

Bayesian hypothesis testing for the distribution of insurance claim counts using the Gibbs sampler

Athanassios Katsis^{a,*} and Ioannis Ntzoufras^b

^a*Department of Social & Education Policy, University of Peloponnese, Korinthos, Greece*

^b*Department of Statistics, Athens University of Economics & Business, Athens, Greece*

Abstract. We construct and present a Markov Chain Monte Carlo (MCMC) algorithm for the estimation of posterior odds and probabilities of alternative models used to evaluate competing hypotheses regarding three common discrete distributions involved in the modeling of the outstanding claim counts in actuarial science. The proposed methodology involves advanced statistical techniques of Bayesian modeling which make use of the Gibbs sampling variable selection algorithm. One of the main advantages of this approach over the popular reversible jump algorithm [12] is its straightforward implementation using the MCMC language tool of WINBUGS software [17]. The methodology is applied to a real data set. Directions regarding the implementation in WINBUGS are provided at the Appendix. It is worth noting that although the context of the problem is actuarial, the methodology can be applied to any field of science where the aim is the comparison or selection of discrete distributions of counts.

Keywords: Bayesian point of view, Gibbs sampling, hypothesis tests, Markov Chain Monte Carlo, reversible jump, WINBUGS

1. Introduction

Modeling random events has always generated a great deal of research interest in Science, Economics and Engineering. Relevant topics range from the level of rainfall and car traffic to market penetration of a certain commodity and the pricing of an option. However, the researcher is often uncertain about the appropriate statistical representation of the phenomenon under study. Initially, the most pressing statistical question she/he faces is

Which is the best statistical distribution to use?

which usually leads to the more specific one:

How do we evaluate the available ‘candidate’ distributions and choose the best one(s)?

*Corresponding author. E-mail: katsis@uop.gr.

To this end, Bayesian statistics seems like a natural approach since any prior beliefs about the unknown parameters may affect the final decision. Therefore, the researcher must incorporate these beliefs while conducting the analysis of the experiment. However, most of the Bayesian algorithms on the topic of Bayesian model comparison are so computationally intensive that effectively forbid most practitioners from applying them to their work. Hence, the need for a simpler and software-friendlier approach is apparent.

In actuarial science particularly, the choice of the distribution for the modeling and prediction of outstanding claims incurred in an insurance company has been extremely important to both practitioners and academics. It is common practice for these companies not to pay the outstanding claims immediately but with some delay. Hence, an accurate representation of the number of outstanding claims is of important practical significance. In addition to that, the Bayesian paradigm has become an integral research tool in the actuarial discipline (see [15] for a review). However, in the outstanding claims problem, the Bayesian approach has been mostly confined to parameter estimation (see [5,16,20]) rather than to testing hypotheses about distributions.

The contribution of this paper is twofold. Firstly, we develop a new algorithm for the estimation of posterior model odds based on the idea of Gibbs variable selection algorithm of Dellaportas et al. [6]. The algorithm is used to evaluate and compare three discrete statistical distributions for the modeling of the outstanding claim counts in actuarial science using the Bayesian approach. The proposed methodology is computationally simpler and can be generalized to a larger number of candidate distributions as well as to any other similar research problem in another field (e.g. compare the distributions of arrivals in a queue). Secondly, we implement our algorithm on the popular, freely available, software WINBUGS [17]. Competing methods such as RJMCMC can not be implemented via WINBUGS. This is an important advantage since many researchers favoring the Bayesian approach are familiar with WINBUGS. Hence, they can now implement our proposed methodology directly following the detailed description provided at the Appendix of this paper.

The paper is organized into five further sections. Section 2 reviews the basic formulation of Bayesian model comparison. A description of the distributions for the modeling of the outstanding claims is provided in Section 3. In Section 4 we describe in detail the Gibbs sampling algorithm constructed for our hypothesis tests and its advantages over the existing techniques. Section 5 analyzes the implementation of the method in a specific example of insurance claims data using WINBUGS. Finally, conclusions and closing remarks are presented in Section 6.

2. Bayesian inference

Generally, Bayesian inference is based on constructing a model m , its likelihood $f(\mathbf{y}|\boldsymbol{\theta}_m, m)$ and the corresponding prior distribution $f(\boldsymbol{\theta}_m|m)$, where $\boldsymbol{\theta}_m$ is a parameter vector and \mathbf{y} is the data vector. Although, inference is primarily based on the posterior distribution $f(\boldsymbol{\theta}_m|\mathbf{y}, m)$, we may also be interested in quantifying model uncertainty by estimating the posterior model probability $f(m|\mathbf{y})$.

Let us consider two competing models m_0 and m_1 . If $f(m)$ is the prior probability of model m , then, using the Bayes theorem, the posterior odds PO_{01} of model m_0 versus model m_1 are given by

$$PO_{01} = \frac{f(m_0|\mathbf{y})}{f(m_1|\mathbf{y})} = \frac{f(\mathbf{y}|m_0)}{f(\mathbf{y}|m_1)} \times \frac{f(m_0)}{f(m_1)} = B_{01} \times \frac{f(m_0)}{f(m_1)} \quad (1)$$

where B_{01} and $\frac{f(m_0)}{f(m_1)}$ are the ‘Bayes factor’ and the ‘prior model odds’ of model m_0 against model m_1 , respectively. The quantity $f(\mathbf{y}|m)$ involved in the Bayes Factor is defined as the marginal likelihood of

model m and is given by $f(\mathbf{y}|m) = \int f(\mathbf{y}|\boldsymbol{\theta}_m, m)f(\boldsymbol{\theta}_m|m)d\boldsymbol{\theta}_m$. The Bayes factor B_{10} , of model m_1 against m_0 , evaluates the evidence *against* the null hypothesis which is a familiar concept to classical significance tests. Thus, large values of, say, B_{01} (usually greater than 12) indicate strong posterior support of model m_0 against model m_1 . For more details on Bayes factors see Kass and Raftery [13]. Alternatively, when we consider a set of competing models $\mathcal{M} = \{m_1, m_2, \dots, m_{|\mathcal{M}|}\}$, then we focus our interest on the posterior probability of model $m \in \mathcal{M}$, defined as

$$f(m|\mathbf{y}) = \frac{f(\mathbf{y}|m)f(m)}{\sum_{m_l \in \mathcal{M}} f(\mathbf{y}|m_l)f(m_l)} = \left(\sum_{m_l \in \mathcal{M}} PO_{m_l, m} \right)^{-1}$$

where \mathcal{M} and $|\mathcal{M}|$ denote the set and the number of models under consideration respectively.

The integrals involved in the computation of the posterior model probabilities are mostly analytically intractable. Therefore asymptotic approximations or alternative computational methods must be frequently employed. One of the most popular techniques for calculation of these quantities is the Markov Chain Monte Carlo (MCMC) methodology (see [11]) and its recent extensions (reversible jump MCMC algorithm or RJMCMC) in varying dimension models [12]. Moreover, RJMCMC methodology helps us to account for model uncertainty using Bayesian model averaging techniques (see [1,3,9,13]). However, this methodology is demanding in both the design stage and the implementation.

3. Distributions for claim counts

In this section we focus on three popular distributions for modeling the marginal claim counts, more specifically the simple Poisson distribution [18], the negative binomial [21] and the Generalized Poisson distribution (see [4]) denoted by m_1, m_2 and m_3 respectively. The Generalized Poisson distribution is also known as Lagrangian Poisson distribution. Consequently, the simple Poisson model can be regarded as a special case of either the negative binomial or the Generalized Poisson distribution.

Let us assume data $y_i, i = 1, \dots, n$. Consequently, the simple Poisson model is given by

$$f(y_i|\lambda, m_1) = \frac{\lambda^{y_i} \exp(-\lambda)}{y_i!}.$$

It is well known that for the Poisson distribution the mean equals the variance. This property is not common in real data sets where the sample variance usually exceeds the sample mean. For this reason alternative models that allow for over-dispersion relative to the simple Poisson model have been considered. The variance to the mean ratio, called Dispersion Index (DI), is usually calculated as a measure for data dispersion; see for example Douglas [8]. For the Poisson random variable the Dispersion Index is equal to one.

An alternative model for this type of data is the negative binomial distribution given by

$$f(y_i|\lambda, \vartheta, m_2) = \frac{\Gamma(y_i + \vartheta)}{\Gamma(y_i + 1)\Gamma(\vartheta)} \left(\frac{\lambda}{\lambda + \vartheta} \right)^{y_i} \left(\frac{\vartheta}{\lambda + \vartheta} \right)^\vartheta \tag{2}$$

where $\vartheta > 0$. Although the mean of the negative binomial model is the same as in the Poisson model, $E(y_i) = \lambda$, the variance now also depends on the parameter ϑ since $V(y_i) = \lambda + \lambda^2/\vartheta$. The Poisson model is a limiting distribution of Eq. (2) for $\vartheta \rightarrow \infty$. We may adopt the parameterization $\vartheta = \lambda/\phi$. For

the negative binomial distribution, the Dispersion Index is equal to $DI = 1 + \phi$. For $\phi \rightarrow 0$ the above distribution reduces to the simple Poisson distribution.

Finally, the Generalized Poisson model with parameters λ and ω , is defined in the following way:

$$f(y_i|\lambda, \omega, m_3) = (1 - \omega)\lambda_i \frac{\{(1 - \omega)\lambda_i + \omega y_i\}^{y_i - 1}}{y_i!} e^{-\{(1 - \omega)\lambda_i + \omega y_i\}}. \quad (3)$$

According to Ter Berg [19], valid values for ω are within the interval $[0, 1)$. Typically, the distribution can be defined for $|\omega| < 1$ but negative values lead to under-dispersion which is not a common property of claim count data. For this reason we will not consider this case in the present paper. For $\omega = 0$, the above distribution reduces also to the simple Poisson model with mean λ . The mean of y_i is given by $E(y_i) = \lambda$ while the variance and the dispersion index are obtained by $V(y_i) = \lambda(1 - \omega)^{-2}$ and $DI = (1 - \omega)^{-2}$ respectively. This parameterization is beneficial for the interpretation of the parameters of the three models and also simplifies the implementation of the MCMC algorithm described below.

4. Gibbs sampling for testing hypotheses

Gibbs sampling has been widely used for Bayesian model, variable selection and hypothesis tests; see George and McCulloch [10], Carlin and Chib [2], Kuo and Mallick [14] and Dellaportas et al. [6]. The Gibbs sampling approach presented here is mainly based on the approach of Gibbs variable selection of Dellaportas et al. [6] and although some concepts are similar to the approach of Carlin and Chib [2], it is more flexible since it allows common parameters among different models avoiding over-parameterization and generation of nuisance parameters.

Without loss of generality, we describe the methodology for a general setup of comparing two nested models m_1 and m_0 . We wish to test that a parameter sub-vector of the super-model m_1 is constrained to a sub-model m_0 . Incorporation of more models (or even non-nested models) is possible in a similar manner.

Let us denote the parameter vectors of models m_0 and m_1 by θ_{m_0} and θ_{m_1} respectively. Then we have the parameter vector $\theta_{m_1}^T = (\theta_{m_0}^T, \theta_{\setminus m_0}^T)$; where $\theta_{\setminus m_0}$ denotes the parameters of model m_1 that do not appear in the sub-model m_0 . We should further note that common parameters should have similar interpretation otherwise posterior distributions will be different and the algorithm will fail (for example in our distributions the parameter λ has exactly the same interpretation since it is the mean of y_i in all models).

The hypothesis we wish to test takes the form $H_0 : \theta_{\setminus m_0} = \mu_0$ (model m_0) versus $H_1 : \theta_{\setminus m_0} \neq \mu_0$ (model m_1). Model m_0 is exactly the same (in terms of likelihood) with model m_1 with parameter vector $(\theta_{m_0}^T, \theta_{\setminus m_0}^T = \mu_0^T)$.

In order to construct our algorithm we employ a latent binary indicator γ taking values zero and one when supporting the null or the alternative hypothesis. The posterior distribution of this indicator will give us the posterior probability of each model (or hypothesis) and/or the posterior odds for each model (or hypothesis) comparison. When no prior information concerning the prior model weights $f(m_i)$ is provided then we use $f(\gamma) = 1/2$ for $\gamma = 0, 1$.

For $\gamma = 0, 1$ the model likelihood is now rewritten as

$$f(\mathbf{y}|\theta_{m_\gamma}, m_\gamma) = f(\mathbf{y}|\theta_{m_0}, m_0)^{1-\gamma} f(\mathbf{y}|\theta_{m_1}, m_1)^\gamma$$

while the prior distributions are given by $f(\boldsymbol{\theta}_{m_i}, m_\gamma, \gamma)$ for $i = 0, 1$. If $\gamma = i$ then we have the usual prior distributions

$$f(\boldsymbol{\theta}_{m_\gamma}, m_\gamma, \gamma) = f(\boldsymbol{\theta}_{m_\gamma} | m_\gamma) f(\gamma).$$

If $\gamma \neq i$, the resulting distributions are called pseudo-prior distributions since they do not affect the posterior distributions; for details see Carlin and Chib [2] and Dellaportas et al. [6]. More specifically,

$$\begin{aligned} f(\boldsymbol{\theta}_{m_1}, m_0, \gamma = 0) &= f(\boldsymbol{\theta}_{m_1} | m_0) f(\gamma = 0) \\ &= f(\boldsymbol{\theta}_{m_0}, \boldsymbol{\theta}_{\setminus m_0} | m_0) f(\gamma = 0) \\ &= f(\boldsymbol{\theta}_{m_0} | m_0) f(\boldsymbol{\theta}_{\setminus m_0} | \boldsymbol{\theta}_{m_0}, m_0) f(\gamma = 0). \end{aligned} \tag{4}$$

Thus, the pseudo-prior distribution in Eq. (4) can be rewritten as a product of an actual prior distribution of simpler model m_0 given by $f(\boldsymbol{\theta}_{m_0} | m_0)$ and an additional pseudo-prior distribution $f(\boldsymbol{\theta}_{\setminus m_0} | \boldsymbol{\theta}_{m_0}, m_0)$ for the non-common parameters. This distribution is specified via a pilot run of model m_1 . Similarly, the pseudo-prior

$$f(\boldsymbol{\theta}_{m_0}, m_1, \gamma = 1) = f(\boldsymbol{\theta}_{m_0} | m_1) f(\gamma = 1)$$

is the prior of model m_1 for the common parameters $\boldsymbol{\theta}_{m_0}$. Using this setup, the resulting Gibbs Sampler is given by

1. Sample model parameters:

– If $\gamma = 0$ then

(a) Sample $\boldsymbol{\beta}_{m_0}$ from the conditional posterior distribution

$$f(\boldsymbol{\beta}_{m_0} | m_0, \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}_{m_0}, m_0) f(\boldsymbol{\beta}_{m_0} | m_0).$$

(b) Sample $\boldsymbol{\beta}_{\setminus m_0}$ from the pseudo-prior distribution $f(\boldsymbol{\beta}_{\setminus m_0} | \boldsymbol{\beta}_{m_0}, m_1)$.

– If $\gamma = 1$ then sample $\boldsymbol{\beta}_{m_1} = (\boldsymbol{\beta}_{m_0}^T, \boldsymbol{\beta}_{\setminus m_0}^T)$ from the conditional posterior distribution

$$f(\boldsymbol{\beta}_{m_1} | m_1, \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}_{m_1}, m_1) f(\boldsymbol{\beta}_{m_1} | m_1).$$

2. Sample γ from the posterior distribution $f(\gamma | \boldsymbol{\theta}_{m_0}, \boldsymbol{\theta}_{\setminus m_0}, \mathbf{y}) = \text{Bernoulli}\left(\frac{\Omega}{1+\Omega}\right)$; where Ω is given by

$$\Omega = LR \times PR_0 \times PR_{\setminus 0} \times \frac{f(m_1)}{f(m_0)}$$

where LR, PR_0 and $PR_{\setminus 0}$ denote the likelihood ratio, and the prior ratios given by

$$LR = \frac{f(\mathbf{y} | \boldsymbol{\theta}_{m_1}, m_1)}{f(\mathbf{y} | \boldsymbol{\theta}_{m_0}, m_0)}, \quad PR_0 = \frac{f(\boldsymbol{\theta}_{m_0} | m_1)}{f(\boldsymbol{\theta}_{m_0} | m_0)} \quad \text{and} \quad PR_{\setminus 0} = \frac{f(\boldsymbol{\theta}_{\setminus m_0} | \boldsymbol{\theta}_{m_0}, m_1)}{f(\boldsymbol{\theta}_{\setminus m_0} | \boldsymbol{\theta}_{m_0}, m_0)}.$$

Note that common parameters are assumed to have the same interpretation hence the prior distributions for the two models can be set equal resulting to $PR_0 = 1$.

The approach described above can be easily extended for non-nested models say, m_1 and m_2 much like our comparison between the negative binomial and generalized Poisson models. Let us assume a model m_0 with parameter vector $\boldsymbol{\theta}_{m_0}$ including all parameters common in both model m_1 and m_2 (if

no common parameters can be identified, then m_0 can be set to the null model with no parameters). Then we may rewrite the parameters $\boldsymbol{\theta}_{m_i}^T = (\boldsymbol{\theta}_{m_0}^T, \boldsymbol{\theta}_{m_i/m_0}^T)$ for $i = 1, 2$ and follow the same approach as in the previous section. In limiting cases, we may identify links between parameters and use suitable transformations in order to automatically specify the pseudo-priors and avoid having to use pilot-run estimates (the approach is analogous to setting suitable transformations in reversible jump MCMC algorithm). Such a case is here where $\boldsymbol{\theta}_{m_1} = (\lambda, \vartheta)$ and $\boldsymbol{\theta}_{m_2} = (\lambda, \omega)$. The non-common parameters ϑ and ω can be efficiently linked by equating the dispersion indexes of the two models.

Unlike reversible jump MCMC, the proposed algorithm is simpler because it is based on Gibbs sampler rather than Metropolis Hastings algorithm. Since it is a Gibbs sampling based algorithm it can be implemented in a straightforward manner using the freely available MCMC software of WINBUGS. It is flexible enough to handle nested models or models with common and non-common parameters. For this reason, it can be used in a wide variety of similar problems and can be easily extended to accommodate a larger number of distributions. An application of the aforementioned methodology is presented in Section 5.

5. Implementation in insurance claims data

In this section we demonstrate the proposed model formulation in the insurance claim data of Belgium for the year 1993 [7]. The determination of the prior distributions, algorithm specifications and the results are presented below. Details for the implementation on the WINBUGS software are given at the Appendix. The full code is available from the authors upon request.

5.1. Prior distributions

The specification of the prior distributions is very important in Bayesian model comparison. As we have already mentioned, we examine three models: m_1 (Poisson), m_2 (negative binomial) and m_3 (generalized Poisson) with parameter vectors $\boldsymbol{\theta}_{m_1} = (\lambda)$, $\boldsymbol{\theta}_{m_2} = (\lambda, \vartheta)^T$ and $\boldsymbol{\theta}_{m_3} = (\lambda, \omega)^T$ respectively.

In order to be consistent across models we must specify the same prior distributions over the common parameter λ . Hence, we consider a Gamma distribution, that is,

$$f(\lambda|m_i) = \text{Gamma}(0.01, 0.01) \text{ for } i = 1, 2, 3.$$

The effect of this choice on the model comparisons will be minimal since λ is a common parameter in all models.

The prior distributions $f(\vartheta|\lambda, m_2)$ and $f(\omega|\lambda, m_3)$ will be determined in a way such that the Dispersion Index will induce the same a priori information for both models. Hence, for the prior of ω we use the Uniform non-informative prior distribution which gives the equal probability to any interval of the same range, that is,

$$f(\omega|m_3) = \text{Uniform}(0, 1)$$

while the prior $f(\vartheta|m_2)$ is constructed by setting the dispersion indexes equal for the two distributions resulting to

$$\vartheta = \lambda \frac{(1 - \omega)^2}{\omega(2 - \omega)}. \quad (5)$$

Assuming uniform prior for ω , the resulting distribution for ϑ is a scaled Beta type II prior distribution given by

$$f(\vartheta|\lambda, m_2) = \frac{1}{2}\lambda\vartheta^{-2}(1 + \lambda/\vartheta)^{-3/2}$$

5.2. Algorithm specifications: the pseudo-prior distributions

The pseudo-prior distributions are defined by pilot runs estimates. Hence for model m_3 we use a Beta distribution, that is,

$$f(\omega|m_1) = \text{Beta}(\bar{a}, \bar{b})$$

where the parameters \bar{a}, \bar{b} of the Beta distribution are matched with the posterior mean and variance of a pilot run of model m_2 . Hence \bar{a} and \bar{b} are obtained by

$$\bar{\omega} = \frac{\bar{a}}{\bar{a} + \bar{b}}, \quad \bar{\sigma}_\omega^2 = \frac{\bar{a}\bar{b}}{(\bar{a} + \bar{b})^2(\bar{a} + \bar{b} + 1)}$$

which leads to

$$\bar{a} = \bar{\omega} \left(\frac{\bar{\omega}(1 - \bar{\omega})}{\bar{\sigma}_\omega^2} - 1 \right), \quad \bar{b} = \bar{a} \frac{1 - \bar{\omega}}{\bar{\omega}}.$$

where $\bar{\omega}$ and $\bar{\sigma}^2$ are the posterior mean and variance of ω estimated by a pilot run. The efficiency of the chain can be improved by suitably increasing or decreasing the variance of the pseudo-prior distribution in order to achieve high acceptance rates. In our example the pseudo-priors were estimated from a pilot run of 1000 iterations and were found $\bar{\omega} = 0.039$ and $\bar{\sigma} = 0.0036$ resulting to $f(\omega|m_1) = \text{Beta}(112.7, 2778.1)$.

The pseudo-prior $f(\vartheta|m_1)$ is defined indirectly by specifying a pseudo-prior for ω and using Eq. (5). When comparing negative binomial and generalized Poisson models then pseudo-priors are indirectly defined using Eq. (5). This results to pseudo-priors which take values directly from the posterior distribution of the competing model and transforming it appropriately.

Finally, if we wish to estimate the log-Bayes posterior odds or the Bayes factor with increased precision, we may use initial model probabilities $f^{(0)}(m_i)$ which will result to posterior weights in the interval (0.40–0.60) and then recalculate the desired posterior model odds using the following equation:

$$\log PO_{ij} = \log PO_{ij}^{(0)} - \log \left(\frac{f^{(0)}(m_i)}{f^{(0)}(m_j)} \right) + \log \left(\frac{f(m_i)}{f(m_j)} \right) \tag{6}$$

where $f^{(0)}(m_i)$ are the initial model probabilities only used to estimate posterior odds with precision, $PO_{ij}^{(0)}$ are the posterior model odds estimated using the initial model probabilities while $f(m_i)$ and $\log PO_{ij}$ are the desired prior probabilities and posterior model odds respectively. When there is no prior information concerning the model space, the log Bayes factor is estimated by the above equation eliminating the log-ratio of the actual prior model probabilities which is equal zero. In our example we have used $\log f^{(0)}(m_2)/f^{(0)}(m_1) = 81$, $\log f^{(0)}(m_3)/f^{(0)}(m_1) = 82$ and $\log f^{(0)}(m_3)/f^{(0)}(m_2) = 0$. Initial values of $f^{(0)}(m)$ can be obtained using simple estimators of the marginal likelihood (for example using Laplace approximation; for details see [13]).

Table 1

Posterior Odds and Probabilities for the Belgium 1993 dataset; 1 = Poisson, 2 = Negative Binomial, 3 = Generalized Poisson Model; [$PrM_{ji}^{(0)} = f^{(0)}(m_j)/f^{(0)}(m_i)$, $f^{(0)}(m_i|y)$ posterior weights obtained from MCMC using $PrM_{ji}^{(0)}$]

Compared Models	i, j	MCMC Results			Estimates
		$\log PrM_{ji}^{(0)}$	$f^{(0)}(m_i y)$	$\log PO_{ij}^{(0)}$	$\log B_{ij}$
Neg.Bin. vs. Poisson	2,1	81	0.610	0.447	81.447
Gen.Poisson vs. Poisson	3,1	82	0.533	0.134	82.134
Gen.Poisson vs. Neg.Bin.	3,2	0	0.671	0.714	0.714

5.3. Results

Results for the Belgium 1993 dataset are provided in Table 1. Trace, density and probability plots of the dispersion index and the model indicator for each pairwise comparison are also provided in Figs 1–3. All results have been generated using 1,000 iterations as an initial burn-in period while keeping additional 10,000 iterations for the estimation of the posterior distribution.

Regarding Table 1, the first two columns provide details of the compared models. In the third column we present the initial model odds used to make the chain mobile (i.e. ‘jump’ from one model to the other) while the next two columns display the MCMC results directly from the WINBUGS output. The last column presents the final estimate of the logarithm of the Bayes Factor for each model comparison using Eq. (6), the initial model odds (third column) and the MCMC results. For example, the second line of the Table depicts the comparison of the Generalised Poisson versus the Poisson model (model m_3 vs. model m_2). The initial model odds was set equal to 82 in favor of the Poisson model while the Gibbs sampling algorithm yielded the posterior model probability for the Generalised Poisson model equal to 0.533. Furthermore, the logarithm of the posterior model odds, $\log PO_{31}^{(0)}$, is equal 0.134. Using Eq. (6), we derive the final estimate for the logarithm of the posterior model odds equal to 82.134 which strongly supports the Generalised Poisson model.

Figures 1–3 are produced directly from WINBUGS software and are provided to give some insight for the results and the convergence of the algorithm. In all figures, plot (a) depicts the marginal posterior distribution of the Dispersion index for each comparison using the corresponding initial model odds. In Figs 1 and 2 we clearly see a spike at the value of one. This is natural since the Poisson distribution is compared with the Negative Binomial and Generalised Poisson, respectively. When the chain supports the Poisson model (in both cases) then the produced DI is equal to one. In most cases, when two models are compared using MCMC, the marginal posterior distribution of a parameter of interest will produce two modes (one for each model) unless the parameters have similar behavior in both models (this is the case in Fig. 3(a) or one model is not supported at all (i.e. has low posterior probability). Figs 1(b)–3(b) present a graph of the posterior distribution of γ , $f(\gamma|y)$, that is the posterior model weights for each comparison (also provided in the fourth column of Table 1). Figures 1(c)–3(c) are Trace plots of the DI for each comparison. In common MCMC (when only parameter estimation in a single model is considered) we use such graphs to monitor the convergence of the chain. Plots similar to Fig. 3(c) indicate convergence. In MCMC constructed for model comparison, it is natural to expect violent jumps from the posterior distribution of one model to the other as in plots Figs 1(c) and 2(c). Such jumps are not observed in the comparison of Generalised Poisson and Negative Binomial model because in both models the distributions of the dispersion index are quite close (as plot Fig. 3(a) also indicated). Finally, Figs 1(d)–3(d) are also trace plots of the model indicator γ . These plots are used to monitor the convergence of γ . A large number of jumps (changes from one model to the other) indicates that the

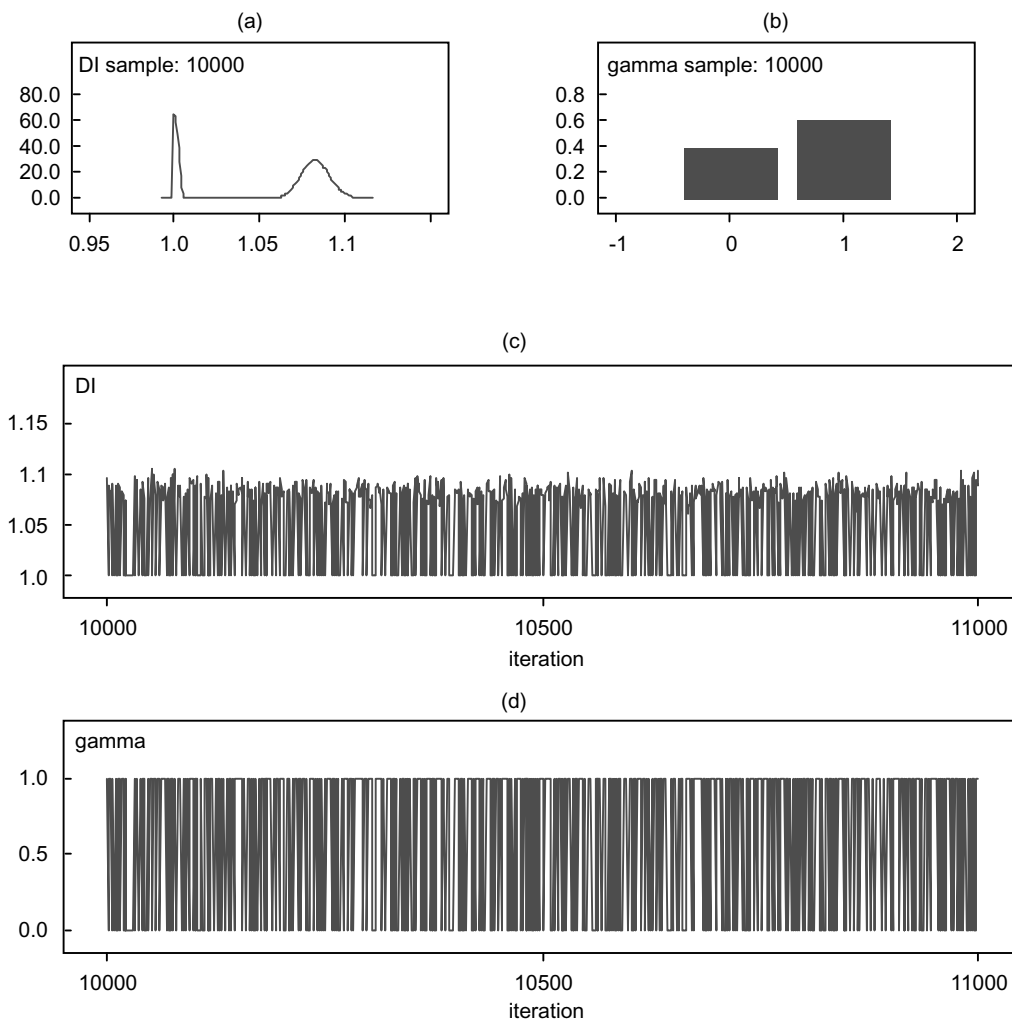


Fig. 1. Plots from MCMC Output for Comparison of Negative Binomial vs. Poisson model: (a) Histogram of the Marginal Posterior Distribution of DI; (b) Posterior Model Weights (0 = Poisson, 1 = Neg. Binomial); (c) Trace Plot of DI; (d) Trace Plot of Model Indicator γ .

algorithm works efficiently. When the chain remains in one model for many iterations then the graph presents a gap of white area. In all comparisons, the chains were highly mobile and this is also depicted in the corresponding graphs with few short white sequences.

As a conclusion, we observe that the data strongly support the negative binomial and the generalized Poisson models in favor of the simple Poisson model (log-Bayes factors equal to 81.45 and 82.13 respectively). For the comparison between generalized Poisson and negative binomial models we may calculate the corresponding log-Bayes factor directly by the difference of the above log-Bayes factors (equal to 0.69 in favor of the generalized Poisson model) or by generating results from an MCMC directly sampling from these two models (log Bayes factor equal to 0.71). This leads to a slight advantage of the Generalized Poisson Model over the Negative binomial one ($B_{32} \approx 2$ and posterior probability $f(m_3|\mathbf{y}) = 0.671$).

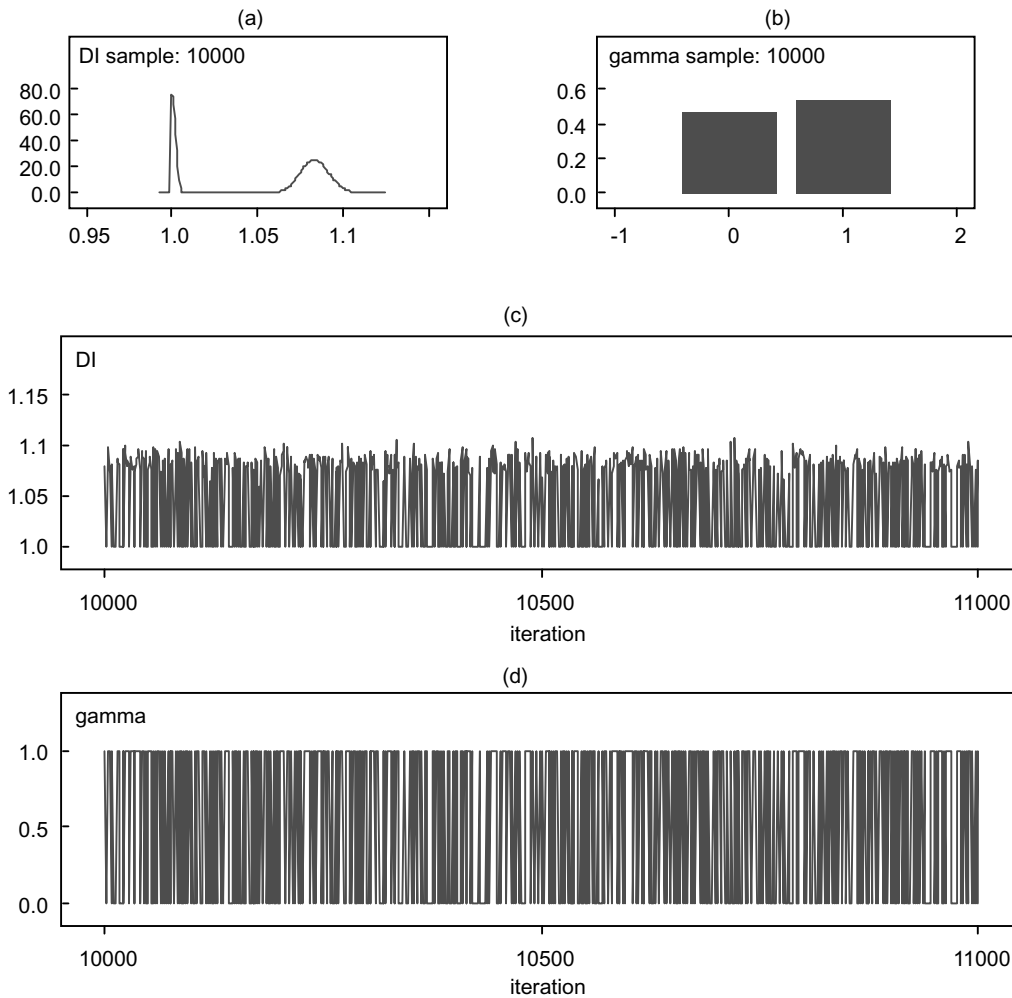


Fig. 2. Plots from MCMC Output for Comparison of Generalized Poisson vs. Poisson model: (a) Histogram of the Marginal Posterior Distribution of DI; (b) Posterior Model Weights (0 = Poisson, 1 = Gen.Poisson); (c) Trace Plot of DI; (d) Trace Plot of Model Indicator γ .

6. Conclusions

In this paper, we have developed an advanced MCMC algorithm based on the idea of Gibbs variable selection in order to compare three statistical distributions that model the marginal claim counts in actuarial science. The proposed algorithm is simpler than the existing ones such as RJMCMC and can be implemented in the software-friendly environment of WINBUGS. Furthermore, it can be extended to a wide variety of applications that include model comparisons. A larger number of distributions may be examined as well.

The results from our case study demonstrate the superiority of the Negative Binomial and the Generalized Lagrangian Poisson distributions over the simpler Poisson. Between the first two distributions, the Generalized Lagrangian distribution seems to be slightly supported a-posteriori. Further research may include covariates on λ in order to treat more sophisticated cases such as the prediction of outstanding

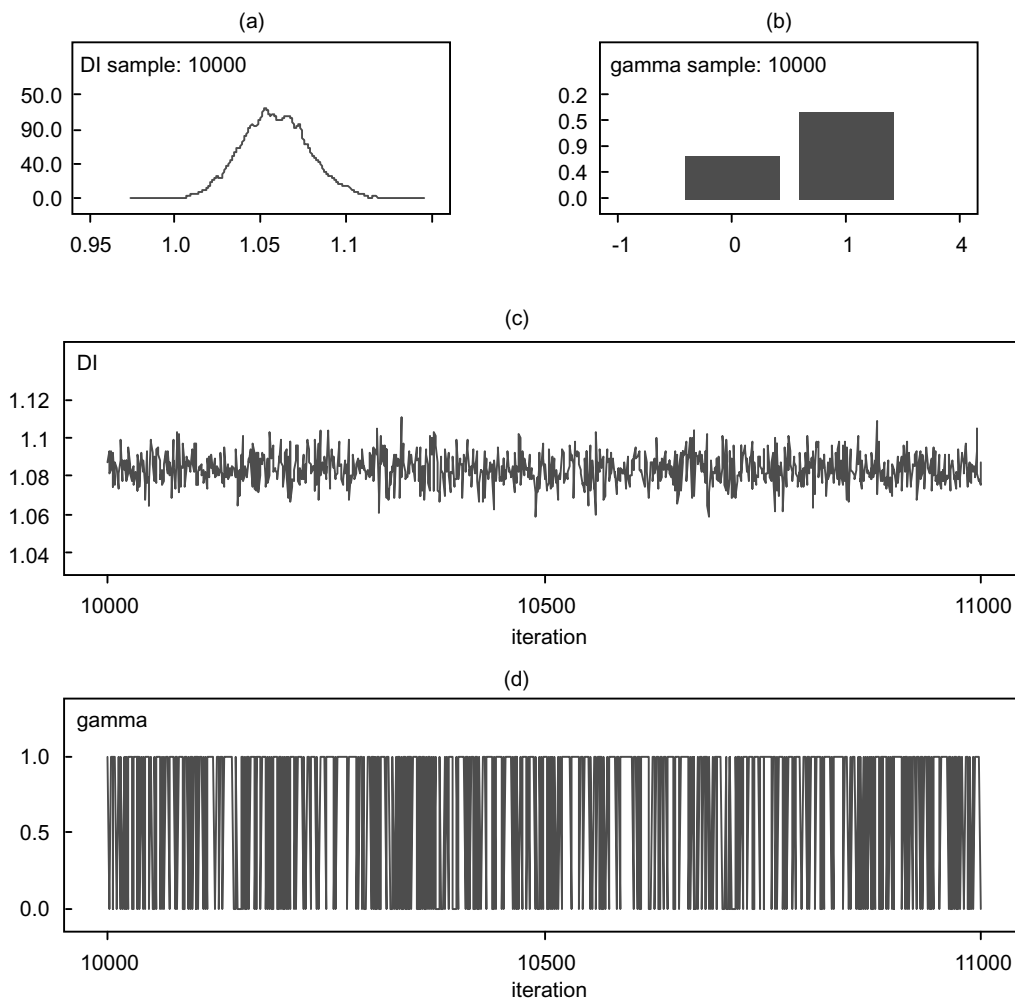


Fig. 3. Plots from MCMC Output for Comparison of Generalized Poisson vs. Negative Binomial model: (a) Histogram of the Marginal Posterior Distribution of DI; (b) Posterior Model Weights (0 = Neg. Binomial, 1 = Gen. Poisson); (c) Trace Plot of DI; (d) Trace Plot of Model Indicator γ .

claim reserves. Another important issue is the possible extension of our proposed methodology to a larger variety of claim distributions. Using this approach, we may compare different models involved in the claim counts and amounts literature. Implementing Gibbs sampling in a wider variety of actuarial models and problems will enable us to use Bayesian model averaging techniques which increase the predictive ability of any quantity of interest.

Appendix: Implementation Using WINBUGS

Data: Data are in the form of two vectors: $y[]$ for the values of y and $w[]$ the frequency of each value of y .

Likelihood: The likelihood is defined using the method of zeros and ones (see [17], in section: Tricks:

Advanced Use of the BUGS Language). This allows us to use any form likelihood and does not restrict us in the limited number of distributions available in WINBUGS. If K is the number of observed values of y then the likelihood is defined as:

```
for (i in 1:K) {
ones[i]<-w[i];
ones[i]~dbin( p[i], w[i] );
p[i]<-gamma*exp(loglikel[i])+ (1-gamma)*exp(loglike0[i])
}
```

where $\text{likelihood1}[i]$ and $\text{likelihood0}[i]$ are the densities of the distribution models m_1 and m_0 , respectively, evaluated at y_i . In our examples $\text{likelihood0}[i]$ is set equal to the Poisson log-density while $\text{likelihood1}[i]$ is set equal to the negative binomial or generalized Poisson log-density depending on the comparison we wish to implement. The WINBUGS code for setting the distribution used in this paper follows:

```
# likelihood of poisson
loglike0[i] <- -lambda + y[i] * log( lambda ) - logfact( y[i] );
#likelihood of negative binomial
loglikel[i] <- loggam( y[i] + theta ) - loggam( y[i]+1 )
               - loggam(theta)
               + y[i]*log( lambda ) + theta* log( theta )
               -(y[i]+theta)*log(lambda+theta);
#likelihood of generalized poisson
loglikel[i] <- log(1-omega)+log(lambda)
               +(y[i]-1)*log( (1-omega)*lambda+omega*y[i] )
               -( (1-omega)*lambda+omega*y[i] ) -logfact(y[i]) ;
```

Prior Distribution of λ : The prior distribution on λ is common for all models. Since no information is available we use a $\text{Gamma}(0.01, 0.01)$ prior distribution defined in WINBUGS by

```
lambda~dgamma(0.01,0.01);
```

Prior and Pseudo-prior Distributions on ϑ : Here we set ϑ indirectly as a function of ω which follows $\text{Uniform}(0, 1)$ – or equivalently $\text{Beta}(1, 1)$ - prior distribution when $\gamma = 1$. When $\gamma = 0$ then we have the pseudo-prior for ω which is set as a beta distribution with parameters specified by a pilot run. The WINBUGS variables $w\text{mean}$ and $s2$ are the posterior mean and variance of ω when we run the negative binomial model. Then the parameters $a\text{bar}$ and $b\text{bar}$ of the beta pseudo-prior are matched using the moments estimates. The parameter $c2$ is used to adjust the variance of the pseudo-prior. The relevant code for the Belgium 1993 example is the following:

```
theta<-lambda*(1-omega)*(1-omega)/(omega*(2-omega))
omega~dbeta(a,b);
a<-gamma*1 +(1-gamma)*abar;
b<-gamma*1 +(1-gamma)*bbar;
abar<-wmean*( wmean*(1-wmean)/(c2*s2)-1);
```

```
bbar<-abar*(1-wmean)/wmean;
wmean<-0.039;
s2<-pow(0.0036,2);
c2<-0.20;
```

Prior and Pseudo-prior Distributions on ω : The approach similar as above with the difference now we directly estimate ω . The WINBUGS code for example one is given by

```
omega~dbeta(a,b);
a<-gamma*1+(1-gamma)*abar;
b<-gamma*1+(1-gamma)*bbar;
abar<-wmean*(wmean*(1-wmean)/(c2*s2)-1);
bbar<-abar*(1-wmean)/wmean;
wmean<-0.039;
s2<-pow(0.036,2);
c2<-1.0;
```

Prior on Model Indicator γ : When no prior information is available and we wish to roughly estimate the posterior model probabilities then we simply set that γ to follow a Bernoulli distribution with probability $1/2$ defined by

```
gamma~dbern(0.5);
```

When we wish to estimate the posterior model odds with higher precision then we may use a prior such that the posterior model weights are close to 0.5 (this can be achieved by repeated pilot runs) and then calculate backwards the Bayes factor using Eq. (6). In such case we use the code for example one are given by

```
gamma~dbern( pmp );
# values for the comparison Poisson vs. Negative Binomial
pmp2<- -81;
pmp<-exp(pmp2)/(1+exp(pmp2))
# values for the comparison Poisson vs. Generalized Poisson
pmp2<- -82;
pmp<-exp(pmp2)/(1+exp(pmp2));
```

Comparison of Negative Binomial and Generalized Poisson: The approach is similar as above but no pseudo-prior is needed since we link the parameters by equating the dispersion indexes of the two models. Hence the WINBUGS code for ϑ and ω is simply given by

```
theta<-lambda*(1-omega)*(1-omega)/(omega*(2-omega))
omega~dbeta(1,1);
```

References

- [1] A.J.G. Cairns, A discussion of parameter and model uncertainty in insurance, *Insurance: Mathematics and Economics* **27** (2000), 313–350.
- [2] B.P. Carlin and S. Chib, Bayesian Model Choice via Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society, B* **157** (1995), 473–484.
- [3] C. Chatfield, Model uncertainty, data mining and statistical inference, (with discussion) *Journal of the Royal Statistical Society, A* **158** (1995), 419–466.
- [4] P.C. Consul, *Generalized Poisson Distribution: Properties and applications*, New York, Decker, 1989.
- [5] E. de Alba, Bayesian estimation of outstanding claim reserves, *North American Actuarial Journal* **6**(4) (2002).
- [6] P. Dellaportas, J.J. Forster and I. Ntzoufras, On Bayesian Model and Variable Selection Using MCMC, *Statistics and Computing* **12** (2002), 27–36.
- [7] M. Denuit, A new distribution of Poisson-type for the number of claims, *Astin Bulletin* **27** (1997), 229–242.
- [8] J.B. Douglas, *Analysis of Standard Contagious Distributions*, (Vol. 4), Statistical Distributions in Science Work Series, International Cooperative Publishing House, Fairland, Maryland, USA, 1980.
- [9] D. Draper, Assessment and propagation of model uncertainty (with discussion), *Journal of the Royal Statistical Society, B* **57** (1995), 47–97.
- [10] E.I. George and R.E. McCulloch, Variable Selection via Gibbs Sampling, *Journal of the American Statistical Association* **88** (1993), 881–889.
- [11] W.R. Gilks, S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, UK, 1996.
- [12] P. Green, Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82** (1995), 711–732.
- [13] R.E. Kass and A.E. Raftery, Bayes factors, *Journal of the American Statistical Association* **90** (1995), 773–795.
- [14] L. Kuo and B. Mallick, Variable Selection for Regression Models, *Sankhyā* **60** (1998), 65–81.
- [15] U.E. Makov, Principal applications of Bayesian methods in Actuarial science: A perspective, *North American Actuarial Journal* **5**(4) (2001), 53–73.
- [16] I. Ntzoufras and P. Dellaportas, Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty, (with discussion) *North American Actuarial Journal* **6** (2002), 113–128.
- [17] D. Spiegelhalter, A. Thomas, N. Best and D. Lunn, *WinBUGS User Manual, Version 1.4*, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine, available at <http://www.mrc-bsu.cam.ac.uk/bugs>, 2003.
- [18] P. Ter Berg, On the loglinear Poisson and Gamma model, *ASTIN Bulletin* **11** (1980), 35–40.
- [19] P. Ter Berg, A loglinear Lagrangian Poisson model, *ASTIN Bulletin* **26** (1996), 123–129.
- [20] R. Verrall, Bayes and Empirical Bayes estimation of the chain-ladder model, *ASTIN Bulletin* **20**(2) (1990), 217–243.
- [21] R. Verrall, An investigation into stochastic claims reserving models and the chain-ladder technique, *Insurance: Mathematics and Economics* **26** (2000), 91–99.