Stat Comput (2008) 18: 27–40
DOI 10.1007/s11222-007-9035-x

Exact Bayesian modeling for bivariate Poisson data and extensions

Dimitris Karlis · Panagiotis Tsiamyrtzis

Received: 2 June 2006 / Accepted: 23 July 2007 / Published online: 11 September 2007
© Springer Science+Business Media, LLC 2007

Abstract Bivariate count data arise in several different disciplines (epidemiology, marketing, sports statistics just to name a few) and the bivariate Poisson distribution being a generalization of the Poisson distribution plays an important role in modelling such data. In the present paper we present a Bayesian estimation approach for the parameters of the bivariate Poisson model and provide the posterior distributions in closed forms. It is shown that the joint posterior distributions are finite mixtures of conditionally independent gamma distributions for which their full form can be easily deduced by a recursively updating scheme. Thus, the need of applying computationally demanding MCMC schemes for Bayesian inference in such models will be removed, since direct sampling from the posterior will become available, even in cases where the posterior distribution of functions of the parameters is not available in closed form. In addition, we define a class of prior distributions that possess an interesting conjugacy property which extends the typical notion of conjugacy, in the sense that both prior and posteriors belong to the same family of finite mixture models but with different number of components. Extension to certain other models including multivariate models or models with other marginal distributions are discussed.

Keywords Bayesian sequentially updated mixtures · Conjugacy · Direct sampling · Gamma mixtures

D. Karlis (✉) · P. Tsiamyrtzis
Department of Statistics, Athens University of Economics and Business, 76 Patission Str., 10434 Athens, Greece
e-mail: karlis@aueb.gr

P. Tsiamyrtzis
e-mail: pt@aueb.gr

1 Introduction

Bivariate count data arise in certain circumstances. For example, in medicine, we may have pretreatment and post-treatment measurements at the same individuals, or we may consider the incidence of two diseases in certain sites. In marketing we are interested in the joint modeling of the purchases of two products or product categories (e.g. food and non-food products). Usual asymptotic arguments based on approximation through bivariate normal distributions fail, especially when the counts are relatively small. Thus, bivariate Poisson models are considered as a useful starting point for studying and modeling such data. The random variables X, Y follow jointly a bivariate Poisson distribution, denoted as $BP(\theta_1, \theta_2, \theta_3)$, if their joint probability function is given by

$$\begin{aligned}P_{X,Y}(x, y) &= P(X = x, Y = y)\\ &= e^{-(\theta_1+\theta_2+\theta_3)} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!}\\ &\quad\times \sum_{k=0}^{\min(x,y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\theta_3}{\theta_1\theta_2}\right)^k \qquad (1)\end{aligned}$$

where $\theta_i > 0$ and $x, y = 0, 1, \ldots$. Marginally each random variable follows a Poisson distribution with $E(X) = \theta_1 + \theta_3$ and $E(Y) = \theta_2 + \theta_3$. Moreover, $\mathrm{Cov}(X, Y) = \theta_3$, and hence θ_3 is a measure of dependence between the two random variables. If $\theta_3 = 0$ then the two variables are independent. For a comprehensive treatment of the bivariate Poisson distribution and its multivariate extensions the reader can refer to Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997).

Inference for the bivariate Poisson model is not an easy task, due to the complicated form of the likelihood (it involves the product of n summations, where n is the sample size). Usually, special cases like the product of independent Poisson distributions (i.e. the case $\theta_3 = 0$) are considered, see, e.g. DeSouza (1992), while any correlation between the counts is induced via mixing (Aitchison and Ho 1989; Chib and Winkelmann 2001; Dabney and Wakefield 2005). Bayesian estimation for the model in (1) relies on computationally intensive numerical methods (e.g. Markov Chain Monte Carlo, MCMC) as proposed in Tsionas (1999). However, using an MCMC approach, when the correlation is very small, it will most likely cause the chain to have bad mixing conditions and probably it will be trapped for a long period to certain values. In the present paper we overcome the need of MCMC schemes by deriving the posteriors in closed form expressions.

The aim of the present paper is twofold: firstly to provide relatively easy exact Bayesian inference for the bivariate Poisson model with $\theta_3 > 0$. The posterior distributions are derived in full detail and thus direct simulation from them will be inexpensive. The marginal posteriors belong to the family of finite mixtures of conditionally independent gamma distributions, from which moments, quantiles and other quantities of interest can be deduced in closed form expressions. In addition joint posterior distributions are easy to extract. In cases of certain function of the parameters that may be of interest one can simulate directly from the posteriors avoiding time consuming MCMC schemes. Our derivation relies on a recursive scheme for updating the posterior density, similar to the one proposed in Fearnhead (2005).

Secondly, we base our inference upon a special prior distribution that allows to incorporate external (subjective) information regarding the correlation of the parameters. This is further motivated for the parametrization in (1) since θ_3 appears in both marginal means and it is also the covariance of the model. Dependent prior distributions can be quite flexible since they offer the opportunity to represent special patterns. We define and base our inference on a prior which is a finite mixture of conditionally independent gamma densities and extend the idea of mixtures of conjugate priors (Dalal and Hall 1983). Both the joint and marginal posterior distributions can be obtained as a finite mixture of conditionally independent gamma distributions, i.e. they belong to the same family as the prior but with a different (but known) number of components.

We would like to emphasize that the proposed approach can be generalized to the multivariate Poisson and other multivariate discrete model cases relatively easy. However, in order to keep the notation handy, we will only present the bivariate case. Later on we will (briefly) provide the multivariate extension.

The remaining of the paper proceeds as follows: Sect. 2 provides all the technical ingredients to carry out the Bayesian estimation, namely we derive the likelihood of the model, we define the prior and we provide the posterior. Section 3 contains details about inference, including properties of the posterior distributions and the predictive distribution. In Sect. 4, we provide an application to illustrate the feasibility of the model. Section 5 provides some details on the extension of the proposed method to the multivariate Poisson case. Concluding remarks and possible extensions of this work can be found in Sect. 6.

2 Bayesian modeling

2.1 The likelihood

The main obstacle for both classical and Bayesian inference concerning the bivariate Poisson model is that the likelihood is too complicated. Suppose data $\mathbf{z}_n = (z_1, \ldots, z_n)$ are available, where each $z_i = (x_i, y_i)$, $i = 1, \ldots, n$ is a pair of observations. Then the likelihood will be given by:

$$L_n(\theta; \mathbf{z}_n) = \prod_{i=1}^{n} e^{-(\theta_1+\theta_2+\theta_3)} \frac{\theta_1^{x_i}}{x_i!} \frac{\theta_2^{y_i}}{y_i!} \times \sum_{k=0}^{\min\{x_i, y_i\}} \binom{x_i}{k} \binom{y_i}{k} k! \left(\frac{\theta_3}{\theta_1\theta_2}\right)^k \tag{2}$$

where $\theta = (\theta_1, \theta_2, \theta_3)$. This (awkward) form of the likelihood involves product of summations which makes it rather inappropriate for efficient calculations in classical statistics and quite cumbersome to be used in Bayes theorem. We will rewrite this likelihood in a polynomial form, where the coefficients can be obtained recursively. Namely we will prove the following:

Lemma 1 *Given a random sample $\mathbf{z}_n$ of size n from a bivariate Poisson distribution, the likelihood can be written in the form*:

$$L_n(\theta; \mathbf{z}_n) = \exp(-n(\theta_1 + \theta_2 + \theta_3)) \times \theta_1^{\sum x_i} \theta_2^{\sum y_i} \sum_{k=0}^{S_n} c_k^{(n)} \left(\frac{\theta_3}{\theta_1\theta_2}\right)^k \tag{3}$$

where $S_n = \sum_{i=1}^{n} \min\{x_i, y_i\}$ and $c_k^{(n)}$ are coefficients that can be obtained recursively by:

$$c_k^{(n)} = \sum_{r=\max\{0,k-s_n^*\}}^{\min\{k,s_n^*\}} v_r^{(n)} c_{k-r}^{(n-1)} \tag{4}$$

where:

$$s_i = \min\{x_i, y_i\}, \qquad S_k = \sum_{i=1}^{k} s_i, \qquad s_n^* = \min\{s_n, S_{n-1}\},$$

$$v_r^{(n)} = \frac{1}{(x_n - r)!(y_n - r)!r!} \quad \textit{and} \quad c_k^{(1)} = v_k^{(1)}.$$

A useful result involving product of polynomials that will be used in the proof of Lemma 1 (and in other proofs as well) is the following:

$$\left[\sum_{i=0}^{n_1} a_i \theta^i\right] \times \left[\sum_{j=0}^{n_2} b_j \theta^j\right] = \sum_{k=0}^{n_1+n_2} \left[\sum_{\ell=\alpha}^{\beta} a_\ell b_{k-\ell}\right] \theta^k$$

where $\alpha = \max\{0, k - \min\{n_1, n_2\}\}$ and $\beta = \min\{k, \min\{n_1, n_2\}\}$.

The proof of the preceding lemma which is a simple induction problem can be found in Appendix 1. This lemma allows us to write the likelihood as a polynomial with respect to the parameters and thus simplify matters in Bayes theorem.

2.2 The prior distribution

In the framework of the Bayesian modeling, we need to define a prior distribution for the parameters $(\theta_1, \theta_2, \theta_3)$ of the bivariate Poisson model. The natural (conjugate) choice would be to consider independent gamma distributions. *A priori* independence might be convenient but not optimal to use always. For example it might be the case that some prior elicitation procedure provided some dependent structure regarding the θ's. In such a case we would prefer to have a prior that will provide us the flexibility to incorporate the dependence structure among the parameters. At the same time we would like to keep the computational complexity relative low by using some form of conjugate prior, which will allow us to have in closed form the posterior distribution.

It is well known that mixtures of conditionally independent densities gives a standard way to impose forms of dependence (see for example, Joe 1997, p. 112). Given that the gamma densities are the conjugate choice for each parameter we are driven to choose as a prior a mixture of conditionally independent gamma densities:

$$\pi(\theta_1, \theta_2, \theta_3) = \sum_{j=0}^{r} p_j G(\theta_1; \alpha_1 - j, \beta_1) \times G(\theta_2; \alpha_2 - j, \beta_2) G(\theta_3; \alpha_3 + j, \beta_3)$$

where $\alpha_1 > r$, $\alpha_2 > r$, $\alpha_3 > 0$, $\beta_i > 0$, $i = 1, 2, 3$, $\sum_{j=0}^{r} p_j = 1$ and $G(x; \alpha, \beta) = \beta^\alpha x^{\alpha-1} \exp(-\beta x)/\Gamma(\alpha)$ denotes the gamma distribution with parameters α and β.

Appropriate choice of the mixing probabilities and the gamma density parameters can provide arbitrarily close approximation (Dalal and Hall 1983) to a wide range of possible scenarios regarding the dependence among θ's. In general the literature regarding multivariate mixtures of gamma densities is sparse (we are aware of the work of Gaver 1970 in the general setting).

It must be noted that the above prior, having a finite mixture representation, can also be seen as a synthesis of more priors elicited by different experts with different opinions on the parameters of interest.

In this section we justified the use of conditional independent gamma mixtures as prior, to allow for incorporation of available prior information. In the next section we will further facilitate our choice by providing the exact posterior distribution in the case of bivariate Poisson data.

To provide an idea on the flexibility of the proposed prior we have depicted contour plots of the joint density of the joint prior for θ_1 and θ_3. We have used various parameter settings. The different plots of Fig. 1 correspond to

$$\mathrm{a}: \alpha_1 = \alpha_2 = 8,\ \beta_1 = \beta_2 = 1,\quad p = (1/5, 1/5, 1/5, 1/5, 1/5),$$

$$\mathrm{b}: \alpha_1 = \alpha_2 = 8,\ \beta_1 = \beta_2 = 1,\quad p = (0.1, 0, 0, 0, 0.9),$$

$$\mathrm{c}: \alpha_1 = \alpha_2 = 8,\ \beta_1 = \beta_2 = 1,\quad p = (0.5, 0, 0, 0, 0, 0.5),$$

$$\mathrm{d}: \alpha_1 = \alpha_2 = 7,\ \beta_1 = \beta_2 = 1,\quad p = (0.5, 0, 0, 0, 0, 0.5),$$

$$\mathrm{e}: \alpha_1 = 6,\ \alpha_2 = 8,\ \beta_1 = 1,\ \beta_2 = 3,\quad p = (1/5, 1/5, 1/5, 1/5, 1/5) \quad \text{and}$$

$$\mathrm{f}: \alpha_1 = \alpha_2 = 10,\ \beta_1 = \beta_2 = 1,\quad p = (0.05, 0.1, 0.2, 0.3, 0.2, 0.1, 0.05).$$

One can see that the joint density can have a wide range of shapes, including large tails, bimodal shapes, skewness etc. As far as the moments of the prior are concerned we discuss this issue later on when treating the posterior moments, since they belong to the same family.

2.3 The posterior distribution

Once the likelihood and the prior has been specified use of Bayes theorem will provide the posterior distribution for the vector of $\theta = (\theta_1, \theta_2, \theta_3)$. More precisely we have the following:

Therem 2 *Given a random sample $\mathbf{z}_n$ of size n from a bivariate Poisson distribution with*:

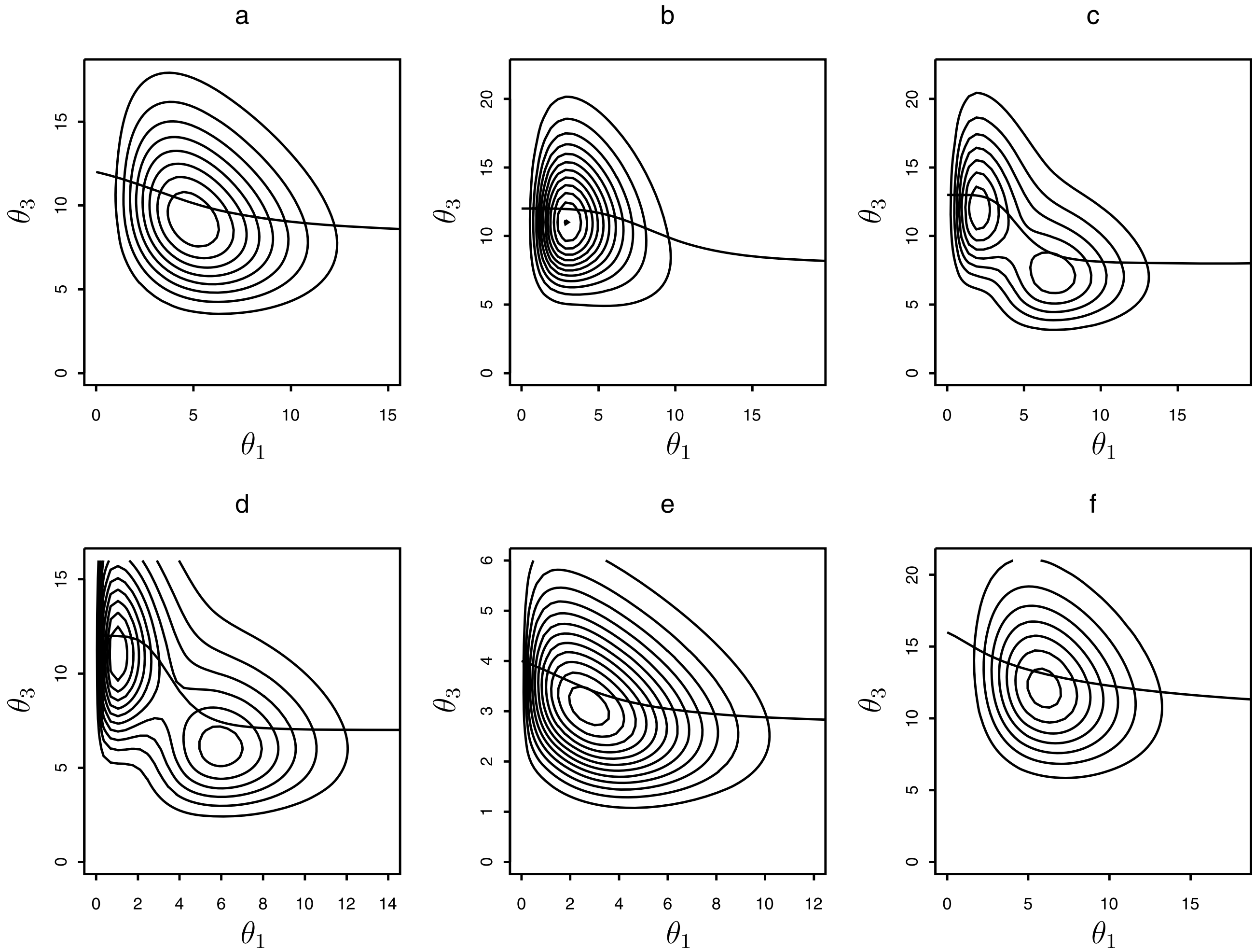

Fig. 1 The form of the joint prior distribution for θ_1 and θ_3 for certain different choices of parameters

$$f(\mathbf{z}_n|\theta) = \exp(-n(\theta_1+\theta_2+\theta_3)) \times \theta_1^{\sum x_i}\theta_2^{\sum y_i}\sum_{i=0}^{S_n} c_i^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i$$

where the prior distribution is given by:

$$\pi(\theta) = \sum_{j=0}^{r} p_j G(\theta_1; \alpha_1 - j, \beta_1) \times G(\theta_2; \alpha_2 - j, \beta_2) G(\theta_3; \alpha_3 + j, \beta_3)$$

we get the posterior distribution of θ *to be*:

$$p(\theta|\mathbf{z}_n) = \sum_{k=0}^{S_n+r} w_k G\left(\theta_1; \sum x_i + \alpha_1 - k, n+\beta_1\right) \times G\left(\theta_2; \sum y_i + \alpha_2 - k, n+\beta_2\right) \times G(\theta_3; \alpha_3 + k, n+\beta_3)$$

where

$$w_k = \tilde{p}_k \Big/ \sum_{m=0}^{S_n+r} \tilde{p}_m,$$

$$\tilde{p}_k = \left[\sum_{l=A}^{B} c_l^{(n)} p_{k-l}^*\right]\Gamma\left(\sum x_i + \alpha_1 - k\right) \times \Gamma\left(\sum y_i + \alpha_2 - k\right)\Gamma(\alpha_3+k)\left[\frac{(n+\beta_1)(n+\beta_2)}{n+\beta_3}\right]^k$$

with $A = \max\{0, k - \min\{S_n, r\}\}$, $B = \min\{k, \min\{S_n, r\}\}$ *and*

$$p_j^* = p_j \frac{(\beta_1)^{\alpha_1-j}(\beta_2)^{\alpha_2-j}(\beta_3)^{\alpha_3+j}}{\Gamma(\alpha_1-j)\Gamma(\alpha_2-j)\Gamma(\alpha_3+j)}.$$

The proof of the preceding theorem can be found in Appendix 2. The posterior is provided in closed form and is (like the prior) a mixture of conditionally independent gamma distributions. This implies a conjugacy result in a broad sense, since the posterior belongs to the same mixture family with the prior but with an increased number of components.

Provided that the data are conditionally independent given θ one could get the same result for the posterior in a sequential manner, where we start with a single data point and the prior to obtain the posterior which then we use as a prior for the upcoming data point and so on. This provides an alternative tool for the required computations where instead of computing the overall likelihood we can proceed sequentially. More importantly though offers a useful tool to handle scenarios where the data are obtained sequentially and especially when we are interested in drawing inference even with very few data available. Therefore one can view the proposed prior as a dynamically updated mixture. It is worth mentioning here that even in the case where the initial prior distribution is just independent gamma distribution (special case of the prior where $r = 0$) we will have the posterior to be a mixture of conditionally independent gammas.

3 Inference

Once the posterior distribution is obtained point estimates and hypothesis testing issues regarding the θ's can be addressed within the Bayesian decision theory framework. Towards this direction the marginal distributions are helpful and easy to obtain from the joint posterior. More precisely if we will denote $G^{(1)}$, $G^{(2)}$ and $G^{(3)}$ the first, second and third gamma component in the posterior mixture we will have:

$$p(\theta_i, \theta_j | \mathbf{z}_n) \sim \sum_{k=0}^{S_n+r} w_k G_k^{(i)} G_k^{(j)} \quad \text{and}$$

$$p(\theta_m | \mathbf{z}_n) \sim \sum_{k=0}^{S_n+r} w_k G_k^{(m)}$$

where $i, j, m \in \{1, 2, 3\}$ and $i < j$.

Certain properties of the gamma distribution allow us to obtain the exact distribution of various functions of θ. However, in cases where we are interested in more complicated functions of the parameters, then Monte Carlo techniques can be applied on (computationally inexpensive) samples from the exact posterior distribution to obtain the required inference.

3.1 Posterior moments

Let $p(k)$ denotes the probability function of a discrete random variable K, which takes values in a set $\mathcal{S}$ subset of $\{0, 1, \ldots\}$. In general the joint posterior will be of the form

$$p(\theta_1, \theta_2, \theta_3 | \mathbf{z}_n) = \sum_{k \in \mathcal{S}} p(k) \prod_{i=1}^{3} G(\theta_i; \gamma_i + c_i k, \delta_i),$$

for known constants $c_1 = c_2 = -1$ and $c_3 = 1$, where the set $\mathcal{S}$ is defined so that $\gamma_i + c_i k > 0$ for all i. Then the posterior expectations, variances and covariances are:

$$E(\theta_i | \mathbf{z}_n) = \frac{\gamma_i + c_i E(k)}{\delta_i},$$

$$\mathrm{Var}(\theta_i | \mathbf{z}_n) = \frac{\gamma_i + c_i E(k) + c_i{}^2 \mathrm{Var}(k)}{\delta_i{}^2}, \quad i = 1, 2, 3,$$

$$\mathrm{Cov}(\theta_i, \theta_j | \mathbf{z}_n) = \frac{c_i c_j \mathrm{Var}(k)}{\delta_i \delta_j}, \quad i, j = 1, 2, 3, \quad i \neq j.$$

Thus, the correlation can be obtained as:

$$\rho(\theta_i, \theta_j) = \frac{c_i c_j \mathrm{Var}(k)}{\sqrt{\gamma_i + c_i E(k) + c_i{}^2 \mathrm{Var}(k)}} \times \frac{1}{\sqrt{\gamma_j + c_j E(k) + c_j{}^2 \mathrm{Var}(k)}},$$
$$i, j = 1, 2, 3, \quad i \neq j.$$

Note that in our case we have $\gamma_1 = \sum x_i + \alpha_1$, $\delta_1 = n + \beta_1$, $\gamma_2 = \sum y_i + \alpha_2$, $\delta_2 = n + \beta_2$, $\gamma_3 = \alpha_3$, $\delta_3 = n + \beta_3$. In our case, and since $c_1 = c_2 = -c_3$ we will have positive correlation between θ_1, θ_2 and negative correlation between any of θ_1, θ_2 and θ_3.

3.2 An interesting special case

An interesting special case arises when $\beta_1 = \beta_2 = \beta_3$. In this case and using the well known result that the sum of two gamma variates with the same scale parameter is also a gamma variate, one can deduce that the distribution of $\theta_1 + \theta_3$ (marginal mean of X) and similarly that of $\theta_2 + \theta_3$ (marginal mean of Y) are simple gamma distributions.

On the other hand consider the case when one proceeds to the analysis by assuming two independent Poisson variates, i.e. assume that $\theta_3 = 0$. The standard approach is to consider independent gamma priors which lead to independent gamma posteriors for the two marginal means in consistence with the case with $\beta_1 = \beta_2 = \beta_3$ described above.

Thus, both approaches coincide to a gamma posterior density for the marginal mean. Our approach however assumes also dependence between the two variables and thus the joint distribution of $(\theta_1 + \theta_3, \theta_2 + \theta_3)$ is not the product of two independent gammas, as the one we would obtain assuming independence. This implies that correlation between the marginal means is present, leading to different inferences for certain functions of the marginal means (like the ratio of the two means).

3.3 Predictive distribution

Another interesting issue that can be developed in the framework of the Bayesian modeling is the ability to do forecasting. Specifically having seen the data $\mathbf{z}_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ we can derive the predictive distribution of the next data point $z_{n+1} = (x_{n+1}, y_{n+1})$. The predictive distribution (Geisser 1993) will be given by:

$$P(z_{n+1}|\mathbf{z}_n) = \int f(z_{n+1}|\theta)p(\theta|\mathbf{z}_n)d\theta$$

where $f(z_{n+1}|\theta)$ refers to the likelihood of the next observable and $p(\theta|\mathbf{z}_n)$ is the posterior once the data $\mathbf{z}_n$ have been observed. More precisely we will have the following:

Lemma 3 *The predictive distribution of $z_{n+1}|\mathbf{z}_n$ will be given by*:

$$\sum_{k=0}^{S_{n+1}+r} \sum_{l=A}^{B} v_l^{(n+1)} w_{k-l} \frac{(n+\beta_1)^{\sum_{i=1}^n x_i+\alpha_1-k+l}(n+\beta_2)^{\sum_{i=1}^n y_i+\alpha_2-k+l}(n+\beta_3)^{\alpha_3+k-l}}{(n+\beta_1+1)^{\sum_{i=1}^{n+1} x_i+\alpha_1-k}(n+\beta_2+1)^{\sum_{i=1}^{n+1} y_i+\alpha_2-k}(n+\beta_3+1)^{\alpha_3+k}}$$
$$\times \frac{\Gamma(\sum_{i=1}^{n+1} x_i+\alpha_1-k)\Gamma(\sum_{i=1}^{n+1} y_i+\alpha_2-k)\Gamma(\alpha_3+k)}{\Gamma(\sum_{i=1}^{n} x_i+\alpha_1-k+l)\Gamma(\sum_{i=1}^{n} y_i+\alpha_2-k+l)\Gamma(\alpha_3+k-l)}$$

where $A = \max\{0, k - \min\{s_{n+1}, S_n + r\}\}$ *and* $B = \min\{k, \min\{s_{n+1}, S_n + r\}\}$.

The proof of this lemma can be found in Appendix 3. Just like before within the decision theory framework we can provide point/interval estimates and/or hypothesis testing for the next observable. The previous lemma defines the one step ahead forecasting distribution. Use of Lemma 1 allows a straightforward generalization of the above lemma to the more general case of a k-step ahead forecast distribution.

4 Illustration

In order to illustrate the proposed approach, we use the data related to the demand for health care in Australia, taken by Cameron and Trivedi (1998). The data refer to the Australian Health survey for 1977–1978. The sample size was quite large ($n = 5190$). We will use two variables, namely the number of consultations with a doctor or a specialist (X) and the total number of prescribed and non-prescribed medications used in past 2 days (Y). The data can be seen in Table 1. It is interesting that the data are correlated (the Pearson correlation coefficient is 0.27). A bivariate Poisson model is plausible due to the correlation.

We applied the Bayesian approach discussed in previous sections. For sensitivity checking purposes we used five different priors. Three of them were independent gamma priors Gamma(a_i, b_i) for each parameter θ_i, $i = 1, 2, 3$, with hyperparameters $a_i = b_i = 0.1, 1, 10$ respectively, for $i = 1, 2, 3$. We denote those priors by π_1, π_2, π_3 respectively. All of them correspond to unit prior mean and varying variance. The first choice provides diffused prior while for the last one the prior is quite informative as the variance is small. Furthermore, dependent priors were used. The first one assumes

Table 1 Data from the Australian Health Survey (Cameron and Trivedi, 1998)

Number of consultations (X)	Number of prescribed and nonprescribed medications (Y)								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	2009	1144	492	262	124	47	28	18	17
1	164	193	178	98	65	41	21	8	14
2	39	30	40	17	21	10	9	4	4
3	5	12	4	4	1	1	2	1	0
4	6	7	3	4	3	1	0	0	0
5	2	2	4	0	1	0	0	0	0
6	2	0	0	4	1	2	1	1	1
7	1	0	2	3	1	1	1	0	3
8	1	1	0	1	1	0	1	0	0
9	0	0	0	0	0	0	0	0	1

$$\pi_4(\theta) = \sum_{k=0}^{9} \frac{1}{10} G(\theta_1; 10-k, 10)$$
$$\times G(\theta_2; 10-k, 10)G(\theta_3; 10+k, 10)$$

i.e. a dependent prior, which is a 10-component finite mixture of conditionally independent gamma densities and

$$\pi_5(\theta) = \sum_{k=0}^{9} p(k) G(\theta_1; 10-k, 10)$$
$$\times G(\theta_2; 10-k, 10)G(\theta_3; 10+k, 10)$$

where $p(k) = Ce^{-1}/k!$, $k = 0, 1, \ldots, 9$, with C being the normalizing constant. The two distributions impose a dif-

Table 2 Description of the characteristics of the priors used in the illustration

Prior	Mean			Variance			Correlations		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	$r(\theta_1,\theta_2)$	$r(\theta_1,\theta_3)$	$r(\theta_2,\theta_3)$
π_1	1	1	1	10	10	10	0	0	0
π_2	1	1	1	1	1	1	0	0	0
π_3	1	1	1	0.1	0.1	0.1	0	0	0
π_4	0.55	0.55	1.45	0.1375	0.1375	0.2275	0.1	−0.0912	−0.0912
π_5	0.9	0.9	1.1	0.1	0.1	0.12	0.6	−0.4664	−0.4664

Table 3 Posterior summaries for the health data using the five different priors

		π_1	π_2	π_3	π_4	π_5
θ_1	Mean	0.1768	0.1777	0.1774	0.1775	0.1767
	Std dev	0.0076	0.0075	0.0075	0.0075	0.0076
	95% Cr.I.	(0.165, 0.189)	(0.166, 0.190)	(0.165, 0.190)	(0.165, 0.190)	(0.165, 0.189)
θ_2	Mean	1.0932	1.0925	1.0922	1.0923	1.0932
	Std dev	0.0153	0.0152	0.0152	0.0152	0.0153
	95% Cr.I.	(1.068, 1.119)	(1.068, 1.118)	(1.067, 1.117)	(1.067, 1.118)	(1.068, 1.119)
θ_3	Mean	0.1253	0.1273	0.1276	0.1275	0.1251
	Std dev	0.0069	0.0068	0.0069	0.0068	0.0069
	95% Cr.I.	(0.114, 0.137)	(0.116, 0.139)	(0.117, 0.139)	(0.116, 0.139)	(0.114, 0.137)

ferent level of correlation between the parameters. For the second one the correlation is much higher. Summary for all the five priors used can be seen in Table 2.

One can easily verify that $\sum_{i=1}^{5190}\min(x_i, y_i) = 1075$. According to the findings of Sect. 2.3, the posterior distribution is a finite mixture with 1076 components for priors π_1, π_2, π_3 and 1085 components for priors π_4, π_5, respectively. Since $\sum_{i=1}^{n} X_i = 1566$ and $\sum_{i=1}^{n} Y_i = 6323$ each one of the components of the joint posterior is the product of three gamma distributions of the form $G(\theta_1; 1566 + \alpha_1 - k, 5190 + b_1)$, $G(\theta_2; 6323 + \alpha_2 - k, 5190 + b_2)$ and $G(\theta_3; a_3 + k, 5190 + b_3)$ respectively. The mixing weights can be found as described in Sect. 2.3. The marginal posteriors are of the form:

$$P^j(\theta_1) = \sum_{k=0}^{1074+d_j} p_j(k) G(\theta_1; a_{1j} + 1566 - k, b_{1j} + 5190)$$

$$P^j(\theta_2) = \sum_{k=0}^{1074+d_j} p_j(k) G(\theta_2; a_{2j} + 6323 - k, b_{2j} + 5190)$$

$$P^j(\theta_3) = \sum_{k=0}^{1074+d_j} p_j(k) G(\theta_3; a_{3j} + k, b_{3j} + 5190)$$

where $P^j(\cdot)$ is the marginal posterior based on prior π_j, p_j's are the mixing weights derived from the jth prior ($j = 1, \ldots, 5$) and parameters a_{ij}, b_{ij} correspond to the ith parameter ($i = 1, 2, 3$) for the jth prior. Similarly d_j denotes the number of components of the prior.

Bayesian point and interval estimates regarding the parameters can be found in Table 3.

As expected there are minor differences between the estimates for different prior choices (due to the large sample size which minimizes the effect of the prior). Mean and variances were obtained using formulas provided in Sect. 3.1. Percentile points were obtained through direct evaluation of the cumulative distribution. Plots of the marginal posteriors can be seen in Fig. 2 for all priors.

For parameter θ_2 we observe that the five posteriors are visually obscured. However, there are some differences in the posteriors of θ_1 and θ_3. Bivariate contour plots for the three pairs of parameters, namely (θ_1, θ_2), (θ_1, θ_3) and (θ_2, θ_3) are depicted in Fig. 3 based on the prior π_4. From the plots it is apparent that the parameters are correlated a posteriori.

In general a quantity of interest is $\phi = (\theta_1 + \theta_3)/(\theta_2 + \theta_3)$ (the ratio of the marginal means). This quantity can be of special importance as this ratio in pre and after studies reveals the improvement or deterioration of the mean (i.e. the treatment effect). In our case it provides an estimate of the number of medications per doctor visit. The posterior of ϕ is not easy to be derived in closed form but direct simulation from it is straightforward. It is also important to note that when β_j prior parameters are the same, one can derive the posterior of ϕ as a finite mixture of Beta distributions type I.

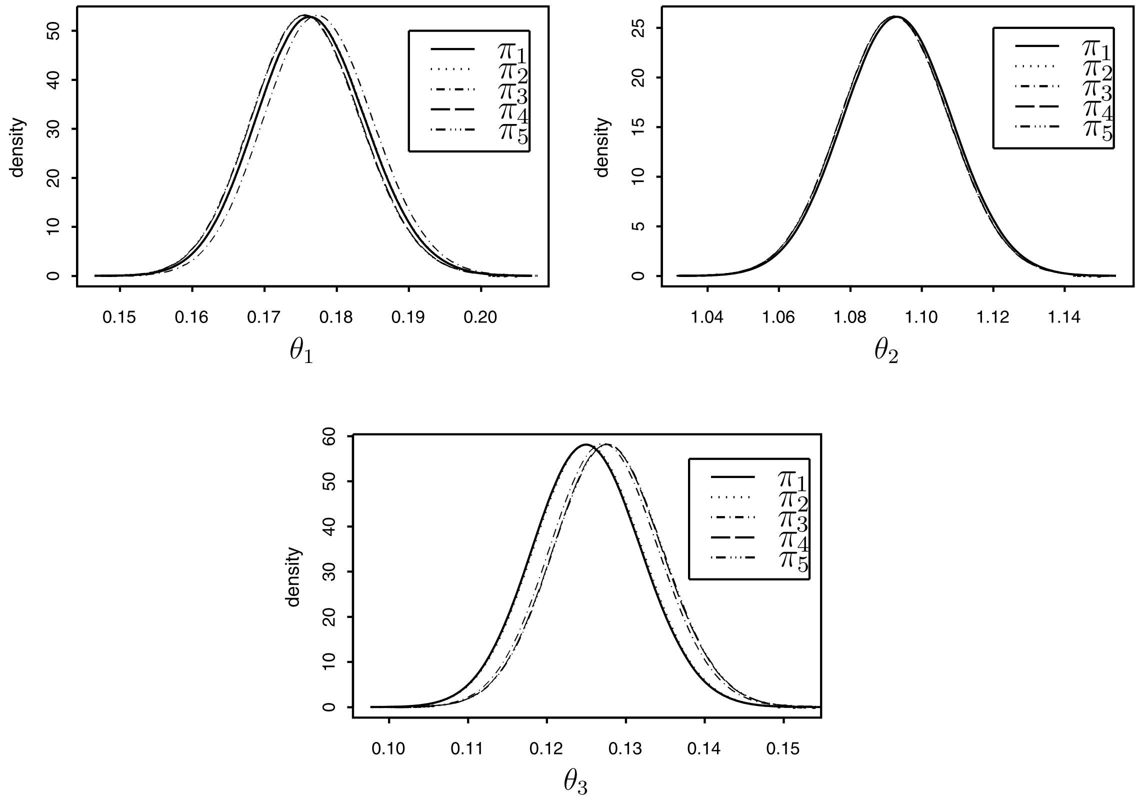

Fig. 2 Posterior densities for the parameters

Since our priors have common β_i's , the marginal posterior densities for $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$ are gamma (see Sect. 3.2), while the posterior of ϕ will be a finite mixture of beta densities. We would like to compare this posterior to the one derived when one ignores the correlation of the data (i.e. assumes $\theta_3 = 0$). The standard conjugate approach with gamma priors to the marginal means leads to a beta posterior for ϕ. We have selected prior parameters so as to much the prior means.

Our approach, assuming the existence of correlation between the variables, ends up with a distribution which is different from the assumed beta as one can see in Fig. 4c. This is a Q-Q plot of the derived sample form the posterior of ϕ (its histogram is depicted in Fig. 4a while the scatterplot of the two means is in Fig. 4b) with the corresponding beta density if $\theta_3 = 0$. There is a clear departure from the assumed beta density. In fact the mean is smaller and the tails much different. This illustrates that inferences based on a model that ignores the existing correlation of the data may be wrong.

5 Multivariate extension

Generalizing the bivariate to multivariate case is straightforward. If $(X_1, X_2, \ldots, X_N)$ follows jointly the multivariate Poisson distribution with (vector of) parameters $(\theta_1, \ldots, \theta_N, \theta_{N+1})$ and probability function given by:

$$\begin{aligned} &P(X_1 = x_1, \ldots, X_N = x_N) \\ &\quad = \exp\{-(\theta_1 + \cdots + \theta_N + \theta_{N+1})\} \frac{\theta_1^{x_1}}{x_1!} \cdots \frac{\theta_N^{x_N}}{x_N!} \\ &\qquad \times \sum_{k=0}^{\min\{x_1,\ldots,x_N\}} \binom{x_1}{k} \cdots \binom{x_N}{k} k! \left(\frac{\theta_{N+1}}{\theta_1 \cdots \theta_N} \right)^k \end{aligned}$$

we have that marginally each X_i will be a Poisson $(\theta_i + \theta_{N+1})$ and $\operatorname{Cov}(X_i, X_j) = \theta_{N+1}$, for $i \neq j$. So extending the bivariate to multivariate case, keeping the analogous structure (common pairwise covariance), will cause the likelihood to have an analogous (multivariate) form.

5.1 Multivariate likelihood

We will use superscripts in parenthesis to denote the sample point, i.e. the ith observation $(i = 1, \ldots, n)$ will be denoted by: $(x_1^{(i)}, \ldots, x_N^{(i)})$. Then the joint likelihood function will be given by:

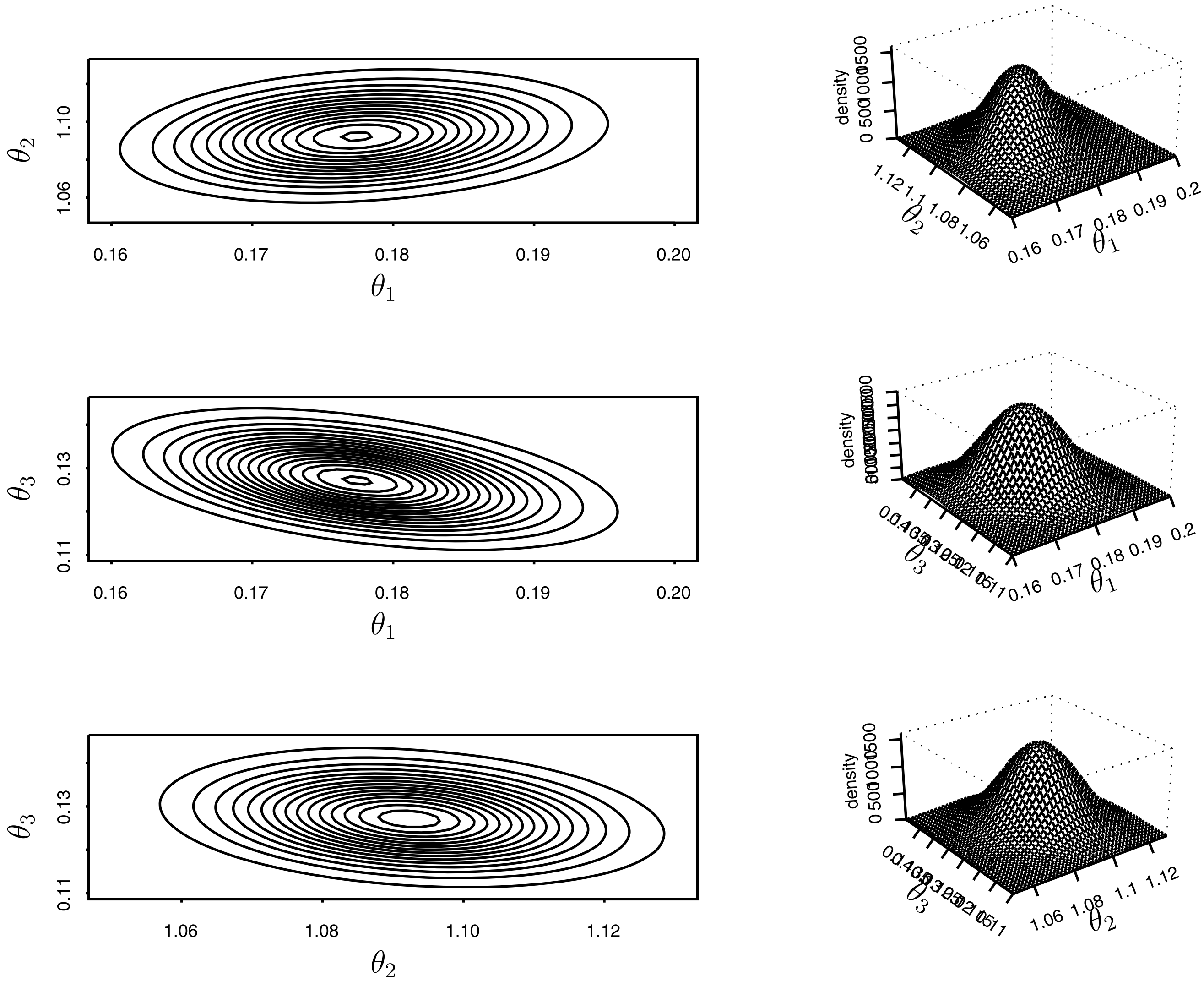

Fig. 3 Bivariate contour plot and density plots for the bivariate joint posterior densities when prior π_4 is used

$$L_n(\theta) = \exp\left\{-n\sum_{j=1}^{N+1}\theta_j\right\}\theta_1^{\sum_{i=1}^n x_1^{(i)}}\cdots\theta_N^{\sum_{i=1}^n x_N^{(i)}} \times \sum_{k=0}^{S_n} c_k^{(n)}\left(\frac{\theta_{N+1}}{\theta_1\cdots\theta_N}\right)^k$$

where $S_n = \sum_{i=1}^n \min\{x_1^{(i)},\ldots,x_N^{(i)}\}$ and $c_k^{(n)}$ are coefficients that can be obtained recursively by:

$$c_k^{(n)} = \sum_{r=\max\{0,k-s_n^*\}}^{\min\{k,s_n^*\}} v_r^{(n)} c_{k-r}^{(n-1)}$$

where:

$$s_i = \min\{x_1^{(i)},\ldots,x_N^{(i)}\},\quad S_k = \sum_{i=1}^k s_i,$$

$$s_n^* = \min\{s_n, S_{n-1}\}$$

$$v_r^{(n)} = \frac{1}{(x_1^{(i)}-r)!\cdots(x_N^{(i)}-r)!(r!)^{N-1}} \quad \text{and} \quad c_k^{(1)} = v_k^{(1)}.$$

5.2 Multivariate prior

The conjugate choice for the prior parameter vector $(\theta_1,\ldots,\theta_N,\theta_{N+1})$ is a mixture of conditionally independent gamma densities:

$$\begin{aligned}\pi(\theta_1,\ldots,\theta_N,\theta_{N+1}) &= \sum_{j=0}^r p_j G(\theta_1;\alpha_1-j,\beta_1)\cdots \\ &\quad\times G(\theta_N;\alpha_N-j,\beta_N)G(\theta_{N+1};\alpha_{N+1}+j,\beta_{N+1})\end{aligned}$$

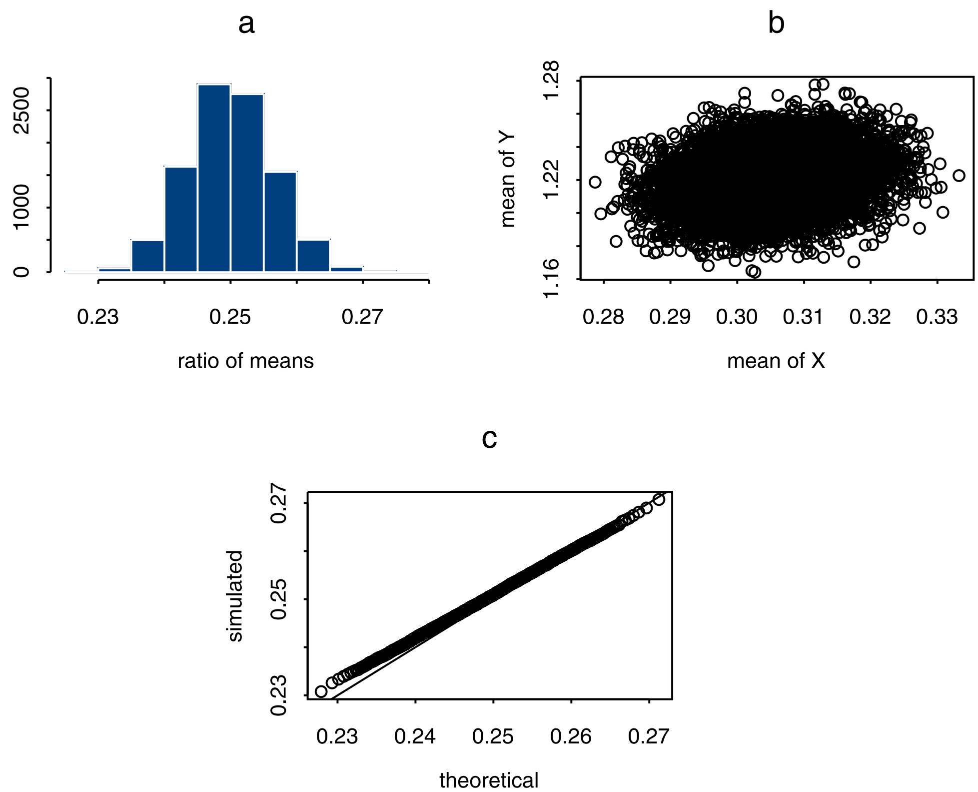

Fig. 4 Posterior summaries for the marginal means. Plot **a** is a histogram of ϕ, **b** is a scatterplot of the simulated values for the two means and **c** is the Q-Q plot of the simulated values with the assumed Beta distribution under an independence assumption and same priors

where $\min\{\alpha_1,\dots,\alpha_N\} > r$, $\alpha_{N+1} > 0$, $\beta_i > 0$, $i = 1,\dots, N+1$ and $\sum_{j=0}^{r} p_j = 1$.

5.3 Multivariate posterior

Using the likelihood and the prior distribution given above we obtained (in closed form) the posterior distribution as a mixture of conditionally independent gamma densities:

$$\sum_{k=0}^{S_n+r} w_k G\left(\theta_1; \sum_{i=1}^{n} x_1^{(i)} + \alpha_1 - k, n+\beta_1\right)$$
$$\times G\left(\theta_N; \sum_{i=1}^{n} x_N^{(i)} + \alpha_N - k, n+\beta_N\right)$$
$$\times G(\theta_{N+1}; \alpha_{N+1} + k, n + \beta_{N+1})$$

where

$$w_k = \tilde{p}_k \Big/ \sum_{m=0}^{S_n+r} \tilde{p}_m$$

$$\tilde{p}_k = \left[\sum_{l=A}^{B} c_l^{(n)} p_{k-l}^*\right]\left[\frac{(n+\beta_1)\cdots(n+\beta_N)}{n+\beta_{N+1}}\right]^k$$
$$\times \Gamma\left(\sum_{i=1}^{n} x_1^{(i)} + \alpha_1 - k\right)\cdots$$
$$\times \Gamma\left(\sum_{i=1}^{n} x_N^{(i)} + \alpha_N - k\right)\Gamma(\alpha_{N+1}+k)$$

with $A = \max\{0, k - \min\{S_n, r\}\}$, $B = \min\{k, \min\{S_n, r\}\}$ and

$$p_j^* = p_j \frac{(\beta_1)^{\alpha_1 - j}\dots(\beta_N)^{\alpha_N - j}(\beta_{N+1})^{\alpha_{N+1}+j}}{\Gamma(\alpha_1 - j)\cdots\Gamma(\alpha_N - j)\Gamma(\alpha_{N+1}+j)}$$

for $j = 0, 1, \dots, r$.

6 Concluding remarks

We described Bayesian inference for the bivariate Poisson model that does not rely on MCMC schemes. We derived the joint posterior in closed form and thus, either quantities of interest are easy to calculate exactly, or direct sampling form the posterior is inexpensive. However, the ideas discussed in the present paper can be extended beyond this model towards certain directions. Fearnhead (2005) described a similar idea for mixtures of discrete distributions. In certain other circumstances where the likelihood involves sums our approach is applicable, such as other discrete models like the generalized Poisson model, bivariate binomial models etc. For example, Winkelmann (2000) derived a multivariate negative binomial model based on a reduction technique similar to the one used to create the bivariate (multivariate)

Poisson model. Our approach is extendable to this case by considering mixtures of conditionally independent beta distributions as priors and deriving closed form posteriors of the same family.

Another interesting point is that the family of prior densities considered allows for correlation between the parameters while as a mixture of conjugate priors lead to posteriors that belong to the same family of finite mixtures, extending the notion of conjugacy.

Our approach can be suitable for modelling bivariate count data much easier than other existing models. We note that in certain applications, like pre and after treatment measurements, or counts in a dichotomized population, bivariate models are themselves of particular interest. DeSouza (1992) developed a bivariate model starting by independent Poisson variates, i.e. assuming $\theta_3 = 0$ and proceeded with approximate Bayesian inference assuming multivariate normal priors for a reparametrization of the model. Chib and Winkelmann (2001) based on the work of Aitchinson and Ho (1989) defined a multivariate Poisson lognormal distribution which has negative correlation. This correlation was incorporated by the mixing multivariate lognormal distribution. They also proposed Bayesian estimation through MCMC. Our model generalizes this idea in the sense that we start from a bivariate model while their model was based on two independent Poisson distributions. Following the terminology of Subrahmaniam (1966), our model assumes intrinsic correlation as well (measured by θ_3) and thus it is much more general. Our advance is that we also assume a multivariate prior allowing for correlation between the parameters but in our case the posterior can be obtained in closed form without the need of MCMC or approximations.

We also presented the extension to the multivariate Poisson setting. Extending our approach to multivariate models is more demanding in the sense that while similar results can be deduced in practice it is not so easy to calculate the quantities needed. However for dimensions up to four computations involved are not prohibitively large and extensions are rather straightforward. Another extension with great practical potential is the introduction of covariate information, i.e. bivariate Poisson regression models as in Jung and Winkelmann (1993). The use of categorical covariates allows for a relatively easy to handle extension with closed form expression, but the use of continuous covariate information is not easy and generalizations towards this are not simple.

Appendix 1

The proof will be given by induction: Clearly it holds for $n = 2$ since

$$
\begin{aligned}
&L_2(\theta; \mathbf{z}_n)\\
&\quad= \left[\exp(-(\theta_1+\theta_2+\theta_3))\theta_1^{x_1}\theta_2^{y_1}\sum_{i=0}^{s_1} v_i^{(1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right]\\
&\qquad\times\left[\exp(-(\theta_1+\theta_2+\theta_3))\theta_1^{x_2}\theta_2^{y_2}\sum_{j=0}^{s_2} v_j^{(2)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right]\\
&\quad= \exp(-2(\theta_1+\theta_2+\theta_3))\theta_1^{x_1+x_2}\theta_2^{y_1+y_2}\\
&\qquad\times\left[\sum_{i=0}^{s_1} v_i^{(1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right]\left[\sum_{j=0}^{s_2} v_j^{(2)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right]\\
&\quad= \exp(-2(\theta_1+\theta_2+\theta_3))\theta_1^{x_1+x_2}\theta_2^{y_1+y_2}\\
&\qquad\times\sum_{k=0}^{s_1+s_2}\left\{\left[\sum_{r=\max\{0,k-s_2^*\}}^{\min\{k,s_2^*\}} v_r^{(1)}v_{k-r}^{(2)}\right]\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k\right\}
\end{aligned}
$$

and thus writing

$$
c_k^{(2)} = \sum_{r=\max\{0,k-s_2^*\}}^{\min\{k,s_n^*\}} v_r^{(2)}v_{k-r}^{(1)} \tag{5}
$$

where $s_2^* = \min\{s_2, s_1\}$ we see that it is true.

Assume that it holds for $n-1$, i.e. that

$$
\begin{aligned}
&L_{n-1}(\theta; \mathbf{z}_{n-1})\\
&\quad= \prod_{i=1}^{n-1} P((x_i, y_i) \mid \theta_1, \theta_2, \theta_3)\\
&\quad= \exp(-(n-1)(\theta_1+\theta_2+\theta_3))\theta_1^{\sum_{i=1}^{n-1} x_i}\theta_2^{\sum_{i=1}^{n-1} y_i}\\
&\qquad\times\left[\sum_{k=0}^{S_{n-1}} c_k^{(n-1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k\right]
\end{aligned}
$$

we will show that it holds for n. We obtain

$$
\begin{aligned}
&L_n(\theta; \mathbf{z}_n)\\
&\quad= L_{n-1}(\theta; \mathbf{z}_{n-1})P((x_n, y_n) \mid \theta_1, \theta_2, \theta_3)\\
&\quad= \exp(-(n-1)(\theta_1+\theta_2+\theta_3))\theta_1^{\sum_{i=1}^{n-1} x_i}\theta_2^{\sum_{i=1}^{n-1} y_i}\\
&\qquad\times\left[\sum_{k=0}^{S_{n-1}} c_k^{(n-1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k\right]\\
&\qquad\times\left[\exp(-(\theta_1+\theta_2+\theta_3))\theta_1^{x_n}\theta_2^{y_n}\sum_{j=0}^{s_n} v_j^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right]\\
&\quad= \exp(-n(\theta_1+\theta_2+\theta_3))\theta_1^{\sum_{i=1}^{n} x_i}\theta_2^{\sum_{i=1}^{n} y_i}\\
&\qquad\times\left[\sum_{k=0}^{S_{n-1}} c_k^{(n-1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k\right]\left[\sum_{j=0}^{s_n} v_j^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right]
\end{aligned}
$$

$$= \exp(-n(\theta_1+\theta_2+\theta_3))\theta_1^{\sum_{i=1}^n x_i}\theta_2^{\sum_{i=1}^n y_i} \times \sum_{k=0}^{S_n}\left[\sum_{r=\max\{0,k-s_n^*\}}^{\min\{k,s_n^*\}} u_r^{(n)} c_{k-r}^{(n-1)}\right]\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k$$

but by definition

$$c_k^{(n)} = \sum_{r=\max\{0,k-s_n^*\}}^{\min\{k,s_n^*\}} u_r^{(n)} c_{k-r}^{(n-1)}. \tag{6}$$

This completes the proof.

Appendix 2

We will apply Bayes theorem to obtain the posterior distribution. For the prior distribution we have:

$$\begin{aligned}\pi(\theta_1,\theta_2,\theta_3) &= \sum_{j=0}^r p_j G(\alpha_1-j,\beta_1)G(\alpha_2-j,\beta_2)G(\alpha_3+j,\beta_3)\\ &= \sum_{j=0}^r p_j^*(\theta_1^{\alpha_1-j-1}\exp\{-\theta_1\beta_1\})(\theta_2^{\alpha_2-j-1}\exp\{-\theta_2\beta_2\})\\ &\quad\times(\theta_3^{\alpha_3+j-1}\exp\{-\theta_3\beta_3\})\end{aligned}$$

where

$$p_j^* = p_j\frac{(\beta_1)^{\alpha_1-j}(\beta_2)^{\alpha_2-j}(\beta_3)^{\alpha_3+j}}{\Gamma(\alpha_1-j)\Gamma(\alpha_2-j)\Gamma(\alpha_3+j)} \quad \text{for } j=0,1,\ldots,r$$

$\alpha_1 > r$, $\alpha_2 > r$, $\alpha_3 > 0$, $\beta_i > 0$, $i = 1,2,3$, $\sum_{j=0}^r p_j = 1$ and for the likelihood we have

$$f(\mathbf{z}_n|\theta) = \exp(-n(\theta_1+\theta_2+\theta_3))\theta_1^{\sum x_i}\theta_2^{\sum y_i} \times \sum_{i=0}^{S_n} c_i^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i.$$

Then we get:

$$\begin{aligned}p(\theta|\mathbf{z}_n) &\propto f(\mathbf{z}_n|\theta)\pi(\theta)\\ &= \left[\exp(-n(\theta_1+\theta_2+\theta_3))\theta_1^{\sum x_i}\theta_2^{\sum y_i} \times \sum_{i=0}^{S_n} c_i^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right]\\ &\quad\times\left[\sum_{j=0}^r p_j^*\theta_1^{\alpha_1-j-1}\theta_2^{\alpha_2-j-1}\theta_3^{\alpha_3+j-1} \times \exp\{-\theta_1\beta_1-\theta_2\beta_2-\theta_3\beta_3\}\right]\\ &= \exp\{-(n+\beta_1)\theta_1-(n+\beta_2)\theta_2-(n+\beta_3)\theta_3\}\\ &\quad\times\theta_1^{\sum x_i+\alpha_1-1}\theta_2^{\sum y_i+\alpha_2-1}\theta_3^{\alpha_3-1}\\ &\quad\times\left[\sum_{i=0}^{S_n} c_i^{(n)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right]\left[\sum_{j=0}^r p_j^*\left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right]\\ &= \exp\{-(n+\beta_1)\theta_1-(n+\beta_2)\theta_2-(n+\beta_3)\theta_3\}\\ &\quad\times\theta_1^{\sum x_i+\alpha_1-1}\theta_2^{\sum y_i+\alpha_2-1}\theta_3^{\alpha_3-1}\\ &\quad\times\sum_{k=0}^{S_n+r}\left[\sum_{l=\max\{0,k-\min\{S_n,r\}\}}^{\min\{k,\min\{S_n,r\}\}} c_l^{(n)} p_{k-l}^*\right]\left(\frac{\theta_3}{\theta_1\theta_2}\right)^k\\ &= \sum_{k=0}^{S_n+r}\left[\sum_{l=\max\{0,k-\min\{S_n,r\}\}}^{\min\{k,\min\{S_n,r\}\}} c_l^{(n)} p_{k-l}^*\right]\\ &\quad\times(\theta_1^{\sum x_i+\alpha_1-k-1}\exp\{-(n+\beta_1)\theta_1\})\\ &\quad\times(\theta_2^{\sum y_i+\alpha_2-k-1}\exp\{-(n+\beta_2)\theta_2\})\\ &\quad\times(\theta_3^{\alpha_3+k-1}\exp\{-(n+\beta_3)\theta_3\}).\end{aligned}$$

If for $k = 0,1,\ldots,S_n+r$ we will define the appropriate normalizing constants $\tilde{p}_k$ as:

$$\left[\sum_{l=\max\{0,k-\min\{S_n,r\}\}}^{\min\{k,\min\{S_n,r\}\}} c_l^{(n)} p_{k-l}^*\right]\Gamma\left(\sum x_i+\alpha_1-k\right) \times\Gamma\left(\sum y_i+\alpha_2-k\right)\Gamma(\alpha_3+k)\left[\frac{(n+\beta_1)(n+\beta_2)}{n+\beta_3}\right]^k$$

and subsequently for $k = 0,1,\ldots,S_n+r$ normalize them (to be weights):

$$w_k = \tilde{p}_k / \sum_{m=0}^{S_n+r}\tilde{p}_m$$

we will have that:

$$\begin{aligned}p(\theta|\mathbf{z}_n) &= \sum_{k=0}^{S_n+r} w_k G\left(\sum x_i+\alpha_1-k, n+\beta_1\right)\\ &\quad\times G\left(\sum y_i+\alpha_2-k, n+\beta_2\right)G(\alpha_3+k, n+\beta_3).\end{aligned}$$

Appendix 3

Regarding the likelihood and the posterior we have respectively:

$$f(z_{n+1}|\theta) = \exp(-(\theta_1+\theta_2+\theta_3))\theta_1^{x_{n+1}}\theta_2^{y_{n+1}} \times\sum_{i=0}^{s_{n+1}} v_i^{(n+1)}\left(\frac{\theta_3}{\theta_1\theta_2}\right)^i$$

$$p(\theta|\mathbf{z}_n) = \sum_{j=0}^{S_n+r} w_j G\left(\sum x_i + \alpha_1 - j, n+\beta_1\right) \times G\left(\sum y_i + \alpha_2 - j, n+\beta_2\right) G(\alpha_3 + j, n+\beta_3).$$

Then:

$$\begin{aligned} P(z_{n+1}|\mathbf{z}_n) &= \int f(z_{n+1}|\theta) p(\theta|\mathbf{z}_n) d\theta \\ &= \int \Bigg[\exp(-(\theta_1+\theta_2+\theta_3))\theta_1^{x_{n+1}} \theta_2^{y_{n+1}} \\ &\quad \times \sum_{i=0}^{s_{n+1}} v_i^{(n+1)} \left(\frac{\theta_3}{\theta_1\theta_2}\right)^i \Bigg] \\ &\quad \times \Bigg[\sum_{k=0}^{S_n+r} w_j^* (\theta_1^{\sum_{i=1}^n x_i + \alpha_1 - j - 1} \exp\{-(n+\beta_1)\theta_1\}) \\ &\quad \times (\theta_2^{\sum_{i=1}^n y_i + \alpha_2 - j - 1} \exp\{-(n+\beta_2)\theta_2\}) \\ &\quad \times (\theta_3^{\alpha_3 + j - 1} \exp\{-(n+\beta_3)\theta_3\}) \Bigg] d\theta = (\mathbf{I}) \end{aligned}$$

where

$$w_j^* = w_j \frac{(n+\beta_1)^{\sum_{i=1}^n x_i + \alpha_1 - j}}{\Gamma(\sum_{i=1}^n x_i + \alpha_1 - j)} \times \frac{(n+\beta_2)^{\sum_{i=1}^n y_i + \alpha_2 - j} (n+\beta_3)^{\alpha_3 + j}}{\Gamma(\sum_{i=1}^n y_i + \alpha_2 - j)\Gamma(\alpha_3 + j)}$$

thus:

$$\begin{aligned} (\mathbf{I}) &= \int \exp\{-(n+\beta_1+1)\theta_1 \\ &\quad - (n+\beta_2+1)\theta_2 - (n+\beta_3+1)\theta_3\} \\ &\quad \times \theta_1^{(\sum_{i=1}^{n+1} x_i + \alpha_1 - 1)} \theta_2^{(\sum_{i=1}^{n+1} y_i + \alpha_2 - 1)} \theta_3^{(\alpha_3 - 1)} \\ &\quad \times \left[\sum_{i=0}^{s_{n+1}} v_i^{(n+1)} \left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right] \left[\sum_{j=0}^{S_n+r} w_j^* \left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right] d\theta \\ &= (\mathbf{II}) \end{aligned}$$

but

$$\left[\sum_{i=0}^{s_{n+1}} v_i^{(n+1)} \left(\frac{\theta_3}{\theta_1\theta_2}\right)^i\right] \left[\sum_{j=0}^{S_n+r} w_j^* \left(\frac{\theta_3}{\theta_1\theta_2}\right)^j\right] = \sum_{k=0}^{S_{n+1}+r} \left[\sum_{l=A}^{B} v_l^{(n+1)} w_{k-l}^*\right] \left(\frac{\theta_3}{\theta_1\theta_2}\right)^k$$

where $A = \max\{0, k - \min\{s_{n+1}, S_n + r\}\}$ and $B = \min\{k, \min\{s_{n+1}, S_n + r\}\}$ and therefore

$$\begin{aligned} (\mathbf{II}) &= \sum_{k=0}^{S_{n+1}+r} \left[\sum_{l=A}^{B} v_l^{(n+1)} w_{k-l}^*\right] \int [\theta_1^{(\sum_{i=1}^{n+1} x_i + \alpha_1 - k - 1)} \exp\{(n+\beta_1+1)\theta_1\}] \\ &\quad \times [\theta_2^{(\sum_{i=1}^{n+1} y_i + \alpha_2 - k - 1)} \exp\{(n+\beta_2+1)\theta_2\}][\theta_3^{(\alpha_3+k-1)} \exp\{(n+\beta_3+1)\theta_3\}] d\theta \\ &= \sum_{k=0}^{S_{n+1}+r} \left[\sum_{l=A}^{B} v_l^{(n+1)} w_{k-l}^*\right] \frac{\Gamma(\sum_{i=1}^{n+1} x_i + \alpha_1 - k)\Gamma(\sum_{i=1}^{n+1} y_i + \alpha_2 - k)\Gamma(\alpha_3 + k)}{(n+\beta_1+1)^{(\sum_{i=1}^{n+1} x_i + \alpha_1 - k)} (n+\beta_2+1)^{(\sum_{i=1}^{n+1} y_i + \alpha_2 - k)} (n+\beta_3+1)^{(\alpha_3+k)}} \end{aligned}$$

and replacing w_{k-l}^* we obtain:

$$\begin{aligned} (\mathbf{II}) &= \sum_{k=0}^{S_{n+1}+r} \sum_{l=A}^{B} v_l^{(n+1)} w_{k-l} \frac{(n+\beta_1)^{\sum_{i=1}^n x_i + \alpha_1 - (k-l)} (n+\beta_2)^{\sum_{i=1}^n y_i + \alpha_2 - (k-l)} (n+\beta_3)^{\alpha_3 + k - l}}{(n+\beta_1+1)^{\sum_{i=1}^{n+1} x_i + \alpha_1 - k} (n+\beta_2+1)^{\sum_{i=1}^{n+1} y_i + \alpha_2 - k} (n+\beta_3+1)^{\alpha_3 + k}} \\ &\quad \times \frac{\Gamma(\sum_{i=1}^{n+1} x_i + \alpha_1 - k)\Gamma(\sum_{i=1}^{n+1} y_i + \alpha_2 - k)\Gamma(\alpha_3 + k)}{\Gamma(\sum_{i=1}^{n} x_i + \alpha_1 - (k-l))\Gamma(\sum_{i=1}^{n} y_i + \alpha_2 - (k-l))\Gamma(\alpha_3 + k - l)}. \end{aligned}$$

References

Aitchison, J., Ho, C.H.: The multivariate Poisson log-normal distribution. Biometrika **76**, 643–653 (1989)

Cameron, A.C., Trivedi, P.K.: Regression Analysis of Count Data. Oxford University Press, Oxford (1998)

Chib, S., Winkelmann, R.: Markov chain Monte Carlo analysis of correlated count data. J. Bus. Econ. Stat. **19**, 428–435 (2001)

Dabney, A.R., Wakefield, J.C.: Issues in the mapping of two diseases. Stat. Method. Med. Res. **14**, 83–112 (2005)

Dalal, S.R., Hall, W.J.: Approximating priors by mixture of natural conjugate priors. J. Roy. Stat. Soc. B **45**, 278–286 (1983)

DeSouza, C.M.: An approximate bivariate Bayesian method for analyzing small frequencies. Biometrics **48**, 1113–1130 (1992)

Fearnhead, P.: Direct simulation for discrete mixture distributions. Stat. Comput. **15**, 125–133 (2005)

Gaver, D.P.: Multivariate Gamma distributions generated by mixture. Sankhya **32**, 123–126 (1970)

Geisser, S.: Predictive Inference: An Introduction. Chapman and Hall, London (1993)

Joe, H.: Multivariate Models and Dependence Concepts. Chapman and Hall, New York (1997)

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

Jung, R., Winkelmann, R.: Two aspects of labor mobility: a bivariate Poisson regression approach. Empir. Econ. **18**, 543–556 (1993)

Kocherlakota, S., Kocherlakota, K.: Bivariate Discrete Distributions. Marcel Decker, New York (1992)

Subrahmaniam, K.: A test for intrinsic correlation in the theory of accident proneness. J. Roy. Stat. Soc. B **28**, 180–189 (1966)

Tsionas, E.G.: Bayesian analysis of the multivariate Poisson distribution. Commun. Stat. Theory Method. **28**, 431–451 (1999)

Winkelmann, R.: Seemingly unrelated negative binomial regression. Oxf. Bull. Econ. Stat. **62**, 553–560 (2000)