

Bayesian multivariate Poisson mixtures with an unknown number of components

Loukia Meligkotsidou

Received: March 2006 / Accepted: October 2006 / Published online: 8 February 2007
© Springer Science + Business Media, LLC 2007

Abstract In this paper we present Bayesian analysis of finite mixtures of multivariate Poisson distributions with an unknown number of components. The multivariate Poisson distribution can be regarded as the discrete counterpart of the multivariate normal distribution, which is suitable for modelling multivariate count data. Mixtures of multivariate Poisson distributions allow for overdispersion and for negative correlations between variables. To perform Bayesian analysis of these models we adopt a reversible jump Markov chain Monte Carlo (MCMC) algorithm with birth and death moves for updating the number of components. We present results obtained from applying our modelling approach to simulated and real data. Furthermore, we apply our approach to a problem in multivariate disease mapping, namely joint modelling of diseases with correlated counts.

Keywords Bayesian inference · Disease mapping · Mixture models · Multivariate Poisson distribution · Reversible jump MCMC

1 Introduction

Finite mixture models have attracted a vast amount of attention over the past decades, as a flexible way of modelling heterogeneity or describing non-standard distributions. For general background on mixture models, see Titterington et al. (1985), Lindsay (1995) and McLachlan and Peel (2000). The missing data representation of mixture models and recent advances in computational methods have helped

in developing inferential procedures for these models, both in a classical and in a Bayesian framework. Maximum likelihood estimation (ML) of mixture distributions was proposed via the expectation-maximisation (EM) algorithm of Dempster et al. (1977), based on a data augmentation scheme. A similar scheme was used by Diebolt and Robert (1994) to construct a Gibbs sampler for Bayesian analysis of mixture distributions.

Much work on finite mixture models with an unknown number of components, K , considers assessment of the number of components after performing inference with fixed K . In particular, assessment of the number of components in a classical framework is most commonly done using non-parametric maximum likelihood (Lindsay, 1995; Bohning and Seidel, 2003). Various Bayesian methods have been proposed for inferring K , for example based on a Bayesian non-parametric approach (Escobar and West, 1995), Bayes factors (Chib, 1995) or distributional distances (Mengensen and Robert, 1996). More recently, fully Bayesian approaches to inference have been developed; namely reversible jump MCMC (Richardson and Green, 1997) and birth-and-death MCMC (Stephens, 2000), which treat K as a further parameter to be estimated together with the model parameters via trans-dimensional MCMC schemes. The latter schemes have been extremely popular in applications of mixture models (see Nobile and Green, 2000; Robert et al., 2000; Viallefont et al., 2000; Fernández and Green, 2002; Dellaportas and Papageorgiou, 2006, among others).

In general, most Bayesian applications of mixture models deal with univariate distributions, with a few exceptions concerning the multivariate normal mixture model (for example, see Dellaportas and Papageorgiou, 2006; Nobile and Fearnside, 2007). This model, although extremely useful, can not serve as an approximate model for multivariate count data sets, especially if they contain many small and/or zero

L. Meligkotsidou (✉)
Department of Mathematics and Statistics, Lancaster University,
Lancaster LA1 4YF
e-mail: l.meligkotsidou@lancaster.ac.uk

counts. Multivariate count data appear in a wide range of real life applications; for example, counts of different related types of diseases, accidents, crimes, product purchases, and different kinds of animals or plants among others. Karlis and Meligkotsidou (2007) have proposed finite mixtures of multivariate Poisson distributions to describe such data, as a natural extension of the univariate Poisson finite mixture model. The model parameters for any given number of components were estimated via ML using an EM type algorithm and the number of components was then assessed using various information based criteria.

In this paper we consider mixtures of the multivariate Poisson distribution with two-way covariance structure, which has a quite simple form and can be regarded as the discrete counterpart of the multivariate normal distribution. While the simple multivariate Poisson model allows only for positive correlations and can not account for overdispersion, its finite mixtures allow for both positive and negative correlations between pairs of variables as well as for overdispersion and are, in addition, computationally tractable. Alternative mixed Poisson models, like for example the multivariate Poisson-lognormal model of Aitchinson and Ho (1989) also exist, but they require a large computational burden for parameter estimation (see, Chib and Winkelmann, 2001; Wedel et al., 2003).

We perform full Bayesian analysis of multivariate Poisson mixtures with an unknown number of components. We construct a reversible jump MCMC algorithm (Green, 1995) consisting of updates of the model parameters and of K . The main obstacle in using the multivariate Poisson distribution for applications stems from the complicated form of the joint probability mass function (pmf), which causes difficulties in classical inferential procedures (see Johnson et al., 1997). The definition of the multivariate Poisson model through auxiliary variables suggests a data augmentation scheme for classical or Bayesian inference on the model parameters (Karlis and Meligkotsidou, 2005). Often, evaluation of the pmf can be accomplished using some efficient recurrence relationships (Kano and Kawamura, 1991; Tsiamyrtzis and Karlis, 2004). In a Bayesian framework, calculating the pmf can be avoided by constructing a simple Gibbs sampler to simulate the auxiliary variables from their full conditional posterior distributions (for details, see Section 2.2).

An important problem in Bayesian analysis of mixture distributions is the non identifiability of the components due to the invariance of the posterior distribution to permutations of the component labels. This problem occurs when exchangeable priors are used for the component specific parameters, in which case the arising marginal posterior distributions of these parameters are identical for all of the components. Jasra et al. (2005) provide a review of the existing solutions to this problem. However, in many applications interest lies in inferring quantities that, unlike the component specific

parameters, are invariant to permutations of the component labels (see, for example, Section 6). In the present paper we focus our attention on modelling overdispersed and correlated multivariate count data via multivariate Poisson mixtures and on inference about K and other quantities invariant to permutations of the component labels.

We present results obtained from applying our modelling approach to simulated data and to a real data set; the crime data of Karlis and Meligkotsidou (2005) which consist of observed counts of four different types of crime in the Greek prefectures. The aim is to identify the number of clusters comprising the studied population and to describe the correlation structure of the data. Furthermore, we apply our method to a problem in multivariate disease mapping. Our application concerns counts of deaths from three different types of cancer in the mainland French départements (the data set is taken from Rezvani et al. (1997)). Our objective is to estimate the risk surface for each type of cancer, taking into account information from the others.

The paper proceeds as follows. In Section 2 we introduce the multivariate Poisson distribution and describe Bayesian estimation of the model. In Section 3 we consider full Bayesian inference for the finite multivariate Poisson mixture model with unknown number of components. Section 4 discusses some computational issues regarding mixture modelling. Section 5 presents results from the application of our method to simulated and real multivariate count data, while Section 6 concerns an application of our approach to multivariate disease mapping. Finally, Section 7 concludes the paper.

2 The multivariate Poisson model with two-way covariance structure

2.1 The multivariate Poisson distribution

The derivation of multivariate Poisson models is based on a general multivariate reduction scheme. In this paper we only consider the multivariate Poisson model with two-way covariance structure. This can be regarded as the discrete analog of the multivariate normal model. This discrete counterpart, however, allows only for positive correlations between the variables, an assumption that is often but not always fulfilled by multivariate count data. The definition of this particular case of the multivariate Poisson distribution is given below. For further details on the general case see Johnson et al. (1997) and Karlis and Meligkotsidou (2005).

Consider the discrete-valued vector $\mathbf{Y} = (Y_1, \dots, Y_m)^T$, $\mathbf{Y} \in (\mathbb{N}_0)^m$. The definition of the m -variate Poisson distribution with two-way covariance structure is based on an augmented vector $\mathbf{Y}^* = (\mathbf{Y}_1^{*T}, \mathbf{Y}_2^{*T})^T$ of length $m^* = m(m+1)/2$ and a set of m many to one mappings. The

elements of the m -vector \mathbf{Y}_1^* , denoted by Y_{jj}^* , $j = 1, \dots, m$, and the elements of the $(m^* - m)$ -vector \mathbf{Y}_2^* , denoted by $Y_{j\ell}^*$, $j = 1, \dots, m - 1$, $j < \ell \leq m$ are assumed to be independent univariate Poisson random variables, i.e. $Y_{j\ell}^* \sim Poisson(\theta_{j\ell})$, $j = 1, \dots, m$, $j \leq \ell \leq m$. Let $\mathbf{A} = [\mathbf{I}_m, \mathbf{C}]$ be an $m \times m^*$ binary matrix, where \mathbf{I}_m is the identity matrix of size $m \times m$ and \mathbf{C} is an $m \times m(m-1)/2$ matrix; each column of \mathbf{C} has exactly 2 ones and $m-2$ zeros and there are no duplicate columns. Then the vector $\mathbf{Y} = \mathbf{AY}^*$, i.e. the vector whose elements are given by

$$Y_j = \sum_{\ell=1}^j Y_{\ell j}^* + \sum_{\ell=j+1}^m Y_{j\ell}^*, \quad j = 1, \dots, m, \quad (1)$$

follows the m -variate Poisson distribution with two-way covariance structure. We will denote this distribution by $\mathbf{Y} \sim MP_m(\theta)$, where θ is a parameter vector of length m^* with elements $\theta_{j\ell}$, $j = 1, \dots, m$, $j \leq \ell \leq m$. Throughout the paper the characterisation *with two-way covariance structure* will be implied and therefore omitted.

Under this model, the j th element of \mathbf{Y} marginally follows a univariate Poisson distribution with mean $\lambda_j = \sum_{\ell=1}^j \theta_{\ell j} + \sum_{\ell=j+1}^m \theta_{j\ell}$. Furthermore, $Cov[Y_j, Y_\ell] = \theta_{j\ell}$, for $j < \ell$. Hence, the parameter $\theta_{j\ell}$, has the natural interpretation of being the covariance between the variables Y_j and Y_ℓ . The mean and the covariance matrix of \mathbf{Y} are $E(\mathbf{Y} | \theta) = \mathbf{A}\theta = (\lambda_1, \dots, \lambda_m)^T$, and $Var(\mathbf{Y} | \theta) = \mathbf{A}Diag(\theta)\mathbf{A}^T$, where $Diag(\theta)$ is the diagonal covariance matrix of \mathbf{Y}^* . The multivariate Poisson model is a natural extension of the univariate Poisson distribution and, as its univariate case, can not account for overdispersion.

The joint probability of the m -vector $\mathbf{y} = (y_1, \dots, y_m)^T$ is given by the sum of the joint probabilities of all m^* -vectors \mathbf{y}^* which are consistent with all of the m mappings in (1). We denote the latter set by $S(\mathbf{y})$. Hence, the probability mass function (pmf) of \mathbf{Y} is defined by

$$\begin{aligned} P_m(\mathbf{y}; \theta) &= \Pr(\mathbf{Y} = \mathbf{y}; \theta) = \sum_{\mathbf{y}^* \in S(\mathbf{y})} \Pr(\mathbf{Y}^* = \mathbf{y}^*; \theta) \\ &= \sum_{\mathbf{y}^* \in S(\mathbf{y})} \prod_{1 \leq j \leq \ell \leq m} Po(y_{j\ell}^*; \theta_{j\ell}), \end{aligned} \quad (2)$$

where $Po(y; \theta)$ denotes the pmf of the univariate Poisson distribution. Analytical calculation of the multivariate Poisson probabilities via (2) is computationally expensive and increases both with m and with the values of the elements of \mathbf{y} . If the calculation of the probabilities is necessary, a solution can be the use of recurrence relationships (Kano and Kawamura, 1991; Tsiamirtzis and Karlis, 2004). However, an MCMC scheme for the analysis of the multivariate Poisson model can be based on the Gibbs sampler with data augmentation, avoiding the calculation of the pmf.

2.2 Bayesian estimation via data augmentation

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be independent and identically distributed (i.i.d) observations from the multivariate Poisson model, that is $\mathbf{Y}_i \sim MP_m(\theta)$, $i = 1, \dots, n$. The augmented vectors \mathbf{Y}_i^* , $i = 1, \dots, n$, introduced in Section 2.1 for the definition of the model can be regarded as a set of auxiliary variables. The joint distribution of $(\mathbf{Y}_i^T, \mathbf{Y}_i^{*T})$ is a product of m^* independent univariate Poisson distributions, with parameters $\theta_{j\ell}$. The point mass conditional distribution of $\mathbf{Y}_i^* | \mathbf{Y}_i$, obtained from (1) is

$$p(\mathbf{y}_i^* | \mathbf{y}_i, \theta) = \frac{\prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; \theta_{j\ell}) I[y_i^* \in S(\mathbf{y}_i)]}{\sum_{\mathbf{y}_i^* \in S(\mathbf{y}_i)} \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; \theta_{j\ell})}. \quad (3)$$

Although the vector \mathbf{y}_i^* has m^* components, there are only $m^* - m$ random variables, the remaining components being uniquely determined by the restriction that $\mathbf{y}_i^* \in S(\mathbf{y}_i)$. Thus, we can use (3) to draw $(\mathbf{y}_{2i}^* : y_{ij\ell}^*, \ell > j)$ and then set

$$y_{ijj}^* = y_{ij} - \sum_{\ell=1}^{j-1} y_{i\ell j}^* - \sum_{\ell=j+1}^m y_{ij\ell}^*, \quad j = 1, \dots, m, \quad (4)$$

according to (1). Simulating the vector \mathbf{y}_{2i}^* jointly (as in Karlis and Meligkotsidou (2005)) can be slow, or even exhausting for large counts and large m . Instead, one may use Gibbs sampling by drawing the elements of \mathbf{y}_{2i}^* one at a time conditional on the values of the others (at the cost of increasing the autocorrelations in the sampler). That is, each element $y_{ij\ell}^*$, $j = 1, \dots, m - 1$, $\ell > j$, of \mathbf{y}_{2i}^* can be drawn from its full conditional posterior distribution, which is proportional to $p(\mathbf{y}_i^* | \mathbf{y}_i, \theta)$ in (3) regarded as a function of $y_{ij\ell}^*$ only. The terms in the numerator of (3) which involve $y_{ij\ell}^*$ are $Po(y_{ijj}^*; \theta_{jj})$, $Po(y_{i\ell\ell}^*; \theta_{\ell\ell})$ and $Po(y_{ij\ell}^*; \theta_{j\ell})$, since y_{ijj}^* and $y_{i\ell\ell}^*$ are the only components of \mathbf{y}_{1i}^* involving $y_{ij\ell}^*$ according to (4). Therefore, the full conditional posterior distribution of $y_{ij\ell}^*$, $j = 1, \dots, m - 1$, $\ell > j$, is given by

$$\begin{aligned} p(y_{ij\ell}^* | \mathbf{y}_i, \mathbf{y}_{2i,(-j\ell)}, \theta) \\ \propto Po(y_{ijj}^*; \theta_{jj}) Po(y_{i\ell\ell}^*; \theta_{\ell\ell}) Po(y_{ij\ell}^*; \theta_{j\ell}) I[y_i^* \in S(\mathbf{y}_i)], \end{aligned}$$

where $\mathbf{y}_{2i,(-j\ell)}$ denotes the vector consisting of all the elements of \mathbf{y}_{2i}^* except for element $j\ell$. Given the augmented data \mathbf{y}_i^* simulation from the conditional posterior distributions of the model parameters, $\theta_{j\ell}$, is straightforward.

3 Finite multivariate Poisson mixtures

Let $\mathbf{y} = (y_1, \dots, y_m)^T$ be a multivariate observation from the multivariate Poisson mixture model

$$\mathbf{y} \sim \sum_{k=1}^K w_k P_m(\mathbf{y}; \theta_k), \quad (5)$$

where $P_m(\mathbf{y}; \theta)$ denotes the pmf of the multivariate Poisson distribution, and w_k , $k = 1, \dots, K$ is the weight of the k th component with parameter θ_k . The weights satisfy $w_k > 0$ and $\sum_{k=1}^K w_k = 1$. The expectation of the multivariate vector \mathbf{Y} under this mixture model is given by

$$\mathbb{E}(\mathbf{Y}) = \sum_{k=1}^K w_k \mathbf{A} \theta_k, \quad (6)$$

while its covariance matrix is given by

$$\text{Var}(\mathbf{Y}) = \mathbf{A} \left[\sum_{k=1}^K w_k \left(\text{Diag}(\theta_k) + \theta_k \theta_k^T \right) - \left(\sum_{k=1}^K w_k \theta_k \right) \left(\sum_{k=1}^K w_k \theta_k \right)^T \right] \mathbf{A}^T, \quad (7)$$

where \mathbf{A} is defined in 2.1. Note that the correlations between pairs of variables can be calculated as functions of the θ_k 's. An interesting feature of multivariate Poisson mixture models is that, unlike the single multivariate Poisson distribution, they can incorporate negative correlations between variables. This is an important extension of the multivariate Poisson model which, together with the ability of mixtures to capture overdispersion in the data, widens the applicability of these models. For details on the correlation structure of multivariate Poisson mixtures and on identifiability issues, see Karlis and Meligkotsidou (2007).

In the applications in this paper we consider m -variate observations \mathbf{y}_i , $i = 1, \dots, n$ where i denotes a region of population size N_i , from the multivariate Poisson mixture

$$\mathbf{y}_i \sim \sum_{k=1}^K w_k P_m(\mathbf{y}_i; N_i \theta_k), \quad (8)$$

where the population sizes are used as offsets.

3.1 Bayesian analysis

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be a sample of independent m -variate observations from model (8). According to the standard latent structure of finite mixture models (Titterington et al., 1985; Lindsay, 1995; McLachlan and Peel, 2000), we introduce n independent discrete variables z_i , $i = 1, \dots, n$, representing the allocations of the corresponding observations

\mathbf{y}_i to the mixture components. The hidden allocation variables have the multinomial distribution $p(z_i = k|K, \mathbf{w}) = w_k$, $k = 1, \dots, K$, where $\mathbf{w} = (w_1, \dots, w_K)$. Then, given the value of z_i , each observation \mathbf{y}_i is considered to be drawn from the respective mixture component, i.e. $\mathbf{y}_i|z_i \sim MP_m(N_i \theta_{z_i})$, independently for $i = 1, \dots, n$. The latent variables $\mathbf{z} = (z_1, \dots, z_n)^T$ are regarded as missing data. Hence, the complete data needed for inference consist of the augmented data \mathbf{y}_i^* , $i = 1, \dots, n$ and \mathbf{z} , and the complete data likelihood is

$$\begin{aligned} L(\mathbf{y}, \mathbf{y}^*, \mathbf{z} | K, \mathbf{w}, \theta) \\ = p(\mathbf{z} | K, \mathbf{w}) L(\mathbf{y}, \mathbf{y}^* | K, \mathbf{z}, \mathbf{w}, \theta) \\ = \prod_{i=1}^n \left\{ w_{z_i} \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{z_i j \ell}) I[y_i^* \in S(\mathbf{y}_i)] \right\}, \end{aligned}$$

where $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, $\mathbf{y}^* = (\mathbf{y}_1^*, \dots, \mathbf{y}_n^*)$ and $\theta = (\theta_1, \dots, \theta_K)$.

In the Bayesian framework, K is regarded as a further parameter of the model and prior distributions are specified for all the unknown parameters, K , \mathbf{w} and θ , in order to construct a hierarchical model (Richardson and Green, 1997; Nobile and Green, 2000). For the number of components, K , we adopt a Poisson(β) prior distribution, truncated over the set $\{1, \dots, K_{\max}\}$, where K_{\max} is the prespecified maximum value of K . Given K , we introduce independent random variables c_k , $k = 1, \dots, K$, with prior distribution $c_k \sim Gamma(\delta, 1)$. Let $S_K = \sum_{\ell=1}^K c_\ell$. Then, the mixture weights are defined by $w_k = c_k/S_K$. Note that the vector of weights \mathbf{w} follows *a-priori* a symmetric Dirichlet distribution $Dir(\delta, \dots, \delta)$, while the prior distribution of S_K is $Gamma(K\delta, 1)$. In what follows we will work with two equivalent parameterisations, namely with the variables w_1, \dots, w_{K-1}, S_K or with the reparametrisation c_1, \dots, c_K . For the component specific multivariate Poisson parameters θ_k we assume exchangeable priors given by the product of m^* independent Gamma distributions, that is

$$p(\theta_k) \equiv \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{j\ell}; a_{j\ell}, b_{j\ell}), \quad k = 1, \dots, K, \quad (9)$$

where $Ga(x; a, b)$ denotes the pdf of the Gamma distribution with mean a/b . If $p(K)$, $p(\mathbf{w}, S_K | K)$ and $p(\theta | K)$ denote the prior distributions for the model parameters, then the joint posterior distribution is given by

$$\begin{aligned} p(K, \mathbf{w}, S_K, \theta, \mathbf{z}, \mathbf{y}^* | \mathbf{y}) &\propto K! p(K) p(\mathbf{w}, S_K | K) p(\theta | K) \\ &\quad \times p(\mathbf{z} | K, \mathbf{w}) L(\mathbf{y}, \mathbf{y}^* | K, \mathbf{z}, \mathbf{w}, \theta), \end{aligned}$$

where the term $K!$ arises due to the exchangeability of the priors of the θ_k 's (see Cappé et al., 2003).

3.2 The sampler

An MCMC algorithm can be constructed to iteratively sample from the conditional posterior distributions of the model parameters. Denoting the current state of the chain by $(K, \mathbf{y}^*, \mathbf{z}, \theta, \mathbf{w})$, the sampler consists of the following steps (a)–(e). At the end of a complete sweep through the five steps the state is recorded. Following Fernández and Green (2002), we integrate the allocation variables out of the posterior distributions of the number of components and the data augmentation parameters in steps (a) and (b). For details on the sampler, see the Appendix.

- (a) Update of the number of components K : we use a reversible jump move to update K to K' . We propose to add a component, i.e. $K' = K + 1$, with probability $b_K \in [0, 1]$ and to remove a component, i.e. $K' = K - 1$, with probability $1 - b_K$. In applications in this paper the proposal probability for moving from K to $K + 1$ components is $b_K = 1$ if $K = 1$, $b_K = 1/2$ if $1 < K < K_{\max}$, and $b_K = 0$ if $K = K_{\max}$. Updating K implies a change in dimensionality for $\theta = (\theta_1, \dots, \theta_K)^T$ and $\mathbf{w} = (w_1, \dots, w_K)$. If $K' = K + 1$, we simulate the new component's parameter vector θ_* from the prior distribution (9) and the respective variable c_* from the prior $Ga(\delta, 1)$. We randomly place the new component in one of the $K + 1$ possible positions. Then we form θ' and \mathbf{c}' accordingly and set $\mathbf{w}' = \frac{1}{\sum_{k=1}^{K+1} c'_k} \mathbf{c}'$. If $K' = K - 1$, we randomly choose a component (*) to remove and form θ' and \mathbf{c}' by removing the respective values θ_* and c_* from θ and \mathbf{c} . We then set $\mathbf{w}' = \frac{1}{\sum_{k=1}^{K-1} c'_k} \mathbf{c}'$. The move that adds one component is accepted with probability $\min\{1, A\}$, where

$$A = \frac{1 - b_{K+1}}{b_K} \frac{\beta}{K + 1} \\ \times \prod_{i=1}^n \left\{ \frac{\sum_{k=1}^{K+1} w'_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta'_{k\ell})}{\sum_{k=1}^K w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{k\ell})} \right. \\ \left. \times I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \right\},$$

whereas the acceptance probability for the move that removes one component only requires obvious changes. If the proposed move is not accepted then the chain remains at the current state.

- (b) Update of the data augmentation parameters \mathbf{y}_i^* : we simulate the data augmentation parameters independently from the mixture, that is according to their conditional posterior distributions $p(\mathbf{y}_i^* | \mathbf{y}_i, N_i, \mathbf{w}, \theta)$, $i = 1, \dots, n$.
- (c) Update of the allocations z_i : we draw the allocation variables from their mutually independent full con-

ditional distributions $p(z_i = k | \mathbf{y}_i, \mathbf{y}_i^*, N_i, K, \mathbf{w}, \theta)$, $i = 1, \dots, n$.

- (d) Update of the mixture weights w_k , $k = 1, \dots, K$: we simulate the mixture weights from their joint posterior distribution, which is $Dir(\delta + n_1, \dots, \delta + n_K)$, where $n_k = \sum_{i=1}^n I[z_i = k]$, for $k = 1, \dots, K$. Then, we draw S_K from the $Gamma(\delta K, 1)$ density and set $\mathbf{c} = S_K \mathbf{w}$.
- (e) Update of the multivariate Poisson parameters θ_k : we simulate from the full conditional posterior distribution of each θ_k , that is $p(\theta_k | \mathbf{y}, \mathbf{y}^*, \mathbf{N}, \mathbf{z}, \mathbf{w})$ independently for $k = 1, \dots, K$.

The above reversible jump MCMC algorithm consists of Gibbs steps for updating the model parameters and of simple birth/death moves for updating K . The advantage of this updating scheme is that the formula for the acceptance probability of the reversible jump move is simple and there is no need for calculating Jacobians.

4 Computational issues

4.1 Prior specification

Here we discuss some issues related to the specification of the priors in the absence of prior information on the mixture parameters. If prior information is available, it can be used to specify informative prior distributions of the form discussed in Section 3.1. However, in applications in this paper we assume that there is no prior information available and, therefore, we attempt to specify prior distributions which are only weakly informative. Note that, in mixture modelling, it is not possible to assign improper non-informative prior distributions on the component specific parameters (see Diebolt and Robert, 1994). Since there is always the possibility that empty components arise for which the data will be uninformative, using improper non-informative priors would lead to improper posteriors for these components' parameters. Furthermore, if the priors of the component specific parameters are exchangeable, then the prior distribution of K must be also proper. According to the results of Nobile (2004) this is a necessary and sufficient condition for the posterior of K to be proper.

We use exchangeable priors for the component specific multivariate Poisson parameters and a truncated Poisson (or discrete uniform) prior distribution on K . We use weakly informative proper prior distributions for the θ_k 's, which may or may not be data dependent, as suggested by Richardson and Green (1997). In this line we adopt a prior of the form (9) which is centered at some ‘good’ value and has moderately large variance. One such choice is to use $a_{j\ell} = \kappa \hat{\theta}_{j\ell}$ and $b_{j\ell} = \kappa$, where $\hat{\theta}_{j\ell}$ is the MLE of $\theta_{j\ell}$ under the simple multivariate Poisson model ($K = 1$) and κ is a tuning constant which ensures that the variance of the prior is comparable to

the magnitude of the observed data. Under this specification, the prior mean of θ_{je} is $\hat{\theta}_{je}$, while its prior variance is $\hat{\theta}_{je}/\kappa$.

It is well known that the choice of prior for the component specific parameters in mixture models affects the posterior distribution of K . According to Richardson and Green (1997), as the prior on the component means becomes less informative (that is more disperse) at first the posterior distribution of K tends to favour a higher number of components. This is attributed to the fact that a more disperse prior is more permissive of components with close means. However, as the prior on these parameters becomes even more disperse, the posterior of K favours lower values of K , that is more parsimonious models (see Jennison, 1997; Stephens, 1997). Under a highly dispersed prior for the component specific parameters $\theta_k, k = 1, \dots, K$, the probability that all the θ_k 's take values consistent with a subset of the data becomes smaller as K increases.

In our setting, the prior on the θ_k 's becomes less informative as the value of the hyperparameter κ decreases. Note that, under the multivariate Poisson mixture model, both the component means and the component variances are functions of the θ_k 's. We have found that as κ increases, at first the posterior of K tends to put more mass to larger values of K . However, as κ increases further, the posterior of K tends to put less mass to larger values of K . This is consistent with the findings of Richardson and Green (1997). The sensitivity of the posterior of K to the choice of prior for $\theta_k, k = 1, \dots, K$, as well as for K will be illustrated in the examples (see Section 5).

4.2 Mixing and convergence

Taking into account the complexity of our model, we conclude that the mixing of our algorithm is in general acceptably good. If the prior distribution of the component specific parameters, which also serves as a proposal for the reversible jump move, is weakly informative, then the acceptance rate of the reversible jump step is high enough for the algorithm to mix quite well. In the examples in this paper the acceptance rates obtained range from 0.044 to 0.185. We have tried to use different proposals for the reversible jump step, other than drawing independently from the prior, but in all cases the acceptance rates achieved were lower. To account for the slow mixing we had long runs of our algorithm and we based inferences on 5000 equally spaced draws from the output. To ensure convergence we run several independent chains and compared the results. Also, convergence was monitored using standard approaches such as ergodic averaging for the parameters, or functions of parameters, which are not affected by label switching (results not shown).

5 Examples

5.1 Simulated data

The aim of our simulation experiment is firstly to examine the performance of our method at identifying the true number of mixture components and at uncovering the covariance structure of the data. Secondly, we are interested in investigating how the posterior distribution of K depends on the prior of the component specific parameters $\theta_k, k = 1, \dots, K$, and in particular on the choice of κ .

We have simulated data sets of size $n = 50$ and $n = 100$ from the 3-variate Poisson mixture model (5) with (a) $K = 2$ and (b) $K = 3$ components. Let $\theta_k = (\theta_{k11}, \theta_{k22}, \theta_{k33}, \theta_{k12}, \theta_{k13}, \theta_{k23})^T$ denote the parameter vector associated with the k th component. For $K = 2$ we have simulated data from the mixture with parameters $\theta_1 = (3.2, 1.1, 2.5, 6.3, 5.8, 10.2)^T$, $\theta_2 = (1.2, 0.8, 5.1, 4, 2.2, 3.8)^T$, and $p_1 = 0.4$, $p_2 = 0.6$. For $K = 3$ we have simulated data from the mixture with parameters θ_1 and θ_2 as above, $\theta_3 = (0, 8, 1, 8.6, 1.1, 6.7)^T$, and $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.3$. The expectation of \mathbf{Y} under model (a) is $(8.76, 10.3, 12.6)^T$ and under model (b) is $(11.02, 7.78, 13.14)^T$, while the respective covariance matrices are

$$(a) \begin{bmatrix} 13.838 & 5.712 & 5.505 \\ 5.712 & 12.460 & 6.492 \\ 5.506 & 6.492 & 16.194 \end{bmatrix}, \text{ and}$$

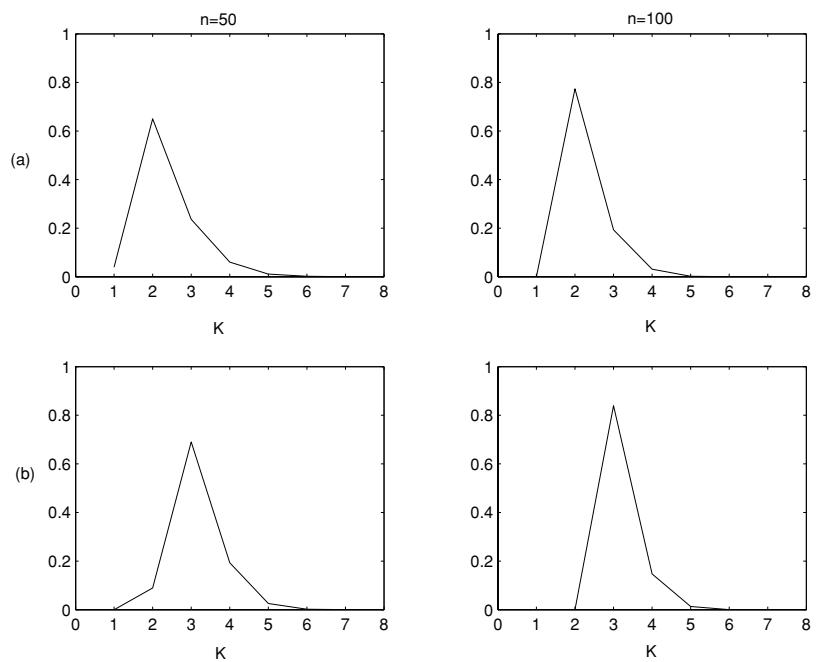
$$(b) \begin{bmatrix} 27.992 & -9.578 & 12.435 \\ -9.578 & 23.149 & -1.293 \\ 12.435 & -1.293 & 18.830 \end{bmatrix}.$$

We used a truncated Poisson prior on K with rate $\beta = 3$ and $K_{\max} = 8$, and a symmetric Dirichlet prior on the mixture weights with parameter $\delta = 1$. For all the component specific parameters $\theta_k, k = 1, \dots, K$, we used the prior distribution given by (9) with $a_{je} = \kappa \hat{\theta}_{je}$ and $b_{je} = \kappa$, where $\hat{\theta}_{je}$ is the MLE of θ_{je} under the simple multivariate Poisson model and κ has some prespecified value. In each case we run an MCMC algorithm with 110000 iterations discarding the first 10000 iterations as burn-in. The acceptance rates of these runs varied between 0.044 and 0.126.

5.1.1 Performance of the approach

Here we present results for a fixed value of κ , namely for $\kappa = 0.5$. A sensitivity analysis has shown that posterior inferences are quite robust to the choice of κ for these examples (see Section 5.1.2). In Fig. 1 is shown the posterior distribution of the number of components K , for each simulated data set.

Fig. 1 Posterior distribution of K for the simulated data sets of size $n = 50$ and $n = 100$ from the 3-variate Poisson mixture model (5) with (a) $K = 2$ and (b) $K = 3$ components



The mode of the posterior is always equal to the true value of K and has posterior probability grater than 0.5.

For each example, at the end of every iteration of the MCMC algorithm we calculated the expectation and the covariance matrix of \mathbf{Y} (based on the draws of the θ_k 's) using (6) and (7). In case (a), the posterior mean of the expectation for the sample of size $n = 50$ was $(9.391, 10.838, 13.443)^T$, while for the sample of size $n = 100$ it was $(9.025, 10.273, 12.032)^T$. The respective posterior means of the covariance matrices were

$$\begin{bmatrix} 11.468 & 3.727 & 4.154 \\ 3.727 & 11.945 & 5.111 \\ 4.154 & 5.111 & 18.957 \end{bmatrix}, \begin{bmatrix} 14.154 & 5.794 & 5.597 \\ 5.794 & 13.122 & 7.049 \\ 5.597 & 7.049 & 17.255 \end{bmatrix}.$$

In case (b), the posterior mean of the expectation for the sample of size $n = 50$ was $(11.646, 7.076, 13.418)^T$, while for the sample of size $n = 100$ it was $(10.810, 8.116, 13.498)^T$. The respective posterior means of the covariance matrices were

$$\begin{bmatrix} 32.881 & -10.721 & 15.539 \\ -10.721 & 21.403 & -3.529 \\ 17.539 & -3.529 & 23.703 \end{bmatrix},$$

$$\begin{bmatrix} 26.915 & -6.685 & 13.927 \\ -6.685 & 22.148 & -1.028 \\ 13.927 & -1.028 & 21.058 \end{bmatrix}.$$

It can be seen that the mean and the covariance structure of the simulated data is accurately reproduced using our approach. As expected, the accuracy of our estimates increases with the sample size.

5.1.2 Sensitivity analysis

Here we examine how the posterior distribution of the number of components depends on the prior of the component specific parameters. We have analysed the simulated data set of size 50 from the 2-component mixture and the simulated data set of size 100 from the 3-component mixture using the prior specification of Section 5.1 for several values of κ . The posterior distributions of K for different κ 's, for the two examples, are shown in Table 1. Note that a highly concentrated prior ($\kappa = 10$) favours a low value of K . Then, as the value of κ decreases, i.e. the prior becomes less informative, at first the posterior of K assigns more mass to higher numbers of components, while as it decreases further, the posterior tends to assign less mass to larger values of K . In practice, we advocate the use of a weakly informative prior on the component specific parameters, that is with prior variance which is equal or about two times the value of the prior mean ($\kappa \in [0.5, 1]$).

5.2 Crime data

We consider the crime data of Karlis and Meligkotsidou (2005). The data concern counts of four different types of crime in 49 Greek prefectures for year 1998. The different

Table 1 Posterior distribution of K for different values of κ for the simulated data set (a) of size 50 from the 2-component 3-variate Poisson mixture, and (b) of size 100 from the 3-component 3-variate Poisson mixture

K	1	2	3	4	5	≥ 6
Case (a)						
10	0.0795	0.8488	0.0691	0.0024	0.0000	0.0000
5	0.0100	0.8018	0.1647	0.0210	0.0020	0.0004
2	0.0049	0.7925	0.1774	0.0224	0.0024	0.0004
1	0.0047	0.6761	0.2727	0.0422	0.0030	0.0010
0.5	0.0013	0.6979	0.2583	0.0357	0.0045	0.0022
0.1	0.0025	0.7120	0.2489	0.0326	0.0040	0.0000
0.05	0.0030	0.8076	0.1657	0.0217	0.0022	0.0000
Case (b)						
10	0.0000	0.0000	0.8339	0.1487	0.0157	0.0018
5	0.0000	0.0000	0.8007	0.1728	0.0163	0.0003
2	0.0000	0.0000	0.7945	0.1742	0.0298	0.0015
1	0.0000	0.0000	0.8055	0.1721	0.0195	0.0029
0.5	0.0000	0.0000	0.8144	0.1669	0.0173	0.0014
0.1	0.0000	0.0000	0.8193	0.1625	0.0174	0.0008
0.05	0.0000	0.0000	0.8274	0.1556	0.0156	0.0014

crimes considered are rapes (Y_1), arsons (Y_2), manslaughter (Y_3) and smuggling of antiquities (Y_4). The population (in millions) of each prefecture is also known. The data were collected from the relevant publications of the National Statistical Service of Greece. This data set has been previously analysed by Karlis and Meligkotsidou (2005) via multivariate Poisson regression and by Karlis and Meligkotsidou (2007) using multivariate Poisson mixtures. In the latter paper, an EM type algorithm was used for ML inference on mixture models with a given number of components. Then various information based criteria were considered for assessing the number of components, under a non-parametric ML approach. Here we perform Bayesian analysis of this data set under a multivariate Poisson mixture model with an unknown number of components.

The focus of our analysis was on modelling the correlation structure of the data and on inference about K . We used a truncated Poisson prior on K with rate $\beta = 3$ and $K_{\max} = 12$, and a symmetric Dirichlet prior on the weights with parameter $\delta = 1$. For all the component specific parameters θ_k , $k = 1, \dots, K$, we used the prior distribution given by (9) with $a_{je} = \kappa \hat{\theta}_{je}$ and $b_{je} = \kappa$, where $\kappa = 0.5$ and $\hat{\theta}_{je}$ is the MLE of θ_{je} under the simple multivariate Poisson model (see Karlis and Meligkotsidou, 2005).

We run an MCMC algorithm with 110000 iterations discarding the first 10000 iterations as burn-in. The acceptance rate of this run was 0.185. In Fig. 2 is shown the posterior distribution of the number of components K . The mode of this posterior is $K = 4$ and has posterior probability 0.403. For comparison, it is also shown the posterior of K under two other priors; a truncated Poisson prior with rate $\beta' = 2$ and

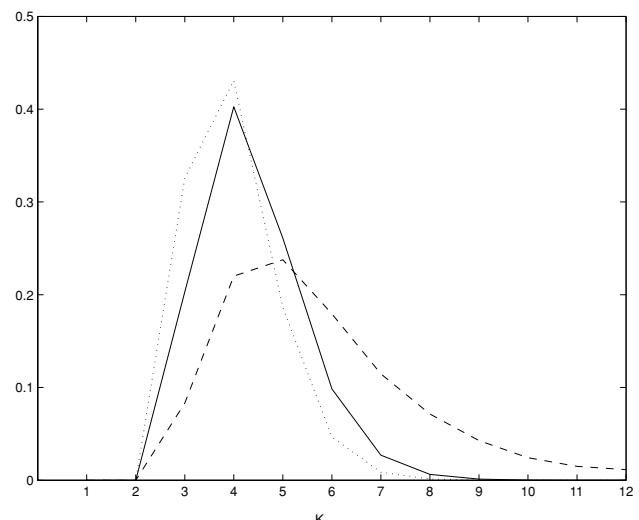


Fig. 2 Posterior distribution of K . The solid line corresponds to the truncated Poisson prior with rate $\beta = 3$, the dotted line corresponds to the truncated Poisson prior with rate $\beta' = 2$, and the dashed line corresponds to the discrete uniform prior over the set $\{1, \dots, 12\}$

a discrete uniform prior over the set $\{1, \dots, 12\}$. Note that the optimal number of components indicated by the Akaike Information Criterion (AIC) for this data set, as reported by Karlis and Meligkotsidou (2007), is $\hat{K} = 4$. However, different criteria advocate different optimal values for K under the non-parametric ML approach. On the other hand, our fully Bayesian approach, instead of providing a ‘best estimate’ of K , bases inferences on the posterior distribution of K which conveys the uncertainty about the number of mixture components.

Our analysis has shown that there exist significant correlations between different types of crime, both positive and negative, which are captured by our multivariate Poisson mixture model. Specifically, the correlation of the pair rapes-manslaughter is positive, as expected, and quite high (the posterior mode of this correlation is 0.43). On the other hand, the correlation of the pair arsons-manslaughter is negative (the posterior mode is -0.4). This can be explained by the fact that arsons occur mostly in non-inhabited regions.

To examine how good our multivariate Poisson mixture model is in terms of fitting the data we perform the following comparative study. We compare our model (denoted by M_1) to a simple mixture of independent univariate Poisson distributions for the different variables (denoted by M_2). The comparison is in terms of the in-sample predictive performance of the two models. For each model we estimate the set of the conditional predictive ordinates, $\Pr(\mathbf{y}_i | \mathbf{y}_{-i}, M_j)$, $i = 1, \dots, n$, $j = 1, 2$, where \mathbf{y}_{-i} denotes the set of the observed data except for the i th observation (see Pettit and Young, 1990). For both models the estimation is based on n independent MCMC runs, each excluding the i th observation, $i = 1, \dots, n$. A large value of $\Pr(\mathbf{y}_i | \mathbf{y}_{-i}, M_j)$

indicates that the i th observation supports the model. The ratio

$$R = \frac{\prod_{i=1}^n \Pr(\mathbf{y}_i | \mathbf{y}_{-i}, M_1)}{\prod_{i=1}^n \Pr(\mathbf{y}_i | \mathbf{y}_{-i}, M_2)}$$

can be regarded as a surrogate for the Bayes factor (Gelfand et al., 1992). If $R > 1$ the data support M_1 , otherwise M_2 is supported. We have estimated the pseudo-Bayes factor as $R = 10.7$. Therefore, we conclude that the multivariate Poisson mixture offers better model fit than the mixture of independent univariate Poisson distributions.

6 Application to multivariate disease mapping

In recent years there has been increased interest in the study of variation in rates of the incidence of disease or mortality within a geographical area. Several applications in disease mapping concerning one particular disease can be found in the literature, including Besag et al. (1991), Knorr-Held and Rasser (2000), Biggeri et al. (2000), Fernández and Green (2002), and Green and Richardson (2002). However, the applications in multivariate disease mapping are rather sparse (see, for example, Knorr-Held and Best, 2001).

We model different types of diseases jointly using multivariate Poisson mixtures, which take into account the correlations between pairs of diseases. In this way we can study the co-variation in rates of incidence of several diseases and may detect certain common causes of different types of diseases. The mixture models also account for the extra-Poisson variation in the data, i.e. the heterogeneity in the population that could not be captured by a single multivariate Poisson distribution.

If we also incorporate spatial correlation in the model, we can further uncover spatial patterns in disease spread (see, Fernández and Green, 2002, for the univariate case). This could help to formulate aetiological hypotheses towards certain common geographical or environmental causes of the diseases under consideration. We do not consider spatial multivariate Poisson mixtures in the present paper, we note however that this is an interesting and very useful extension of our work.

Our analysis is based on multivariate count data consisting of the numbers of observed cases of m diseases, $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$, $i = 1, \dots, n$, for n regions that constitute a geographical partition of the area of interest. For the j th disease, a preliminary analysis of the counts y_{ij} , usually under a population stratification on age and sex, yields an expected count E_{ij} for each region i . Our objective is to estimate the risk surface for each disease, taking into account the information from the others. We assume that the vectors \mathbf{y}_i have conditionally independent distributions given in (8),

with N_i being the population at risk in region i , which should be the same for all diseases under consideration. This leads to natural expressions for the expectations

$$E(y_{ij}) = \text{Var}(y_{ij}) = N_i \left(\sum_{\ell=1}^j \theta_{z_i \ell j} + \sum_{\ell=j+1}^m \theta_{z_i \ell j} \right), \\ j = 1, \dots, m, \quad (10)$$

$$\text{Cov}(y_{ij}, y_{i\ell}) = N_i \theta_{z_i j \ell}, \quad 1 \leq j < \ell \leq m.$$

The relative risk R_{ij} of the j th disease for region i with allocation $z_i \in \{1, \dots, K\}$ is then given as

$$R_{ij} = \frac{N_i}{E_{ij}} \left(\sum_{\ell=1}^j \theta_{z_i \ell j} + \sum_{\ell=j+1}^m \theta_{z_i \ell j} \right). \quad (11)$$

Note that the expressions in (10) suggest that the population in each region is homogeneous. If we work with age and/or sex specific populations at risk, then this assumption is fulfilled. If not, the estimates of the risks given in (11) are still valid, since the expected counts E_{ij} , calculated under the population stratification, can be thought of as a correction for heterogeneity.

Obviously, while the populations at risk are the same for all diseases, the relative risks are not. This is the reason why we choose to work with the N_i 's instead of the R_{ij} 's in the multivariate Poisson mixture setting, unlike what is usually done in univariate Poisson disease mapping applications, where the model used for univariate counts y_i is

$$y_i \sim \sum_{j=1}^k w_j \text{Po}(y_i; E_i R_i).$$

In order to be consistent with the prior specification in those applications, if the risk for the j th disease R_{ij} would follow a $\text{Gamma}(a_{jj}, a_{jj})$ prior so as the mean to be unity, then we choose

$$\theta_{j\ell} \sim \text{Gamma}(a_{j\ell} b_{j\ell}, a_{j\ell}), \quad 1 \leq j < \ell \leq m \quad \text{and} \\ \theta_{jj} \sim \text{Gamma}\left(a_{jj} \left[\frac{\sum_{i=1}^n y_{ij}}{\sum_{i=1}^n N_i} - \sum_{r \neq j} b_{jr} \right], a_{jj} \right). \quad (12)$$

Under this prior for the multivariate Poisson parameters, the prior means of the risks of all diseases are again equal to 1.

Since the relative risks are the real quantities of interest, they are calculated via (11) at the end of each sweep of the MCMC algorithm, given K , θ and the allocations of the regions. Hence, the labeling of the mixture components does

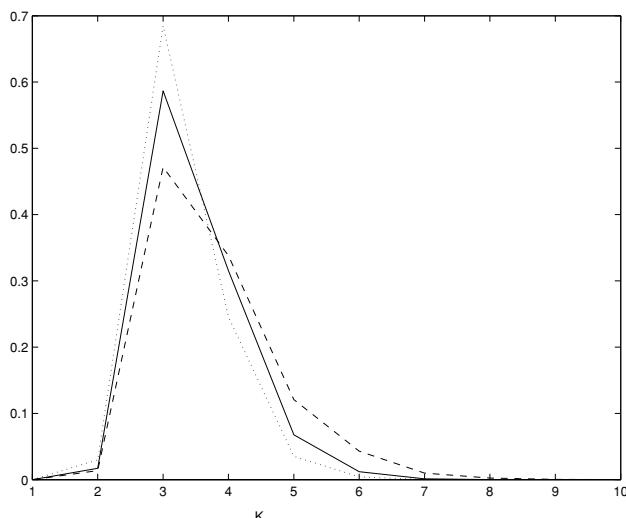


Fig. 3 Posterior distribution of K . The solid line corresponds to the truncated Poisson prior with rate $\beta = 3$, the dotted line corresponds to the truncated Poisson prior with rate $\beta' = 2$, and the dashed line corresponds to the discrete uniform prior over the set $\{1, \dots, 10\}$

not really matter for this application, as the summary of our analysis consists of the relative risks for all diseases and all regions, the correlations between each pair of diseases and the number of mixture components K .

6.1 Cancer data

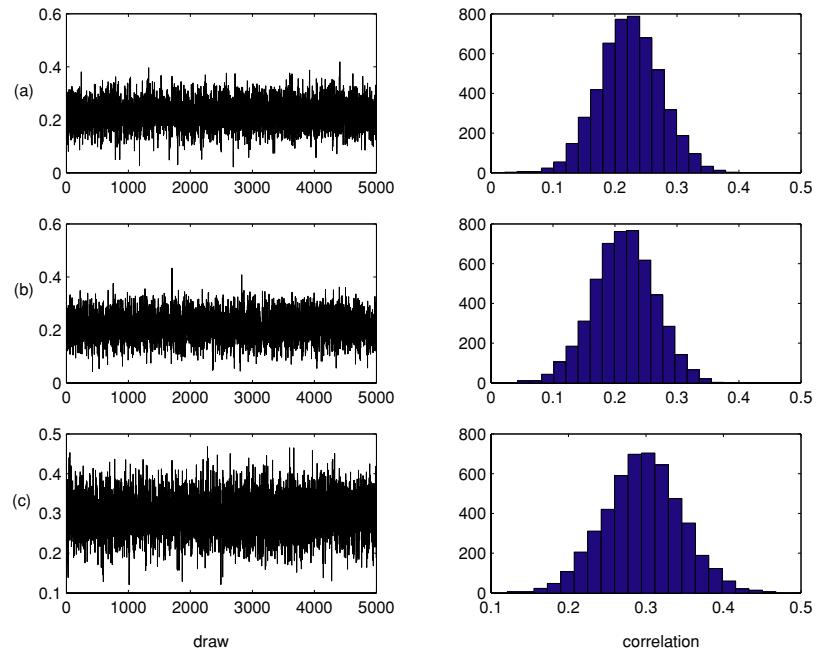
We consider cancer data consisting of counts of deaths from three different types of cancer (larynx, pharynx and oral cavity) for females in the 94 mainland French départements. The data set is taken from the “Atlas de la mortalité par Cancer

en France, Période 1986–1993” (Rezvani et al., 1997). The female populations at risk are available and the expected numbers of cases E_{ij} are calculated under the population stratification on age. The observed counts of deaths in our data set range from 0 (Lozère) to 75 (Paris) for larynx cancer, from 1 (Ariège and Lozère) to 155 (Paris) for pharynx cancer and from 5 (Ariège and Haute-Saône) to 152 (Paris) for oral cavity cancer. The total observed and expected counts are equal for each type of cancer and they are given by (1339, 2379, 2849) for larynx, pharynx and oral cavity cancer, respectively. The standardised mortality ratios (SMRs) vary from 0 (Lozère) to 2.12 (Hautes-Alpes) for larynx cancer, from 0.14 (Ariège) to 1.72 (Seine-Saint-Denis) for pharynx cancer and from 0.41 (Haute-Saône) to 1.63 (Ardennes) for oral cavity cancer.

The larynx cancer data set has been analysed by Fernández and Green (2002) using spatial univariate Poisson mixtures. Here we model jointly different types of cancer, using finite multivariate Poisson mixtures. The quantities of ultimate interest are the region-specific relative risks. Unlike the component-specific parameters $\theta_1, \dots, \theta_K$ which can only be inferred conditional on a fixed number of components, the relative risks have an unequivocal interpretation regardless of the value of K .

For the prior of the multivariate Poisson parameters in (12), which is also the proposal for the reversible jump step, we chose the hyperparameter values $a_{jj} = a_{j\ell} = 3$ for all j and ℓ , $b_{12} = b_{13} = 1$ and $b_{23} = 3$. These values were chosen after tuning the variance of the proposal to achieve reasonable acceptance rates for the reversible jump step. The prior on K was truncated Poisson with rate $\beta = 3$ and

Fig. 4 Histograms and trace plots of the unconditional correlations between (a) larynx and pharynx cancer, (b) larynx and oral cavity cancer, and (c) pharynx and oral cavity cancer



$K_{\max} = 10$, while the value of the parameter of the Dirichlet prior on the mixture components was $\delta = 1$.

We run an MCMC algorithm with 220000 iterations discarding the first 20000 iterations as burn-in. The acceptance rate of this run was 0.0548. In Fig. 3 is shown the posterior distribution of the number of components K . The mode of this posterior is $K = 3$ and has posterior probability 0.587. For comparison, it is also shown the posterior of K under two other priors; a truncated Poisson prior with rate $\beta' = 2$ and a discrete uniform prior over the set $\{1, \dots, 10\}$.

Figure 4 presents histograms and trace plots of the unconditional correlations between the three different pairs of diseases. It can be seen that all of the correlations are positive and significant. In Figs. 5–7 are displayed the pos-

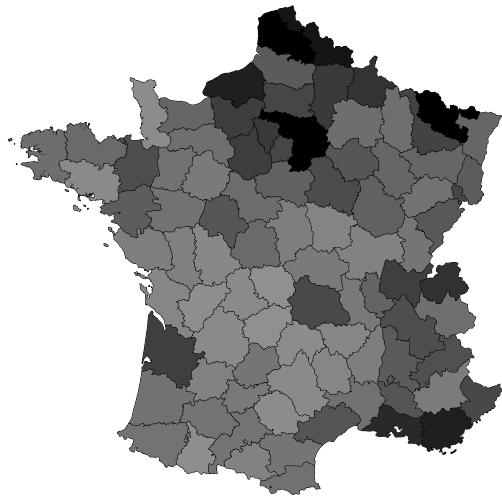


Fig. 5 Posterior means of the relative risks in the mainland French départements for larynx cancer. The posterior mean relative risks range from 0.692 (lightest) to 1.536 (darkest)

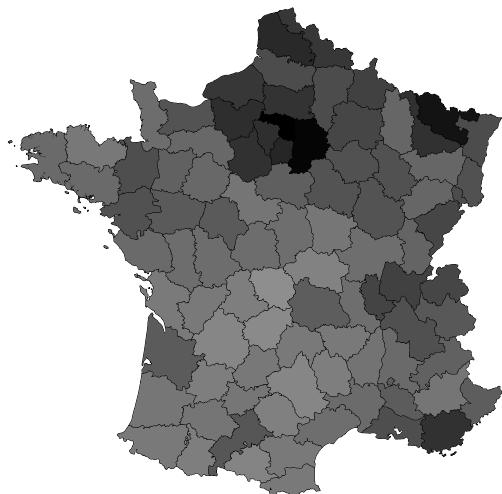


Fig. 6 Posterior means of the relative risks in the mainland French départements for pharynx cancer. The posterior mean relative risks range from 0.701 (lightest) to 1.421 (darkest)

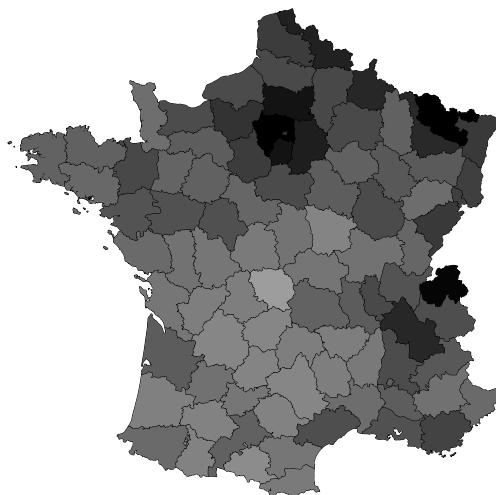


Fig. 7 Posterior means of the relative risks in the mainland French départements for oral cavity cancer. The posterior mean relative risks range from 0.627 (lightest) to 1.365 (darkest)

terior means of the relative risks in the mainland French départements for the three types of cancer. The posterior mean relative risks range from 0.692 (Corrèze) to 1.536 (Seine-Saint-Denis) for larynx cancer, from 0.701 (Creuse) to 1.421 (Seine-Saint-Denis) for pharynx cancer, and from 0.627 (Creuse) to 1.365 (Val d'Oise) for oral cavity cancer. The pattern in the three maps is similar, which suggests that there exist some common causes of the three different types of cancer considered. This can be explained by the correlation structure between diseases. Figure 8 presents scatter plots of the relative risk of one type of cancer versus another for the three pairs of diseases at certain French départements. In all cases the relative risks are positively correlated.

Furthermore, the multivariate Poisson mixture model seems to uncover the spatial correlation structure of the data. Comparing Fig. 5 to Fig. 5 of Fernández and Green (2002) it can be seen that the same spatial pattern is reproduced. Figure 9 displays the scatter plot of the relative risk of larynx cancer at Aisne versus the relative risk of larynx cancer at its neighbouring département of Marne. Again these results are similar to those obtained by Fernández and Green (2002) which suggests that spatial clusters are captured by our independent mixture model. However, the predictive performance of a spatial model (like, for example those considered by Fernández and Green (2002), and Green and Richardson (2002)) would be better. Such a model would be able to use information from neighbouring départements in order to predict the relative risk of a given département. Therefore, more sophisticated models should be used in order to incorporate spatial correlation, however this is beyond the scope of the present paper.

Fig. 8 Scatter plots of the relative risk at certain French départements of (a) larynx cancer versus pharynx cancer, (b) larynx cancer versus oral cavity cancer, and (c) pharynx cancer versus oral cavity cancer

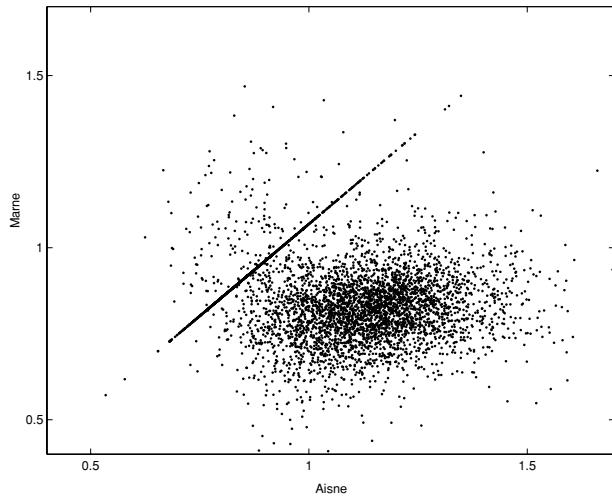
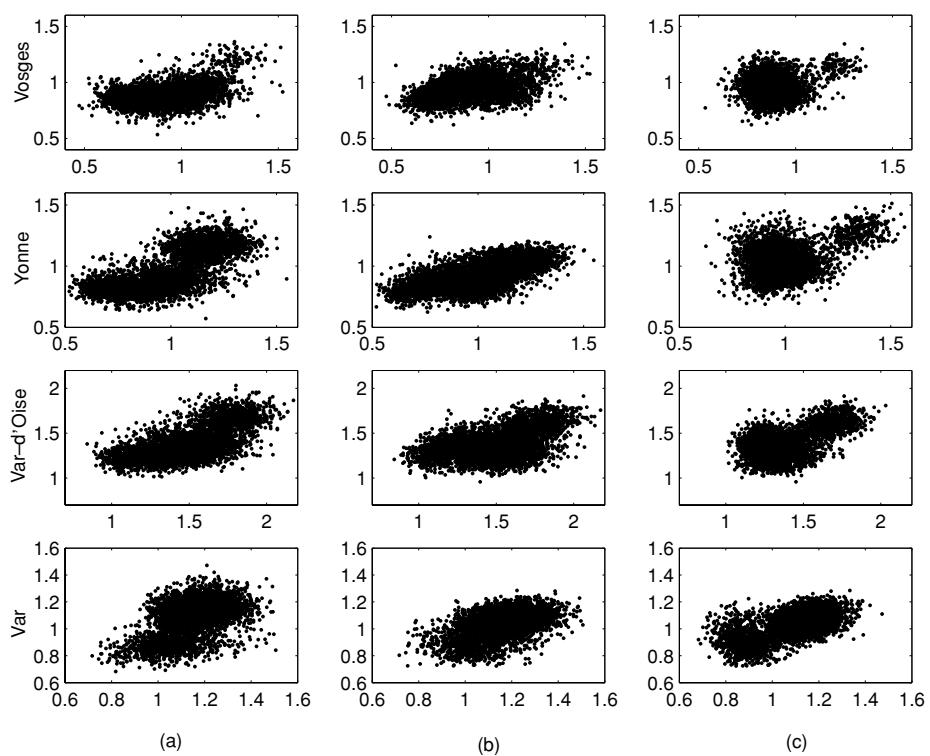


Fig. 9 Scatter plot of the relative risk of larynx cancer at Aisne versus the relative risk of larynx cancer at its neighbouring département of Marne

7 Conclusions

The multivariate Poisson distribution is an appropriate model for correlated multivariate count data, since it models separately the covariance of each different pair of variables. While this simple model only allows for positive correlations between variables, multivariate Poisson mixtures also allow for negative correlations as well as for overdispersion. Therefore, such mixtures provide a flexible and useful

tool for modelling overdispersed and correlated multivariate count data.

In this paper we have considered full Bayesian inference for multivariate Poisson mixtures with an unknown number of components. We have overcome the difficulty of calculating the pmf of the multivariate Poisson distribution by constructing a Gibbs sampler to simulate the auxiliary variables needed for the definition of this distribution. We have adopted a reversible jump MCMC algorithm with simple birth and death moves for updating the number of mixture components.

An important problem in Bayesian analysis of mixture distributions is the non identifiability of the components due to the invariance of the posterior distribution to permutations of the component labels. However, in the present paper we have focused our attention on modelling the correlation structure of overdispersed and correlated multivariate count data and on inference about the number of mixture components as well as about quantities which are invariant to permutations of the component labels.

We have applied our approach to simulated data and to a real data set consisting of counts of four different types of crime. The approach seems to work well in terms of uncovering the correlation structure of the data and inferring the number of mixture components. Furthermore, we have considered joint modelling of different types of diseases via multivariate Poisson mixtures. We have analysed a data set consisting of counts of three different types of cancer. Our aim was to study the co-variation in rates of incidence of the

three diseases. The quantities of interest in our analysis were the relative risks of the diseases and the correlations between pairs of diseases, therefore the label switching problem did not affect inferences in this application.

Although both data sets under consideration are likely to be spatially correlated, we did not take into account spatial correlation under our model specification. Our independent multivariate Poisson mixture model was shown to capture spatial patterns in the data. However, incorporating spatial correlation in the model could help to formulate aetiological hypothesis towards common geographical causes of crime or disease occurrence. We believe that extending our model in the spirit of Fernández and Green (2002) should be an interesting topic of future research.

Appendix: Details on the sampler

The joint posterior distribution of the model's unknown quantities, including the model parameters and the auxiliary variables, is given by

$$\begin{aligned} p(K, \mathbf{w}, S_K, \theta, \mathbf{z}, \mathbf{y}^* | \mathbf{y}) \\ \propto K! p(K) p(\mathbf{w}, S_K | K) p(\theta | K) p(\mathbf{z} | K, \mathbf{w}) \\ \times L(\mathbf{y}, \mathbf{y}^* | K, \mathbf{z}, \mathbf{w}, \theta) \\ \propto K! Po(K; \beta) Dir(\mathbf{w}; \delta, \dots, \delta) Ga(S_K; K\delta, 1) \\ \times \prod_{k=1}^K \left\{ \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{kj\ell}; a_{j\ell}, b_{j\ell}) \right\} \\ \times \prod_{i=1}^n \left\{ w_{z_i} \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{z_i j \ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \right\}, \end{aligned}$$

or, equivalently, by

$$\begin{aligned} p(K, \mathbf{c}, \theta, \mathbf{z}, \mathbf{y}^* | \mathbf{y}) \\ \propto K! Po(K; \beta) \prod_{k=1}^K \left\{ Ga(c_k; \delta, 1) \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{kj\ell}; a_{j\ell}, b_{j\ell}) \right\} \\ \times \prod_{i=1}^n \left\{ w_{z_i} \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{z_i j \ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \right\}. \end{aligned}$$

Now, integrating out the allocation variables \mathbf{z} , we obtain

$$\begin{aligned} p(K, \mathbf{c}, \theta, \mathbf{y}^* | \mathbf{y}) \\ \propto K! Po(K; \beta) \prod_{k=1}^K \left\{ Ga(c_k; \delta, 1) \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{kj\ell}; a_{j\ell}, b_{j\ell}) \right\} \end{aligned}$$

$$\times \prod_{i=1}^n \sum_{k=1}^K \left\{ w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{kj\ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \right\}.$$

In the following MCMC algorithm the allocation variables are integrated out of the posterior distributions of K and \mathbf{y}_i^* , $i = 1, \dots, n$ (steps (a) and (b)).

- (a) We update K to K' via a reversible jump move as described in Section 3.2. Let $q(K, K')$ denote the proposal probability of moving from K to K' . Then, the move that adds one component is accepted with probability $\min\{1, A\}$, where

$$\begin{aligned} A &= \frac{(K+1)! p(K+1, \mathbf{c}', \theta', \mathbf{y}^* | \mathbf{y}) q(K+1, K)}{K! p(K, \mathbf{c}, \theta, \mathbf{y}^* | \mathbf{y}) q(K, K+1)} \\ &= \frac{Po(K+1; \beta) \prod_{k=1}^{K+1} \{ Ga(c'_k; \delta, 1) \prod_{1 \leq j \leq \ell \leq m} Ga(\theta'_{kj\ell}; a_{j\ell}, b_{j\ell}) \}}{Po(K; \beta) \prod_{k=1}^K \{ Ga(c_k; \delta, 1) \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{kj\ell}; a_{j\ell}, b_{j\ell}) \}} \\ &\quad \times \frac{\prod_{i=1}^n \sum_{k=1}^{K+1} \{ w'_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta'_{kj\ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \}}{\prod_{i=1}^n \sum_{k=1}^K \{ w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{kj\ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \}} \\ &\quad \times \frac{1 - b_{K+1}}{b_K} \frac{(K+1)^{-1}}{Ga(c_*; \delta, 1) \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{*j\ell}; a_{j\ell}, b_{j\ell})} \\ &= \frac{1 - b_{K+1}}{b_K} \frac{\beta}{K+1} \\ &\quad \times \prod_{i=1}^n \left\{ \frac{\sum_{k=1}^{K+1} w'_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta'_{kj\ell})}{\sum_{k=1}^K w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{kj\ell})} \right. \\ &\quad \left. \times I[\mathbf{y}_i^* \in S(\mathbf{y}_i)] \right\}. \end{aligned}$$

- (b) We simulate \mathbf{y}_i^* , $i = 1, \dots, n$, independently from their conditional posterior distributions

$$\begin{aligned} p(\mathbf{y}_i^* | \mathbf{y}_i, N_i, \mathbf{w}, \theta) \\ \propto \sum_{k=1}^K w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{kj\ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)], \\ i = 1, \dots, n. \quad (13) \end{aligned}$$

We construct a Gibbs sampler to draw the $m^* - m$ random elements of \mathbf{y}_i^* (i.e. the elements of \mathbf{y}_{2i}^*) one at a time, and then use (1) to obtain the non-random ones, as stated in Section 2.2. The extra complication here is that the \mathbf{y}_{2i}^* 's have to be drawn from the mixture distribution (13), since the allocations have been integrated out. Therefore, each element $y_{ij\ell}^*$, $j = 1, \dots, m-1$, $\ell < j$, of \mathbf{y}_{2i}^* , $i = 1, \dots, n$, will be drawn from its full conditional posterior distribution which is proportional to $p(\mathbf{y}_i^* | \mathbf{y}_i, N_i, \mathbf{w}, \theta)$ in (13) regarded as a function of $y_{ij\ell}^*$ only.

(c) We draw z_i , $i = 1, \dots, n$, from their mutually independent full conditional distributions

$$\begin{aligned} p(z_i = k | \mathbf{y}_i, \mathbf{y}_i^*, N_i, K, \mathbf{w}, \theta) \\ \propto w_k \prod_{1 \leq j \leq \ell \leq m} Po(y_{ij\ell}^*; N_i \theta_{kj\ell}) I[\mathbf{y}_i^* \in S(\mathbf{y}_i)], \\ i = 1, \dots, n. \end{aligned}$$

(d) We simulate the mixture weights, \mathbf{w} , from their joint posterior distribution

$$\begin{aligned} p(\mathbf{w} | \mathbf{z}) &\propto \left\{ \prod_{k=1}^K w_k^{\delta-1} \right\} \left\{ \prod_{i=1}^n w_{z_i} \right\} = \prod_{k=1}^K w_k^{\delta+n_k-1} \\ &\equiv Dir(\delta + n_1, \dots, \delta + n_K), \end{aligned}$$

where $n_k = \sum_{i=1}^n I[z_i = k]$, for $k = 1, \dots, K$. Given K and the $Gamma(\delta, 1)$ prior for the variables c_k , the distribution of the sum $S_K = \sum_{\ell=1}^K c_\ell$ is $Gamma(\delta K, 1)$ and it is independent from the posterior distribution of \mathbf{w} . Hence, we draw S_K from this Gamma density and we then update \mathbf{c} via $\mathbf{c} = S_K \mathbf{w}$.

(e) We simulate each θ_k , $k = 1, \dots, K$, independently from its full conditional posterior, that is

$$\begin{aligned} p(\theta_k | \mathbf{y}, \mathbf{y}^*, \mathbf{N}, \mathbf{z}, \mathbf{w}) \\ \propto \prod_{1 \leq j \leq \ell \leq m} \theta_{kj\ell}^{a_{j\ell}-1} e^{-b_{j\ell}\theta_{kj\ell}} \\ \times \prod_{i=1}^n \left\{ \prod_{1 \leq j \leq \ell \leq m} \theta_{z_i j \ell}^{y_{ij\ell}^*} e^{-\theta_{z_i j \ell} N_i} I[z_i = k] \right\} \\ = \prod_{1 \leq j \leq \ell \leq m} \theta_{kj\ell}^{a_{j\ell} + \sum_{i=1}^n y_{ij\ell}^* I[z_i=k]-1} e^{-\{b_{j\ell} + \sum_{i=1}^n N_i I[z_i=k]\} \theta_{kj\ell}} \\ \equiv \prod_{1 \leq j \leq \ell \leq m} Ga(\theta_{kj\ell}; a_{j\ell} + \sum_{i=1}^n y_{ij\ell}^* I[z_i = k], \\ b_{j\ell} + \sum_{i=1}^n N_i I[z_i = k]). \end{aligned}$$

Acknowledgment I would like to thank Carmen Fernández for numerous helpful discussions. I am also grateful to the editor, the associate editor and two referees for many useful comments and suggestions which improved the paper.

References

- Aitchinson J. and Ho C.H. 1989. The multivariate Poisson-log normal distribution. *Biometrika* 75: 621–629.
- Besag J., York J., and Mollié A. 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* 43: 1–59.
- Biggeri A., Marchi M., Lagazio C., Martuzzi M., and Bohning D. 2000. Non-parametric maximum likelihood estimators for disease mapping. *Statistics in Medicine* 19: 2539–2554.
- Bohning D. and Seidel W. 2003. Editorial: Recent developments in mixture models. *Computational Statistics and Data Analysis* 41: 349–357.
- Cappé O., Robert C.P., and Rydén T. 2003. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B* 65: 679–700.
- Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90: 1313–1321.
- Chib S. and Winkelmann R. 2001. Markov chain Monte Carlo analysis of correlated count data. *Journal of Business and Economic Statistics* 19: 428–435.
- Dellaportas P. and Papageorgiou I. 2006. Multivariate mixtures of normals with unknown number of components. *Statistics and Computing* 16: 57–68.
- Dempster A.P., Laird N.M., and Rubin D.B. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39: 1–38.
- Diebolt J. and Robert C. 1994. Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society B* 56: 363–375.
- Escobar M.D. and West M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90: 577–588.
- Fernández C. and Green P.J. 2002. Modelling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society, Series B* 64: 805–826.
- Gelfand A.E., Dey D.K., and Chang H. 1992. Model determination using predictive distributions with implementation via sampling-based methods. In: Bernardo J.M., Berger J.O., Dawid A.P., and Smith A.F.M. (Eds.), *Bayesian Statistics 4*, Oxford University Press, Oxford, pp. 147–167.
- Green P.J. 1995. Reversible jump Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Biometrika* 82: 711–732.
- Green P.J. and Richardson S. 2002. Hidden Markov models and disease mapping. *Journal of the American Statistical Association* 97: 1055–1070.
- Jasra A., Holmes C.C., and Stephens D.A. 2005. Markov chain Monte Carlo and the label switching problem in Bayesian mixture modelling. *Statistical Science* 20: 50–67.
- Jennison C. 1997. Discussion of ‘On Bayesian Analysis of Mixtures with an Unknown Number of Components’. *Journal of the Royal Statistical Society, Series B* 59: 731–779.
- Johnson N., Kotz S., and Balakrishnan N. 1997. *Discrete Multivariate Distributions*. Wiley, New York.
- Kano K. and Kawamura K. 1991. On recurrence relations for the probability function of multivariate generalized poisson distribution. *Communications in Statistics -Theory and Methods* 20: 165–178.
- Karlis D. and Meligkotsidou L. 2005. Multivariate Poisson Regression with Covariance Structure. *Statistics and Computing* 15: 255–265.
- Karlis D. and Meligkotsidou L. 2007. Finite mixtures of multivariate Poisson distributions with Application. *Journal of Statistical Planning and Inference*, (to appear).
- Knorr-Held L. and Best G. 2001. A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society, Series A* 164: 73–85.

- Knorr-Held L. and Rasser G. 2000. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* 56: 13–21.
- Lindsay B.G. 1995. Mixture models: Theory, geometry and applications. In: *Regional Conference Series in Probability and Statistics*, Vol 5, (Inst. of Mathem. Stat. and Amer. Stat. Assoc.).
- McLachlan G.J. and Peel D. 2000. *Finite Mixture Models*, Wiley, NY.
- Mengersen K.L. and Robert C.P. 1996. Testing for mixtures: A Bayesian entropic approach. In: Bernardo J.M., Berger J.O., Dawid A.P., and Smith A.F.M. (Eds.), *Bayesian Statistics 5*, Oxford University Press, pp. 255–276.
- Nobile A. and Fearnside A. 2007. Bayesian finite mixtures with an unknown number of components: The allocation sampler. To appear in *Statistics and Computing*.
- Nobile A. 2004. On the posterior distribution of the number of components in a finite mixture. *The Annals of Statistics* 32: 2044–2073.
- Nobile A. and Green P.J. 2000. Bayesian analysis of factorial experiments by mixture modelling. *Biometrika* 87: 15–35.
- Pettit L.I. and Young K.D.S. 1990. Measuring the effect of observations on Bayes factors. *Biometrika* 77: 455–466.
- Rezvani A., Mollié A., Doyon F., and Sancho-Garnier H. 1997. *Atlas de la mortalité par cancer en France, Période 1986–1993*. Les éditions INSERM.
- Richardson S. and Green P. 1997. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B* 59: 731–79.
- Robert C.P., Ryden T., and Titterington D.M. 2000. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B* 52: 57–75.
- Stephens M. 1997. Bayesian methods for mixtures of normal distributions. PhD thesis, University of Oxford.
- Stephens M. 2000. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Journal of the Royal Statistical Society, Series B* 62: 795–809.
- Titterington D.M., Smith A.F.M., and Makov U.E. 1985. *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.
- Tsiampyrtzis P. and Karlis D. 2004. Strategies for efficient computation of multivariate probabilities. *Communications in Statistics - Simulation and Computation* 33: 271–293.
- Viallefont V., Richardson S., and Green P.J. 2000. Bayesian analysis of Poisson mixtures. *Nonparametric Statistics* 14: 181–202.
- Wedel M., Bockenholt U., and Kamakura W.A. 2003. Factor models for multivariate count data. *Journal of Multivariate Analysis* 87: 356–369.