Journal of Applied Statistics, Vol. 30, No. 1, 2003, 63–77

An EM algorithm for multivariate Poisson distribution and related models

DIMITRIS KARLIS, *Department of Statistics, Athens University of Economics and Business, Athens, Greece*

ABSTRACT *Multivariate extensions of the Poisson distribution are plausible models for multivariate discrete data. The lack of estimation and inferential procedures reduces the applicability of such models. In this paper, an EM algorithm for Maximum Likelihood estimation of the parameters of the Multivariate Poisson distribution is described. The algorithm is based on the multivariate reduction technique that generates the Multivariate Poisson distribution. Illustrative examples are also provided. Extension to other models, generated via multivariate reduction, is discussed.*

1 Introduction

While univariate discrete distributions have been studied and applied extensively, multivariate versions of these distributions have not been used to a similar extent. This shortage of multivariate applications of discrete distributions can be attributed to computational difficulties associated with applications. In addition, multivariate models are complicated enough, so that standard inferential procedures are not applicable.

A large amount of multivariate inference is essentially based on multi-normal approximations, even when the data are, in nature, discrete. Multi-normal inference has been exploited in depth and it is available in almost any statistical package. However, such approaches are not appropriate in many circumstances, because they imply restrictions that are not met in practice.

Among the existing multivariate discrete distributions, the multivariate Poisson distribution provides a natural generalization of the simple Poisson distribution in many dimensions. Multivariate Poisson models can be used in a wide range of disciplines. In market research, one might be interested in modelling the number

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

ISSN 0266-4763 print; 1360-0532 online/03/010063-15
DOI: 10.1080/0266476022000018510

of purchases of a series of different products that a household makes in a time period. Similarly, actuaries are interested in examining the number of accidents of an insured person over a series of consecutive time periods. In epidemiology, attention is focused on the number of incidents of different but related diseases in a given area. Lastly, in sports modelling—an increasing area due to the huge amounts invested in such activities as well as the amounts spent for betting purposes—one might be interested in modelling jointly the number of occurrences of several events, e.g. in soccer the number of fouls committed, the number of goals, the number of red cards, etc. The common element in the aforementioned examples is that one deals with count data with some kind of dependence and, hence, the multivariate Poisson model seems plausible. Derivations of bivariate and multivariate Poisson distributions can be found in Kocherlakota & Kocherlakota (1992), Johnson *et al.* (1997) and Krummenauer (1998). However, due to the complicated form of the multivariate Poisson distribution, inference is not straightforward and the need for special methods is obvious.

Usually, multivariate models are based on certain structures that generate such models from univariate ones. In many cases, special effort is made so that the multivariate models constitute generalizations of the univariate models. The EM algorithm can be quite useful for multivariate distributions because of the underlying structure that allows a missing data interpretation. The EM algorithm is based on the missing data principle. Many multivariate models have been constructed using simpler conditions that are not, in fact, observable. Such an example is the multivariate reduction technique that generates multivariate models by considering a certain function of univariate variates. Therefore, the EM algorithm is suitable for such models. The derivation of the EM algorithm for the multivariate Poisson distribution, which will be presented in the following, stems from such a multivariate reduction technique.

The EM algorithm (Dempster *et al.*, 1977, McLachlan & Krishnan, 1997) is a powerful algorithm for ML estimation of data containing missing values, or which can be considered as containing missing values. An important feature of the EM algorithm is that it is not merely a numerical technique but that it also offers useful statistical insight (Meng & Van Dyk, 1997).

Suppose that the complete data $Y_i = (X_i, Z_i)$ consist of an observable part X_i and an unobservable part Z_i. When the direct maximization of $\log p(X \mid \phi)$ with respect to the parameter ϕ is not easy, the algorithm augments the observed data to a set of complete data, which can be reduced to the observed data via a many to one mapping. The EM maximizes $\log p(X \mid \phi)$ by iteratively maximizing $E(\log p(Y \mid \phi))$. At the E-step of the $(k+1)$th iteration, the expected log-likelihood of the complete data model is calculated as $Q(\phi \mid \phi^{(k)}) = E(\log p(Y \mid \phi) \mid X, \phi^{(k)})$ where the expectation is taken with respect to the conditional distribution $f(Y \mid X, \phi^{(k)})$ and then, at the M-step, $Q(\phi \mid \phi^{(k)})$ is maximized over ϕ. When the complete model is from the exponential family, then the E-step computes the conditional expectations of its sufficient statistics. This is quite helpful in our case.

The purpose of the present article is to provide Maximum Likelihood estimation for the multivariate Poisson distribution via the EM algorithm. The multivariate Poisson distribution is briefly reviewed in Section 2. The derivation of the EM algorithm is described in Section 3. Certain examples concerning sports, crime and accident data are provided in Section 4. Another bivariate discrete distribution arising from a similar reduction scheme is also demonstrated. It is particularly interesting that the method is applicable to other multivariate distributions that arise from similar models. This is discussed in Section 5 where concluding remarks can also be found.

2 The multivariate Poisson distribution

2.1 The probability distribution

It is customary to regard as multivariate extensions of a univariate distribution, multivariate distributions with marginals of the univariate form. For this reason, the same name has been used for different multivariate distributions. It must be noted that such multivariate generalizations are not unique in the sense that different multivariate distributions may have marginal distributions of the same family (see for example the book of Arnold *et al.*, 1999).

A useful derivation of the multivariate Poisson distribution is via multivariate reduction. This technique has been used extensively for the construction of bivariate models (see for example, Mardia, 1971). The idea is to start with some independent random variables, which are generally elementary ones, and to create new ones by considering some functions of the original variables. Then, since the new variables contain jointly some of the original ones, a kind of structure is imposed creating multivariate models.

Suppose that Y_i are independent Poisson random variables with mean θ_i for $i = 0, \ldots, m$. Then the random variables $X_i = Y_0 + Y_i$, $i = 1, \ldots, m$ follow jointly an m-variate Poisson distribution, where m denotes the dimension of the distribution.

The joint probability function is given by

$$\begin{aligned} P(X) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) \\ &= \exp\left(-\sum_{i=1}^{m} \theta_i\right) \prod_{i=1}^{m} \frac{\theta_i^{x_i}}{x_i!} \sum_{i=0}^{s} \prod_{j=1}^{m} \binom{x_j}{i} i! \left(\frac{\theta_0}{\prod_{k=1}^{m} \theta_i}\right)^i \end{aligned} \tag{1}$$

where $s = \min(x_1, x_2, \ldots, x_m)$. This distribution is denoted as $m - P(\theta_0, \theta_1, \theta_2, \ldots, \theta_m)$. Marginally, each X_i follows a Poisson distribution with parameter $\theta_0 + \theta_i$. Parameter θ_0 is the covariance between all the pairs of random variables. If $\theta_0 = 0$ then the variables are independent and the multivariate Poisson distribution reduces to the product of independent Poisson distributions.

The above-defined version of multivariate Poisson distribution is different from a more general model used by many authors (Mahamunulu, 1967; Loukas & Kemp, 1983; Johnson *et al.*, 1997), which assumes another more complicated structure. The multivariate Poisson distribution defined in (1) is, in fact, a reduced version of the general multivariate Poisson distribution used in the above publications. The literature for the multivariate Poisson distribution is large and many references, as well as historical remarks, can be found in Johnson *et al.* (1997) and Krummenauer (1998). The case of bivariate Poisson distribution has been studied in more depth. The reader can refer to the book of Kocherlakota & Kocherlakota (1992).

A main obstacle that limits the usage of multivariate distributions is the complexity of calculating the probability function. The summations needed might be exhausting in some cases, especially when the number of dimensions is large. Computation of the probabilities can be accomplished via recursive schemes. Kano

& Kawamura (1991) provided a general scheme for constructing recurrence relations for multivariate Poisson distributions.

For the multivariate Poisson distribution defined in (1) a recursive scheme is the following

$$\begin{aligned} x_i P(X) &= \theta_i P(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = x_i - 1, X_{i+1} = x_{i+1}, \ldots, X_m = x_m) \\ &\quad + \theta_0 P(X_1 = x_1 - 1, X_2 = x_2 - 1, \ldots, X_m = x_m - 1), \qquad i = 1, \ldots, m \\ P(X_1 &= x_1, \ldots, X_k = x_k, 0, 0, \ldots, 0) \\ &= P(X_1 = x_1 - 1, \ldots, X_k = x_k - 1, 0, 0, \ldots, 0) \prod_{i=1}^{k} \frac{\theta_i}{x_i} \end{aligned}$$

for $k = 1, \ldots, m-1$, where the order of X_is must be interchanged in order to cover all the cases. while, $P(0,0,\ldots,0) = \exp(-\Sigma_{i=0}^{m} \theta_i)$.

It can be seen that, since at every case at least one of the values equals 0, i.e. $\min(x_1, x_2, \ldots, x_m) = 0$, the sum appearing in the joint probability function vanishes and hence the joint probability function takes the useful form $P(X) = \exp(-\theta_0) \Pi_{i=1}^{m} \mathrm{Po}(x_i \mid \theta_i)$, where $\mathrm{Po}(x \mid \theta) = \exp(-\theta)\theta^x / x!$ denotes the probability function of the simple Poisson distribution with parameter θ. This expression then arises by using the recurrence relation for the univariate Poisson distribution.

It is clear that as m increases, the computational effort increases. In fact, for a multivariate Poisson model one has to initialize in this way

$$\sum_{i=1}^{m-1} \binom{m}{i}$$

probabilities of the form $P(X_1 = x_1, \ldots, X_k = x_k, 0, 0, \ldots, 0)$, for $k = 1, \ldots, m$ interchanging the order of non-zero x_is. For example, if $m = 4$, 14 different recursive schemes must be run. The computational burden is large. Note also that errors due to recursion are accumulated and thus, for large m the scheme can be unstable. Fortunately, as it will be shown in the next section, one can proceed with ML estimation without needing to calculate the probability function.

3 Maximum likelihood estimation via the EM algorithm

In this section, we focus on the estimation of the parameters θ_i, $i = 0, \ldots, m$ of a multivariate Poisson distribution. Let the vector $X_i = (x_{i1}, x_{i2}, \ldots, x_{im})$, $i = 1, \ldots, n$ denote the ith observation. Applying the standard ML technique, one can see that $\hat{\theta}_i = \bar{x}_i - \theta_0$, $i = 1, \ldots, m$, where $\bar{x}_i = n^{-1} \Sigma_{j=1}^{n} x_{ij}$, i.e. the sample mean of the ith variable. For $\hat{\theta}_0$, a numerical technique has to be used replacing the rest parameters.

An alternative EM scheme will be given using the missing data derivation of the multivariate Poisson distribution via the multivariate reduction scheme. The complete data for the ith observation are $Y_i = (Y_{0i}, Y_{1i}, Y_{2i}, \ldots, Y_{mi}), i = 1, \ldots, n$. ML estimation for the complete data is straightforward since we have $m + 1$ independent Poisson variates.

Hence, at the E-step one has to calculate the conditional expectations of Y_i given the data and the current values of the estimates, while the M-step merely calculates the means of these pseudo values. In addition, only the derivation of $E(Y_{0i} \mid X_i, \theta)$, where $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$, is needed since the rest can be easily found via subtraction.

Consider the more general model where $X_i \sim m - P(t_i\theta_0, t_i\theta_1, t_i\theta_2, \ldots, t_i\theta_m)$, where t_i is an offset such as a population or an area. This implies that the random variables Y_{ji}, $j = 0, \ldots, m$, $i = 1, \ldots, n$ (j denotes the variable and i the observation) follows a $Po(t_j\theta_i)$ distribution.

Thus, the EM algorithm is as follows.

E-step: Using the data and the current estimates after the kth iteration $\theta^{(k)}$, calculate the pseudo-values

$$s_i = E(Y_{0i} \mid X_i, t_i, \theta^{(k)}) =$$

$$= \theta_0 t_i \frac{P(X_1 = x_{1i} - 1, X_2 = x_{2i} - 1, \ldots, X_m = x_{mi} - 1)}{P(X_i)}$$

M-step: Update the estimates by

$$\theta_0^{(k+1)} = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} t_i}, \qquad \theta_i^{(k+1)} = \frac{\bar{x}_i}{\bar{t}} - \theta_0^{(k+1)} \qquad i = 1, \ldots, m$$

If some convergence criterion is satisfied, stop iterating otherwise return to the E-step for one more iteration. If the initial values for θ are in the admissible range (i.e. $\theta_i > 0$, $i = 0, \ldots, m$) then the final estimates will also be in the admissible range.

Remark 1: It is quite interesting to show that the conditional distribution of Y_{0i} is given by

$$f(Y_{0i} = y \mid x_i, \theta^{(k)}) = P(y) = \frac{\frac{\theta_0^y}{y!} \prod_{j=1}^{m} \frac{\theta_j^{-y}}{(x_{ji} - y)!}}{\sum_{y=0}^{s} \frac{\theta_0^y}{y!} \prod_{j=1}^{m} \frac{\theta_j^{-y}}{(x_{ji} - y)!}}$$

Since this distribution has finite support, the probabilities can be obtained rather easily via a simple recursive scheme. Moreover, note that, in fact, the distribution is truncated at the right, and the quantity in the denominator is the cumulative distribution up to the truncation point. Therefore, the expectation needed in the E-step can be calculated without computing the probabilities of the multivariate Poisson distribution. This overcomes problems occurring with the computation of the probability function using the recurrence relations in equation (2).

The full recurrence scheme is given as: set as initial value $P(y = 0) = 1$. Since the quantities will be rescaled at the end, this does not cause any problem and helps to avoid overflow errors. Then, calculate

$$P(y) = P(y-1) \left(\frac{\theta_0}{\prod_{i=1}^{m} \theta_i} \right) \frac{\prod_{i=1}^{m} (x_i - y + 1)}{y}$$

for $y = 1, \ldots, s$. Then one has to rescale the probabilities in order to sum to 1, dividing each one by their sum, namely $\Sigma_{y=0}^{s} P(y)$, to obtain the probabilities for the conditional distribution of $Y_{0i} = y \mid X_i, \theta$.

Remark 2: Note that if $s=0$, i.e. the smallest value is 0 then the expectation at the E-step equals 0. Thus, for data sets where there is at least one 0 at each observation the ML estimate of θ_0 necessarily equals 0.

The algorithm described above generalizes the results of Adamidis & Loukas (1994) where an EM scheme for the bivariate Poisson distribution was given. In addition, the multivariate reduction derivation of the multivariate Poisson distribution has been used by Tsionas (1999) for Bayesian estimation of the parameters of the multivariate Poisson distribution.

4 Examples

4.1 Bivariate Poisson distribution

Soccer data. The first example concerns the simplest case of a bivariate Poisson distribution. The probability function is given by equation (1) for $m=2$. It allows for positive dependence between the two random variables. Marginally, each random variable follows a Poisson distribution with parameters $\theta_1+\theta_0$ and $\theta_2+\theta_0$ respectively. Parameter θ_0 is the covariance between the two variables and hence it can be considered as a parameter that incorporates dependence between the two random variables. If $\theta_0=0$ then the two variables are independent and the bivariate Poisson distribution reduces to the product of two independent Poisson distributions. The conditional distribution of X_1 given X_2 is the convolution of a Poisson with a binomial distribution (see Kocherlakota & Kocherlakota, 1992). ML estimation is discussed in Kocherlakota & Kocherlakota (1992) and Adamidis & Loukas (1994). It is interesting to note that the M-step described above for the general multivariate Poisson distribution, can be written as

$$\theta_0^{(k+1)}=\frac{1}{n}\sum_{i=1}^{n}\frac{P(x_{1i}-1,x_{2i}-1)}{P(x_{1i},x_{2i})}\theta_0^{(k)}$$

which is similar to the equation derived in Kocherlakota & Kocherlakota (1992). This verifies the interesting feature of the EM algorithm of not being merely a numerical technique but a technique that provides statistical insight (Meng & Van Dyk, 1997).

In analysing soccer data, it is usually assumed that the number of goals scored by each team follows a Poisson distribution (see e.g. Maher, 1982). Such an assumption has been criticized by some authors due to the over-dispersion present in the data (see, for example, Baxter & Stevenson, 1988), but it has been found that the over-dispersion is small and the Poisson distribution is still applicable. Suppose that the two variables X_1 and X_2 denote the number of goals scored by the home team and the guest team, respectively. It seems natural that since the two teams interact in the game, the random variables are dependent. In fact, this dependence between the goals scored by each team has been ignored mainly due to the empirical evidence that the dependence is rather small (see Lee, 1997, Karlis & Ntzoufras, 2000). Bivariate models, for jointly fitting the two variables, have been considered by Maher (1982) and Dixon & Coles (1997). In the latter paper, an extended Bivariate Poisson distribution was used.

In the present paper, data concerning the number of goals scored by each team for 24 European Leagues were used and a Bivariate Poisson distribution was fitted for each league applying the EM algorithm. The results appear in Table 1. As one can see, $\hat{\theta}_3$ is usually rather small, verifying the small covariance that has been

TABLE 1. Table of fitted bivariate Poisson distribution for 24 European leagues

League	Observed mean number of goals		Estimated parameters			Sample covariance	p-value
	Home team	Guest team	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$		
Dutch 93	1.73	1.21	1.73	1.21	0.00	−0.38	0.01
Dutch 94	1.90	1.31	1.90	1.31	0.00	−0.26	0.00
Dutch 95	1.75	1.25	1.75	1.25	0.00	−0.29	0.00
Ger 89	1.68	0.91	1.68	0.91	0.00	−0.04	0.03
Ger 93	1.75	1.17	1.75	1.17	0.00	−0.15	0.08
Ger2 93	1.53	0.97	1.76	1.24	0.00	0.04	0.03
Ger 94	1.76	1.24	1.53	1.17	0.00	−0.14	0.40
Ger 95	1.54	1.17	1.22	0.72	0.15	0.01	0.38
Italy 89	1.37	0.87	1.38	0.81	0.05	0.12	0.70
Italy 90	1.43	0.86	1.24	0.84	0.10	0.05	0.01
Italy 91	1.34	0.93	1.50	0.86	0.22	0.11	0.03
Italy 92	1.72	1.08	1.45	0.91	0.03	0.24	0.10
Italy 93	1.48	0.94	1.57	0.96	0.00	0.03	0.07
Italy 94	1.57	0.96	1.59	0.90	0.07	−0.05	0.00
Italy 95	1.66	0.97	1.49	1.04	0.04	0.09	0.12
France 93	1.42	0.81	1.41	0.80	0.01	0.01	0.49
France 94	1.59	0.92	1.58	0.91	0.01	0.01	0.07
France 95	1.44	0.84	1.41	0.81	0.03	0.03	0.07
Spain 93	1.59	1.01	1.50	0.93	0.04	0.10	0.00
Spain 94	1.54	1.00	1.51	0.92	0.08	−0.09	0.00
Spain 95	1.56	1.14	1.54	1.00	0.00	0.06	0.00
Engl 93	1.44	1.15	1.51	1.09	0.05	0.02	0.05
Engl 94	1.51	1.08	1.42	1.13	0.02	0.04	0.01
Engl 95	1.53	1.07	1.47	1.04	0.04	0.04	0.02

observed with soccer data. In order to test the hypothesis that the bivariate Poisson distribution fits the data well, a variant of the index of dispersion for bivariate data (proposed by Loukas & Kemp, 1986), was applied.

The test statistic is

$$I_B = \frac{n(\bar{x}_2 s_1^2 - 2\text{cov}(x_1, x_2) + \bar{x}_1 s_2^2)}{\bar{x}_1 \bar{x}_2 - \text{cov}(x_1, x_2)^2}$$

Asymptotically, this test statistic follows a chi-square distribution with $2n$ degrees of freedom. The p-value reported in Table 1 was calculated via this chi-square result. One can see that the fit is poor for the majority of the leagues, supporting the evidence that the Poisson distribution does not fit the data well.

An interesting feature is that, for every case where the sample covariance was negative, the ML estimate of θ_0 was equal to 0. This was verified for a large number of simulated data sets.

An alternative bivariate model. The bivariate Poisson distribution arises, by considering the random variables

$$X_1 = Y_0 + Y_1$$
$$X_2 = Y_0 + Y_2$$

where each Y_i, $i = 0, 1, 2$, follows a Poisson distribution with parameter θ_i.

If we now assume that Y_0 follows a binomial distribution instead of a Poisson distribution, then the bivariate discrete Charlier Series Distribution arises. Now, Y_0 follows a binomial distribution with parameter N and probability of success p, and (X_1, X_2) jointly follow the bivariate discrete Charlier Series distribution. More details about the distribution can be found in Papageorgiou & Loukas (1995). The distribution arises as the conditional distribution from a trivariate Poisson distribution studied by Loukas & Papageorgiou (1991) and Loukas (1993). Denoting as $\mathrm{Po}(x \mid \theta)$ and $\mathrm{Bin}(x \mid N, p)$ the probability functions of a Poisson and a Binomial distribution with the appropriate parameters, respectively, it can be easily seen that a simple EM scheme for ML estimation of the parameters of the bivariate discrete Charlier Series distribution is given by:

E-step: Using the data and the current estimates $\theta_1^{(k)}$, $\theta_2^{(2)}$ and $p^{(k)}$ after the kth iteration, calculate the pseudo-values.

$$s_i = \frac{\sum_{z=0}^{\min(x_{1i}, x_{2i})} z \mathrm{Po}(x_{1i} - z \mid \theta_1^{(k)}) \, Po(x_{2i} - z \mid \theta_2^{(k)}) \, \mathrm{Bin}(z \mid N, p^{(k)})}{\sum_{z=0}^{\min(x_{1i}, x_{2i})} \mathrm{Po}(x_{1i} - z \mid \theta_1^{(k)}) \, Po(x_{2i} - z \mid \theta_2^{(k)}) \, \mathrm{Bin}(z \mid N, p^{(k)})}$$

M-step: Update the estimates by

$$p^{(k+1)} = (Nn)^{-1} \sum_{i=1}^{n} s_i, \qquad \theta_1^{(k+1)} = \bar{x}_1 - Np^{(k+1)} \qquad \text{and} \qquad \theta_2^{(k+1)} = \bar{x}_2 - Np^{(k+1)}$$

If some convergence criterion is satisfied stop iterating otherwise return to the E-step for one more iteration.

It is clear that the only change relative to the EM scheme for the bivariate Poisson distribution is the appearance of the binomial probability function in the E-step instead of the probability function of the Poisson distribution. Note again that the estimate for $\hat{p}$ equals 0 if $\forall i$, $\min(x_{1i}, x_{2i}) = 0$ and there is no need to calculate the probability function in order to find the ML estimates. If the parameter N of the binomial distribution is not known, then one can start with a small value of N and, by successively maximizing the likelihood using the EM algorithm, obtain the ML estimate for N.

The model proposed by Papageorgiou & Loukas (1995) assumed that $Y_i \sim \mathrm{Po}(\theta_i p)$, $i = 1, 2$ and $Y_0 \sim \mathrm{Bin}(N, p)$. This model is only a re-parameterization of the above Bivariate Charlier Series model. Not surprisingly, the update for the parameter p for each iteration can be written in the form

$$p^{(k+1)} = \frac{1}{nN} \sum_{i=1}^{n} \frac{\sum_{z=0}^{\min(x_{1i}, x_{2i})} z \mathrm{Po}(x_{1i} - z \mid \theta_1^{(k)} p^{(k)}) \, \mathrm{Po}(x_{2i} - z \mid \theta_2^{(k)} p^{(k)}) \, \mathrm{Bin}(z \mid N, p^{(k)})}{\sum_{z=0}^{\min(x_{1i}, x_{2i})} \mathrm{Po}(x_{1i} - z \mid \theta_1^{(k)} p^{(k)}) \, \mathrm{Po}(x_{2i} - z \mid \theta_2^{(k)} p^{(k)}) \, \mathrm{Bin}(z \mid N, p^{(k)})}$$

which is similar to the likelihood equation given in Papageorgiou & Loukas (1995, p. 111).

For illustration, the Bivariate Charlier Series distribution was fitted to the data

TABLE 2. The number of absences of 113 students from a lecture in two successive semesters

	y									
x	0	1	2	3	4	5	6	7	8	Total
0	15	10	4	4	2	0	0	0	0	35
1	6	11	9	4	2	0	0	0	0	33
2	5	7	6	5	0	0	0	0	0	25
3	1	3	2	4	3	1	0	0	0	17
4	1	0	2	0	1	0	0	0	0	8
5	0	0	0	0	0	1	1	0	0	7
6	0	0	0	0	0	0	2	0	0	8
7	0	0	0	0	0	0	0	0	0	7
8	0	0	0	0	0	0	0	0	0	8
9	1	0	0	0	0	0	0	0	0	10
Total	29	31	23	17	8	2	3	0	0	113

TABLE 3. The results when fitting the bivariate Charlier Series distribution to data of Table 2

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{p}$	$\hat{N}$	Log-likelihood
	1.176	1.362	0.150	2	−379.290
	1.098	1.284	0.126	3	−377.872
	1.050	1.236	0.106	4	−376.942
	1.021	1.207	0.091	5	−376.319
	0.979	1.165	0.062	8	−375.327
	0.973	1.158	0.056	9	−375.140
	0.967	1.153	0.051	10	−374.991
	0.954	1.140	0.037	14	−374.609
	0.952	1.138	0.035	15	−374.546
	0.945	1.131	0.026	20	−374.326
	0.942	1.127	0.021	25	−374.195
	0.939	1.125	0.017	30	−374.109
	0.934	1.120	0.010	50	−373.937
	0.932	1.118	0.007	75	−373.853
	0.931	1.117	0.005	100	−373.810
	0.930	1.116	0.003	150	−373.768
	0.929	1.115	0.002	250	−373.735
	0.929	1.115	0.001	350	−373.720
Biv. Pois.	0.92835	1.11419	$\hat{\theta}_0 = 0.54953$	∞	−373.6844

in Table 2 concerning the number of absences of 113 students from a lecture course in two successive semesters. The parameter N of the binomial distribution was considered unknown and it was attempted to estimate this. The results for certain choices of the value of N are reported in Table 3. The likelihood is unbounded with respect to N, and hence it can be considered that the binomial distribution approaches the Poisson distribution. The fitted bivariate Poisson distribution is also reported, corresponding to a value of $N \to \infty$. This implies that the distribution of X_0 is not binomial but Poisson.

4.2 Crime data—multivariate models

The second example concerns crime data taken from the National Statistical Service of Greece for 1997. Five different kinds of crimes were examined for 50 prefectures of Greece. The different crimes considered were rapes (X_1), arson (X_2), manslaughter (X_3), smuggling of antiquities (X_4) and general smuggling (X_5). The data are presented in Table 4 along with the population of each prefecture in millions. It is assumed that the whole vector $(X_1, \ldots, X_5)$ follows a $5 - P(\theta_0 t_i, \theta_1 t_i, \ldots, \theta_5 t_i)$ distribution. The parameters θ_i, $i = 1, \ldots, 5$ can be considered as measuring the specific type of crime, while the parameter θ_0 measures the general crime tendency.

In order to examine the way that the different crimes vary together, we fitted multivariate Poison models for all the pairs, triples, quadruples of different types as well as a general 5-variate model using the EM algorithm. The results are presented in Tables 5–7. In Table 5, one can see the bivariate models fitted to the pairs of different crimes. It is interesting to note that, for each specific crime, the estimate of the corresponding parameter is the mean minus a covariance term. If the specific covariance is zero then the estimate equals the sample mean, as expected.

Examining the results, some interesting issues arise. The stronger covariance term appears between rapes and manslaughter. These types of crimes are directly against another person's life. However, when considering trivariate models that include these two types of crimes the covariance terms are not large, or the covariance is cancelled out. Note also that the triple 'arson', 'manslaughter' and 'smuggling' has a zero covariance for both the trivariate model or the bivariate models. The same holds for the 4-variate models that include them. The investigation of the covariances appearing in tables illustrates the relationships among different types of crime. The more general model, considering all five types of crime together, does not exhibit a common covariance term.

Finally, an interesting feature of the EM algorithm is the fact that the 'missing' information calculated at each step can be useful for further examination of the data. For example, in this dataset, the unobservable quantity Y_{0i} reflects the common element for the prefecture i of the five different types of crime. Then one would like to measure this quantity as a general measure of crime tendency. It is clear that the quantity of interest can be estimated as $E(Y_{0i} | data)$, which is the pseudo-values calculated at the last iteration of the EM algorithm. Therefore, another advantage of the EM algorithm is that the byproducts of the algorithm may be useful for further analysis, and that they are readily available after the termination of the algorithm. For our example, the large proportion of zero values results in a large number of prefectures with zero conditional expectation and thus comparison of the prefectures via this quantity is not applicable.

4.3 Accident Data

The third example concerns the number of accidents in 24 central roads of Athens for a time period of 5 years (see Table 8). Only accidents that caused injuries are considered. We assume that for each road i, $i = 1, \ldots, 24$, the observations constitute a vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{i5})$ that follows a $5 - P$ distribution with parameter vector $\theta = (\theta_0 t_i, \theta_1 t_i, \ldots, \theta_5 t_i)$, where t_i represents the length of each road. The parameter θ_0 can be considered as a covariance factor that measures the risk of the area, in the sense that if $\theta_0 = 0$ then a simple Poisson distribution could be fitted to each year separately. Thus, this parameter is common to all time periods.

The EM algorithm was applied to the data. Note that the time series nature of the data has been ignored.

TABLE 4. Different types of crimes committed in 50 prefectures of Greece for 1997

Prefecture	Rapes	Arson	Manslaughter	Smuggling of antiquities	Smuggling	Population (in millions)
Thessaloniki	21	29	32	9	58	0.98
Aitoloakarnania	6	2	7	1	3	0.23
Evros	7	0	4	0	2	0.13
Argolida	1	0	4	0	0	0.10
Arkadia	3	4	0	3	0	0.12
Arta	1	1	1	0	0	0.08
Ahaia	6	0	12	2	0	0.32
Viotia	1	0	11	1	0	0.16
Grevena	0	0	0	0	0	0.04
Drama	0	1	0	1	0	0.10
Dodekanisa	8	3	0	11	4	0.17
Evoia	5	0	4	1	0	0.23
Evritania	0	0	0	0	0	0.03
Zakinthos	5	0	4	0	0	0.03
Ilia	4	0	16	0	0	0.18
Imathia	1	2	3	0	2	0.15
Irakleio	9	3	9	8	0	0.28
Thesprotia	0	1	0	0	3	0.05
Ioannina	2	1	3	0	2	0.17
Kavala	0	1	8	1	1	0.14
Karditsa	0	0	2	1	0	0.13
Kastoria	0	0	2	0	7	0.05
Corfu	7	1	6	0	0	0.11
Kefallonia	0	0	1	0	0	0.03
Kilkis	0	3	1	0	1	0.08
Kozani	1	1	7	0	2	0.15
Korinthia	2	0	7	3	1	0.17
Cyclades	4	13	1	0	1	0.10
Lakonia	3	0	5	1	0	0.11
Larisa	3	5	12	5	0	0.27
Lasithi	1	1	2	1	0	0.07
Lesvos	6	4	2	3	0	0.10
Lefkada	1	0	0	0	0	0.02
Magnisia	4	0	6	2	0	0.20
Messinia	5	3	11	5	1	0.18
Xanthi	2	0	3	4	0	0.09
Pella	2	0	4	1	6	0.02
Pieria	1	0	3	1	0	0.13
Preveza	2	0	0	0	0	0.06
Rethimno	4	0	3	0	0	0.07
Rodopi	0	1	0	0	0	0.10
Samos	1	3	1	0	1	0.04
Serres	1	0	3	9	1	0.20
Trikala	0	1	0	0	0	0.14
Fthiotida	0	5	3	3	0	0.19
Florina	1	0	0	0	4	0.05
Fokida	0	0	0	4	0	0.06
Halkidiki	1	0	6	4	1	0.11
Hania	1	13	5	0	0	0.14
Hios	1	1	1	0	0	0.05
Total	134	103	215	85	101	6.92

TABLE 5. Bivariate models for crime data; estimated parameters for each type of crime

Rapes	Firing	Manslaughter	Smuggling of antiquities	Smuggling	Covariance term
18.46	13.95	—	—	—	0.88
15.04	—	26.83	—	—	4.30
17.82	—	—	10.84	—	1.52
18.26	—	—	—	13.60	1.08
—	14.84	31.13	—	—	0.00
—	14.84	—	12.36800	—	0.00
—	12.49	—	—	12.34	2.34
—	—	31.13	12.36	—	0.00
—	—	29.82	—	13.38	1.31
—	—	—	12.32	14.65	0.04

TABLE 6. Trivariate models for crime data; estimated parameters for each type of crime

Rapes	Firing	Manslaughter	Smuggling of antiquities	Smuggling	Covariance term
19.35	14.84	31.13	—	—	0.00
18.70	14.19	—	11.72	—	0.64
17.89	13.38	—	—	13.24	1.45
18.89	—	30.68	11.90	—	0.45
18.44	—	30.23	—	13.78	0.90
18.74	—	—	11.76	14.09	0.60
—	14.84	31.13	12.36	—	0.00
—	13.79	30.09	—	13.65	1.04
—	14.63	—	12.16	14.49	0.20
—	—	30.86	12.09	14.42	0.26

TABLE 7. Models for crime data with 4 and 5 variables; estimated parameters for each type of crime

Rapes	Firing	Manslaughter	Smuggling of antiquities	Smuggling	Covariance term
19.35	14.88	31.06	12.28	—	0
18.52	14.04	30.22	—	13.75	0.83
18.83	14.35	—	11.75	14.06	0.52
19.04	—	30.74	11.96	14.27	0.31
—	14.88	31.06	12.28	14.59	0
19.35	14.88	31.06	12.28	14.59	0

It was found that $\hat{\theta}_0 = 3.753$, which is a large value compared with the other estimates. This can be explained as there is a strong common factor in all years that has to do with the hazardness of these roads. It should be kept in mind that, due to the inhomogeneity of the roads, a multivariate mixed Poisson model would be preferable.

5 Concluding remarks

In this paper, ML estimation for the multivariate Poisson distribution was described using the derivation of the multivariate Poisson distribution via a multivariate reduction technique. The latent structure implied by such a technique, makes the EM algorithm easily applicable. It must be pointed out that such an approach can

TABLE 8. The number of accidents for 24 roads of Athens for the period 1987–1991

Road	Year					Length (km)
	1987	1988	1989	1990	1991	
Akadimias	11	33	25	23	6	1.2
Alexandras	41	63	91	77	29	2.6
Amfitheas	5	35	44	21	13	2.4
Aharnon	44	79	91	88	33	5.5
Vas. Olgas	5	3	4	4	0	0.5
Vas. Konstantinou	8	15	26	13	7	1.3
Vas. Sofias	34	63	81	67	23	2.6
Vouliagmenis	17	16	24	24	4	2.1
G' Septemvriou	16	24	30	30	13	1.7
Galatsioy	13	13	15	17	9	1.1
Iera Odos	7	15	20	19	8	2.7
Kalirois	15	24	39	32	7	2.6
Katehaki	2	3	27	24	7	1.4
Kifisias	22	23	38	22	11	1.4
Kifisou	38	48	60	53	24	7.9
Leof. Kavalas	4	6	12	9	3	2.0
Lenorman	19	30	37	48	22	2.0
Leof. Athinon	15	11	16	21	28	6.1
Mesogeion	20	30	33	28	9	1.5
P. Ralli	13	14	13	17	9	2.6
Panepistimiou	24	58	40	36	5	1.1
Patision	80	108	114	113	86	4.1
Peiraios	86	89	109	90	49	8.0
Sigrou	60	61	87	86	29	4.8
Estimated parameters	$\hat{\theta}_1 = 4.902$	$\hat{\theta}_2 = 8.731$	$\hat{\theta}_3 = 11.795$			
	$\hat{\theta}_4 = 10.147$	$\hat{\theta}_5 = 2.517$	$\hat{\theta}_0 = 3.753$			

be used for several other multivariate distributions arising from similar schemes. The bivariate discrete Charlier Series distribution was briefly discussed.

The bivariate exponential distribution of Marshall & Olkin (1967) can be derived by assuming that the random variables Y_1, Y_2, Y_3 follow independently an exponential distribution with parameter λ_i, $i = 1, 2, 3$ and then we define the random variables $X_1 = \min(Y_1, Y_3)$ and $X_2 = \min(Y_2, Y_3)$ which jointly follow the Bivariate Marshall Olkin exponential distribution.

Stein & Juritz (1987) described bivariate mixed Poisson distributions based on trivariate reduction. Famoye & Consul (1995) described bivariate generalized Poisson distributions while Mardia (1971) dealt with certain bivariate continuous distributions arising from a trivariate scheme. Zheng & Matis (1993) and Lai (1995) generalized trivariate reduction schemes.

A more general multivariate Poisson distribution has also been described by many authors. Its derivation is based on a general multivariate reduction scheme where there are terms for all the pairwise covariances, covariances among three variables and so on. For example, for the trivariate case (see for example Loukas & Kemp, 1983) we have that

$$\begin{aligned} X_1 &= Y_1 + Y_{12} + Y_{13} + Y_{123} \\ X_2 &= Y_2 + Y_{12} + Y_{23} + Y_{123} \\ X_3 &= Y_3 + Y_{13} + Y_{23} + Y_{123} \end{aligned} \tag{3}$$

where all Y_is, $i \in (\{1\},\{2\},\{3\},\{12\},\{13\},\{23\},\{123\})$ are independently Poisson distributed random variables. The above scheme assumes two-way covariances, as well as three-way covariance terms in an ANOVA-like fashion. For more dimensions, this scheme implies the existence of all order covariances and hence it leads to a rather complicated structure. The above scheme can be put in a more formal basis by assuming Y_i, $i = 1, \ldots, k$ independent Poisson random variables and $\mathbf{A}$ is a $m \times k$ matrix with zero and ones, and then the vector $X = (X_1, X_2, \ldots, X_m)$ defined as $X = \mathbf{A}Y$ follows a multivariate Poisson distribution. The most general form assumes that $\mathbf{A}$ is a matrix of size $m \times (2^m - 1)$ of the form

$$\mathbf{A} = [A_1, A_2, A_3, \ldots, A_m] \tag{4}$$

where A_i, $i = 1, \ldots, m$ are matrices with m rows and (m/i) columns. The matrix A_i contains columns with exactly i ones and $(m - i)$ zeros, with no duplicate columns, for $i = 1, 2, \ldots, m$. Thus, A_m is the column vector of ones while A_1 becomes the identity matrix of size $m \times m$.

The structure examined in Section 2 uses $\mathbf{A} = [A_1, A_m]$ or

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 1 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 1 & \ldots & 0 & 1 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & 1 & 1 \end{bmatrix}$$

and $\mathbf{Y} = (Y_0, Y_1, \ldots, Y_m)$, while for the multivariate distribution defined by equation (4), the matrix $\mathbf{A}$ takes the form $\mathbf{A} = [A_1, A_2, A_3]$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

and $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_{12}, Y_{13}, Y_{23}, Y_{123})$.

This representation allows for easy derivation of properties of multivariate distributions. It is known that $E(X) = \mathbf{AM}$ and $\mathrm{Var}(X) = \mathbf{A\Sigma A}^{\mathrm{T}}$, where $\mathbf{M}$ and $\boldsymbol{\Sigma}$ are the mean vector and the variance-covariance matrix for the variables $Y_0, Y_1, \ldots, Y_k$ respectively. Note that $\boldsymbol{\Sigma}$ is diagonal because of the independence of Y_is. Moreover, if each Y_i follows a Poisson distribution, the diagonal elements are simply the parameters θ_i.

The EM algorithm can be also applied to the more complicated structure, given in equation (4). Unfortunately, the E-step needs the conditional distribution of the vector Y given the observed data. This conditional distribution is now multivariate and hence the expectations need summation, which can be exhaustive in many dimensions. In such schemes, versions of the EM algorithm that use simulation rather than exhausting summation could be helpful. Thus, schemes such as the simulated maximum likelihood or the Stochastic EM algorithm may overcome problems arising because of the complicated form of the probability function, and thus they may be applicable. In particular the Stochastic EM algorithm is easily applicable since only generation from the joint conditional probability function is needed. A Gibbs sampler, which generates from each conditional recursively, can reduce the complexity of such generation.

REFERENCES

ADAMIDIS, K. and LOUKAS, S. (1994) ML estimation in the bivariate poisson distribution in the presence of missing values via the EM algorithm, *Journal of Statistical Computation and Simulation*, 50, pp. 163–172.

ARNOLD, B., CASTILLO, E. & SARABIA, J. M. (1999) *Conditional Specification of Statistical Models: Models and Applications* (Springer).

BAXTER, M. & STEVENSON, R. (1988) Discriminating between the Poisson and the negative binomial distributions: an application to goal scoring in association football, *Journal of Applied Statistics*, 15, pp. 347–348.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, B 39, pp. 1–38.

DIXON, M. J. & COLES, S. G. (1997) Modelling association football scored and inefficiencies in football betting market, *Applied Statistics*, 46, pp. 265–280.

FAMOYE, F. & CONSUL, P. C. (1995) Bivariate generalized Poisson distribution with some applications, *Metrika*, 42, pp. 127–138.

JOHNSON, N., KOTZ, S. & BALAKRISHNAN, N. (1997) *Discrete Multivariate Distributions* (New York, Wiley).

KANO, K. & KAWAMURA, K (1991) On recurrence relations for the probability function of multivariate generalized Poisson distribution, *Communications in Statistics—Theory and Methods*, 20, pp. 165–178.

KARLIS, D. & NTZOUFRAS, J. (2000) Discrete distributions with applications in sports. Department of Statistics, Athens University of Economics, Technical Report, p. 101.

KOCHERLAKOTA, S. & KOCHERLAKOTA, K. (1992) *Bivariate Discrete Distributions* (New York, Marcel Dekker).

KRUMMENAUER, F. (1998) Limit theorems for multivariate discrete distributions, *Metrika*, 47, pp. 47–69.

LAI, C. D. (1995) Construction of bivariate distributions by a generalized trivariate reduction technique, *Statistics and Probability Letters*, 25, pp. 265–270.

LEE, A. J. (1997) Modeling scores in the premier league: is Manchester United really the best? *Chance*, 10, pp. 15–19.

LOUKAS, S. (1993) Some methods of estimation for a trivariate Poisson distribution, *Zastoswania Mathematica*, 21, pp. 503–510.

LOUKAS, S. & KEMP, C. D. (1983) On computer sampling from trivariate and multivariate discrete distributions, *Journal of Statistical Computation and Simulation*, 17, pp. 113–123.

LOUKAS, S. and KEMP, C. D. (1986) The index of dispersion test for the bivariate Poisson distribution, *Biometrics*, 42, pp. 941–948.

LOUKAS, S. & PAPAGEORGIOU, H. (1991) On a trivariate Poisson distribution, *Applications in Mathematics*, 36, pp. 432–439.

MAHAMUNULU, D. M. (1967) A note on regression in the multivariate Poisson distribution, *Journal of the American Statistical Association*, 61, pp. 251–258.

MAHER, M. J. (1982) Modelling association football scores, *Statistica Neerlandica*, 36, pp. 109–118.

MARDIA, K. V. (1971) *Families of Bivariate Distributions*. Griffin's Statistical Monograph No. 27 (London, Charles Griffin).

MARSHALL, A. W. & OLKIN, I. (1967) A multivariate exponential distribution, *Journal of the American Statistical Association*, 62, pp. 30–44.

MCLACHLAN, G. & KRISHNAN, N. (1997) *The EM Algorithm and Extensions* (Chichester, Wiley).

MENG, X. L. & VAN DYK, D. (1997) The EM algorithm: an old folk song sung to a fast new tune (with discussion), *Journal of the Royal Statistical Society, B 59*, pp. 511–567.

PAPAGEORGIOU, H. & LOUKAS, S. (1995) A bivariate Charlier Series distribution, *Biometrical Journal*, 37, pp. 105–117.

STEIN, G. Z. & JURITZ, J. (1987) Bivariate compound Poisson distributions, *Communications in Statistics—Theory and Methods*, 16, pp. 3591–3607.

TSIONAS, E. (1999) Bayesian analysis of the multivariate Poisson distribution, *Communications in Statistics—Theory and Methods*, 28, pp. 431–451.

ZHENG, Q. & MATIS, J. H. (1993) Approximating discrete multivariate distributions from known moments, *Communications in Statistics—Theory and Methods*, 22, pp. 3553–3567.