Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2604–2622
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Efficient and accurate approximate Bayesian inference with an application to insurance data

George Streftaris[a,*], Bruce J. Worton[b]

[a] School of Mathematical and Computer Sciences and Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK
[b] School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, James Clerk Maxwell Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK

Received 5 December 2006; received in revised form 7 September 2007; accepted 8 September 2007
Available online 18 September 2007

Abstract

Efficient and accurate Bayesian Markov chain Monte Carlo methodology is proposed for the estimation of event rates under an overdispersed Poisson distribution. An approximate Gibbs sampling method and an exact independence-type Metropolis–Hastings algorithm are derived, based on a log-normal/gamma mixture density that closely approximates the conditional distribution of the Poisson parameters. This involves a moment matching process, with the exact conditional moments obtained employing an entropy distance minimisation (Kullback–Liebler divergence) criterion. A simulation study is conducted and demonstrates good Bayes risk properties and robust performance for the proposed estimators, as compared with other estimating approaches under various loss functions. Actuarial data on insurance claims are used to illustrate the methodology. The approximate analysis displays superior Markov chain Monte Carlo mixing efficiency, whilst providing almost identical inferences to those obtained with exact methods.

Keywords: Bayes risk; Entropy distance; Effective sample size; Hierarchical Bayesian analysis; Insurance claims; Markov chain Monte Carlo; Mixture distribution; Monte Carlo error

1. Introduction

Simultaneous inference for several Poisson distributions has attracted much attention, especially in the case of additional variation caused by the dependence among the Poisson parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^{\mathrm{T}}$. Applications involving such inference have emerged in various areas, including actuarial science (Haberman and Renshaw, 1996; Makov et al., 1996) and epidemiology (Clayton and Kaldor, 1987; Ainsworth and Dean, 2006). The problem has been tackled in the past using various shrinkage estimating approaches, aiming to exploit the information provided in the entire vector of the Poisson parameters $\boldsymbol{\theta}$ (e.g. Morris, 1983). Bayesian methodology provides a natural framework for exploiting the relation between the components of $\boldsymbol{\theta}$ through the prior distribution, thus also dealing with the problem of overdispersion. The use of Bayes and empirical Bayes methods for the analysis of Poisson data in actuarial science, including the consideration of Poisson/gamma models for insurance claims, is discussed by Makov et al. (1996),

* Corresponding author. Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Colin Maclaurin Building, Heriot-Watt University, Edinburgh EH14 4AS, UK. Tel.: +44 131 451 3679; fax: +44 131 451 3249.
E-mail address: G.Streftaris@hw.ac.uk (G. Streftaris).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2007.09.006

Haastrup (2000), Czado et al. (2005) and Ntzoufras et al. (2005) among others. Advances in Markov chain Monte Carlo (MCMC) stochastic integration methodology (e.g. Tierney, 1994) have facilitated the generic implementation of full Bayesian analysis in related problems.

In this paper we work under a hierarchical Bayesian framework assuming a log-normal prior distribution for the Poisson parameters, and develop very efficient and accurate MCMC methodology for posterior analysis. The motivation for the work is that generally, with non-conjugate models, efficiency in the mixing behaviour of a Markov chain usually depends on the choice and construction of suitable proposal distributions, and this is often not given sufficient consideration in easily implemented MCMC algorithms employed in the analysis of hierarchical models. We propose a method which improves on the efficiency of commonly used Gibbs and Metropolis–Hastings schemes, while retaining the accuracy of the posterior inference. The presented approach may also be extended to a larger class of related models, where other prior distributions are assumed for the Poisson parameters.

We investigate the use of a close approximation to the conditional distribution of the Poisson rates $\theta_1, \theta_2, \ldots, \theta_m$, given all other model parameters and the data. The proposed approximation is based on a log-normal/gamma mixture density which matches the first three moments of the original distribution. For the computation of the moments of the posterior distribution we use a method relying on entropy distance (Kullback–Liebler divergence) minimisation. The resulting density is then employed in a Gibbs sampling scheme which mixes more efficiently than traditionally used approaches, and provides very accurate posterior inference that also performs favourably in terms of Bayes risk. We also employ the approximate density as the proposal distribution for an independence-type Metropolis–Hastings step (Tierney, 1994), which gives a very efficient exact algorithm with acceptance rate close to 1. In contrast, standard MCMC approaches often incorporate Gibbs sampling algorithms using rejection sampling techniques (e.g. as implemented in the *WinBUGS* software, http://www.mrc-bsu.cam.ac.uk/bugs), or various Metropolis–Hastings schemes that rely on fine-tuning the variance of candidate distributions (Gelfand and Smith, 1990; George et al., 1994; Damien et al., 1999). Although the implementation of such approaches appears to be straightforward, mixing efficiency is not necessarily guaranteed. Our results demonstrate that ease of implementation can be offset by an increase in the number of iterations required to achieve a certain level of estimation precision, which can be important in problems with slow chain mixing. Despite the fact that rapid advance of computing power continuously changes the balance among the developing, computing and running time of algorithms, there is arguably still scope for improved efficiency.

The proposed methodology is applied to the estimation of insurance claim intensities. The analysis confirms the accuracy of the methodology, and additionally demonstrates its convergence efficiency in terms of mixing of the Markov chain.

In Section 2 we introduce the Poisson/log-normal hierarchical model, while the derivation of the MCMC methods for the analysis is outlined in Sections 3–5. The results of an extensive Monte Carlo simulation study are presented in Section 6 to assess the risk properties of the proposed estimators, and compare them with those of other Bayes and classical methods under various scenarios concerning the prior distribution and loss function. The insurance data application is discussed in Section 7, where we also compare the efficiency of the considered MCMC schemes.

2. The model

We assume that given the parameters $\theta_1, \theta_2, \ldots, \theta_m$, the counts $Y_1, Y_2, \ldots, Y_m$, are conditionally independent Poisson random variables with respective means $\theta_i E_i$, $i = 1, \ldots, m$, i.e.

$$Y_i|\theta_i \sim \text{Poisson}(\theta_i E_i), \quad i = 1, \ldots, m, \tag{1}$$

where E_i, $i = 1, \ldots, m$, represent different exposure times. The parameters θ_i, $i = 1, \ldots, m$, give the rate of occurrence of events and depend on p explanatory variables in a log-linear regression structure expressed as

$$\log(\theta_i) = \mathbf{x}_i^{\mathrm{T}}\mathbf{b} + \varepsilon_i, \quad i = 1, \ldots, m, \tag{2}$$

where $\mathbf{x}_i^{\mathrm{T}} = (x_{1,i}, x_{2,i}, \ldots, x_{p,i})$, for $i = 1, \ldots, m$, are known values of the explanatory variables, $\mathbf{b} = (b_1, b_2, \ldots, b_p)^{\mathrm{T}}$ is a vector of regression coefficients and ε_i, $i = 1, \ldots, m$, are random error terms.

In actuarial science (1) and (2) can be used to model the occurrence of insurance claims. In this context, Y_i represents the number of actual claims in group $i = 1, \ldots, m$, $\mathbf{x}_i$ are covariates related to group i (e.g. age), and E_i is the total