The Bayesian Analysis of Generalized Poisson Models for Claim Frequency Data Utilising Markov Chain Monte Carlo Methods

David P. M. Scollnik , Ph.D. , A.S.A.

Department of Mathematics and Statistics

University of Calgary

Calgary , Alberta , T2N 1N4

E-mail address : scollnik@acs.ucalgary.ca

Key words and phrases :

Abstract :

A fairly popular and well-studied alternative to the standard Poisson distribution is the Lagrangian-Poisson distribution (Johnson, Kotz and Kemp 1992, p. 189), also known as the generalized Poisson distribution (GPD), which was introduced to the statistical literature by Consul and Jain (1973). The GPD is preferred over the Poisson distribution when the data exhibit substantial extra-Poisson variation, or overdispersion, relative to a Poisson model. This distribution has previously been considered within the actuarial context by Consul (1990), Gerber (1990), Goovaerts and Kass (1991), Kling and Goovaerts (1993), and Ambagaspitiya and Balakrishnan (1994). This paper considers the Bayesian analysis of the GPD and the generalized Poisson regression (GPR) model. Markov chain Monte Carlo (MCMC) methods and random variate generation strategies such as adaptive rejection sampling (ARS) for log-concave densities and adaptive rejection Metropolis sampling (ARMS) are discussed and then utilised in order to advance the Bayesian analysis of these generalized Poisson models.

1. Introduction.

The assumption of a Poisson distribution is a popular one when it is necessary to analyse a set of random count data. But since the equality of the mean and variance characterizes the Poisson distribution within the class of power series distributions with non-zero probabilities over all non-negative integers, it is sadly the case that this distribution is oftentimes a poor choice when it is known or suspected that the random counts exhibit substantial extra-Poisson variation, or overdispersion, relative to a Poisson model. This shortcoming of the Poisson distribution is also a concern in the context of Poisson regression, when the mean of the response variable is affected by a number of explanatory variables. For this reason, numerous authors have proposed tests for detecting overdispersion in Poisson models (e.g. Collings and Margolin 1985; Dean and Lawless 1989; Dean 1992; Ganio and Schafer 1992; Lambert and Roeder 1993) and others have proposed models accommodating overdispersion (eg. Lawless 1987; Dean, Lawless and Willmot 1989; Consul and Famoye 1992; Famoye 1993).

A fairly popular and well-studied alternative to the standard Poisson distribution is the Lagrangian-Poisson distribution (Johnson, Kotz and Kemp 1992, p. 189), also known as the generalized Poisson distribution (GPD), which was introduced to the statistical literature by Consul and Jain (1973). Consul (1989) provided a guide to the current state of modeling with the GPD at that time, and documented many real life examples. Consul and Famoye (1992) considered a generalized Poisson regression (GPR) model. Recently, the GPD has been making appearances in the actuarial literature. Consul (1990) used it in order to model the six data sets found in Gossiaux and Lemaire (1981) relating to the number of injuries in automobile accidents, and found it to perform at least as well as a number of traditional distributional alternatives. Gerber (1990), Goovaerts and Kass (1991), Kling and Goovaerts (1993), and Ambagaspitiya and Balakrishnan (1994) have considered the properties of compound generalized Poisson models.

One version of the GPD has a probability function given by

$$P (Y = y \mid \theta , \lambda) = \theta (\theta + y \lambda)^{y-1} (y !)^{-1} \exp (-\theta - y \lambda) \qquad (1.1)$$

where $\theta > 0$ and $0 \le \lambda < 1$, for those values of y on the non-negative integers, and zero elsewhere. When $\lambda = 0$ this distribution reduces to the standard Poisson. It is well-known that the GPD has mean $\theta (1 - \lambda)^{-1}$ and variance $\theta (1 - \lambda)^{-3}$ (Consul 1989), and so this distribution may be suitable when count data is observed with a sample variance considerably larger than the sample mean.

Often, the random count data are affected by a number of explanatory regressor variables. For instance, this is the case when insurance policies are grouped according to the different levels of various risk factors, and we are modelling the observed number of claims by class. In this case it is easy to define a GPR model based upon the GPD. For a sample of size n, let the i th response variable be denoted Y_i and let $\mathbf{x}_i$ denote the associated $p \times 1$ vector of explanatory variables. Given its covariate vector $\mathbf{x}_i$, let the distribution of Y_i be that of the GPD with probability function (1.1) with mean $E(Y_i \mid \mathbf{x}_i ; \boldsymbol{\beta}, \lambda) = \mu(\mathbf{x}_i ; \boldsymbol{\beta}) = \mu_i > 0$, and with secondary parameter λ. Our assumption is that $\mu(\mathbf{x}_i ; \boldsymbol{\beta})$ is a known function of $\mathbf{x}_i$ and an associated $p \times 1$ vector $\boldsymbol{\beta}$ of regression parameters. Taking into account that the mean of the GPD is given by $\mu_i = \theta(1-\lambda)^{-1} = \theta \phi$, the corresponding GPR model for the response variable Y_i may be written as

$$P(Y_i = y_i \mid \mathbf{x}_i ; \boldsymbol{\beta}, \lambda) \tag{1.2}$$

$$= \mu_i [\mu_i + (\phi - 1) y_i]^{y_i - 1} \phi^{-y_i} (y_i !)^{-1} \exp\{ -[\mu_i + (\phi - 1) y_i]/\phi \}$$

where $\mu_i > 0$ and $0 \leq \lambda < 1$, for those values of y_i on the non-negative integers. We note in passing that the parameter $\phi = (1 - \lambda)^{-1}$ represents the square root of the index of dispersion, and that the variance of Y_i is $\mathrm{Var}(Y_i \mid \mathbf{x}_i ; \boldsymbol{\beta}, \lambda) = \phi^2 \mu_i > 0$.

When we are faced with a random sample that has been generated according to either the GPD or the GPR model, it is evident that the likelihood function will be formed as a product of terms of form (1.1) or (1.2), respectively. If we simply combine this likelihood with the Bayesian practitioner's prior for the model parameters using Bayes' theorem, then the result will be the posterior distribution. In general, unless the size of the data set is very small and the prior density also has a very convenient form, the resulting posterior will not have a particularly tractable form for analytical analysis. This is illustrated by Shoukri and Consul (1989), who considered a Bayesian analysis of the basic GPD model. At the time, the best those authors could supply was a limited form of approximate Bayesian analysis for the GPD model requiring the use of Pearson curves along with the assumption that the ratio λ/θ was supported on a finite number of values.

The purpose of the present paper will be twofold. In the first place, we wish to demonstrate how a fully Bayesian analysis of the models presented above may proceed using today's very popular Markov chain Monte Carlo (MCMC) procedures. In the second place, we wish to take this opportunity to introduce a number of these MCMC procedures to the actuarial readership at large. The format of the paper is as follows. In Section 2, basic

elements of MCMC methods will be reviewed. In Section 3 we will apply MCMC methods in order to conduct illustrative Bayesian analyses of two insurance data sets, one data set for each of the GPD and the GPR model. In Section 4 we conclude our presentation, and provide directions to a useful storehouse of information relating to MCMC methods.

2. Markov chain Monte Carlo and the Gibbs sampler.

A MCMC method is a sampling-based procedure that may be used in order to generate a dependent sample from a certain distribution of interest. Such a method proceeds by first specifying an irreducible and aperiodic Markov chain with unique invariant distribution $\pi(x)$ equal to the desired distribution of interest (or *target* distribution), and then simulating one or more realisations of this Markov chain on a fast computer. Each path will form a dependent random sample from the distribution of interest, provided that certain regularity conditions are satisfied. These dependent sample paths may then be utilised for inferential purposes in a variety of ways. Specifically, if the Markov chain is aperiodic and irreducible, with unique invariant distribution $\pi(x)$, and X^1, X^2, ... , is a realisation of this chain, then available asymptotic results (see Tierney 1994 or Roberts and Smith 1994, for example) tell us that

$$X^t \xrightarrow{d} X \sim \pi(x) \quad \text{as} \quad t \to \infty \tag{2.1}$$

and

$$\frac{1}{t} \sum_{i=1}^{t} h(X^t) \to E_\pi[h(X)] \quad \text{as} \quad t \to \infty , \text{ almost surely} . \tag{2.2}$$

In the second result above, h is an arbitrary π-integrable real-valued function. Notice that if $h(X)$ is taken to be the conditional density for some random variable Y given X, then (2.2) suggests that the marginal density of Y may be estimated at the point y by averaging the conditional density $f(y \mid X)$ over the realised values X^t (see Gelfand and Smith 1990, pages 402-403).

The Gibbs sampler is a special kind of MCMC method. It was introduced by Geman and Geman (1984) in the context of image restoration, and its suitability for a wide range of problems in the field of Bayesian inference was recognised by Gelfand and Smith (1990). An elementary introduction to the Gibbs sampler is given in Casella and George (1992), and

those readers unfamiliar with the methodology are encouraged to peruse this reference. More sophisticated discussions of the Gibbs sampler and MCMC methods in general are given in Chib and Greenberg (1994b), Neal (1993), Smith and Roberts (1993), and Tierney (1994). Basically, the Gibbs sampler constructs a Markov chain by using the full conditional distributions associated with the target distribution to define the transition probabilities in the Markov chain. Specifically, let us now interpret the target distribution as a joint distribution for k blocks of random variables, that is, $\pi(x)$ corresponds to $\pi(x_1, x_2, \cdots, x_k)$. Let the notation $\pi(x_j \mid x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k)$ represent the full conditional distribution of the jth block of variables, x_j, given the remaining blocks. Then when we speak of a Gibbs sampler, we are actually referring to an implementation of the following iterative sampling scheme :

1. Select initial values $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_k^{(0)})$.

   Set $i = 0$.

2. Simulate the sequence of random draws

$$x_1^{(i+1)} \sim \pi(x_1 \mid x_2^{(i)}, x_3^{(i)}, \cdots, x_k^{(i)})$$

$$x_2^{(i+1)} \sim \pi(x_2 \mid x_1^{(i+1)}, x_3^{(i)}, \cdots, x_k^{(i)}).$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$x_k^{(i+1)} \sim \pi(x_k \mid x_1^{(i+1)}, x_2^{(i+1)}, \cdots, x_{k-1}^{(i+1)})$$

3. Set $i \leftarrow i + 1$ and return to step 2.

It may be shown to follow directly from the definition of the Gibbs sampling algorithm that the target distribution $\pi(x)$ is an invariant distribution of the Markov chain so defined by any implementation of the Gibbs sampler (see Tierney 1994 or Chib and Greenberg 1994b, for example). In order to ensure that a particular implementation of the Gibbs sampler is such that results (2.1) and (2.2) hold, any one of a number of sets of sufficient conditions may be checked. One convenient such set is provided by *Theorem 2* in Roberts and Smith (1994). This theorem is essentially as follows below :

Suppose that (i) $\pi(x) \equiv \pi(x_1, x_2, \cdots, x_k)$ *is dominated with respect to n-dimensional Lebesgue measure,* $1 < k \leq n$ *; (ii)* $\pi(x) > 0$ *implies that there exists an open neighbourhood* N_x *containing* x *and* $\varepsilon > 0$ *such that, for all* $y \in N_x$, $\pi(y) \geq \varepsilon > 0$ *; (iii)* $\int \pi(x) dx_j$ *is locally bounded for* $j = 1, \cdots, k$ *; and (iv) the support of* x *is arc connected. Then results (2.1) and (2.2) obtain.*

Although the result in the paragraph above allows us to check the theoretical convergence of a Gibbs sampler, there are important practical considerations which must be taken into account when designing and implementing a Gibbs sampler. For one, it is apparent that the blocking of the variables is dependent upon the choice made by the practitioner. Intuitively, the simplest and most naive implementation of a Gibbs sampler would be obtained by taking each block of random variables to be a single variable, so that each required draw from a full conditional distribution is a univariate one. However, the simplest implementation is not necessarily the best. When a number of variables are strongly correlated, for example, this naive implementation will result in a Markov chain with autocorrelations that are slow to die out. In this case, convergence is improved by blocking highly correlated variables together (see Liu *et al* 1994). Another useful concept when implementing a Gibbs sampler is the idea of data augmentation (Tanner and Wong 1987) which amounts to adding variables to those originally in the model specification in order to simplify the form of otherwise intractable full conditional distributions. Finally, when implementing a Gibbs sampler, or any other MCMC method, it is of paramount importance to monitor the convergence of the algorithm. This may be done using either *ad hoc* methods or rather formalized techniques. As examples of the former, we might monitor empirical moments or quantiles for the sample paths of several variables or functions thereof until these quantities stabilize. Examples of the latter include the convergence diagnostics of Gelman and Rubin (1992), Raftery and Lewis (1992), Ritter and Tanner (1992), and Zellner and Min (1993). The review paper by Cowles and Carlin (1994) describe these convergence diagnostics along with numerous others.

In passing, we remark that the Gibbs sampler has already made several appearances within the actuarial literature to date. Carlin (1992a) utilised the Gibbs sampler in order to study the Bayesian state space modeling of non-standard actuarial time series, and Carlin (1992b) utilised the Gibbs sampler in order to develop various Bayesian approaches to graduation. Klugman and Carlin (1993) also utilised the Gibbs sampler in the arena of Bayesian

graduation, this time concentrating on a hierarchical version of Whittaker-Henderson graduation. Scollnik (1993) studied a simultaneous equations model for insurance rate-making, and conducted a Bayesian analysis of this model by making use of the Gibbs sampler. Discussions of specific convergence diagnostics appear in Carlin (1992a) and Scollnik (1993). The idea of data augmentation plays a central role in Scollnik (1993) as well.

3. Two Illustrative Analyses.

At this juncture, we return to a consideration of the generalized Poisson models presented in Section 1. We will examine two actuarial data sets assumed to have been generated according to these models, and discuss their respective Bayesian analyses using MCMC methods.

3.1 Number of Injuries in Automobile Accidents.

First of all, we turn our attention to one of the six data sets presented by Gossiaux and Lemaire (1981) relating to the number of injuries in automobile accidents. These data sets are all overdispersed, and Consul (1990) previously determined that model (1.1) fit these data sets at least as well as a number of traditional distributional alternatives. In this section, we consider a Bayesian analysis of this model for the Zaire (1974) data. This data is presented in Table 1 for the reader's convenience, along with the first two sample moments and Consul's (1990) maximum likelihood estimates.

The likelihood function for a sample $Y = (y_1 , \cdots , y_n)$ from (1.1) will be of form

$$l (\theta , \lambda \mid Y) \propto \theta^n \exp (- n \theta - Z \lambda) \prod_{i = 1}^{n} \frac{(\theta + y_i \lambda)^{y_i - 1}}{y_i !} \qquad (3.1.1)$$

where $Z = y_1 + \cdots + y_n$. A Bayesian analysis proceeds in the obvious manner by first assigning a prior distribution to θ and λ, and then deriving the posterior distribution for these parameters by means of Bayes' theorem. If we desire our Bayesian analysis to proceed under a diffuse but proper prior density specification, then we might proceed as follows. Since the parameter θ has as its support the real line from 0 to ∞, we may assume that the prior information available for this parameter can be well modelled by a *gamma (a , b)* distribution, for some values of a and b. This is reasonable, since the shape of the *gamma* density function is very flexible. If we happen to have very little prior information concerning θ available, then we note that the selection ($a = 2$, $b = 1$) will result in a fairly satisfactory and relatively diffuse prior for θ. For most analyses it will be reasonable

Table 1. Zaire (1974) automobile accident injury counts.

k	Observations
0	3719
1	232
2	38
3	7
4	3
5	1
6	0

Total 4000 ; $\bar{x} = 0.08650$; $s^2 = 0.12255$

ML estimates for the GPD parameters : $\hat{\theta} = 0.072808$, $\hat{\lambda} = 0.158290$

to assume that θ and λ are *a priori* independent with $\lambda \sim$ *uniform* $(0, 1)$. If we adopt this specification for the prior distribution, then we will have

$$p(\theta, \lambda) \propto \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \qquad (3.1.2)$$

for $\theta > 0$ and $0 \le \lambda < 1$. Combining (3.1.1) and (3.1.2) by means of Bayes' theorem yields the posterior distribution

$$p(\theta, \lambda \mid \mathbf{Y}) \propto \theta^{n+a-1} \exp(-[n+b]\theta - Z\lambda) \prod_{i=1}^{n} (\theta + y_i \lambda)^{y_i - 1} \qquad (3.1.3)$$

for $\theta > 0$ and $0 \le \lambda < 1$. It will prove advantageous to make the transformation in variable $\lambda = \theta \beta$, so that we are left with

$$p(\theta, \beta \mid \mathbf{Y}) \propto \theta^{Z+a} \exp(-[n+b]\theta - Z\theta\beta) \prod_{i=1}^{n} (1 + y_i \beta)^{y_i - 1} \qquad (3.1.4)$$

for $\theta > 0$ and $0 \le \beta < \theta^{-1}$. The form of (3.1.4) is such that its associated full conditional distributions are particularly well suited for an application of the Gibbs sampler, as we now explain below.

In order to implement the Gibbs sampling algorithm, we need to identify the full conditional posterior distribution for each parameter, and determine whether random draws from these distributions are easily obtained. We observe from (3.1.4) that the full conditional distribution for θ is described by

$$p(\theta \mid \mathbf{Y};\beta) \propto \theta^{Z+a} \exp(-[n+b+Z\beta]\theta) \qquad (3.1.5)$$

for $0 \leq \theta < \beta^{-1}$, which is the same as saying that $p(\theta \mid \mathbf{Y};\beta) \sim$ gamma $(Z+a+1, n+b+Z\beta)$ but restricted to the interval $(0, \beta^{-1})$. Generating a random draw from this distribution is a straightforward procedure. As for β, it is apparent from (3.1.4) that this parameter's full conditional distribution is such that

$$p(\beta \mid \mathbf{Y};\theta) \propto \exp(-Z\theta\beta) \prod_{i=1}^{n} (1+y_i\beta)^{y_i-1} \qquad (3.1.6)$$

for $0 \leq \beta < \theta^{-1}$. Generating a random draw from this distribution is a slightly more complicated task, but it may be accomplished by making use of the procedure set out in the paragraph below.
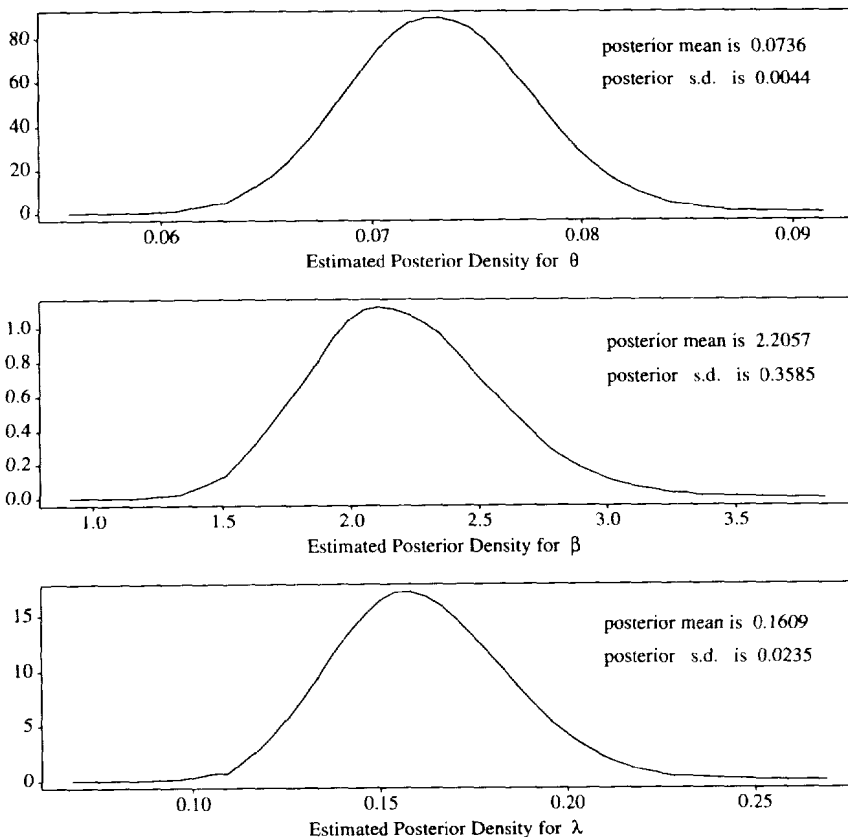
It is readily apparent that $p(\beta \mid \mathbf{Y};\theta)$ is continuous and differentiable throughout the support of β, and is also such that

$$\frac{\partial^2 \ln[p(\beta \mid \mathbf{Y};\theta)]}{\partial \beta^2} = -\sum_{i=1}^{n} \frac{(y_i-1)y_i^2}{(1+y_i\beta)^2} \qquad (3.1.7)$$

for $0 < \beta < \theta^{-1}$. This result tells us that $p(\beta \mid \mathbf{Y};\theta)$ is log-concave on the support of β, and is strictly log-concave provided that all of the y_is do not take on the value 0 or 1. When this is the case, it follows that adaptive rejection sampling (ARS) (Gilks 1992; Gilks and Wild 1992; Wild and Gilks 1993) may be utilized in order to effect random draws from (3.1.6). When it is the case that every y_i takes on the value 0 or 1, so that strict log-concavity fails for (3.1.6), then it is very doubtful that we would be modelling within the GPD framework in the first place. However, let us assume for the sake of argument that we were. Then it is evident from (3.1.6) that $p(\beta \mid \mathbf{Y};\theta) \sim$ *exponential* $(Z\theta)$, but truncated on the interval $(0, \theta^{-1})$. In this case a random draw from (3.1.6) may be obtained by simply making use of the inversion method for truncated distributions (eg. DeVroye 1986).

Having determined how independent random draws may be generated from each of the full conditional distributions $p(\theta \mid \mathbf{Y};\beta)$ and $p(\beta \mid \mathbf{Y};\theta)$, we are now ready to

Figure 1 : Estimated Posterior Marginal Densities for θ, β, and λ.

posterior mean is 0.0736
posterior s.d. is 0.0044

Estimated Posterior Density for θ

posterior mean is 2.2057
posterior s.d. is 0.3585

Estimated Posterior Density for β

posterior mean is 0.1609
posterior s.d. is 0.0235

Estimated Posterior Density for λ

invoke the Gibbs sampling algorithm. For the analysis of the Zaire data, we assumed the prior density specification as described above (with $a = 2$ and $b = 1$) and then proceeded to initiate 20,000 independent runs of the Gibbs sampler based upon the full conditional distributions (3.1.5) and (3.1.6). Each replication of the Gibbs sampler ran for 50 iterations. In this simple two parameter problem, convergence of the Gibbs sampler was nearly immediate. Only the last iteration of each Gibbs sampler was retained, thus providing us with 20,000 independent draws from the posterior distribution of θ and β. This random sample may be used in order to make posterior inference with respect to θ or β or with respect to an

arbitrary function of these parameters, such as $\lambda = \theta \beta$. In order to do the latter, we need simply apply the function of interest to the pairs of sampled values for θ and β, thus generating a random sample from the posterior distribution for the function of interest. In this way, we transformed each of the sampled pairs into a realization on λ as well. Estimated posterior marginal density plots for θ, β, and λ based upon the 20,000 simulated draws are presented in Figure 1.

3.2 Ship Damage Incident Data.

For our next example, we turn to the Lloyd's Register of Shipping ship damage incident data as presented in McCullagh and Nelder (1983). This data were also analysed by Lawless (1987) utilising a negative binomial model; and by Consul and Famoye (1992) employing a GPR model and classical analysis. The ship damage incident data set has as its response variable the number of damage incidents for 34 individual ships over various five year periods. The type of damage in question is a form caused by waves to the forward section of cargo vessels. The qualitative factors are : ship type (A, B, C, D, or E); year of construction (1960-64, 1965-69, 1970-74 or 1975-79); period of operation (1960-74 or 1975-79). An exposure (E_i) relating to the aggregate number of months in service is also available. The GPR model (1.2) is applicable to this data set. Specifically, we consider an additive effects model specification, with binary indicator variables used to represent the main effects, and a log-linear form for μ_i, namely

$$\mu_i = E_i \exp (\mathbf{x}_i^T \boldsymbol{\beta}) \ . \tag{3.2.1}$$

To construct our prior density, we assumed that β and λ were were *a priori* independent, with β assigned a noninformative constant prior and λ assigned a uniform prior on the unit interval.
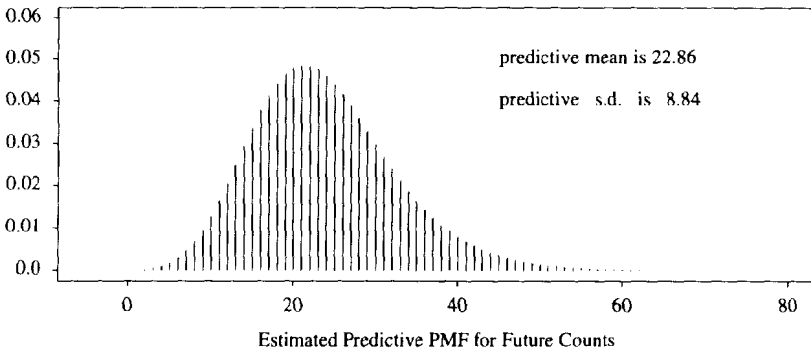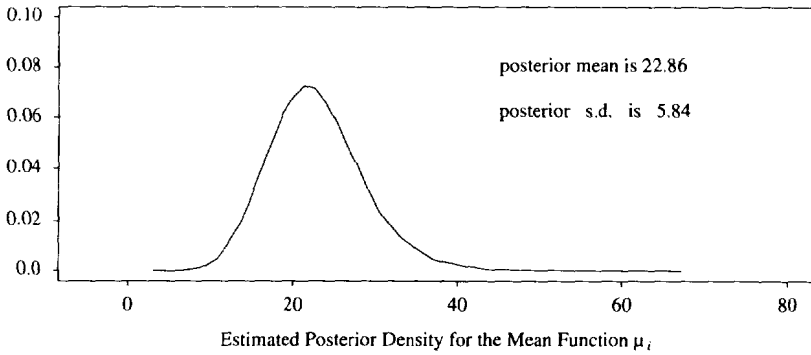
Given the prior density specification described above, our Bayesian analysis then proceeded by using adaptive rejection Metropolis sampling (ARMS) and the ARMS-within-Gibbs strategy described in Gilks *et al* (1993). Our final 'presentation' inferences were based upon single long runs of length 25,000, following a burn-in period consisting of 5,000 iterations. However, these inferences were also compared to, and found to be consistent with, inferences based upon 5 shorter runs of 5,000 iterations each, with burn-in periods consisting of 1,000 iterations. These sampled paths were also monitored for indications of convergence through the use of several graphical aids, and the convergence diagnostic of Gelman and Rubin (1992).

Table 2. Point Estimates for the Ship Damage Incident Data.

Parameter	Posterior Mean	Posterior S.D.
λ	0.25	0.11
Intercept	- 6.38	0.28
Ship Type		
A	-	
B	- 0.57	0.23
C	- 0.76	0.43
D	- 0.22	0.43
E	0.28	0.32
Year of Construction		
1960 - 64	-	
1965 - 69	0.68	0.20
1960 - 74	0.80	0.23
1975 - 79	0.45	0.32
Service Period		
1960 - 64	-	
1975 - 79	0.38	0.16

In Table 2, we provide estimated posterior means and standard deviations for the parameters appearing in the GPR model for the analysis described above. One great advantage and convenience of MCMC based analyses is that the simulated values generated from the target posterior distribution for the model parameters may be re-used in order to make posterior inference with respect to any function of the model parameters. In the context of the present analysis, we might be interested in studying the posterior distribution of the mean level function μ_i defined by (3.2.1) for ships of type A, constructed in 1975-79, operated in 1975-79, with 10,000 aggregate months of service. However, μ_i is merely a function of β, and so inference with respect to its posterior distribution is readily available if we simply transform the realised values of β generated by the MCMC into a series of realisations on (3.2.1). We have done precisely that, and a plot of the resulting estimated

Figure 2 : Plots for the Ship Damage Incident Analysis .

posterior mean is 22.86

posterior s.d. is 5.84

Estimated Posterior Density for the Mean Function μ_i

predictive mean is 22.86

predictive s.d. is 8.84

Estimated Predictive PMF for Future Counts

posterior density for the mean level function μ_i of the ships in question is presented in Figure 2. If we desire to make 'predictive' inference with respect to the number of damage incidents for (presumably unobserved) ships of this same class, still with 10,000 aggregate months of service, then we may proceed as follows. Provided that the damage incident count for these ships is independent of other ship's counts given the model parameters, and that it is generated according to the same GPR model, then an estimate of the predictive probability mass function (pmf) is obtained by simply averaging (1.2) over the realised values of β and λ generated by the MCMC. An estimate of this predictive pmf is presented in Figure 2.

4. Conclusions.

In this paper, we considered how the Bayesian analysis of two overdispersed generalized Poisson models may proceed. The analysis of both models proceeded by way of MCMC, making use of ARMS-within-Gibbs when the sampling densities proved to be inconvenient from which to implement exact draws. Those readers with access to the computer internet may wish to contact the MCMC Preprint Service via ftp (ftp.statslab.cam.ac.uk) or via a World Wide Web interface like Mosaic (www.statslab.cam.ac.uk). Several of the unpublished references mentioned in this paper are presently archived there. An expanded version of this paper, describing ARS and ARMS in more detail, is available upon request from the author. Electronic correspondence may be sent to either *scollnik@acs.ucalgary.ca* or *davids@balducci.math.ucalgary.ca*.

REFERENCES.

Ambagaspitiya, R.S., and N. Balakrishnan, (1994), On the Compound Generalized Poisson Distributions, to appear in *The ASTIN Bulletin.*

Breslow, N., (1984), Extra-Poisson Variation in Log-linear Models, *Applied Statistics*, **33**, 38-44.

Carlin, B.P., and N.G. Polson, (1991), Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler, *The Canadian Journal of Statistics*, **19 (4)**, 399-406.

Carlin, B.P., (1992a), State Space Modeling of Non-Standard Actuarial Time Series, *Insurance : Mathematics and Economics*, **11**, 209-222.

Carlin, B.P., (1992b), A Simple Monte Carlo Approach to Bayesian Graduation, *Transactions of the Society of Actuaries*, **XLIV**, 55-76.

Casella, G., and E.I. George, (1992), Explaining the Gibbs Sampler, *The American Statistician*, **46** , 167-174.

Chib, S., and E. Greenberg, (1994a), Understanding the Metropolis-Hastings Algorithm. Unpublished manuscript, Washington University.

Chib, S., and E. Greenberg, (1994b), Markov Chain Monte Carlo Simulation Methods in Econometrics. Unpublished manuscript, Washington University.

Collings, B.J., and B.H. Margolin, (1985), Testing Goodness-of-Fit for the Poisson Assumption When Observations are Not Identically Distributed, *Journal of the American Statistical Association*, **80**, 411-418.

Consul, P.C., and G.C. Jain, (1973), A Generalization of the Poisson Distribution, *Technometrics*, **15 (4)** , 791-799.

Consul, P.C., (1989), Generalized Poisson Distributions : Properties and Applications, Marcel Dekker, Inc., New York.

Consul, P.C., and F. Famoye, (1992), Generalized Poisson Regression, *Communications in Statistics - Theory and Methods*, **21 (1)** , 89-109.

Cowles, M.K., and B.P. Carlin, (1994), Markov Chain Monte Carlo : A Comparative Review. Research Report 94-008, Division of Biostatistics, School of Public Health, University of Minnesota.

Dean, C., and J.F. Lawless, (1989), Tests for Detecting Overdispersion in Poisson Regression Models, *Journal of the American Statistical Association*, **84**, 467-472.

Dean, C., Lawless, J.F., and G.E. Willmot, (1989), A Mixed Poisson-Inverse Gaussian Regression Model, *Canadian Journal of Statistics*, **17**, 171-182.

Dean, C., (1992), Testing for Overdispersion in Poisson and Negative Binomial Regression Models, *Journal of the American Statistical Association*, **87**, 451-457.

Dellaportas, P., and A.F.M. Smith, (1993), Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling, *Applied Statistics*, **42 (3)** , 443-459.

Devroye, L. (1986), Non-Uniform Random Variate Generation, New York : Springer-Verlag.

Famoye, F., (1993), Restricted Generalized Poisson Regression Model, *Communications in Statistics - Theory and Methods*, **22 (5)** , 1335-1354.

Ganio, L.M., and D.W. Schafer, (1992), Diagnostics for Overdispersion, *Journal of the American Statistical Association*, **87**, 795-804.

Gelfand, A.E., and A.F.M. Smith, (1990), Sampling Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85** , 398-409.

Gelman, A., and D.B. Rubin, (1992), Inference from Iterative Simulation Using Multiple Sequences, with discussion in *Statistical Science*, **7 (4)**, 457-472.

Geman, S., and D. Geman, (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6** , 721-741.

Gerber, H.U., (1990), When Does the Surplus Reach a Given Target ? , *Insurance : Mathematics and Economics*, **9**, 115-119.

Gilks, W.R., (1992), Derivative-free Adaptive Rejection Sampling for Gibbs Sampling, in *Bayesian Statistics 4*, J.M. Bernardo, J.O.Berger, A.P. Dawid and A.F.M. Smith, eds., Oxford : University Press, 641-649.

Gilks, W.R., and P. Wild, (1992), Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics*, **41 (2)**, 337-348.

Gilks, W.R., Best, N.G., and K.K.C. Tan., (1993), Adaptive Rejection Metropolis Sampling for Gibbs sampling. Technical Report, Medical Research Council Biostatistics Unit, Cambridge, UK.

Goovaerts, M.J., and R. Kaas, (1991), Evaluating Compound Generalized Poisson Distributions Recursively, *ASTIN Bulletin*, **21**, 193-197.

Gossiaux, A., and J. Lemaire, (1981), Methodes d'adjustement de distributions de sinistres, *Bulletin of the Association of Swiss Actuaries*, **81**, 87-95.

Hastings, W.K., (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, **57**, 97-109.

Hogg, R.V., and S.A. Klugman, (1984), Loss Distributions, John Wiley & Sons, New York.

Johnson, N.L., Kotz, S., and Kemp, A.W., (1992), Univariate Discrete Distributions, Second Edition, John Wiley & Sons, New York.

Kling, B., and M. Goovaerts, (1993), A Note on Compund Generalized Distributions, *Scandinavian Actuarial Journal*, **1**, 60-72.

Klugman, S.A., and B.P. Carlin, (1993), Hierarchical Bayesian Whittaker Graduation, *Scandinavian Actuarial Journal*, **2**, 183-196.

Lawless, J.F., (1987), Negative Binomial and Mixed Poisson Regression, *The Canadian Journal of Statistics*, **15** (3), 209-225.

Lambert, D., and K. Roeder, (1993), Overdispersion Diagnostics for Generalized Linear Models. AT&T Bell Laboratories Technical Report.

Liu, J.S., Wong, W.W., and A. Kong, (1994), Covariance Structure of the Gibbs Sampler with Applications to the Comparison of Estimators and Augmentation Schemes, *Biometrika*, **81**, 27-40.

McCullagh, P., and Nelder, J.A., (1983), *Generalized Linear Models*, Chapman and Hall, London.

Neal, R.M., (1993), Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

Raftery, A.E., and S. Lewis, (1992), How Many Iterations in the Gibbs Sampler ? , in *Bayesian Statistics 4*, J.M. Bernardo, J.O.Berger, A.P. Dawid and A.F.M. Smith, eds., Oxford : University Press, 763-773.

Ritter, C., and M.A. Tanner, (1992), Facilitating the Gibbs Sampler : The Gibbs Stopper and the Griddy Gibbs Sampler, *Journal of the American Statistical Association*, **87**, 861-868.

Roberts, G.O., and A.F.M. Smith, (1994), Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis- Hastings Algorithms, *Stochastic Processes and their Applications* **49**, 207-216.

Scollnik, D.P.M., (1993), A Bayesian Analysis of a Simultaneous Equations Model for Insurance Rate-Making, *Insurance : Mathematics and Economics*, **12**, 265-286.

Shoukri, M.M., and P.C. Consul, (1989), Bayesian Analysis of a Generalized Poisson Distribution, *Communications in Statistics - Simulation and Computation*, **18** (4) , 1465-1480.

Smith, A.F.M., and G.O. Roberts, (1993), Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society, B*, **55 (1)**, 3-23.

Tanner, M.A., and W.H. Wong, (1987), The Calculation of Posterior Distributions by Data Augmentation (with discussion), *Journal of the American Statistical Association*, **82**, 528-550.

Tanner, M.A., (1993), Tools for Statistical Inference, Methods for the Exploration of Posterior Distributions and Likelihood Functions, second edition, New York : Springer-Verlag.

Tierney, L., (1991), Markov Chains for Exploring Posterior Distributions. Technical Report No. 560, School of Statistics, University of Minnesota.

Tierney, L., (1994), Markov Chains for Exploring Posterior Distributions. to appear in *The Annals of Statistics*.

Wakefield, J.C., Gelfand, A.E., and A.F.M. Smith, (1991), Efficient Generation of Random Variates via the Ratio-of-uniforms Method, *Statist. Comput.*, **1**, 129-133.

Wild, P., and W.R. Gilks, (1993), Adaptive Rejection Sampling from Log-concave Density Functions, *Applied Statistics*, **42 (4)** , 701-709.

Zellner, A., and C-k. Min, (1994), Gibbs Sampler Convergence Criteria (GSC^2), H.C. Alexander Research Foundation Working Paper. Submitted to *Journal of the American Statistical Association*.