STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:1885–1905
Published online 26 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2382

Bayesian analysis of the differences of count data

D. Karlis*,† and I. Ntzoufras

Department of Statistics, Athens University of Economics and Business, 76, Patission Str., 10434 Athens, Greece

SUMMARY

Paired count data usually arise in medicine when before and after treatment measurements are considered. In the present paper we assume that the correlated paired count data follow a bivariate Poisson distribution in order to derive the distribution of their difference. The derived distribution is shown to be the same as the one derived for the difference of the independent Poisson variables, thus recasting interest on the distribution introduced by Skellam. Using this distribution we remove correlation, which naturally exists in paired data, and we improve the quality of our inference by using exact distributions instead of normal approximations. The zero-inflated version is considered to account for an excess of zero counts. Bayesian estimation and hypothesis testing for the models considered are discussed. An example from dental epidemiology is used to illustrate the proposed methodology. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: bivariate Poisson distribution; RJMCMC; Bayes factor; DMFT data; zero-inflated distributions

1. INTRODUCTION

The decayed, missing and filled teeth (DMFT) index is an important indicator in dental epidemiology for the oral health status of a patient. As the name of the index reveals, the number of teeth with problems are counted and the index is treated as a count variable (see for details Reference [1]). Therefore, DMFT measurements before and after a treatment are used to measure the effect of a preventive method in dental caries. Generally, in randomized clinical trials, treatment effects are completely represented by differences over time (before and after the treatment intervention). By using the differences instead of the original data, we eliminate correlation and possible discrepancies observed in the characteristics of the individuals at the beginning of the study (see, for example, Reference [2]). In the case of the DMFT index, the difference is an integer random variable for which normal approximations are not always valid since such data can take values on a small range of integers and may further exhibit skewness. Techniques for continuous or binary paired data are widely available in the literature

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

Received 17 May 2004
Accepted 20 June 2005

while methods suitable for the analysis of paired count data are rare. Such methods are mainly based on normal approximations of discrete distributions explicitly or on modelling the first few moments (e.g. through a GEE approach). It is apparent that methodologies directly based on discrete distributions can improve the inference from paired count data.

Here, we start from the bivariate Poisson distribution as the joint distribution of our data and we develop Bayesian inference and hypothesis tests using the difference of the two correlated Poisson variables. Not surprisingly, the resulting distribution is of the same form as the distribution of the difference of two independent Poisson variates introduced by Skellam [3]. Thus, we recast interest on this ignored distribution. We further consider the case of the zero inflated distribution, which is particularly interesting when an excess of ties is observed. In pre- and post-treatment comparisons, such a model implies a large probability that the patient's condition remains stable.

Our approach differs from other competitive methods like GEE and mixed models. First of all our approach is Bayesian and therefore allows for incorporating prior information. Beyond the philosophical issues for or against the Bayesian approach, our method is relatively simpler in nature than mixed effects models, where specific assumptions must be made concerning the random effects. Moreover, it differs from the GEE approach since we use a well specified model based on a distribution for integer numbers and not on normal approximations or general moment conditions.

The aim of the paper is to introduce the Poisson difference and its zero inflated version and present possible applications of them in the analysis of paired count data in medicine. Further contributions of the paper are: (a) the estimation of the parameters using the Bayesian approach and the development of an MCMC algorithm for this reason; (b) evaluation of certain hypotheses in these distributions using posterior model odds and RJMCMC algorithm for their estimation; (c) development of prior distributions for the hypothesis/model comparison; (d) illustration of their practical use in an example from dental epidemiology. We should further note, the distributions introduced here can be used to analyse discrete distributions which lie in the whole range of integers (positive and negative) even if we cannot assume that they arise as difference of Poisson variables. This set-up is only used for the data augmentation algorithm. In the following, we focus on the distributions introduced rather than in a general log-linear type model in order to underline how these distributions can be used for a direct analysis and comparison of before and after treatment.

The remainder of the paper proceeds as follows: the distribution is derived in Section 2 where Bayesian estimation and hypothesis testing are also considered. An illustration of our methodology is presented using DMFT index data in Section 3. The zero inflated extension of the Poisson difference model is examined in Section 4, followed by its application to DMFT index data. We conclude with a short discussion in Section 6.

2. THE DISTRIBUTION OF THE BIVARIATE POISSON DIFFERENCE

2.1. The probability distribution

Let us consider two count measurements X and Y and their difference $Z = X - Y$. The probability function of the difference Z is a discrete distribution defined on the set of integer numbers $\mathscr{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Although publications on distributions defined on $\mathscr{Z}$ are rare, the

difference of two independent Poisson random variables has been discussed by Irwin [4] for the case of equal means and Skellam [3] for the case of different Poisson means. Suppose that the discrete random variables X, Y jointly follow the bivariate Poisson distribution. The probability function is given by

$$P_{X,Y}(x, y \mid \theta_1, \theta_2, \theta_3) = \mathrm{e}^{-(\theta_1+\theta_2+\theta_3)} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\theta_3}{\theta_1\theta_2}\right)^i \tag{1}$$

$\theta_1, \theta_2, \theta_3 \geqslant 0$, $x, y = 0, 1, \ldots$ with $\mathrm{Cov}(X, Y) = \theta_3$. If $\theta_3 = 0$ then the two variables are independent. Marginally, each random variable follows a Poisson distribution with parameters $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$, respectively. For more details on the bivariate Poisson distribution the reader can refer to Reference [5].

Lemma 1
If (X, Y) jointly follow the bivariate Poisson distribution with probability function (1), then the distribution of the random variable $Z = X - Y$ is given by

$$f_{\mathrm{PD}}(z \mid \theta_1, \theta_2) = P(Z = z \mid \theta_1, \theta_2) = \mathrm{e}^{-(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{z/2} I_{|z|}(2\sqrt{\theta_1\theta_2}) \tag{2}$$

for all $z \in \mathscr{Z}$, where $I_r(x)$ is the modified Bessel function of order r (see Reference [6, p. 375]) defined by

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k!\Gamma(r + k + 1)}$$

Proof
A quick proof of the result can be based on the trivariate reduction of the bivariate Poisson distribution. If one defines three independent random variables W_i, each one following a Poisson distribution with parameter θ_i, $i = 1, 2, 3$, then the random variables $X = W_1 + W_3$ and $Y = W_2 + W_3$, jointly follow a bivariate Poisson distribution. The above implies that $Z = X - Y = W_1 - W_2$ and hence the random variable of the difference does not depend on the random variable W_3 and, hence, its parameter θ_3. □

Moreover, the probability function of Z is identical to that of the difference of two independent Poisson variates with parameters θ_1 and θ_2, respectively, derived by Skellam [3]; see also Reference [7, p. 191] and references therein. Note a misprint in Reference [7] about the order of the modified Bessel function. For simplicity we will refer to this distribution as the Poisson difference (PD) distribution and will be denoted as $\mathrm{PD}(\theta_1, \theta_2)$.

Remark
The distributions of the difference between two independent and two bivariate (correlated) Poisson variates are of the same form. However, the interpretation of the parameters is different. Assuming that the bivariate Poisson distribution is the correct distribution, then the marginal means $\bar{x}$ and $\bar{y}$ will be unbiased estimates of $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$, respectively, instead of the parameters of interest θ_1 and θ_2. Therefore, the parameters of the PD distribution are not directly connected to the marginal means of the actual Poisson distributions.

Before we proceed further, we should underline that the above distribution can be used for modelling any discrete measurement which lie in the whole range of integer numbers (positive and negative). This is important since discrete distributions for such measurements are rare while such data usually arise in medicine as differences of discrete outcomes. There is no need for the original data of the difference to be Poisson. We only use the Poisson augmentation for the Bayesian estimation which follows.

2.2. *Properties of the Poisson difference distribution*

In this section we review existing and provide new properties of the Poisson difference distribution. The expected value of the $\mathrm{PD}(\theta_1, \theta_2)$ distribution is given by $E(Z) = \theta_1 - \theta_2$ while the variance is $\mathrm{Var}(Z) = \theta_1 + \theta_2$. In general, the odd cummulants are equal to $\theta_1 - \theta_2$ while the even cummulants are equal to $\theta_1 + \theta_2$. The skewness is determined by the sign of $\theta_1 - \theta_2$. Thus, if $\theta_1 > \theta_2$ then positive skewness is present. The distribution is symmetric only when $\theta_1 = \theta_2$ (the case discussed by Irwin [4]). For large values of the $\theta_1 + \theta_2$ the distribution can be sufficiently approximated by the normal distribution. If the parameter θ_2 is very close to 0, then the distribution tends to a Poisson distribution. On the contrary if the parameter θ_1 approaches 0, then the distribution is the negative of a Poisson distribution (that is, a Poisson distribution at the negative axis). An interesting property is a 'type' of symmetry given by $f_{\mathrm{PD}}(z|\theta_1, \theta_2) = f_{\mathrm{PD}}(-z|\theta_2, \theta_1)$. This property can be useful for fast probability calculations. The unimodality of the Poisson difference distribution can be proven using the results of Keilson and Gerber [8]. The sum and the difference of two Poisson difference random variables also follow the same distribution as the following lemma shows.

Lemma
Let us consider two random variables say $Z_1 \sim \mathrm{PD}(\theta_1, \theta_2)$ and $Z_2 \sim \mathrm{PD}(\theta_3, \theta_4)$. Then the sum $S_2 = Z_1 + Z_2$ follows a $\mathrm{PD}(\theta_1 + \theta_3, \theta_2 + \theta_4)$ distribution, while the difference $D_2 = Z_1 - Z_2$ follows a Poisson difference distribution with parameters $\theta_1 + \theta_4$ and $\theta_2 + \theta_3$.

Proof
Since Z_1 and Z_2 can be written as $Z_1 = X_1 - X_2$ and $Z_2 = X_3 - X_4$ with $X_i \sim \mathrm{Poisson}(\theta_i)$ independently, for $i = 1, 2, 3, 4$, one obtains that

$$S_2 = (X_1 - X_2) + (X_3 - X_4) = (X_1 + X_3) - (X_2 + X_4)$$

with $(X_1 + X_3) \sim \mathrm{Poisson}(\theta_1 + \theta_3)$ and $(X_2 + X_4) \sim \mathrm{Poisson}(\theta_2 + \theta_4)$. With similar arguments one can show the result for the difference. □

Remark
If we consider a random sample of size n of i.i.d variables $Z_i, i = 1, \ldots, n$ then we can straightforwardly show that $S_n = \sum_{i=1}^{n} Z_i \sim \mathrm{PD}(n\theta_1, n\theta_2)$. This quantity approaches a normal distribution quite quickly.

Lemma 3
Suppose that one has a series of independent Poisson variates, that is $X_i \sim \mathrm{Poisson}(\theta_i)$, $i = 1, \ldots, n$. Further assume a vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^{\mathrm{T}}$, with $\alpha_i \in \{-1, 1\}$. Then the random variable $S_n = \sum_{i=1}^{n} \alpha_i X_i$ follows a Poisson difference distribution with parameters $\lambda_1 = \sum_{i=1}^{n} \theta_i \mathscr{I}(\alpha_i > 0)$ and $\lambda_2 = \sum_{i=1}^{n} \theta_i \mathscr{I}(\alpha_i < 0)$, where $\mathscr{I}(A)$ is the indicator function which

takes value one if A is true and value zero otherwise. Note that if all $\alpha_i = 1$ then the resulting distribution is a simple Poisson distribution with parameter $\lambda_1 = \sum_{i=1}^{n} \theta_i$.

Proof
We set $S_n^{(1)}$ and $S_n^{(2)}$ the sums with $\alpha_i > 0$ and $\alpha_i < 0$, respectively. Therefore

$$S_n^{(1)} = \sum_{i=1}^{n} \mathcal{I}(\alpha_i > 0) X_i \sim \text{Poisson}(\lambda_1 = \sum_{i=1}^{n} \theta_i \mathcal{I}(\alpha_i > 0))$$

and

$$S_n^{(2)} = \sum_{i=1}^{n} \mathcal{I}(\alpha_i < 0) X_i \sim \text{Poisson}(\lambda_2 = \sum_{i=1}^{n} \theta_i \mathcal{I}(\alpha_i < 0))$$

Now, the initial random variable S_n can be expressed as a difference of two Poisson variables, $S_n = S_n^{(1)} - S_n^{(2)}$, and hence is distributed as a Poisson difference distribution with parameters λ_1 and λ_2 as defined above. □

Maximum likelihood estimation can be done via an EM type algorithm using the same data augmentation as the one developed for Bayesian estimation in the next section. More details on classical estimation can be found in Reference [9].

2.3. Bayesian inference

In Poisson models, the gamma conjugate prior distribution is used to facilitate analytic calculations. Here we adopt the same prior structure which will be conjugate only in terms of conditional distributions. Namely we assume independent prior distributions given by

$$\theta_1 \sim \text{Gamma}(a_1, b_1) \quad \text{and} \quad \theta_2 \sim \text{Gamma}(a_2, b_2) \tag{3}$$

Observed differences z_i, $i = 1, 2, \ldots, n$ constitute the data vector $\mathbf{z}$. The full model likelihood is given by

$$f(\mathbf{z} \mid \theta_1, \theta_2) = \mathrm{e}^{-n(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{(1/2)\sum z_i} \prod_{i=1}^{n} I_{|z_i|}(2\sqrt{\theta_1\theta_2})$$

and therefore the posterior distribution $f(\theta_1, \theta_2 \mid \mathbf{z}) \propto f(\mathbf{z} \mid \theta_1, \theta_2) f(\theta_1) f(\theta_2)$ is not analytically tractable. We introduce the latent data

$$v_i \sim \text{Poisson}(\theta_1) \quad \text{and} \quad u_i \sim \text{Poisson}(\theta_2) \tag{4}$$

under the constraint $z_i = v_i - u_i$. Denote $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{u} = (u_1, \ldots, u_n)$. The full joint posterior distribution of the model parameters (θ_1, θ_2) and the latent data $\mathbf{v}, \mathbf{u}$ is given by

$$\begin{aligned} f(\mathbf{v}, \mathbf{u}, \theta_1, \theta_2 | \mathbf{z}) &\propto f(\mathbf{z} | \mathbf{v}, \mathbf{u}, \theta_1, \theta_2) f(\mathbf{v}, \mathbf{u} | \theta_1, \theta_2) f(\theta_1, \theta_2) \\ &\propto \mathrm{e}^{-n(\theta_1+\theta_2)} \theta_1^{\sum_{i=1}^{n} v_i} \theta_2^{\sum_{i=1}^{n} u_i} \left\{ \prod_{i=1}^{n} \frac{\mathcal{I}(z_i = v_i - u_i)}{v_i! u_i!} \right\} f(\theta_1, \theta_2) \end{aligned}$$

where $f(\mathbf{v}, \mathbf{u} | \theta_1, \theta_2)$ is the model likelihood when the latent data are available, and $f(\theta_1, \theta_2)$ is the joint prior of the model parameters.

The gamma prior (3) results in the joint posterior

$$f(\mathbf{v},\mathbf{u},\theta_1,\theta_2|\mathbf{z}) \propto \mathrm{e}^{-(n+b_1)\theta_1-(n+b_2)\theta_2}\theta_1^{n\bar{v}+a_1-1}\theta_2^{n\bar{u}+a_2-1}\left\{\prod_{i=1}^{n}\frac{\mathscr{I}(z_i=v_i-u_i)}{v_i!u_i!}\right\}$$

where $\bar{v}=\sum_{i=1}^{n} v_i/n$ and $\bar{u}=\sum_{i=1}^{n} u_i/n$. Using the above target distribution we construct the following MCMC algorithm:

1. Sample (v_i,u_i) from $f(v_i,u_i|z_i=v_i-u_i,\theta_1,\theta_2)\propto\frac{\theta_1^{v_i}\theta_2^{u_i}}{v_i!u_i!}\mathscr{I}(z_i=v_i-u_i)$, for $i=1,\ldots,n$.
2. Sample θ_1 from $f(\theta_1|\theta_2,\mathbf{v},\mathbf{u})\sim\text{Gamma}(n\bar{v}+a_1,n+b_1)$.
3. Sample θ_2 from $f(\theta_2|\theta_1,\mathbf{v},\mathbf{u})\sim\text{Gamma}(n\bar{u}+a_2,n+b_2)$.

In step 1, for updating the augmented data (v_i,u_i), we use the following Metropolis step:

- If $z_i<0$ and (v_i,u_i) the current values of the augmented data then
  — Propose $v_i'\sim\text{Poisson}(\theta_1)$ and $u_i'=v_i'-z_i$.
  — Accept the proposed move with probability $\alpha=\min\left\{1,\theta_2^{(v_i'-v_i)}\frac{(v_i-z_i)!}{(v_i'-z_i)!}\right\}$.
- If $z_i\geqslant 0$ and (v_i,u_i) the current values of the augmented data then
  — Propose $u_i'\sim\text{Poisson}(\theta_2)$ and $v_i'=u_i'+z_i$.
  — Accept the proposed move with probability $\alpha=\min\left\{1,\theta_1^{(u_i'-u_i)}\frac{(u_i+z_i)!}{(u_i'+z_i)!}\right\}$.

Extensions of the above model and the respective MCMC algorithm can be constructed in a straightforward manner. A direct extension is implied by adopting covariates on the log scale of the parameters θ_1 and θ_2. In such a case, we should use a normal prior distribution on the new parameter space in the same manner as in Poisson log-linear models (see, for example, Reference [10]).

2.4. *Bayesian evaluation of the equality of means*

When paired data are available, the hypothesis of equal means of the two (dependent) measurements is usually under investigation. The equality of θ_1 and θ_2 can be investigated using the posterior distribution of the difference $\theta_1-\theta_2$ (or the ratio θ_1/θ_2) and by checking whether the value of zero (or one, respectively), which corresponds the equality of θ's, lie in the central region or in the tails of the posterior distribution. In this case, the corresponding credible intervals may give us an idea about whether the hypothesized value is plausible or not. But it cannot be used to evaluate evidence in favour or against a specific hypothesis (see Reference [11, p. 262]). Therefore, in this section, we describe how to estimate Bayes factor in order to quantify evidence in favour of the equality of $\theta's$.

Let us consider available data $\mathbf{z}$ and a set of competing models $\mathscr{M}=\{m_1,m_2,\ldots,m_{|\mathscr{M}|}\}$; where $|\mathscr{M}|$ denotes the number of models under consideration. If $f(m)$ is the prior probability of model $m\in\mathscr{M}$ and $\boldsymbol{\theta}_m$ is its corresponding vector of model parameters, then, using the Bayes theorem, the posterior probability of model $m\in\mathscr{M}$ is given by

$$f(m\,|\,\mathbf{z})=\frac{f(m\,|\,\mathbf{z})f(m)}{\sum_{m_k\in\mathscr{M}} f(m_k\,|\,\mathbf{z})f(m_k)}=\frac{f(m)\int f(\mathbf{z}\,|\,\boldsymbol{\theta}_m,m)f(\boldsymbol{\theta}_m\,|\,m)\,\mathrm{d}\boldsymbol{\theta}_m}{\sum_{m_k\in\mathscr{M}} f(m_k)\int f(\mathbf{z}\,|\,\boldsymbol{\theta}_{m_k},m_k)f(\boldsymbol{\theta}_{m_k}\,|\,m_k)\,\mathrm{d}\boldsymbol{\theta}_{m_k}}$$

Alternatively, when we compare two competing models m_1 and m_2 induced by two hypotheses we wish to test, then we focus on the posterior odds PO_{12} of model m_1 *versus* model m_2

defined as

$$\mathrm{PO}_{12}=\frac{f(m_1|\mathbf{z})}{f(m_2|\mathbf{z})}=\frac{f(\mathbf{z}|m_1)}{f(\mathbf{z}|m_2)}\times\frac{f(m_1)}{f(m_2)}=B_{12}\times\frac{f(m_1)}{f(m_2)}$$

where B_{12} and $f(m_1)/f(m_2)$ are the 'Bayes factor' and the 'prior model odds' of model m_1 against model m_2, respectively; for more details see Reference [12].

The integrals involved in the computation of the posterior model probabilities are analytically tractable only in specific examples. Therefore, asymptotic approximations or alternative computational methods are frequently employed. In the following, we facilitate the reversible jump MCMC (RJMCMC) algorithm introduced by Green [13] to evaluate posterior probabilities of competing hypotheses concerning the PD distribution. Under this distribution, a hypothesis of major interest is the equality of means of the latent data. This hypothesis is substantial when we wish to compare the efficiency of a medical treatment on a sample of patients whose performance is measured before and after the treatment. The above hypothesis testing induces the comparison of two models: the Poisson difference model with common parameters denoted by m_1 and parameter $\boldsymbol{\theta}_{m_1}=(\theta)$ and non-common parameters denoted by m_2 with parameter vector $\boldsymbol{\theta}_{m_2}=(\theta_1,\theta_2)^{\mathrm{T}}$.

In order to simplify and accommodate our hypothesis testing in an MCMC set-up we introduce an additional latent binary indicator γ and $m_{1+\gamma}$ model indicator. The prior model probabilities are specified as $f(\gamma)=\frac{1}{2}$, $\gamma=0,1$. The MCMC for model selection can be summarized by the following steps:

1. Generate γ using the following Metropolis step:
   (a) Propose $\gamma'=1-\gamma$ with probability one.
   (b) i. If $\gamma'=1$ then propose θ_1' and θ_2' from a proposal distribution $q(\theta_1',\theta_2'\,|\,\theta)$.
   ii. If $\gamma'=0$ then propose common θ' from $q(\theta'\,|\,\theta_1,\theta_2)$.
   (c) Accept the proposed move with probability $\alpha=\min\left\{1,\frac{O(\theta_1',\theta_2',\theta)^{1-\gamma}}{O(\theta_1,\theta_2,\theta')^{\gamma}}\right\}$ where $O(\theta_1',\theta_2',\theta)$ is given by

$$O(\theta_1',\theta_2',\theta)=\left\{\prod_{i=1}^{n}\frac{f_{\mathrm{PD}}(z_i\,|\,\theta_1',\theta_2')}{f_{\mathrm{PD}}(z_i\,|\,\theta,\theta)}\right\}\frac{f(\theta_1',\theta_2'\,|\,\gamma=1)}{f(\theta\,|\,\gamma=0)}\,\frac{f(\gamma=1)}{f(\gamma=0)}\,\frac{q(\theta\,|\,\theta_1',\theta_2')}{q(\theta_1',\theta_2'\,|\,\theta)}\qquad(5)$$

2. Generate the latent data $(v_i^{(m_{1+\gamma})},u_i^{(m_{1+\gamma})})$ of the current model $m_{1+\gamma}$ from

$$f(v_i^{(m_{1+\gamma})},u_i^{(m_{1+\gamma})}\,|\,z_i,\theta_1^{(m_{1+\gamma})},\theta_2^{(m_{1+\gamma})},\gamma)$$

   following step 1 in Section 2.3; where $\theta_j^{(m_{1+\gamma})}=(1-\gamma)\theta+\gamma\theta_j$ for $j=1,2$ are the parameters of the current model.
3. If $\gamma=1$ generate θ_1 and θ_2 from $f(\theta_1,\theta_2\,|\,\mathbf{v},\mathbf{u})$ as in steps 2 and 3 in the MCMC algorithm of Section 2.3.
   If $\gamma=0$ generate common θ by $f(\theta\,|\,\mathbf{v},\mathbf{u})\sim\mathrm{Gamma}(n\bar{v}+n\bar{u}+a,2n+b)$; where $n\bar{v}=\sum_{i=1}^{n}v_i^{(m_{1+\gamma})}$ and $n\bar{u}=\sum_{i=1}^{n}u_i^{(m_{1+\gamma})}$.

In the above scheme, step 1 refers to the RJMCMC model moves while steps 2 and 3 update the parameters within the model proposed by the RJMCMC step. The latter are optional but are used to improve the mixing and the convergence of the chain.

As proposal distributions we consider gamma distributions for each parameter θ, θ_1 and θ_2 given by $q(\theta_1, \theta_2 \mid \theta) = q(\theta_1 \mid \theta) q(\theta_2 \mid \theta)$ with $q(\theta_j \mid \theta) \sim \text{Gamma}(\bar{a}_j, \bar{b}_j)$ for $j = 1, 2$ and $q(\theta \mid \theta_1, \theta_2) \sim \text{Gamma}(\bar{a}_0, \bar{b}_0)$. The parameters of the proposal distributions are important for the convergence of the RJMCMC algorithm and can be specified using a small pilot run of each model (see Reference [14]). Hence the parameters can be set equal to $\bar{a}_j = \bar{\theta}_j^2 / S_{\theta_j}^2$ and $\bar{b}_j = \bar{\theta}_j / S_{\theta_j}^2$ for $j = 0, 1, 2$; where $\bar{\theta}_j$ and $S_{\theta_j}^2$ are the posterior mean and variance for θ, θ and θ_2 estimated from the small pilot MCMC run. The above-proposed RJMCMC scheme is essentially an independence sampler (or Metropolized Carlin and Chib approach, see Reference [14]). In the above RJMCMC scheme, a natural choice might be to use the current values of θ_1 and θ_2 in order to propose deterministically a value for the parameter of the simpler model, θ. Unfortunately, in this case, the posterior mode of simplest model (with common θ) is far away from the proposed values using either the arithmetic or the geometric mean of θ_1 and θ_2 which are natural choices in similar problems. Another choice could be to consider automatic choices proposed by Brooks *et al.* [15]. This was not straightforward to implement since derivations include the Bessel function and may have complicated considerably the algorithm. For this reason, we have decided that an independence version of RJMCMC would have been a more convenient choice.

After running the RJMCMC algorithm we estimate the posterior model probabilities of models m_k for $k = 1, 2$ by

$$\hat{f}(m_k \mid \mathbf{z}) = \frac{1}{N - B} \sum_{t=B+1}^{N} \mathcal{I}(m^{(t)} = m_k)$$

where N is the total number of iterations considered, B is the number of iterations discarded as burn-in period, $m^{(t)}$ is the value of the model indicator m at iteration t. When we wish to calculate the posterior model odds rather than the posterior model probabilities then we propose to tune the prior model probabilities in such a way that both models are visited and then calculate accordingly the actual posterior model odds.

A popular alternative, for estimating the Bayes factor, is the approach of Chib [16]. Generally the approach of Chib [16] is easier to implement than RJMCMC. On the other hand, the main advantage of RJMCMC is that we can extract results from a single MCMC output while Chib's approach needs output from each model under consideration. This makes Chib's approach difficult to implement when a large number of models is considered. Moreover, in Chib's approach, each model should be treated separately depending on the sampling scheme. Hence, different care should be given if we use Metropolis algorithm instead of Gibbs sampler; see Reference [17]. Finally, some authors have reported that Chib's method (for specific data and models) fails to estimate Bayes factor with precision (see for example Reference [18]). For all the above reasons, we prefer to use RJMCMC which can be adopted even if we use a log-linear set-up or introduce variable selection in the model formulation. In the illustrative example which follows, we have calculated the logarithm of the Bayes factor using also Chib's approach in order to compare its efficiency with that corresponding to the proposed RJMCMC method.

2.5. *Prior distributions for Bayesian model comparison*

One of the difficult tasks in Bayesian model comparison and hypothesis testing is the specification of prior distributions. Difficulties mainly arise due to the behaviour of posterior model odds as noted by Lindley [19] and Bartlett [20]. Essentially, we cannot use priors with large variance which are thought to be non-informative because, in such case the posterior supports the simplest model. Hence, the prior specification for parameters θ_1, θ_2 and θ is difficult since they are defined in $(0, \infty)$. Any prior expressing low information using an extremely large prior variance will activate the Lindley–Bartlett paradox. In order to specify plausible prior distributions for the model comparison of interest, we use ideas induced by Chen *et al.* [21].

Let us assume that we have *a priori* imaginary latent data (v_i^*, u_i^*) of size n^*. Using the imaginary $\mathbf{v}^*$ and $\mathbf{u}^*$, the prior distribution $f(\theta_1, \theta_2 \mid \gamma = 1)$ for model m_2 can be defined by

$$f(\theta_1, \theta_2 \mid \mathbf{v}^*, \mathbf{u}^*, \gamma = 1) \propto f(\mathbf{v}^*, \mathbf{u}^* \mid \theta_1, \theta_2, \gamma = 1)^c f_0(\theta_1, \theta_2 \mid \gamma = 1)$$

where $0 \leqslant c \leqslant 1$ is a parameter controlling the weight of belief on the prior data and $f_0(\theta_1, \theta_2)$ can be considered as the pre-prior distribution of type (3). Here we set $c = 1/(2n^*)$ in order for our prior imaginary data to account for one data point. Standard improper pre-prior can be used by setting $a = a_1 = a_2 = b = b_1 = b_2 = 0$. In this case the above results to the power-prior of Chen *et al.* [21]. We prefer to use proper pre-prior to avoid problems appearing in the computation of Bayes factors (especially when the belief in our imaginary data is weak) which follows. Therefore, for the hyper-parameters we use $a = a_1 = a_2 = b = b_1 = b_2 = 0.01$. Assuming prior data with means $\bar{v}^* = \bar{u}^* = 1$ weighted as one data point, leads us to

$$f(\theta \mid \gamma = 0) \sim \text{Gamma}(1.01, 1.01) \tag{6}$$

$$f(\theta_j \mid \gamma = 1) \sim \text{Gamma}(0.51, 0.51) \quad \text{for } j = 1, 2 \tag{7}$$

3. DECAYED, MISSING AND FILLED TEETH (DMFT) INDEX EXAMPLE

We illustrate our methodology using the DMFT index data of Böhning *et al.* [1]. The data we use here are part of a large prospective study of 797 seven-year old school children from an urban area of Belo Horizonte in Brazil (BELCAP study). Such count data are frequently modelled as Poisson random variables or, in some cases, as zero inflated Poisson random variables (see Reference [1]). Here we consider the difference between the DMFT index before and after a treatment ($Z = \text{DMFT}_1 - \text{DMFT}_2$) to eliminate correlation between measurements (Pearson correlation $= 0.59$). The before and after comparison is performed for the total sample and for six different schools which represent different treatment groups. The available treatment approaches were: oral health education (school 1), enrichment of the school diet with rice bran (school 4), mouthwash with 0.2 per cent sodium fluoride (NaF) solution (school 5) and oral hygiene (school 6). Additionally, in school 2 all the above four methods were used while in school 3 no treatment was used (control group). A histogram of the difference $Z = \text{DMFT}_1 - \text{DMFT}_2$ of the two indexes for each school/treatment group are given in Figure 2.

Here we analyse each group independently in order to keep the conditional conjugacy of the parameters θ_1 and θ_2. Alternatively, a general model can be constructed using the following formulation:

$$Z_{i\kappa} \sim \mathrm{PD}(\theta_{1\kappa}, \theta_{2\kappa}); \quad i = 1, \ldots, n_\kappa, \;\; \kappa = 1, 2, \ldots, 6$$

$$\log(\theta_{j\kappa}) = \mu_j + \alpha_{j\kappa}; \quad j = 1, 2$$

with $\alpha_{11} = \alpha_{21} = 0$; $\theta_{j\kappa}$ is the θ parameter which corresponds to κ school and j period and n_κ is the number of children in κ school/group. Different parameterizations of the above model are possible. No matter which parameterization we use, we can find a 1–1 transformation between θ parameters, calculated from the separate groups analysis, and the parameters of model formulation similar to the above. This implies that the posterior distributions of the parameters of such model can be easily estimated using the MCMC output of the separate group analysis presented in this section. For the model formulation above, we can calculate model parameters using the following transformations:

$$\mu_j = \log \theta_{j1} \quad \text{and} \quad \alpha_{j\kappa} = \log \theta_{j\kappa} - \log \theta_{j1}; \quad j = 1, 2; \;\; \kappa = 2, \ldots, 6$$

Generally, the two sets of parameters will be connected with an invertible matrix $\mathbf{T}$ such that $\boldsymbol{\eta} = \mathbf{T}\boldsymbol{\beta}$; where $\boldsymbol{\eta}^{\mathrm{T}} = (\log \theta_{j\kappa})$ and $\boldsymbol{\beta}$ is the parameter vector of the new model formulation. The new model parameters will be simply given by $\boldsymbol{\beta} = \mathbf{T}^{-1}\boldsymbol{\eta}$.

Initially we provide posterior summaries for the two assumed models. From the posterior summaries presented in Table I, it is evident that the posterior distributions of θ_1 and θ_2 are not close even if we consider the control group. Additionally, for comparison reasons, we provide an error bar plot (Figure 1) with the 2.5 and 97.5 percentiles and the mean of the posterior distributions of $\theta_1 - \theta_2$ which can be thought as estimates of the treatment effect for each school effect. All results are based on 10 000 iterations with additional 1000 iterations as a burn-in. The after treatment DMFT has lower rate measured by θ_2. Posterior summaries for the model with common θ are also provided for comparison reasons. From Figure 1, we can discriminate school 1 as the one with the greatest before and after treatment difference and schools 4 and 6 as the groups with the lower difference. The posterior distributions of the parameters' difference provides sufficient information in favour of the improvement of oral hygiene in all treatment groups. Comparing the posterior distributions of the parameters' difference, we cannot draw a firm conclusion for which treatment efficiency is higher. In the following paragraph, we proceed further and quantify the evidence in favour of the inequality between θ_1 and θ_2 using the Bayes factor.

The logarithm of the Bayes factor of model $m_2 : \mathrm{PD}(\theta_1, \theta_2)$ *versus* model $m_1 : \mathrm{PD}(\theta, \theta)$ is also provided in Table I. The RJMCMC chain was *a priori* calibrated to visit both models under consideration in order to be able to estimate the log-Bayes factors with precision. The evidence in favour of model m_2 is strong even for the control group inducing that the DMFT index was improved in all comparisons. In the same table, Monte Carlo error estimates for the log-Bayes factors using the batch mean method are also given; for more details on the approach see Reference [22] and for similar illustration in variable selection see Reference [14]. A total number of 50 sub samples were used to estimate Monte Carlo error by the standard deviation of the estimates in each sub sample. All Monte Carlo errors are low which means that the Bayes factor is estimated with increased precision. Parameters of the proposal

Table I. Posterior summary results for DMFT data using the Poisson difference distribution (*Priors*: equations (6) and (7); *RJMCMC details*: Burn-in = 1000 iterations, Iterations Kept = 10 000; St. Dev. = Posterior Standard Deviation, $\log B_{21}$: log Bayes factor of m_2 *versus* m_1, MC Error: Monte Carlo error estimated by the standard deviation of the $\log B_{21}$ for 50 sub samples).

		m_1: PD(θ, θ)		m_2: PD(θ_1, θ_2)					
		θ		θ_1		θ_2			
School	n	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	$\log B_{21}$	MC Error
1	124	4.35	0.528	3.28	0.333	1.25	0.304	37.96	0.19
2	127	3.53	0.444	3.11	0.330	1.66	0.310	20.00	0.14
3	136	2.37	0.313	2.10	0.209	0.75	0.181	31.49	0.20
4	132	2.82	0.360	2.71	0.307	1.56	0.294	14.69	0.14
5	155	3.41	0.395	2.65	0.260	0.98	0.239	40.31	0.13
6	123	2.24	0.305	2.13	0.247	1.01	0.227	17.87	0.11
All	797	3.16	0.174	2.72	0.117	1.25	0.108	168.60	0.13

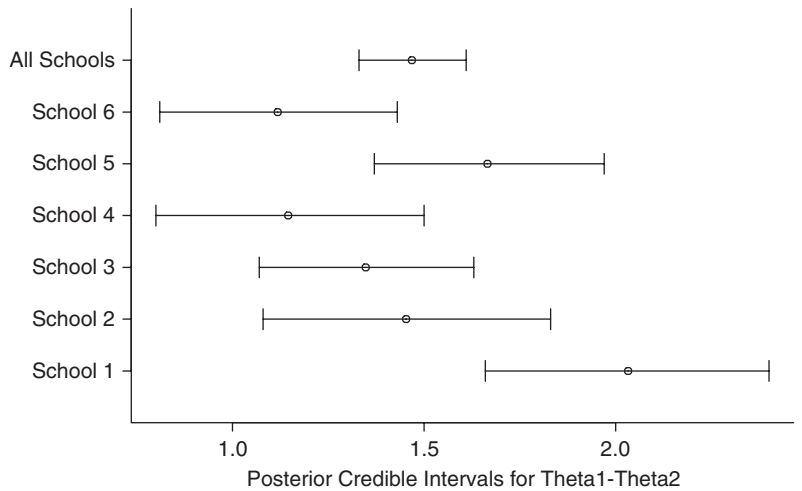

Figure 1. 95 per cent credible intervals of $\theta_1 - \theta_2$ for each school/treatment group using the PD distribution.

distributions were specified following the approach presented in Section 2.4 after running each model for 500 iterations.

We have also calculated $\log B_{21}$ using Chib's marginal likelihood approach using output of length 5000 iterations and additional 500 burn-in for each model. This results in a total of 10 000 iterations kept which is equivalent to the number of iterations we have considered in RJMCMC. Using similar approach as above, we have separated the output in 50 batches and estimated $\log B_{21}$ in each of them using the Chib's marginal likelihood approach. The standard deviation of the estimated quantities in each batch measures the Monte Carlo error and is

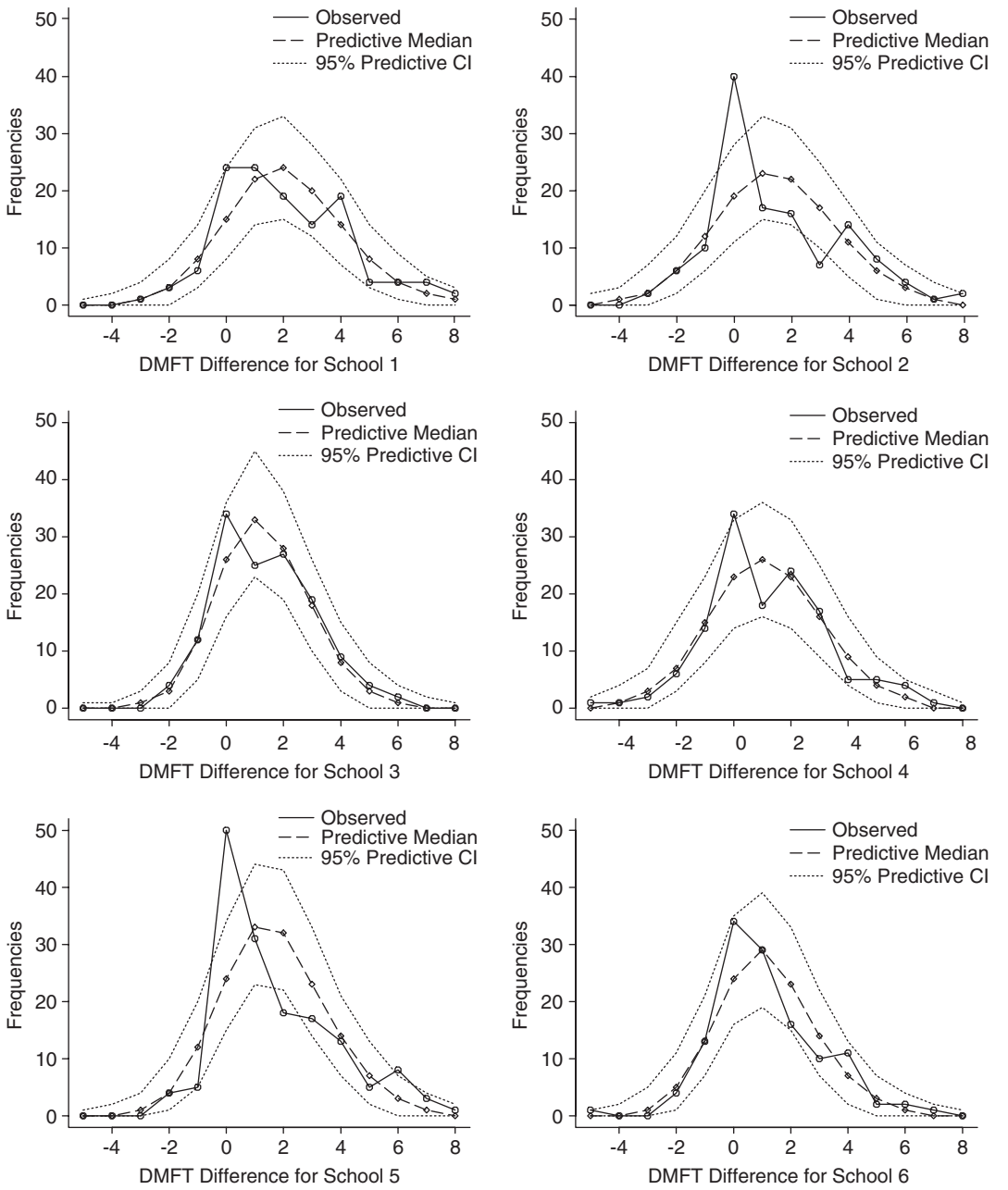

Figure 2. Comparison of observed and median predictive counts of DMFT difference ($DMFT_1 - DMFT_2$) using the Poisson difference distribution for each school/treatment group (dotted lines represent 2.5 and 97.5 per cent quantiles of the predictive distribution of counts).

directly comparable to the corresponding Monte Carlo error calculated for RJMCMC. Chib's method has demonstrated considerably higher Monte Carlo error (about 3–9 times higher).

In all the above comparisons, we used priors (6) and (7). After performing sensitivity analysis for the data of school 2, we found that, for values $\bar{u}^* = \bar{v}^* < 1$, results are quite robust. On the other hand, when we consider increasing the values $\bar{u}^* = \bar{v}^*$ then our priors are quite informative in terms of θ_1 and θ_2 which is also reflected in the values of log-BF giving more support to models with common θ's.

In order to evaluate the fit of our PD model we additionally plotted in Figure 2 the median and the 2.5 and 97.5 percentiles of the predictive counts for each set of the data. This gives us a rough idea of the sufficiency of the PD model concerning the DMFT data. Generally, the PD distribution seems to fit the data with the exception of the zero value. Excess of zeros appears in most of the groups/schools. For this reason, in the following section, we extend our methodology by introducing a zero inflated variation of the PD distribution in order to capture the excess of zeros, which is usually observed in such data.

4. EXTENDED MODELS: THE ZERO INFLATED DISTRIBUTION

Zero-inflated distributions are used when an excess of zeros, relative to the expected frequency, is observed. We extend our model to cover the case of zero inflation. We define the zero inflated Poisson Difference (ZPD) distribution as

$$f_{\mathrm{ZPD}}(0 \mid p, \theta_1, \theta_2) = p + (1 - p) f_{\mathrm{PD}}(0 \mid \theta_1, \theta_2) \quad \text{and}$$

$$f_{\mathrm{ZPD}}(z \mid p, \theta_1, \theta_2) = (1 - p) f_{\mathrm{PD}}(z \mid \theta_1, \theta_2) \tag{8}$$

for $z \in \mathbb{Z} \backslash \{0\}$; where $p \in (0, 1)$ and $f_{\mathrm{PD}}(z \mid \theta_1, \theta_2)$ is given by (2). This distribution will be denoted as ZPD(p, θ_1, θ_2). Zero inflated distributions have been described in Reference [7] and references therein. Recently, Böhning *et al.* [1] recast interest in such distribution proposing zero inflated Poisson distributions allowing for covariates (see also Reference [23]).

The zero-inflated distribution can be thought as a finite mixture distribution with one of the components being a degenerate-at-zero distribution. Thus, the additional parameter p can be considered as the mixing proportion. We introduce a latent binary variable δ_i indicating the component ($\delta_i = 1$ indicates the zero inflated component). The observed data are denoted by $\mathbf{z} = (z_1, \ldots, z_n)^{\mathrm{T}}$. Moreover, we specify

$$P(Z_{0i} = 0 \mid \delta_i = 1, \theta_1, \theta_2) = 1$$

$$P(Z_{0i} = z \mid \delta_i = 0, \theta_1, \theta_2) = f_{\mathrm{PD}}(z \mid \theta_1, \theta_2)$$

where Z_0 is a zero inflated Poisson difference random variable. The parameter vector is now (p, θ_1, θ_2) while the latent data are indicated by $(\boldsymbol{\delta}, \mathbf{v}, \mathbf{u})$. For the parameters θ_1 and θ_2 we use gamma priors as given by (3), while for the mixing proportion p we use a Beta(a_3, b_3) prior distribution; in our illustration we use $a_3 = b_3 = 1$. Then, the target posterior distribution with

the full latent data is given by

$$
\begin{aligned}
f(\boldsymbol{\delta}, \mathbf{v}, \mathbf{u}, p, \theta_1, \theta_2 \mid \mathbf{z}) \propto & \exp\{-(n - n\bar{\delta} + b_1)\theta_1\}\theta_1^{n\bar{v} - \sum_{i=1}^{n} \delta_i v_i + a_1 - 1} \\
& \times \exp\{-(n - n\bar{\delta} + b_2)\theta_2\}\theta_2^{n\bar{u} - \sum_{i=1}^{n} \delta_i u_i + a_2 - 1} \\
& \times \left\{ \prod_{i=1}^{n} [1 - \mathscr{I}(z_i \neq 0)\mathscr{I}(\delta_i = 1)] \left(\frac{\mathscr{I}(z_i = v_i - u_i)}{v_i! u_i!} \right)^{1-\delta_i} \right\} \\
& \times p^{n\bar{\delta} + a_3 - 1}(1 - p)^{n - n\bar{\delta} + b_3 - 1} \qquad (9)
\end{aligned}
$$

where $\bar{\delta} = \sum_{i=1}^{n} \delta_i / n$. Estimation of the above posterior distribution can be obtained using MCMC methods as in Section 2.3; for details see Appendix A.

Moreover, RJMCMC can be used to test two hypotheses of interest: the equality of means (as in Section 2.4) and the existence of excessive zeros ($p > 0$). The second hypothesis is essential when there is an excessive number of ties which implies that ZPD should be used. The combination of the above two hypotheses induces four models: m_1 and m_2 as in Section 2.4 and the ZPD models m_3 and m_4 with parameter vectors $\boldsymbol{\theta}_{m_3} = (\theta, p)^{\mathrm{T}}$ and $\boldsymbol{\theta}_{m_4} = (\theta_1, \theta_2, p)^{\mathrm{T}}$, respectively. In order to incorporate both comparisons in one RJMCMC run we introduce γ and ξ latent binary indicators and m_k model indicator with $k = 1 + \gamma + 2\xi$. The value $\gamma = 0$ implies that $\theta_1 = \theta_2$, while the value $\xi = 0$ implies that $p = 0$. The prior model probabilities are given by $f(\gamma) = f(\xi) = 1/2$, $\gamma, \xi = 0, 1$. The RJMCMC for model selection is summarized by the procedure described in detail in Appendix B.

The ZPD can model the difference of paired data and capture excess of zero-values. The interpretation of zero-values is important when we consider differences of clinical measurements. These zero-values indicate patients whose condition (for several reasons) remains constant and therefore are not affected by the treatment. This percentage may be important for the efficiency of the treatment. In the ZPD model, this percentage can arise from two components. The first component is the PD distribution contributing via the number of observations that we expect to present no change in their measurements under the estimated effects given by θ_2 and θ_1 parameters. The comparison of $\theta_1 = \theta_2$ takes into account this percentage. The second component captures the excess of patients with no improvement which is not predicted by the PD component. In our DMFT example, this may indicate children that are already healthy and their condition cannot be improved further (this cannot be captured by the general PD distribution). Generally, when a treatment is preventive and the sample comes from the general healthy population, then we expect large p to measure a positive effect of the treatment (people are healthy and remain healthy). On the other hand, if the treatment is therapeutic, p may indicate the percentage of patients with non-reversible condition and hence the percentage of patients where the treatment has no effect.

The meaning of the zero difference may vary from problem to problem. Therefore, interpretation may depend upon the measurements we compare. In many cases it might be important to separate zero differences that correspond to 0–0 pairs from the rest. For example, in the DMFT data, the 0–0 pairs correspond to healthy *a priori* children who remain healthy after the treatment. If both before and after data are available, then a possible diagnostic analysis in order to examine the effect of 0–0 counts is to rerun the ZPD model after removing cases

with $X=0$ and compare the estimated mixing proportions. If the mixing proportion is robust, then the deficiency of zeros is similar for $X=0$ and $X>0$. If the proportion decreases, then an important part concerning the excess of zeros might be attributed to the 0–0 case which corresponds to a positive effect of the treatment (patient is healthy and remains healthy).

5. DMFT INDEX EXAMPLE (REVISITED)

We reanalyse the DMFT index data using the ZPD distribution in order to account for the excess of zeros. Posterior means and standard deviations for the ZPD(p,θ_1,θ_2) using 10 000 iterations and 1000 additional iterations as burn-in are given in Table II. Figure 3 depicts error bars of the posterior distribution of the ZPD model for $\theta_1-\theta_2$ and p used to compare differences between schools/treatments. Results are in agreement with Figure 2, where the excess of zeros was clear for schools 2 and 5 and for the aggregated data of all schools, with posterior means of mixing proportions equal to 0.211, 0.229 and 0.148, respectively. Smaller excess of zeros was observed for schools 1, 3, 4 and 6 (posterior means of $p<0.11$); also see Figure 3(b). From Figure 3(a) we observe large differences between θ_1 and θ_2 parameters for schools 1 and 5. On the other hand, for schools 4 and 6 the differences are the lowest.

In order to examine the effect of 0–0 counts differences (recall that in our example the data for both periods are available), we have excluded cases which were healthy at the beginning of the study and we rerun the ZPD model. By this analysis, the effect of removing those observations may indicate more clearly the preventive nature of the treatment, since, healthy patients have now removed. All posterior means of mixing proportion were now found considerably lower (from 0.017 for the aggregated data to 0.030 for school 2). Since in all schools, the mixing proportion was considerably decreased we can infer that the excess of zero differences in the original data was due to 0–0 ties. Such ties here can be considered as in favour of the treatments (which are preventive) since the patient is healthy at the beginning of the study and remains healthy after the treatment.

Following the above analysis, we estimated posterior model probabilities for all four models under consideration using the RJMCMC algorithm. Posterior probabilities for model m_4:

Table II. Posterior summary results for DMFT data using the zero inflated Poisson difference distribution (*Priors*: equations (6) and (7); *RJMCMC details*: Burn-in = 1000 iterations, Iterations Kept = 10 000; St. Dev. = Posterior Standard Deviation, $\log B_{4j}$: log Bayes factor of m_4 *versus* m_j).

	m_4: ZPD(p,θ_1,θ_2)						ZPD(p,θ_1,θ_2) *versus*			
	p		θ_1		θ_2		PD(θ,θ)	PD(θ_1,θ_2)	ZPD(p,θ,θ)	
School	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	$\log B_{41}$	$\log B_{42}$	$\log B_{43}$	$f(m_4 \mid \mathbf{z})$
1	0.083	0.04	3.44	0.34	1.23	0.30	38.43	0.45	39.13	0.596
2	0.211	0.05	3.81	0.42	1.97	0.39	30.00	10.02	20.95	1.000
3	0.076	0.05	2.26	0.23	0.80	0.19	31.04	−0.48	31.68	0.420
4	0.108	0.05	3.00	0.36	1.72	0.32	15.83	1.10	14.54	0.783
5	0.229	0.04	3.21	0.30	1.05	0.25	54.81	14.45	42.25	1.000
6	0.105	0.05	2.37	0.29	1.12	0.25	18.64	0.74	18.11	0.638
All	0.148	0.02	3.10	0.14	1.38	0.12	201.34	32.71	169.73	1.000

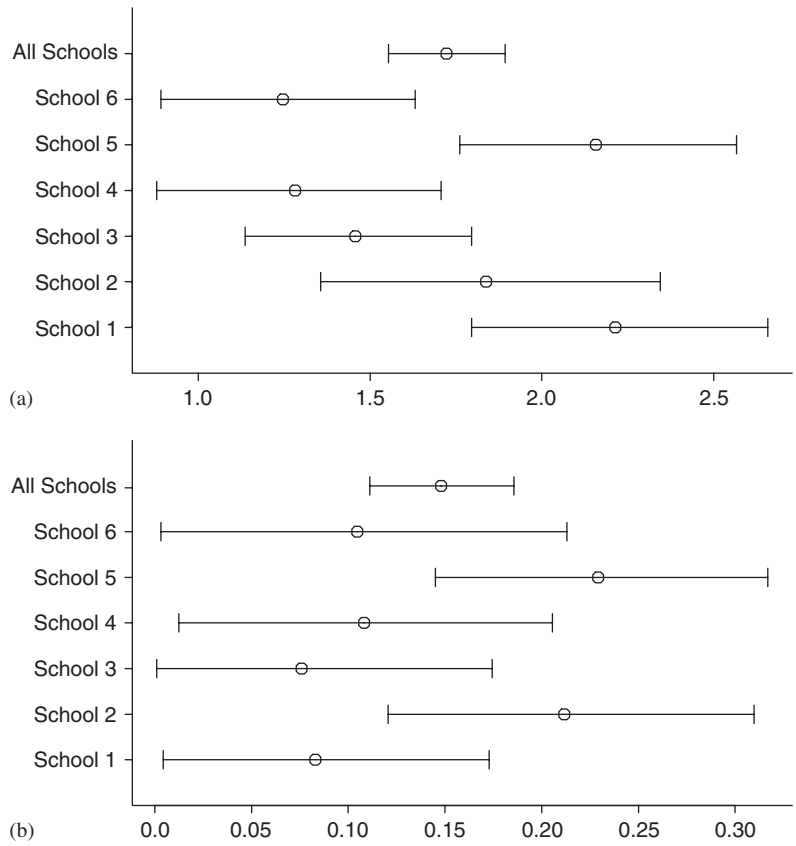

Figure 3. 95 per cent credible intervals of $\theta_1 - \theta_2$ and p for each school/treatment group using the ZPD distribution: (a) posterior credible intervals for $\theta_1 - \theta_2$; and (b) posterior credible intervals for p.

ZPD(p, θ_1, θ_2) are provided in the last column of Table II. In all sets of data, models m_1 and m_3 (with common θ's) were not visited at all. Results confirm our conclusions based on the posterior distributions of the mixing proportion p. Hence, for the aggregated data and schools 2 and 5, the posterior probability of m_4 is found equal to one, indicating strong evidence against both hypotheses tested (equal θ and $p = 0$). For school 6, model m_4 is supported with posterior probabilities equal to 0.78 (Bayes factor equal to 4.6) indicating positive evidence in favour of the hypothesis $p > 0$. For schools 1 and 6, model m_4 is slightly supported with posterior probabilities equal to 0.60 and 0.64 (Bayes factors 1.48 and 1.76), respectively, indicating low evidence against the hypothesis $p = 0$. Finally, for school 3 the hypothesis of zero mixing proportion is slightly supported since m_4 has posterior probability 0.42 and Bayes factor of m_4 *versus* m_3 is equal to 0.75. To complete our analysis, we rerun the RJMCMC algorithm after calibrating prior model probabilities so that all models are visited. This enables us to estimate with precision the logarithm of Bayes factors rather than the posterior probabilities and ensure that the algorithm works efficiently; results are provided in Table II. When running RJMCMC for the data from each school, we used the same proposals for

θ, θ_1 and θ_2 for both cases of $\xi = 0$ (PD) and $\xi = 1$ (ZPD) since their posterior distributions were close, while, for the aggregated data, we used different proposals for each value of ξ to increase the efficiency of the algorithm. The length of all RJMCMC runs was equal to 11 000 iterations discarding the first 1000 iterations as burn-in. In all examples we used priors (6) and (7) for θ, θ_1 and θ_2 and uniform prior for p. After performing sensitivity analysis on the data set of school 2, we found that results concerning the comparison between models with $p > 0$ *versus* models with $p = 0$ are robust to changes of $\vec{v}^* = \vec{u}^*$. Finally, we have used posterior model probabilities to produce predictive median counts and their 2.5 and 97.5 percentiles for each school which are depicted in Figure 4.

To sum up, the ZPD model provides useful information for the DMFT index difference. After the diagnostic excluding prior-to-treatment healthy cases, we have seen that most of the estimated mixing proportion can be attributed to 0–0 differences (healthy cases who remain healthy after the treatment). Hence, here, parameter p may be interpreted as a percentage excess of patients with stable oral hygiene. Large p in ZPD indicates positive evidence in favour of the preventive treatment used. For example, only in the control group (school 3) did we observe (minor) evidence against the excess of patients with stable oral hygiene. On the other hand, in schools 2 and 5 we observe strong evidence in favour of an increased excess of zero DMFT difference. In these groups we expected that the treatments would have been more efficient since a mouthwash with 0.2 per cent sodium fluoride (NaF) was used in school 5 and all four treatments were used in school 2.

6. DISCUSSION

In this paper we examined the distribution of the difference of two correlated Poisson variates and their zero-inflated counterpart. We presented in detail Bayesian estimation and hypothesis testing for the parameters of interest. The PD distribution allows us to test differences on paired count data by eliminating their correlation. Moreover, the ZPD model captures the excess of zeros, which frequently appear in medical data, and introduces over-dispersion on the marginal distributions of counts. Finally, the mixing proportion p in ZPD may be interpreted as an excess of patients with constant condition which, may offer important information concerning the efficiency of the treatment. The interpretation of p depends on the nature of the treatment (preventive or therapeutic). Both PD and ZPD distributions can be used efficiently for modelling the difference of discrete variables even if the original ones are not Poisson distributed. The Poisson data augmentation is used for the estimation procedure in the MCMC algorithm.

An important limitation of the proposed model is the assumption of non-negative correlation between the two measurements under consideration. Such an assumption is reasonable for pre- and post-treatment measurements. If negative correlation is present on our data, then the Bayes factor will support the Poisson model. Before proceeding to the analysis of such data using PD and ZPD distributions we propose to calculate correlation coefficients or use scatter plots to identify the existence of non-positive correlation between the measurements of interest.

Our proposed models can be directly generalized by constructing other distributions based on differences of discrete variables different from the Poisson used in this paper. For example, assume that parameter θ_1 follows a gamma distribution. Then the difference of such variables can be derived by a bivariate model with negative binomial and Poisson marginal distributions. Such model not only allows for modelling dependence between two count variables but also introduces over-dispersion in the marginal distributions.

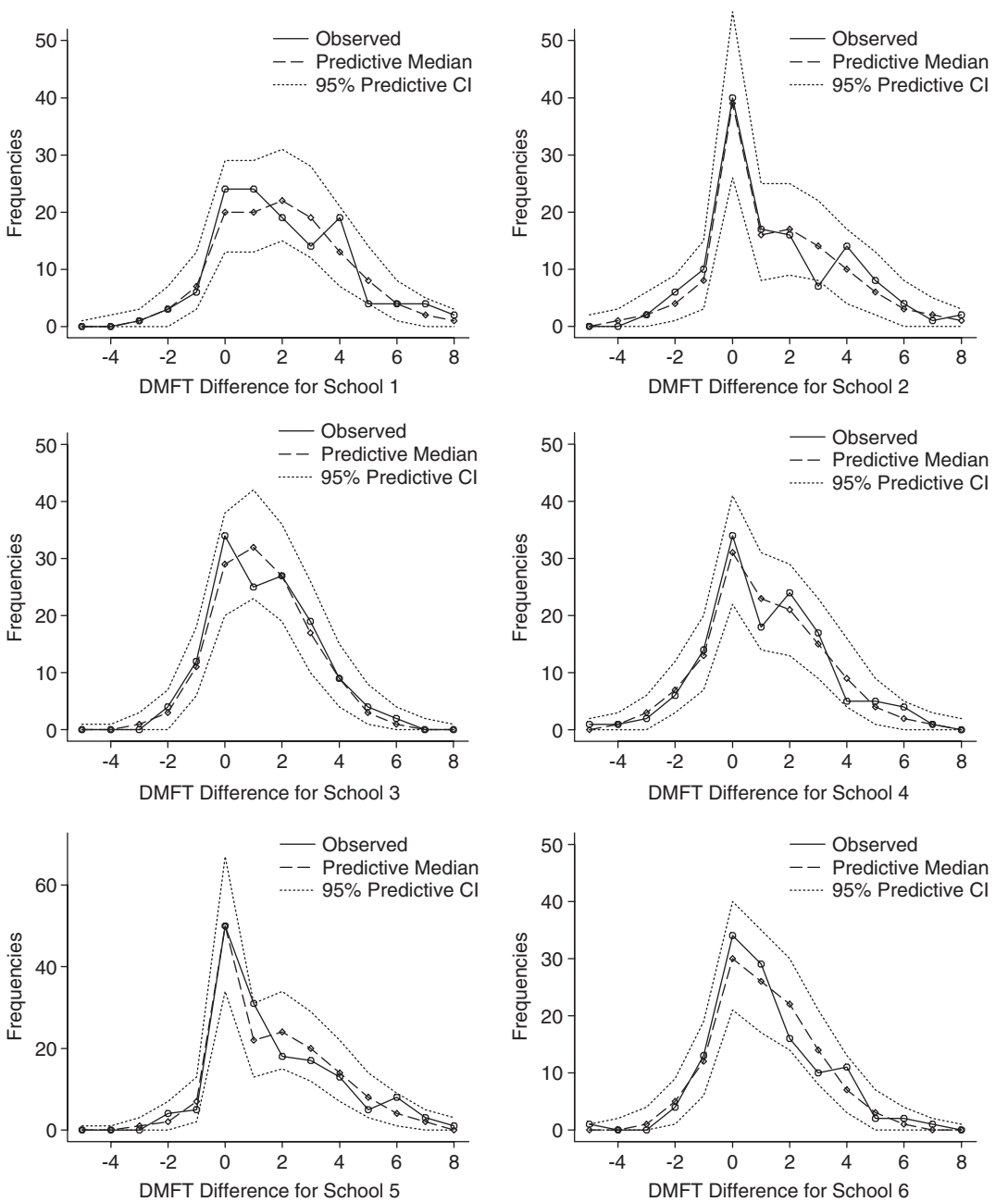

Figure 4. Comparison of observed and model averaged median predictive counts of DMFT difference ($DMFT_1 - DMFT_2$) for each school/treatment group (dotted lines represent 2.5 and 97.5 per cent quantiles of the model averaged predictive distribution of counts); for schools 2 and 5 the predictive values are based on the zero inflated distribution only, since $f(m_4 \mid \mathbf{z}) = 1$.

Finally, our current research involves extending the methodology for models based on PD and ZPD models using a general glm-type model with covariates on model parameters θ_1, θ_2 and possibly p. An interesting problem is to incorporate Bayesian variable selection techniques to identify well fitted models using the data augmentation approach presented in this paper. A further interesting problem is the comparison of θ-differences in the DMFT example using RJMCMC. This application is more complicated than the comparison presented in the paper and also involves consideration of multiple comparisons of groups using the Bayesian approach.

APPENDIX A: MCMC FOR THE ZERO-INFLATED POISSON DIFFERENCE DISTRIBUTION

1. Sample δ_i from Bernoulli($\tilde{p}$) with
$$\tilde{p} = f(\delta_i = 1 \mid p, \theta_1, \theta_2, \mathbf{z}) = \frac{\mathcal{I}(z_i = 0)p}{p + (1-p)f_{\mathrm{PD}}(0 \mid \theta_1, \theta_2)}$$
2. If $\delta_i = 0$ then sample (v_i, u_i) from $f(v_i, u_i \mid z_i = v_i - u_i, \theta_1, \theta_2) \propto \frac{\theta_1^{v_i}\theta_2^{u_i}}{v_i!u_i!}\mathcal{I}(z_i = v_i - u_i)$. If $\delta_i = 1$ we do not need to generate any latent data (v_i, u_i).
3. Sample θ_1 from $f(\theta_1 \mid p, \theta_2, \boldsymbol{\delta}, \mathbf{v}, \mathbf{u}) \sim \mathrm{Gamma}(n\bar{v} - \sum_{i=1}^{n}\delta_i v_i + a_1, n - n\bar{\delta} + b_1)$.
4. Sample θ_2 from $f(\theta_2 \mid p, \theta_1, \boldsymbol{\delta}, \mathbf{v}, \mathbf{u}) \sim \mathrm{Gamma}(n\bar{u} - \sum_{i=1}^{n}\delta_i u_i + a_2, n - n\bar{\delta} + b_2)$.
5. Sample p from $f(p \mid \theta_1, \theta_2, \boldsymbol{\delta}, \mathbf{v}, \mathbf{u}) \sim \mathrm{Beta}(n\bar{\delta} + a_3, n - n\bar{\delta} + b_3)$.

For model $m_3 : \mathrm{ZPD}(\theta, \theta, p)$, with prior $\theta \sim \mathrm{Gamma}(a, b)$, we generate θ from

$$f(\theta \mid p, \boldsymbol{\delta}, \mathbf{v}, \mathbf{u}) \sim \mathrm{Gamma}\left(n\bar{v} + n\bar{u} - \sum_{i=1}^{n}\delta_i(v_i + u_i) + a, 2n(1-\bar{\delta}) + b\right) \tag{A1}$$

instead of generating separate θ_1 and θ_2 in steps 3 and 4 above. In all other steps θ_1 and θ_2 are substituted by the common parameter θ.

APPENDIX B: RJMCMC FOR THE ZERO-INFLATED POISSON DIFFERENCE DISTRIBUTION

1. Generate γ using the following Metropolis step:
   (a) Propose $\gamma' = 1 - \gamma$ with probability one.
   (b) i. If $\gamma' = 1$ then propose θ_j' from $q(\theta_j' \mid \theta, \gamma, \xi)$ for $j = 1, 2$.
   ii. If $\gamma' = 0$ then propose common θ' from $q(\theta' \mid \theta_1, \theta_2, \gamma, \xi)$.
   (c) Accept the proposed move with probability $\alpha = \min\{1, \frac{O_1(\theta_1', \theta_2', \theta, p)^{1-\gamma}}{O_1(\theta_1, \theta_2, \theta', p)^{\gamma}}\}$ with
$$O_1(\theta_1', \theta_2', \theta, p) = \left\{\prod_{i=1}^{n}\frac{f_{\mathrm{ZPD}}(z_i \mid p, \theta_1', \theta_2')}{f_{\mathrm{ZPD}}(z_i \mid p, \theta, \theta)}\right\}\frac{f(\theta_1', \theta_2' \mid \gamma = 1, \xi)}{f(\theta \mid \gamma = 0, \xi)}\left\{\frac{f(p \mid \gamma = 1, \xi)}{f(p \mid \gamma = 0, \xi)}\right\}^{\xi}$$
$$\times \frac{f(\gamma = 1)}{f(\gamma = 0)} \times \frac{q(\theta \mid \theta_1', \theta_2', \gamma = 1, \xi)}{q(\theta_1', \theta_2' \mid \theta, \gamma = 0, \xi)} \tag{B1}$$

If $\xi = 0$ then, in the above equations, p is set equal to zero.

2. Generate ξ using following RJMCMC step:
   (a) Propose $\xi' = 1 - \xi$ with probability one.
   (b) i. If $\xi' = 1$ then propose p from $q(p \mid \gamma)$.
   ii. If $\xi' = 0$ then set $p = 0$.
   (c) Accept the proposed move with probability $\alpha = \min\{1, O_2(p, \theta_1, \theta_2)^{1-2\xi}\}$ with

$$O_2(p, \theta_1, \theta_2) = \left\{ \prod_{i=1}^{n} \frac{f_{\mathrm{ZPD}}(z_i \mid p, \theta_1, \theta_2)}{f_{\mathrm{PD}}(z_i \mid \theta_1, \theta_2)} \right\} \times \frac{f(p \mid \xi = 1)}{q(p \mid \gamma)} \times \frac{f(\xi = 1)}{f(\xi = 0)} \times \left(\frac{f(\theta_1, \theta_2 \mid \xi = 1, \gamma = 1)}{f(\theta_1, \theta_2 \mid \xi = 0, \gamma = 1)} \right)^{\gamma} \times \left(\frac{f(\theta \mid \xi = 1, \gamma = 0)}{f(\theta \mid \xi = 0, \gamma = 0)} \right)^{1-\gamma} \quad \text{(B2)}$$

   If $\gamma = 0$ then θ_1 and θ_2 of the above equation are substituted by θ.

3. If $\xi = 1$ then generate δ_i as in step 1 of the Appendix A; otherwise set $\delta_i = 0$.
4. Generate (v_i, u_i) from $f(v_i, u_i \mid z_i, \gamma, \xi, \delta_i)$. If $\xi = 1$ then follow step 2 of Appendix A; else follow step 1 of Section 2.3.
5. If $\gamma = 1$ generate θ_1, θ_2 as in steps 3 and 4 of Appendix A; else generate θ from (A1).

As a proposal $q(p \mid \gamma)$ in step 2 we consider a Beta$(\bar{a}, \bar{b})$ distribution with parameters calculated using pilot run estimates of model m_4 following Reference [14]. Here, we specify $\bar{a}$ and $\bar{b}$ by matching the moments of the proposal distribution with the posterior mean $\bar{p}$ and variance s_p^2 of the pilot run. Hence, $\bar{a}$ and $\bar{b}$ are given by $\bar{a} = \bar{p}\{\bar{p}(1 - \bar{p})/s_p^2 - 1\}$ and $\bar{b} = \bar{a}(1 - \bar{p})/\bar{p}$. Proposals of θ, θ_1 and θ_2 can be set according to Section 2.4. When the posterior distributions of θ, θ_1 and θ_2 for the ZPD and PD models are far away from each other, we advise to use different proposals depending on the status of ξ.

REFERENCES

1. Böhning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society*, *Series A* 1999; **162**:195–209.
2. Verbeke G, Spiessens B, Lesaffre E. Conditional linear mixed models. *American Statistician* 2001; **55**:25–34.
3. Skellam JG. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, *Series A* 1946; **109**:296.
4. Irwin W. The frequency distribution of the difference between two Poisson variates following the same Poisson distribution. *Journal of the Royal Statistical Society*, *Series A* 1937; **100**:415–416.
5. Kocherlakota S, Kocherlakota K. *Bivariate Discrete Distributions*. Marcel Dekker: New York, 1992.
6. Abramowitz M, Stegun IA. *Handbook of Mathematical Functions*. Dover: New York, 1974.
7. Johnson N, Kotz S, Kemp AW. *Univariate Discrete Distributions*. Wiley: New York, 1992.
8. Keilson J, Gerber H. Some results in discrete Unimodality. *Journal of the American Statistical Association* 1971; **66**:386–389.
9. Karlis D, Ntzoufras I. Distributions based on Poisson differences with applications in sports. *Technical Report 101*, Department of Statistics, Athens University of Economics, 2000.
10. Knuiman MW, Speed TP. Incorporating prior information into the analysis of contingency tables. *Biometrics* 1998; **44**:1061–1071.
11. Bernardo JM, Smith AFM. *Bayesian Theory*. Wiley: Chichester, 1992.
12. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association* 1995; **90**:773–795.
13. Green P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**:711–732.
14. Dellaportas P, Forster JJ, Ntzoufras I. On Bayesian model and variable selection using MCMC. *Statistics and Computing* 2002; **12**:27–36.

15. Brooks SP, Giudici P, Roberts GO. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions (with discussion). *Journal of the Royal Statistical Society*, *Series B* 2003; **65**:3–56 .
16. Chib S. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 1995; **90**:1313–1321.
17. Chib S, Jeliazkov I. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* 2001; **96**:270–281.
18. Lopes HF, West M. Bayesian model assessment in factor analysis. *Statistica Sinica* 2004; **14**:41–67.
19. Lindley DV. A statistical paradox. *Biometrika* 1957; **44**:187–192.
20. Bartlett MS. Comment on D. V. Lindley's statistical paradox. *Biometrika* 1957; **44**:533–534.
21. Chen MH, Ibrahim JG, Shao QM. Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference* 2000; **84**:121–137.
22. Roberts GO. Markov chain concepts related to sampling algorithms. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: New York, 1996.
23. Lambert D. Zero-inflated Poisson regression, with applications to defects in manufacturing. *Technometrics* 1992; **34**:1–14.

