 < y_2$. It is clear that $X_0 \geq t$. Moreover the conditional density of X_0 is given by:

$$f_0(t \mid y_1, y_2, \Theta) = \frac{f_1(s) f_0(t) F_2(t)}{f(y_1, y_2)} = \frac{\theta_0}{\theta_2 + \theta_0}$$

and similarly for $w > t$

$$f_0(w \mid y_1, y_2, \Theta) = \frac{f_1(s) f_0(w) f_2(t)}{f(y_1, y_2)} = \frac{\theta_0 \theta_2}{\theta_2 + \theta_0} \exp(-\theta_0(w - t)).$$

Thus we obtain that

$$(5.1) \qquad f_0(w \mid y_1, y_2, \Theta) = \begin{cases} \dfrac{\theta_0}{\theta_2+\theta_0} & \text{if} \quad w = t \\ \dfrac{\theta_0\theta_2}{\theta_2+\theta_0}\exp(-\theta_0(w-t)) & \text{if} \quad w > t. \end{cases}$$

Thus the conditional expectation can be derived easily as

$$E(X_0 \mid y_1, y_2, \Theta) = \frac{\theta_0}{\theta_2+\theta_0}\left(t + \theta_2 \int_t^\infty w \exp(-\theta_0(w-t))dw\right) = t + \frac{\theta_2}{(\theta_2+\theta_0)}\frac{1}{\theta_0},$$

because the integral merely represents the expectation of a left-truncated at t exponential random variable. In addition it is clear that since $y_1 < y_2$, $E(X_1 \mid y_1, y_2, \Theta) = y_1$ and with arguments similar to those used for the derivation of $E(X_0 \mid y_1, y_2, \Theta)$ one can show that

$$E(X_2 \mid y_1, y_2, \Theta) = t + \frac{\theta_0}{(\theta_2+\theta_0)}\frac{1}{\theta_2}.$$

Moreover, assuming that $y_1 > y_2$ one can derive similar expressions, as those appearing in Table 1.

Consider now the case where $y_1 = y_2$. In this case it is clear that $E(X_0 \mid y_1, y_2, \Theta) = y_1$. Moreover, one obtains that

$$f_1(w \mid y_1, y_2, \Theta) = \frac{f_1(w)f_0(y_1)F_2(y_1)}{f(y_1, y_2)} = \theta_1 \exp(-\theta_1(w - y_1)),$$

$w \geq y_1$ which is the density of a left truncated at y_1 exponential distribution with parameter θ_1. Therefore $E(X_1 \mid y_1 = y_2, \Theta) = y_1 + \frac{1}{\theta_1}$ and in a similar fashion $E(X_2 \mid y_1 = y_2, \Theta) = y_1 + \frac{1}{\theta_2}$. Table 1 summarizes the conditional expectations.

Table 1. The conditional expectations for all cases.

	$E(X_1 \mid y_1, y_2, \Theta)$	$E(X_2 \mid y_1, y_2, \Theta)$	$E(X_0 \mid y_1, y_2, \Theta)$
$y_1 < y_2$	y_1	$y_2 + \frac{\theta_0}{\theta_2+\theta_0}\frac{1}{\theta_2}$	$y_2 + \frac{\theta_2}{\theta_2+\theta_0}\frac{1}{\theta_0}$
$y_1 > y_2$	$y_1 + \frac{\theta_0}{\theta_1+\theta_0}\frac{1}{\theta_1}$	y_2	$y_1 + \frac{\theta_1}{\theta_1+\theta_0}\frac{1}{\theta_0}$
$y_1 = y_2$	$y_1 + \frac{1}{\theta_1}$	$y_2 + \frac{1}{\theta_2}$	y_1

5.2 *The algorithm*

From the aforementioned derivations one can describe the EM algorithm as follows:

E-step: Using the data and the current estimates $\Theta^{(k)} = ({\theta_0}^{(k)}, {\theta_1}^{(k)}, {\theta_2}^{(k)})$ after the k-th iteration, calculate the pseudo-values

$$z_i = E(X_{0i} \mid y_{1i}, y_{2i}, \Theta^{(k)}), \qquad d_i = E(X_{1i} \mid y_{1i}, y_{2i}, \Theta^{(k)}), \qquad \text{and}$$
$$\phi_i = E(X_{2i} \mid y_{1i}, y_{2i}, \Theta^{(k)})$$

for $i = 1, \ldots, n$ (using the entries of Table 1).

M-step: Update the estimates by

$${\theta_0}^{(k+1)} = \frac{n}{\sum_{i=1}^n z_i} \qquad {\theta_1}^{(k+1)} = \frac{n}{\sum_{i=1}^n d_i} \qquad {\theta_2}^{(k+1)} = \frac{n}{\sum_{i=1}^n \phi_i}.$$

If some convergence criterion is satisfied stop iterating otherwise go back to the E-step.

Note that the M-step simply obtains the ML estimates from a random sample from the exponential distribution using the pseudo-values calculated at the E-step.

It is interesting to note that the algorithm can be extended easily to cover the case of covariates at the parameters. Assume that the parameter θ_i, $i = 0, 1, 2$ is associated with a vector of covariates, say z_i, through a log-link function, namely $\theta_i = \exp(\beta_i' z_i)$, $i = 0, 1, 2$, where β_i is a vector of regression coefficients associated to the vector of regressors z_i. Note that z_i may differ among parameters. In such a case, the only amendment in the EM algorithm described above is that at the M-step, the parameter vectors β_i are updated by simply fitting exponential regression models, using as dependent values the pseudo-values of the E-step. This can be done easily by standard statistical packages.

5.3 *Operating characteristics*

The aim of this section is to examine some characteristics of the proposed algorithm. Firstly, the choice of efficient initial values is treated. We conducted a small simulation experiment for examining the plausibility of the moment estimates (MOM) and the INT estimates as initial values for the EM algorithm. Both estimates are given in closed form expressions and thus they can be easily calculated. Table 2 reports the mean number of iterations needed until convergence, when the two different estimates were used as initial values. The convergence criterion used was to stop iterating when the ratio $|(L(k) - L(k-1))/L(k-1)| < tol$, where $L(k)$ denotes the value of the loglikelihood at the k-th iteration and $tol = 10^{-12}$.

Various combinations of parameters, representing situations with large correlation as well as symmetric and asymmetric cases (with respect to θ_1 and θ_2) were selected. Four different sample sizes were used, namely $n = 50, 100, 250, 500$. For each sample size and parametrization 5000 replications were used. Only samples for which all $n_i > 0$ were kept in the analysis as otherwise the ML estimates do not exist or they are inconsistent.

It is evident from Table 2 that the INT estimates are preferable as initial values for all the configurations and sample sizes. As the sample size increases the difference gets smaller but even for the largest sample size ($n = 500$) there is still substantial difference. This indicates that the INT estimators are superior as initial values.

Table 2. The mean number of iterations until convergence for combinations of sample size and parameter values.

			sample size n							
parameter values			50		100		250		500	
θ_1	θ_2	θ_0	MOM	INT	MOM	INT	MOM	INT	MOM	INT
1	1	1	27.9	24.6	24.7	21.9	21.6	19.4	19.9	17.7
2	1	0.5	72.7	49.6	62.1	42.0	54.5	36.9	50.2	33.3
2	1	1	42.1	33.0	37.8	29.1	34.1	26.3	31.8	24.0
2	1	2	36.7	31.0	32.4	26.9	28.3	23.5	26.8	21.8
3	1	0.5	108.2	66.9	91.5	56.7	81.3	48.1	75.7	43.9
3	1	1	62.9	44.6	57.6	40.5	49.4	35.0	47.5	32.1
3	1	2	46.8	36.3	41.1	32.0	36.5	28.2	34.1	25.8
5	1	1	113.4	69.0	96.4	59.4	83.2	51.0	77.5	46.1

This was expected since the derivation of the INT estimators was based on an one-step iteration towards the ML estimator. Note that the stopping rule used was rather strict. Relaxing this criterion, by using a less strict one, like $tol = 10^{-6}$, we got the surprising result that the INT estimators led to ML estimators after quite a few iterations, usually less than 5 for all configurations and sample sizes, as one can see in Table 3. This implies that the INT estimators are quite close to the ML estimators. As the sample size gets larger the number of iterations is reduced.

The above clearly implies that the INT estimates are quite satisfactory for estimation purposes. The relative difference in the loglikelihood calculated for the INT estimates and the maximized loglikelihood was found to be less than 10^{-5}.

Table 3. The mean number of iterations until convergence when the INT estimators were used as initial values for a less strict terminating condition ($tol < 10^{-6}$).

initial values			n			
θ_1	θ_2	θ_0	50	100	250	500
1	1	1	3.25	2.53	1.752	1.308
2	1	0.5	2.082	1.564	1.112	1.018
2	1	1	2.518	1.8	1.288	1.076
2	1	2	2.812	1.98	1.408	1.162
3	1	0.5	1.542	1.212	1.02	1.002
3	1	1	1.902	1.446	1.102	1.018
3	1	2	2.33	1.612	1.234	1.046
5	1	1	1.35	1.168	1.012	1

As far as the convergence is concerned, the algorithm converged after rather few iterations for all the cases. In addition, no multiple solutions were found. Stopping criteria that are not based on the likelihood could be used, like the relative change on the parameter values at successive iterations. However the likelihood can be easily obtained and thus there is no reason for not using it.

6. A censored model

The model described so far, assumes that the two components work in parallel. Considering a model where the two components are connected in series, a censored BMO model arises. In this model the fatal shock may arise from the third process but whenever one component fails the other fails, too. If two components are connected in series, following Pena and Gupta (1990), then the random vector observed on system failure is (Z, δ_1, δ_2), where $Z = \min(Y_1, Y_2)$, $\delta_1 = I(Y_1 < Y_2)$ and $\delta_2 = I(Y_2 < Y_1)$, where $I(A)$ denotes the indicator function of the event A.

One can see that this model implies that one of the two variables of the BMO distribution is censored. The random variable Z follows an $Expo(\theta)$ distribution, while the joint density of the vector (Z, δ_1, δ_2) can be written with respect to the product of the Lebesgue measure on $\Re^+$ and the counting measure on $\mathcal{M} = \{(0,0),(1,0),(0,1)\}$ as

$$f(z, d_1, d_2) = \theta \exp(-\theta z) \left(\frac{\theta_1}{\theta}\right)^{d_1} \left(\frac{\theta_2}{\theta}\right)^{d_2} \left(\frac{\theta_0}{\theta}\right)^{1-d_1-d_2}.$$

Note that the above density can be retrieved as a censored version of the BMO distribution. Direct maximization of the likelihood is not easy but an EM algorithm is applicable. The algorithm is similar to the one used for the uncensored case. The expectations needed are analogous to those in Table 1. The only change is that now we have observed simply z_i which has to substitute in all cases y_1 and y_2 in Table 1. Thus, the conditional expectations can now be seen in Table 4.

Table 4. The conditional expectations for the case of censored observations (model with components in series).

	$E(x_1 \mid z, \delta_1, \delta_2, \Theta)$	$E(x_2 \mid z, \delta_1, \delta_2, \Theta)$	$E(x_0 \mid z, \delta_1, \delta_2, \Theta)$
$\delta_1 = 1,\, \delta_2 = 0$	z	$z + \frac{\theta_0}{\theta_2+\theta_0}\frac{1}{\theta_2}$	$z + \frac{\theta_2}{\theta_2+\theta_0}\frac{1}{\theta_0}$
$\delta_1 = 0,\, \delta_2 = 1$	$z + \frac{\theta_0}{\theta_1+\theta_0}\frac{1}{\theta_1}$	z	$z + \frac{\theta_1}{\theta_1+\theta_0}\frac{1}{\theta_0}$
$\delta_1 = 0,\, \delta_2 = 0$	$z + \frac{1}{\theta_1}$	$z + \frac{1}{\theta_2}$	z

The steps of the EM algorithm are similar with those described for the uncensored case. The only difference is that now the expectations of Table 4 are used.

Pena and Gupta (1990) studied Bayesian estimation for both models (uncensored or censored), corresponding to models with the two components in parallel or in series. They concluded that the case of parallel components could be considerably more efficient than series sample estimates, but from a practical point of view they could be more costly, in order to infer for the parameters of the underlying processes. Lu (1997), Chen and Lu (1998) and Chen *et al.* (2000) exploited the case of alternative models that could balance efficiency and cost.

7. Multivariate Marshall-Olkin distributions

7.1 *Derivation and properties*

Multivariate extension of the BMO distribution can be made via two different approaches. The first one uses the fatal shock derivation of the bivariate model. The idea is that there are m processes along with another, say $(m+1)$-th process that whenever fails all the rest m processes fail, too. Thus the vector $Y = (Y_1, \ldots, Y_m)$ is defined as $Y = (X_0, \ldots, X_0)$, if $X_0 < X_i$, for all $i = 1, \ldots, m$, and $Y = (X_1, X_2, \ldots, X_m)$ otherwise.

Another definition, more consistent with the multivariate reduction technique proposed above, defines the new variables $Y_i = \min(X_i, X_0)$. Such a definition allows for random vectors with some of their elements equal, while the fatal shock model does not allow for ties between subsets of the variables. To this extent, the second model can be considered as more general. In the sequel only the second model will be considered.

The survival function is given as

$$S(y_1, \ldots, y_m) = \exp\left(-\sum_{i=1}^{m} \theta_i y_i - \theta_0 \max(y_1, \ldots, y_m) \right).$$

The joint density function is not absolutely continuous with respect to the Lebesgue measure in $\Re^m$ and has singularities corresponding to the cases where two or more of the y_i's are equal. Thus the density can be written as

$$(7.1)\qquad f(y_1,\dots,y_m)=\begin{cases} S(y_1,\dots,y_m)(\theta_j+\theta_0)\prod_{\substack{i=1\\ i\neq j}}^{m}\theta_i \\ \quad \text{if} \quad y_j=\max(y_1,\dots,y_m),\ j=1,\dots,m \text{ and} \\ \quad y_i\neq y_j, i\neq j \\ S(y_1,\dots,y_m)\theta_0\theta_{i_1}\theta_{i_2}\cdots\theta_{i_k} \\ \quad \text{if} \quad y_{i_1},y_{i_2},\dots,y_{i_k}<y_{j_1}=y_{j_2}=\cdots=y_{j_p} \\ S(y_1,\dots,y_m)\theta_0 \\ \quad \text{if} \quad y_1=y_2=\cdots=y_m>0. \end{cases}$$

It can be verified that the MMO distribution satisfies a multivariate lack of memory property.

7.2 *Derivation of an EM algorithm*

In a similar manner, we can construct an EM algorithm for ML estimation of the parameters of the MMO distribution. Proschan and Sullo (1976) discussed estimation for the MMO distribution. For constructing the EM algorithm, one needs to calculate a series of conditional expectations. Since now the form of the joint density is quite complicated, with a large number of distinct cases, one has to separate all the possible cases. Let Y denote the observed data and Θ the parameters.

Let us start with the simplest case where $y_1=y_2=\cdots=y_m$, i.e. all the values are equal implying the existence of a shock. Therefore $E(X_0\mid Y,\Theta)=y_1$,while $E(X_i\mid Y,\Theta)=y_1+\theta_i{}^{-1}$, $i=1,\dots,m$, corresponding to the expectation of a left-truncated at y_1 exponential distribution for each variable.

The more complicated case is when some of the y_i's are equal but there are some other with smaller values. More formally, $y_{i_1},y_{i_2},\dots,y_{i_k}<y_{j_1}=y_{j_2}=\cdots=y_{j_p}=y^{(0)}$ for some k and p. Clearly for all the subscripts i_b, $b=1,\dots,k$ it holds that $E(X_{i_b}\mid Y,\Theta)=y_{i_b}$ because the observed value corresponds to the unobserved one. For the rest, where the equality holds, it is obvious that the corresponding X_0 is the observed value $y^{(0)}$ and thus $E(X_0\mid Y,\Theta)=y^{(0)}$ while the rest are the expectations from a left-truncated at $y^{(0)}$ exponential distribution and thus $E(X_{j_b}\mid Y,\Theta)=y^{(0)}+\theta_{j_b}{}^{-1}$, $b=1,\dots,p$. The last case is when there is some y_k which is larger than the rest. This implies that the information for X_0 is contained only in this value, and thus $E(X_i\mid Y,\Theta)=y_i$, $i=1,\dots,m$ and $i\neq k$. Then as shown for the bivariate case, it holds

$$E(X_0\mid Y,\Theta)=y_k+\frac{\theta_k}{(\theta_k+\theta_0)}\frac{1}{\theta_0},\qquad \text{and}\qquad E(X_k\mid Y,\Theta)=y_k+\frac{\theta_0}{(\theta_k+\theta_0)}\frac{1}{\theta_k}.$$

7.3 *The EM algorithm*

The conditional expectations needed for the E-step have been defined and, now, the algorithm can be described as:

E-step: Calculate the pseudo-values t_{ji} corresponding to the j-th non observable variable of the i-th observation as

$$t_{ji}=E(X_{ji}\mid Y_i,\Theta^{(k)}),\qquad i=1,\dots,n,\ j=0,\dots,m.$$

M-step: Update the parameters by

$$\theta_j^{(k+1)} = \frac{n}{\sum_{i=1}^n t_{ji}}, \qquad j = 0, \ldots, m.$$

If some criterion is satisfied then stop iterating otherwise go back to the E-step for one more iteration.

Again, one may add covariates to each θ_i, as described for the bivariate case.

8. Illustrative example

In order to illustrate the EM type algorithm derived above, we use a simulated data set from a trivariate Marshall-Olkin distribution. The joint density function of the trivariate Marshall-Olkin distribution, in accordance with (7.1) can be described as

$$f(y_1, y_2, y_3) = \begin{cases} S(y_1, y_2, y_3)\theta_2\theta_3(\theta_1 + \theta_0) & \text{if } y_1 > y_2, y_3 \\ S(y_1, y_2, y_3)\theta_1\theta_3(\theta_2 + \theta_0) & \text{if } y_2 > y_1, y_3 \\ S(y_1, y_2, y_3)\theta_1\theta_2(\theta_3 + \theta_0) & \text{if } y_3 > y_1, y_2 \\ S(y_1, y_2, y_3)\theta_0\theta_1 & \text{if } y_1 < y_2 = y_3 \\ S(y_1, y_2, y_3)\theta_0\theta_2 & \text{if } y_2 < y_1 = y_3 \\ S(y_1, y_2, y_3)\theta_0\theta_3 & \text{if } y_3 < y_1 = y_2 \\ S(y_1, y_2, y_3)\theta_0 & \text{if } y_1 = y_2 = y_3 > 0 \end{cases} \tag{8.1}$$

where $S(y_1, y_2, y_3) = \exp(-\sum_{i=1}^3 \theta_i y_i - \theta_0 \max(y_1, y_2, y_3))$.

The conditional expectations needed for the EM algorithm can be found in Table 5. For convenience we have separated the different types of data that may occur.

Using the expectations of Table 5 the EM algorithm for the trivariate Marshall-Olkin distribution is described as

E-step: Using the data Y and the current estimates $\Theta^{(k)} = (\theta_0{}^{(k)}, \theta_1{}^{(k)}, \theta_2{}^{(k)}, \theta_3{}^{(k)})$ after the k-th iteration calculate the pseudo-values

$$t_{ji} = E(X_{ji} \mid Y_i, \Theta^{(k)}), \qquad i = 1, \ldots, n, \ j = 0, 1, 2, 3.$$

M-step: Update the estimates by

$$\theta_j^{(k+1)} = \frac{n}{\sum_{i=1}^n t_{ji}}, \qquad j = 0, 1, 2, 3.$$

Table 5. The conditional expectations needed for ML estimation for the trivariate Marshall-Olkin distribution.

Type		$E(X_1 \mid y, \Theta)$	$E(X_2 \mid y, \Theta)$	$E(X_3 \mid y, \Theta)$	$E(X_0 \mid y, \Theta)$
I	$y_1 > y_2, y_3$	$y_1 + \frac{\theta_0}{\theta_1+\theta_0}\frac{1}{\theta_1}$	y_2	y_3	$y_1 + \frac{\theta_1}{\theta_1+\theta_0}\frac{1}{\theta_0}$
II	$y_2 > y_1, y_3$	y_1	$y_2 + \frac{\theta_0}{\theta_2+\theta_0}\frac{1}{\theta_2}$	y_3	$y_2 + \frac{\theta_2}{\theta_2+\theta_0}\frac{1}{\theta_0}$
III	$y_3 > y_1, y_2$	y_1	y_2	$y_3 + \frac{\theta_0}{\theta_3+\theta_0}\frac{1}{\theta_3}$	$y_3 + \frac{\theta_3}{\theta_3+\theta_0}\frac{1}{\theta_0}$
IV	$y_1 < y_2 = y_3$	y_1	$y_2 + \frac{1}{\theta_2}$	$y_3 + \frac{1}{\theta_3}$	y_2
V	$y_2 < y_1 = y_3$	$y_1 + \frac{1}{\theta_1}$	y_2	$y_3 + \frac{1}{\theta_3}$	y_1
VI	$y_3 < y_1 = y_2$	$y_1 + \frac{1}{\theta_1}$	$y_2 + \frac{1}{\theta_2}$	y_3	y_1
VII	$y_1 = y_2 = y_3$	$y_1 + \frac{1}{\theta_1}$	$y_2 + \frac{1}{\theta_2}$	$y_3 + \frac{1}{\theta_3}$	y_1

If some convergence criterion is satisfied stop iterating otherwise go back to the E-step.

Consider the data of Table 6 generated from a trivariate Marshall-Olkin distribution. All the four parameters $(\theta_0, \theta_1, \theta_2, \theta_3)$ used to generate the data were set equal to 1. In addition, in Table 6 one can see the Type for each observation (the Type is defined according to the notation used in Table 5), as well as the conditional expectations for the unobserved latent variables. These expectations have been calculated after the termination of the algorithm, and they correspond to the pseudo-values of the E-step.

Table 6. Generated data from a trivariate Marshall-Olkin distribution.

Data				Conditional expectations after the last iteration			
Y_{1i}	Y_{2i}	Y_{3i}	Type	t_{1i}	t_{2i}	t_{3i}	t_{0i}
1.597	1.597	0.150	VI	2.821	2.795	0.150	1.597
1.299	0.398	1.144	I	1.930	0.398	1.144	1.856
0.745	0.745	0.745	VII	1.969	1.943	1.461	0.745
1.227	0.173	1.227	V	2.451	0.173	1.943	1.227
0.086	1.000	0.255	II	0.086	1.612	0.255	1.563
0.360	0.169	0.331	I	0.991	0.169	0.331	0.917
1.400	1.110	0.764	I	2.031	1.110	0.764	1.957
0.192	1.276	0.730	II	0.192	1.888	0.730	1.840
0.024	0.024	0.024	VII	1.248	1.222	0.740	0.024
0.708	0.119	0.190	I	1.339	0.119	0.190	1.266
1.959	1.941	0.692	I	2.590	1.941	0.692	2.516
0.430	0.430	0.256	VI	1.654	1.628	0.256	0.430
0.261	0.345	0.584	III	0.261	0.345	0.859	1.293
0.384	0.440	0.046	II	0.384	1.052	0.046	1.003
0.002	1.218	0.391	II	0.002	1.830	0.391	1.782
0.126	0.126	0.126	VII	1.350	1.324	0.842	0.126
0.379	0.379	0.048	VI	1.603	1.577	0.048	0.379
0.011	0.011	0.011	VII	1.235	1.210	0.727	0.011
0.256	0.288	1.621	III	0.256	0.288	1.896	2.330
0.090	0.145	0.145	IV	0.090	1.344	0.862	0.145

The EM algorithm was applied to the data and the estimates were derived. The initial values were set arbitrarily equal to 1 for all the parameters. The algorithm converged quite quickly, after 23 iterations. The algorithm stopped iterating when the relative change of the loglikelihood became smaller than 10^{-12}. The history of the parameter values with respect to the iterations can be seen in Fig. 1. Note that the parameters required quite few iteration to reach their final values, and the last iterations increased only slightly the likelihood. Several other different initial values were used without change in the final solution. It is quite interesting that even if one starts from a point far away from the ML estimates, like the point $(5,5,5,5)$, the algorithm converged after 25 iteration (with the same stopping criterion). This is an indication that the choice of initial values is not so important.

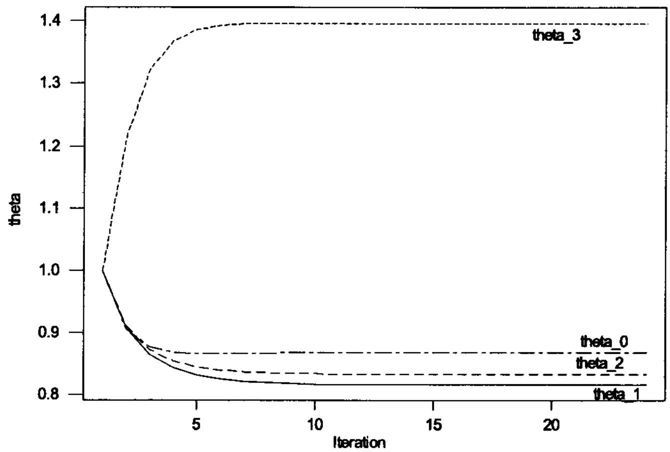

Fig. 1. The history of parameter values across iterations.

Table 7. Results from fitting different models to the data of Table 6.

Model	θ_1	θ_2	θ_3	θ_0	loglikelihood
Full	0.817	0.834	1.396	0.869	−40.7390
$\theta_1 = \theta_2$	0.826	0.826	1.396	0.869	−40.7399
$\theta_1 = \theta_3$	1.091	0.836	1.091	0.869	−41.47594
$\theta_2 = \theta_3$	0.818	1.090	1.090	0.858	−41.45105
$\theta_1 = \theta_2 = \theta_3$	1	1	1	0.849	−41.70794

Reduced models assuming equality of certain parameters were also considered. Results are reported in Table 7. Reduced models can be fitted with minor changes at the M-step of the EM algorithm. For example if we assume that $\theta_1 = \theta_2$ then the only change needed at the M-step of the algorithm is to update the parameters using

$$\theta_1{}^{(k+1)} = \frac{2n}{\sum_{i=1}^{n}(t_{1i} + t_{2i})} \quad \text{and} \quad \theta_2{}^{(k+1)} = \theta_1{}^{(k+1)}$$

retaining the rest of the calculations as described above.

Similar changes can be made for the cases $\theta_1 = \theta_3$ and $\theta_2 = \theta_3$. Finally for the more complicated case $\theta_1 = \theta_2 = \theta_3$ the changes needed at the M-step of the algorithm are to update the parameters using

$$\theta_0{}^{(k+1)} = \frac{n}{\sum_{i=1}^{n} t_{0i}}, \qquad \theta_1{}^{(k+1)} = \frac{3n}{\sum_{i=1}^{n}(t_{1i} + t_{2i} + t_{3i})} \qquad \text{and}$$

$$\theta_2{}^{(k+1)} = \theta_3{}^{(k+1)} = \theta_1{}^{(k+1)}.$$

Looking at the results of Table 7 it is clear that a model that assumes $\theta_1 = \theta_2 = \theta_3$ has the worst loglikelihood. However keeping in mind that this model has fewer parameters than the other models, the symmetry does not seem to be an invalid assumption.

9. Concluding remarks

In this paper, an EM type algorithm for ML estimation of the parameters of the MMO distribution was described. This algorithm made use of the multivariate reduction derivation of the MMO distribution. It is clear that the approach can be expanded to several other models, useful in reliability and survival analysis, that are based on similar shock models as those described for the MMO distribution. For example, allowing the components to follow Weibull distributions one can construct similar algorithms for ML estimation.

Another generalization allows for the use of covariates at the parameters θ_i. Allowing for covariates, the MMO model can be seen as a competing risk model, appropriate for survival data analysis.

Finally, in a more general setting, the EM algorithm can be proven quite helpful for distributions resulting from a multivariate reduction technique.

REFERENCES

Basu, A. P. (1995). Bivariate exponential distributions, *Exponential Distributions: Theory, Methods and Applications* (eds. N. Balakrishnan and A. P. Basu), 326–332, Gordon and Breach Publishers, New Jersey.

Bemis, B. M., Bain, L. J. and Higgins, J. J. (1972). Estimation and hypothesis testing for the parameters of a bivariate exponential distribution, *J. Amer. Statist. Assoc.*, **67**, 927–929.

Bhattacharyya, G. K. and Johnson, R. A. (1973). On a test of independence in a bivariate exponential distribution, *J. Amer. Statist. Assoc.*, **68**, 704–706.

Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension, *J. Amer. Statist. Assoc.*, **69**, 1031–1037.

Brindley, E. C. and Thompson, W. A. (1972). Dependence and aging aspects of multivariate survival, *J. Amer. Statist. Assoc.*, **67**, 822–830.

Chen, D. and Lu, J. C. (1998). The asymptotics of maximum likelihood estimates of parameters based on data type where the failure and the censoring time are dependent, *Statist. Probab. Lett.*, **36**, 379–391.

Chen, D., Li, C. S., Lu, J. C. and Park, J. (2000). Parameter estimation for bivariate shock models with singular distribution for censored data with concomitant order statistics, *Australian and New Zealand Journal of Statistics*, **42**, 323–336.

Cherian, K. C. (1941). A bivariate correlated gamma-type distribution function, *J. Indian Math. Soc.*, **5**, 133–144.

Dowton, F. (1970). Bivariate exponential distributions in reliability theory, *J. Roy. Statist. Soc. Ser. B*, **32**, 408–417.

Freund, J. E. (1961). A bivariate extension of the exponential distribution, *J. Amer. Statist. Assoc.*, **56**, 971–977.

Gumbel, E. J. (1960). Bivariate exponential distributions, *J. Amer. Statist. Assoc.*, **55**, 698–707.

Hanagal, D. D. (1992). Some inference results in bivariate exponential distributions based on censored samples, *Comm. Statist. Theory Methods*, **21**, 1273–1295.

Hanagal, D. D. and Kale, B. K. (1992*a*). Large sample tests for testing symmetry and independence in some bivariate exponential models, *Comm. Statist. Theory Methods*, **21**, 2625–2643.

Hanagal, D. D. and Kale, B. K. (1992*b*). Large sample tests of λ_3 in the bivariate exponential distribution, *Statist. Probab. Lett.*, **12**, 311–313.

Heinrich, G. and Jensen, U. (1995). Parameter estimation for a bivariate lifetime distribution in reliability with multivariate extensions, *Metrika*, **42**, 49–65.

Johnson, N., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, Wiley, New York.

Klein, J. P. (1995). Inference for multivariate exponential distributions, *Exponential Distributions: Theory, Methods and Applications* (eds. N. Balakrishnan and A. P. Basu), 333–342, Gordon and Breach Publishers, New Jersey.

Lai, C. D. (1995). Construction of bivariate distributions by a generalized trivariate reduction technique, *Statist. Probab. Lett.*, **25**, 265–270.

Lu, J. C. (1997). A new plan for life-testing two-components parallel systems, *Statist. Probab. Lett.*, **34**, 19–32.

Mardia, K. V. (1970). *Families of Bivariate Distributions*, Griffin's Statistical Monograph No. 27, Charles Griffin, London.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution, *J. Amer. Statist. Assoc.*, **62**, 30–44.

McLachlan, G. and Krishnan, N. (1997). *The EM Algorithm and Extensions*, Wiley, Chichester.

Papageorgiou, H. and Loukas, S. (1995). A bivariate Charlier series distribution, *Biometrical J.*, **37**, 105–117.

Pena, E. A. and Gupta, A. K. (1990). Bayes estimation for the Marshall Olkin exponential distribution, *J. Roy. Statist. Soc. Ser. B*, **52**, 379–389.

Proschan, F. and Sullo, P. (1976). Estimating the parameters of a multivariate exponential distribution, *J. Amer. Statist. Assoc.*, **71**, 465–472.

Shamseldin, A. A. and Press, S. J. (1984). Bayesian parameter and reliability estimation for a bivariate exponential distribution, *J. Econometrics*, **24**, 363–378.

Stein, G. Z. and Juritz, J. M. (1987). Bivariate compound Poisson distributions, *Comm. Statist. Theory Methods*, **16**, 3591–3607.

Zheng, Q. and Matis, J. H. (1993). Approximating discrete multivariate distributions from known moments, *Comm. Statist. Theory Methods*, **22**, 3553–3567.