

ON THE ANALYSIS OF THE TRUNCATED GENERALIZED POISSON DISTRIBUTION USING A BAYESIAN METHOD

BY

DAVID P M SCOLLNIK

*Department of Mathematics and Statistics
University of Calgary*

ABSTRACT

The generalized Poisson distribution with parameters θ and λ was introduced by Consul and Jain (1973) and has recently found several instances of application in the actuarial literature. The most frequently used version of the distribution assumes that $\theta > 0$ and $0 \leq \lambda < 1$, in which case the mean and variance are $\theta/(1 - \lambda)$ and $\theta/(1 - \lambda)^3$, respectively. These simple moment expressions, along with nearly all of the other theoretical results available for this distribution, fail when $\lambda < 0$ or $\lambda > 1$ (e.g., Johnson, Kotz, and Kemp, 1992, page 397). In these cases, even the definition of the probability mass function usually given in the literature is not properly normalized so that its values do not sum to unity. For this reason, it is common to truncate the support of the distribution and explicitly normalize the probability mass function. In this paper we discuss the estimation of the parameters of this truncated generalized Poisson distribution using a Bayesian method

KEYWORDS

Bayesian; bivariate; generalized Poisson; Langrangian Poisson; truncated; Markov chain Monte Carlo.

1. INTRODUCTION

A great many distributions are available for modelling discrete data arising in the insurance field. A large number of these discrete distributions are described in Chapter 3 of Klugman, Panjer, and Willmot (1997). Recently, some authors have also explored the use of Consul's Generalized Poisson Distribution (GPD) in actuarial settings. Consul (1990) demonstrated that

the GPD, sometimes also known as the Lagrangian Poisson distribution, is a plausible model for claim frequency data, Goovaerts and Kaas (1991) and Ambagaspitaya and Balakrishnan (1994) presented recursive methods to compute the total claims distribution for certain compound GPD models, as did Hesselager (1997) for a class of compound Lagrangian distributions including the compound GPD; Scollnik (1995a) used the GPD, and its extension to a regression context, in order to model various sorts of claim frequency data and showed how Markov chain Monte Carlo (MCMC) methods could be used to implement Bayesian posterior and predictive analyses of these models (see also Scollnik, 1995b and 1995c); Famoye and Consul (1995) introduced a version of bivariate GPD (BGPD), discussed parameter estimation by the method of moments and double zero frequency and by the method of maximum likelihood, and fit the BGPD to a data set on accidents sustained by a group of shunters, Vernic (1997) considered the same BGPD as did Famoye and Consul, and used method of moments estimation to fit this BGPD to the aggregate amount of claims for a compound class of policies submitted to claims of two kinds whose yearly frequencies are *a priori* dependent.

The purpose of this paper is to clarify some points relating to the GPD which are frequently misrepresented in the literature and to discuss how Bayesian posterior and predictive analysis of the truncated GPD and of a truncated BGPD can proceed using MCMC methods. We begin with a discussion of GPD models.

2. GENERALIZED POISSON DISTRIBUTION MODELS

The probability mass function of the basic untruncated GPD is commonly given by

$$Pr(N=n) = p_n(\theta, \lambda, m) = \begin{cases} \theta(\theta+n\lambda)^{n-1} \frac{\exp(-\theta-n\lambda)}{n} & \text{for } n=0, 1, 2, \dots, m \\ 0 & \text{for } n > m \text{ when } \lambda < 0, \end{cases} \quad (1)$$

and zero otherwise, where $\theta > 0$, $\max(-1, -\theta/m) \leq \lambda < 1$, and m is usually taken equal to the largest possible positive integer such that $\theta + m\lambda > 0$ when λ is negative. Often it is explicitly further required that $m \geq 4$ (e.g., as in Vernic, 1997) in order to ensure that there are at least five classes with non-zero probability when λ is negative (see Consul, 1989, page 4), but this obviously need not be the case. At this time, we will review a few of the properties associated with (1). Most of these properties are documented in Consul's (1989) treatment of the GPD. Additional references will be introduced as required. The reader is forewarned that some authors switch the roles of the parameters θ and λ . We have adopted the parametrization found in Consul (1989) and Johnson, Kotz and Kemp (1992, page 396).

To begin with, suppose that $0 \leq \lambda < 1$ and the value of m is taken equal to ∞ . For this case it is known that

$$E(N) = \frac{\theta}{1 - \lambda} \quad \text{and} \quad \text{Var}(N) = \frac{\theta}{(1 - \lambda)^3}, \quad (2)$$

so the variance of the GPD is always larger than or equal to the mean. It is apparent that this instance of the GPD reduces to the standard Poisson with parameter θ when $\lambda = 0$. The two moment expressions in (2), along with simple formulae for skewness and kurtosis and virtually all of the other theoretical results obtained relating to the GPD (e.g., Consul and Jain, 1973; Ambagasptiya and Balakrishnan, 1995; Vernic, 1997), are only valid for the case of the GPD presently under consideration, i.e. when $\theta > 0$, $0 \leq \lambda < 1$, and $m = \infty$.

Henze and Klar (1995, page 1877) make the claim that this fact has not been emphasized enough in the literature, and point to a paper by Alzaid and Al-Osh (1993) in which it is tacitly assumed that (2) also holds for negative values of λ . Famoye and Consul (1995, page 128) recently made the same errant assumption, without alerting the reader as to its nature. It is also very common for authors to estimate the GPD parameters by equating empirical moments to the theoretical moments obtained in the special case described above, even when the sample variance is strictly less than the sample mean so that negative estimates of λ result (e.g., Consul, 1989, see also Vernic, 1997).

Actually, in order to permit cases where the variance is smaller than the mean, Consul and Jain (1973) had proposed to admit negative values of λ . However, when the value of λ is negative the probability mass function (1) is no longer normalized. To see this, suppose that $\theta = 1.6$, $\lambda = -0.75$ and $m = 2$. Then $Pr(N = 0) = 0.2019$, $Pr(N = 1) = 0.6839$, $Pr(N = 2) = 0.0724$, $Pr(N > 2) = 0.0$, and the sum of these supposedly exhaustive 'probabilities' is only 0.9582. This problem was not recognized in the early literature concerning the GPD (e.g., Consul and Jain, 1973) until Nelson (1975) indicated that a cautious approach was warranted in the use of the GPD model with negative values of λ . One solution to this problem is to simply normalize the function in (1) when $\lambda < 0$. In fact, (1) will generally need to be normalized except in the special case that $\theta > 0$, $0 \leq \lambda < 1$, and $m = \infty$. Accordingly, Consul and Famoye (1989) defined the probability mass function of the truncated GPD to be

$$Pr(N = n) = q_n(\theta, \lambda, m) = \frac{p_n(\theta, \lambda, m)}{K(\theta, \lambda, m)} \quad \text{for } n = 0, 1, 2, \dots, m \quad (3)$$

and zero otherwise, where $\theta > 0$, $-\infty < \lambda < \infty$,

$$K(\theta, \lambda, m) = \sum_{n=0}^m p_n(\theta, \lambda, m)$$

and m is any positive integer such that $\theta + m\lambda > 0$. Usually, m is taken equal to the largest such value. Note that the definition of the truncated GPD extends the permitted range of the parameter λ to the entire real line.

When class frequencies are inappropriately calculated using (1) instead of (3), an error of truncation is said to occur Consul and Shoukri (1985) and Consul (1989, Section 9.1.1) have made an analysis of the error of truncation when $-1 < \lambda < 0$. The simulation study they conduct is not exhaustive, but it does appear to indicate that the error of truncation may be serious when the number of non-zero probability classes is 3 or 4 and the value of θ is approximately between 0.7 and 4.5. The reader can easily verify that the error of truncation may also be serious when $\lambda < -1$ or $\lambda > 1$.

Consul and Famoye (1989) studied the truncated GPD in some detail and discussed parameter inference using maximum likelihood (ML) estimation and estimation based upon the empirical mean and the ratio of the first two empirical class frequencies. Their main conclusion was that the ML estimates determined using (3) as the basis of the likelihood function are generally closer to the true values of the population parameters than are the ML estimates determined on the basis of (1). Hence, even though the error of truncation associated with using (1) may be small in some cases, they suggested that one should estimate the values of the parameters θ and λ using the truncated GPD model (3). It should be noted that the estimation methods pursued by Consul and Famoye (1989) are implemented in such a way so as to determine estimates of θ and λ conditional upon a presumed known value of m . Since m is not known, Consul and Famoye (1989) simply set it equal to the value of the largest observation.

Bayesian estimation is a likelihood based style of inference that incorporates prior information on the unknown variables. ML estimates are equivalent to the modes of the Bayesian posterior distribution, when the prior distribution for the unknown variables is flat. However, the goal of a Bayesian analysis is generally not just a point estimate like the posterior mode (or mean or median), but a representation of the entire distribution for the unknown parameter(s) (Gelman, Carlin, Stern, Rubin, 1995, page 301). In the next Section, we discuss how a Bayesian analysis of the truncated GPD with an informative prior distribution can be accomplished using a MCMC approach. We emphasize that the Bayesian estimation method yields a posterior distribution for all of the unknown parameters, including m (cf. Consul and Famoye, 1989).

3. A BAYESIAN ANALYSIS OF THE TRUNCATED GPD MODEL

Consul and Famoye (1989) argue that any discrete probability model for a random variable N defined on the set of non-negative integers is automatically truncated in real life situations because the sample size is always finite and the probabilities for large values of N become so small so as to be unobservable. This is particularly true in an insurance setting when the

number of claims per policy is small. Assuming this context, we suppose that the sampling model is taken to be approximately truncated GPD as in (3) with parameters θ , λ , and m , so that

$$Pr(N = j|\theta, \lambda, m) = q_j(\theta, \lambda, m) = \frac{p_j(\theta, \lambda, m)}{K(\theta, \lambda, m)} \quad \text{for } j = 0, 1, 2, \dots, m, \quad (4)$$

and zero otherwise, with $\theta > 0$ and $-\infty < \lambda < \infty$, with m equal to some positive integer such that $\theta + m\lambda > 0$, and with $1 \leq m \leq M$ so that there is at least one non-zero class with non-zero probability. Setting M equal to a value between 5 & 15, say, will generally suffice when the number of claims per policy or accidents per individual is small. We recognize that the value selected for the parameter M is formally an expression of *a priori* knowledge. This is further discussed in the next paragraph. If the data consists of observed class frequencies $n_j, j = 0, \dots, M$, with $n = n_0 + \dots + n_M$, then the likelihood function is of the form

$$l(\theta, \lambda, m) \propto \prod_{j=0}^m q_j(\theta, \lambda, m)^{n_j} = \frac{\prod_{j=0}^m p_j(\theta, \lambda, m)^{n_j}}{K(\theta, \lambda, m)^n}. \quad (5)$$

If the data includes some grouped class frequencies, then the likelihood function is modified in the obvious way. For example, if we observe the first two class frequencies n_0 and n_1 along with the grouped class frequency $g_2 = n_2 + \dots + n_M$, then the likelihood function is of form

$$l(\theta, \lambda, m) \propto q_0(\theta, \lambda, m)^{n_0} q_1(\theta, \lambda, m)^{n_1} \{1 - q_0(\theta, \lambda, m) - q_1(\theta, \lambda, m)\}^{g_2}.$$

In order to complete the definition of a full probability model, it is now necessary to specify a prior distribution for the unknown parameters θ , λ , and m . The reader is free to use any reasonable prior specification as befits the expert opinion that is available to him or her. For our presentation, we will consider 3 different forms of prior density specification (PDS). For the first PDS, we will assume that the parameters are distributed *a priori* in the following way:

$$p(\theta, \lambda, m) \propto p(\theta)p(\lambda)p(m) \quad \text{when } \theta + m\lambda > 0, \quad (6)$$

and zero otherwise, with

$$p(\theta) \sim \text{Gamma}(1, 2), \quad (7)$$

$$p(\lambda) \sim \text{Normal}(0, 0.1), \quad (8)$$

$$p(m) \sim \text{Uniform}\{1, \dots, M\}. \quad (9)$$

The *Gamma* distribution in (7) is parametrized so as to have mean and standard deviation both equal to 0.5, and the *Normal* distribution in (8) has standard deviation equal to 0.1. With respect to the *Uniform* distribution in (9), we are free to attach a hyper-prior distribution to the parameter M . We

have not pursued this particular avenue, although in Section 5 we will compare the use of several different values of M in the context of a particular data analysis

Another approach is to forgo the introduction of M entirely, and rather specify a distribution $p(m)$ on the entirety of the non-negative integers (in effect, $M = \infty$). In this case, equations (6), (7) and (8) would be unchanged, and (9) might be replaced with

$$p(m) \sim \text{Poisson}(\mu), \quad (10)$$

for some specified value $\mu > 0$. The parameter restrictions in effect would be $\theta > 0$, $-\infty < \lambda < \infty$, and $\theta + m\lambda > 0$. An analysis of the truncated GPD model incorporating this second form of PDS will also follow in Section 5.

Our third PDS will be similar to the two above, with the added restriction that $m = M$, for some specified value $M < \infty$. That is, our third analysis will be conditional on a fixed value of $m < \infty$.

By multiplying the likelihood and prior density functions together, we obtain the form of the posterior distribution up to a normalizing constant, that is

$$p(\theta, \lambda, m | n_0, \dots, n_m) \propto p(\theta)p(\lambda)p(m)l(\theta, \lambda, m) \quad \text{when } \theta + m\lambda > 0, \quad (11)$$

and zero otherwise, with $\theta > 0$ and $-\infty < \lambda < \infty$. If we let n^* denote the value of the largest observation, then we also require that $m \in \{n^*, \dots, M\}$. Here, either the value of $M < \infty$ is known as in the case of our first PDS, or else $M = \infty$ as in the second. In the case of our third PDS, M is assumed to be known and we further condition upon the assumption that $m = M < \infty$. At this stage, the complete probability model can be analysed using a numerical method. We propose the use of a MCMC method in order to complete the analysis of the posterior and predictive distributions.

4. COMPLETING THE BAYESIAN ANALYSIS USING A MCMC METHOD

In order to complete the Bayesian analysis of the truncated GPD model, we adopt a MCMC method. In particular, we implement a 'single-component Metropolis-Hastings' (Gilks, Richardson, and Spiegelhalter, 1996, page 10), or 'variable-at-a-time Metropolis-Hastings' (cf. Chan and Geyer's discussion of Tierney's 1994 paper, page 1748; also, Haastrup and Arjas, 1996, page 156), algorithm. This algorithm simulates a realization of a Markov chain which has the posterior distribution of the unknown parameters θ , λ , and m as its equilibrium distribution. The algorithm generates a sequence of simulated parameter values, $\theta^{(0)}, \lambda^{(0)}, m^{(0)}, \theta^{(1)}, \lambda^{(1)}, m^{(1)}, \dots$, whose empirical distribution converges towards the posterior distribution of the unknown parameters. The posterior distribution can thus be approximated on the basis of these values, and the approximation can be made as exact as we desire by simply increasing the length of the simulation. Note that predictions can also be obtained by simply averaging the truncated GPD

probability mass function over the sampled parameter values. That is, the probability mass function for a future observation N_f , given the observed class frequencies n_0, \dots, n_m , can be estimated using the result that

$$\begin{aligned} Pr(N_f = j | n_0, \dots, n_m) & \qquad \qquad \qquad (12) \\ &= \sum_m \int \int Pr(N_f = j | \theta, \lambda, m) p(\theta, \lambda, m | n_0, \dots, n_m) d\theta d\lambda \\ &\approx \sum_{t=\beta+1}^{B+L} \frac{Pr(N_f = j | \theta^{(t)}, \lambda^{(t)}, m^{(t)})}{L} . \end{aligned}$$

Here, B represents the number of iterations for which the Markov chain is allowed to 'burn-in' and L represents the number of iterations the Markov chain is run thereafter. A method for checking the convergence of the Markov chain by comparing several different and independently simulated sequences is given in Gelman, Carlin, Stern, and Rubin (1995, pages 330-333). If several different and independently simulated sequences are available, then the sample average in (12) should be taken over all of the available sample paths.

There are many ways of implementing the Markov chain described above. We proceed in the following manner. Let $\theta^{(0)}$, $\lambda^{(0)}$, and $m^{(0)}$ denote arbitrary starting values for the 3 random variables under examination. In this context, the i th iteration of the single-component Metropolis-Hastings algorithm consists of 3 updating steps.

Step 1

We enter the first step of the i th iteration with values $\theta^{(i-1)}$, $\lambda^{(i-1)}$, and $m^{(i-1)}$. In this step, we update the value of θ by generating a candidate value θ^* from a proposal distribution indexed by $\theta^{(i-1)}$ with density $q_\theta(\theta | \theta^{(i-1)})$. The candidate value is accepted with probability

$$\min \left(1, \frac{p(\theta^*, \lambda^{(i-1)}, m^{(i-1)} | n_0, \dots, n_m) q_\theta(\theta^{(i-1)} | \theta^*)}{p(\theta^{(i-1)}, \lambda^{(i-1)}, m^{(i-1)} | n_0, \dots, n_m) q_\theta(\theta^* | \theta^{(i-1)})} \right), \qquad (13)$$

where the density $p(\theta, \lambda, m | n_0, \dots, n_m)$ is as given in equation (11). If the candidate value is accepted, we assign $\theta^{(i)}$ equal to θ^* . Otherwise, $\theta^{(i)}$ is set equal to $\theta^{(i-1)}$;

Step 2

We enter the second step of the i th iteration with values $\theta^{(i)}$, $\lambda^{(i-1)}$, and $m^{(i-1)}$. In this step, we update the value of λ by generating a candidate value λ^* from a proposal distribution indexed by $\lambda^{(i-1)}$ with density $q_\lambda(\lambda | \lambda^{(i-1)})$. The candidate value is accepted with probability

$$\min \left(1, \frac{p(\theta^{(i)}, \lambda^*, m^{(i-1)} | n_0, \dots, n_m) q_\lambda(\lambda^{(i-1)} | \lambda^*)}{p(\theta^{(i)}, \lambda^{(i-1)}, m^{(i-1)} | n_0, \dots, n_m) q_\lambda(\lambda^* | \lambda^{(i-1)})} \right), \qquad (14)$$

If the candidate value is accepted, we assign $\lambda^{(i)}$ equal to λ^* . Otherwise, $\lambda^{(i)}$ is set equal to $\lambda^{(i-1)}$,

Step 3

We enter the third and last step of the i th iteration with values $\theta^{(i)}$, $\lambda^{(i)}$, and $m^{(i-1)}$. In this step, we update the value of m by generating a candidate value m^* from a proposal distribution with density $q_m(m|m^{(i-1)})$. The candidate value is accepted with probability

$$\min\left(1, \frac{p(\theta^{(i)}, \lambda^{(i)}, m^* | n_0, \dots, n_m) q_m(m^{(i-1)} | m^*)}{p(\theta^{(i)}, \lambda^{(i)}, m^{(i-1)} | n_0, \dots, n_m) q_m(m^* | m^{(i-1)})}\right), \quad (15)$$

If the candidate value is accepted, we assign $m^{(i)}$ equal to m^* . Otherwise, $m^{(i)}$ is set equal to $m^{(i-1)}$. This concludes the third step of the i th iteration, and we exit from it with the updated values $\theta^{(i)}$, $\lambda^{(i)}$, and $m^{(i)}$.

The specification of the proposal distributions $q_\theta(\cdot|\cdot)$, $q_\lambda(\cdot|\cdot)$, and $q_m(\cdot|\cdot)$ appearing in the steps above still remains. This is discussed in Section 5. It should be emphasized that the algorithm given above describes only one possible implementation of the single-component Metropolis-Hastings algorithm. A fuller discussion of this algorithm and other MCMC methods will not be presented at this time, since several such discussions are readily available in the texts by Carlin and Louis (1996, Section 5.4), Tanner (1996, Chapter 6), and Gelman, Carlin, Stern, and Rubin (1995, Chapter 11). Within the actuarial literature, the recent articles by Haastrup and Arjas (1996) and Scollnik (1995d) may prove instructive to a reader unfamiliar with these methods. Also, Pai (1997) discusses the use of MCMC to perform a Bayesian analysis to scrutinize the compound loss distribution.

5 NUMERICAL ILLUSTRATION

The data we analyse is taken from Adelstein (1949, p. 379) and gives the observed number of accidents in the age-group 26-30 years during the first year of service for a group of railyard shunters. The data appears in Table 1, and is underdispersed with a sample mean of 0.5815 and a sample variance of 0.5719. Consul and Famoye (1989) previously fit a truncated GPD model to this data and obtained the ML estimates $\hat{\theta} = 0.6115$ and $\hat{\lambda} = -0.0676$. However, Consul and Famoye (1989) proceeded by grouping the last three of the class frequencies appearing in Table 1 into a single class of frequencies greater than or equal to 4 and also appear to have set $m = 4$ for the purposes of estimation even though one worker experienced 6 accidents. Consequently, their ML estimates are adversely affected. Our own analysis will use the original form of the data presented by Adelstein.

TABLE I
 ADELSTEIN'S (1949) SHUNTERS ACCIDENTS DATA
 FIRST YEAR OF SHUNTING AGE 26-30 YEARS

<i>Number of Accidents</i>	<i>Number of Men</i>
0	121
1	85
2	19
3	1
4	0
5	0
6	1

We proceed to analyse Adelstein's data using the truncated GPD model along with each PDS introduced in Section 3. We utilise the MCMC method described in Section 4. A few specifics concerning the implementation of the Markov chain are worthy of note. For the univariate proposal distributions associated with the parameters θ and λ , we found that normal distributions centered at the current value of the parameter in question and with standard deviation of 0.05, that is

$$q_{\theta}(\theta|s) \sim \text{Normal}(s, 0.05) \quad \text{and} \quad q_{\lambda}(\lambda|s) \sim \text{Normal}(s, 0.05),$$

yielded acceptance rates in the 50 to 75 per cent range. The proposal distribution for the parameter m was taken to be *Poisson* with mean μ in the case of the analysis incorporating the second PDS, that is

$$q_m(m|s) \equiv q_m(m) \sim \text{Poisson}(\mu).$$

This makes Step 3 of the algorithm an independence sampler (Gilks, Richardson, and Spiegelhalter, 1996, page 9; also, Tierney, 1994, page 1706) since $q_m(m|m^{(t-1)})$ no longer depends on the value of $m^{(t-1)}$. For the analysis incorporating the first PDS, exact draws of m from its full conditional posterior distribution were used. In this case, the acceptance probability (15) is always equal to 1. For the analysis incorporating the third PDS, no draws of m were required since this analysis assumed that the value of m was fixed and known.

For each analysis, four realizations of a Markov chain were simulated. Each chain was permitted to run for 10,000 iterations. The results of the first 5,000 iterations were discarded as 'burn-in', and convergence of the Markov chains for each analysis was formally monitored by applying the diagnostic of Gelman, Carlin, Stern, and Rubin (1995, page 330-333) to the output of iterations 5001 through 10,000. The behaviour of the realised Markov chain sample paths associated with one of the simulations (corresponding to the second PDS with $\mu = 10$) is illustrated in Figures 2, 3 and 4. In these plots, it

is apparent that the simulated Markov chains are well on their way towards convergence by the 100th iteration in each case. Estimated posterior distributions for the parameters θ , λ , and m are presented in Figures 5, 6 and 7. These posterior distributions are estimated on the basis of the 20,000 (4 times 5,000) simulated draws for each parameter from its posterior distribution.

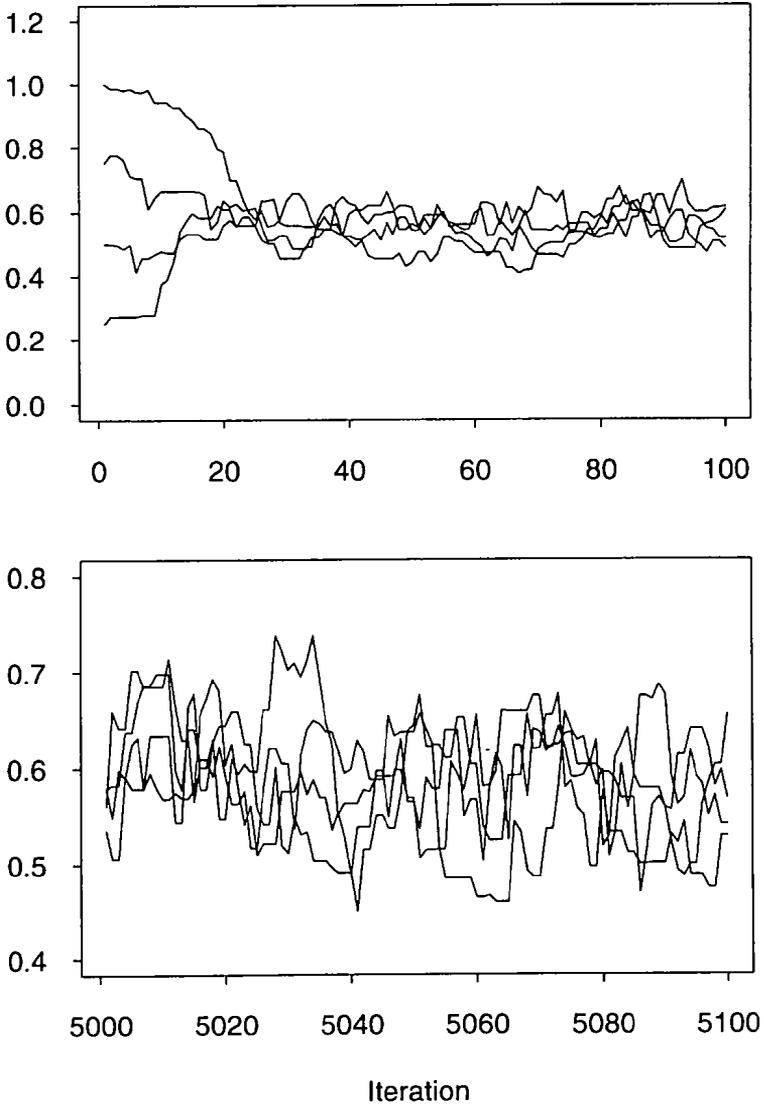


FIGURE 1 Sample Paths for the Parameter θ
Iterations 1 to 100 and 5000 to 5100
(Second PDS with $\mu = 10$)

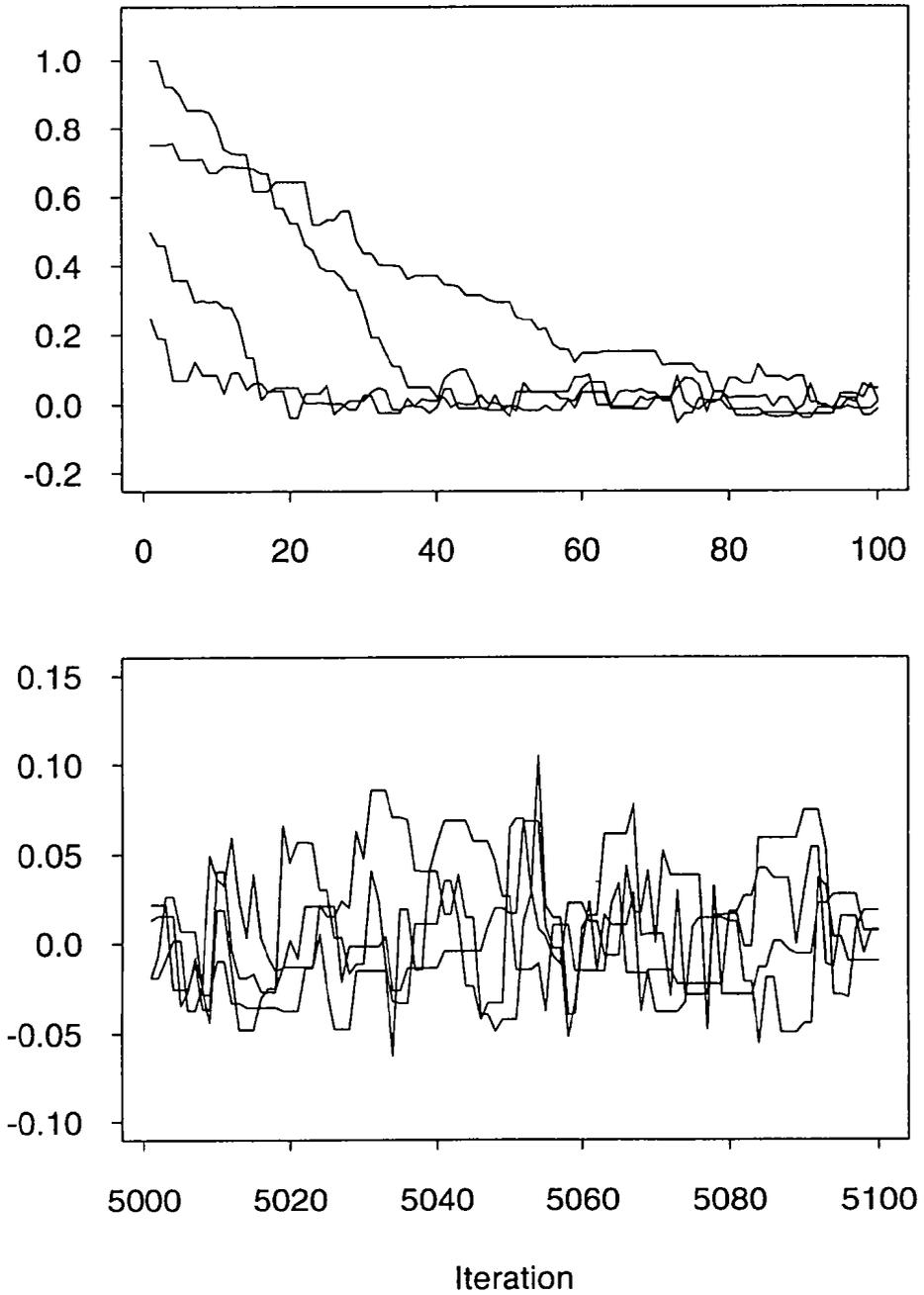
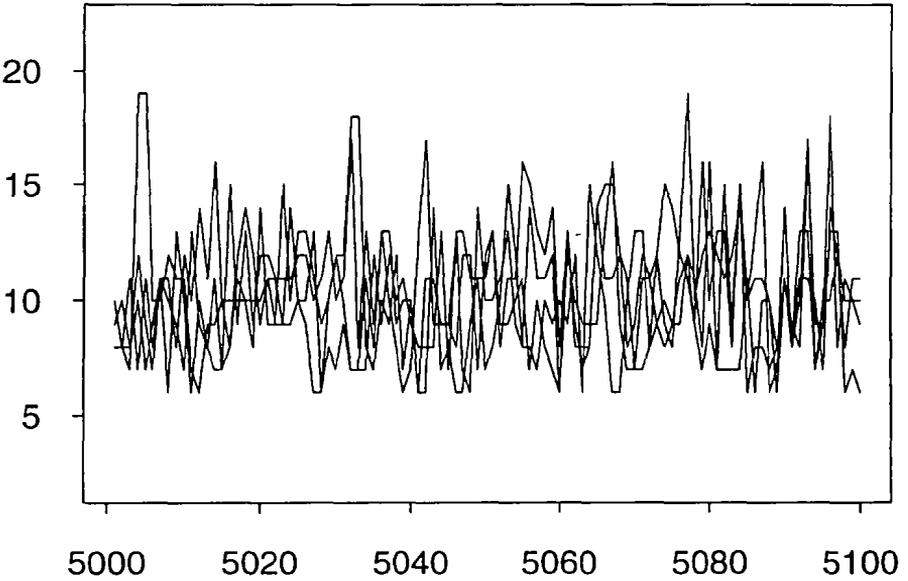
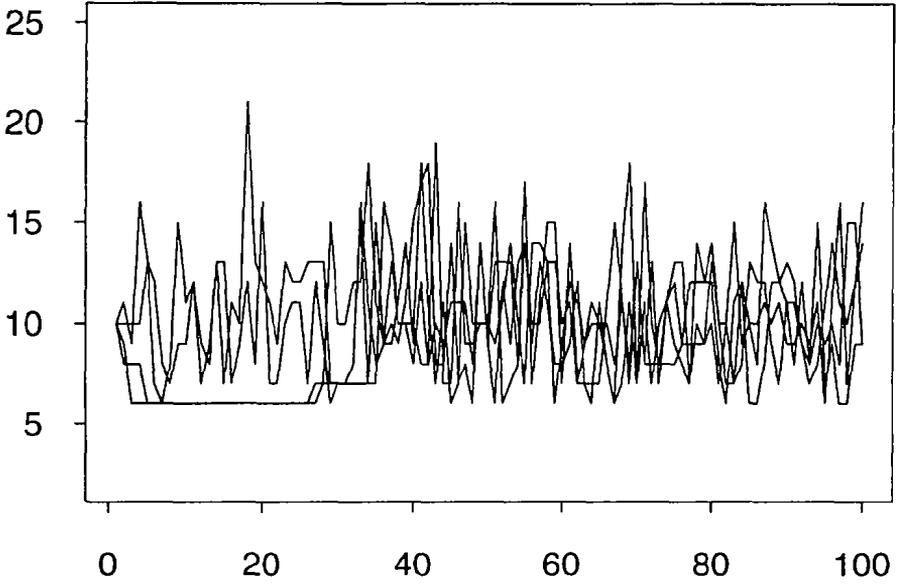


FIGURE 2 Sample Paths for the Parameter λ
Iterations 1 to 100 and 5000 to 5100
(Second PDS with $\mu = 10$)



Iteration

FIGURE 3 Sample Paths for the Parameter m
Iterations 1 to 100 and 5000 to 5100
(Second PDS with $\mu = 10$)

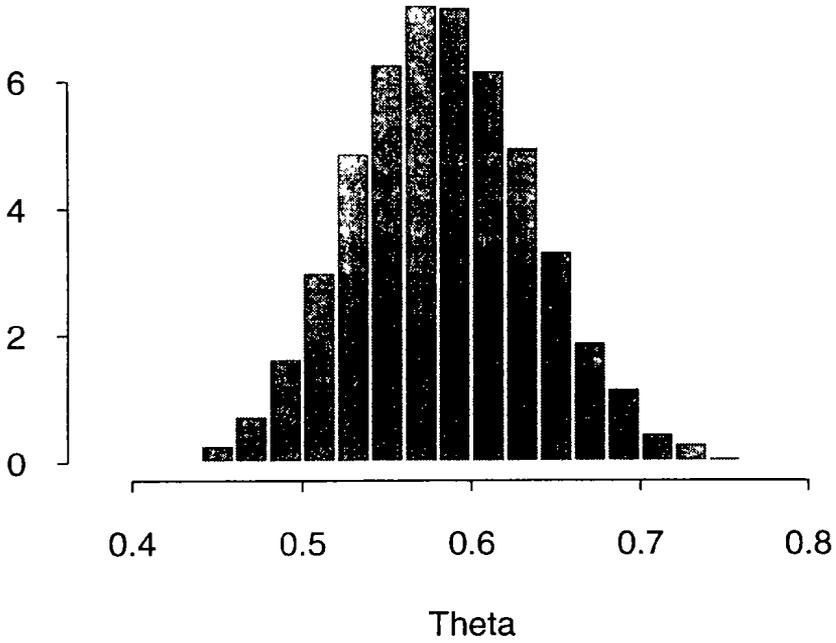


FIGURE 4 Estimated Posterior Density Functions for the Parameter θ
(Second PDS with $\mu = 10$)

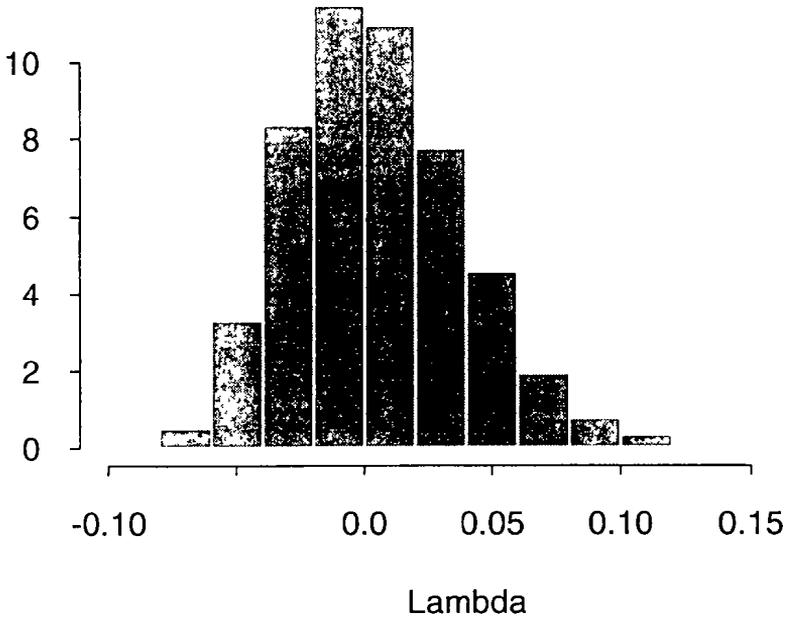


FIGURE 5 Estimated Posterior Density Functions for the Parameter λ
(Second PDS with $\mu = 10$)

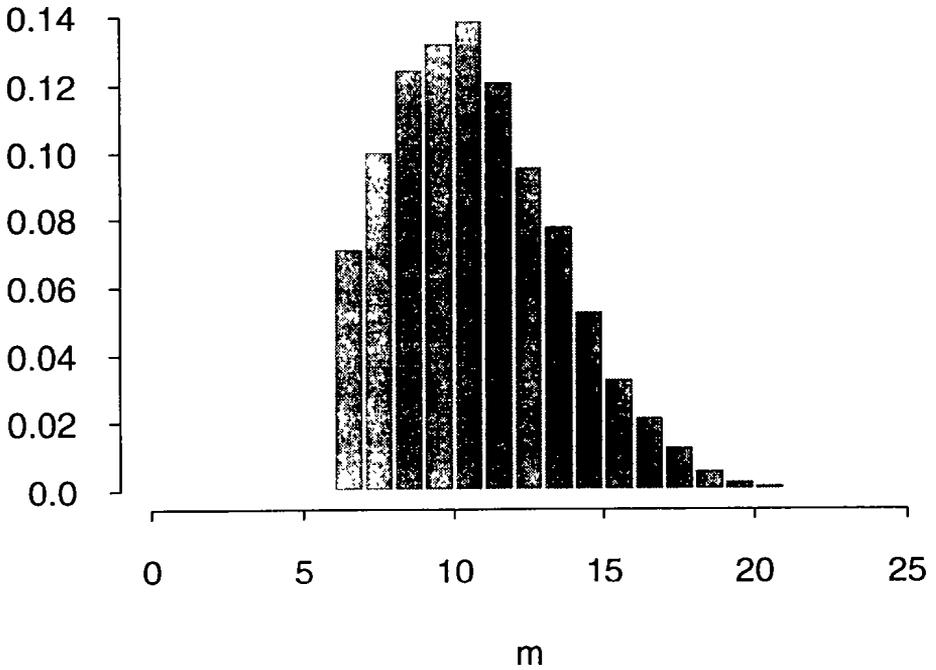


FIGURE 6 Estimated Posterior Density Functions for the Parameter m
(Second PDS with $\mu = 10$)

Summary results for all of our analyses appear in Tables 2 through 7. From Tables 3, 5, and 7, one can observe that predictive inferences are largely unaffected by the particular choice of PDS.

TABLE 2
ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ , λ , AND m
RESULTING UNDER THE FIRST PDS FOR 3 VALUES OF M ($m \leq M$)

<i>Parameter</i>	<i>M</i> = 6	<i>M</i> = 10	<i>M</i> = 25
θ	0.5837 (0.0556)	0.5861 (0.0541)	0.5807 (0.0536)
λ	0.0034 (0.0353)	0.0009 (0.0340)	0.0085 (0.0316)
m	6 (0)	8.0016 (1.4196)	14.9864 (5.7347)

TABLE 3

THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_j = n_j | n_0, \dots, n_m)$
 RESULTING UNDER THE FIRST PDS FOR 3 VALUES OF M ($m \leq M$)

n_j	$M = 6$	$M = 10$	$M = 25$
0	0.5587	0.5573	0.5603
1	0.3237	0.3251	0.3213
2	0.0951	0.0953	0.0952
3	0.0190	0.0189	0.0194
4	0.0030	0.0029	0.0032
5	0.0004	0.0004	0.0005
≥ 6	0.0001	0.0001	0.0001
<i>mean</i>	0.5857	0.5867	0.5858
<i>(sd)</i>	(0.7711)	(0.7697)	(0.7749)

TABLE 4

ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ , λ , AND m
 RESULTING UNDER THE SECOND PDS FOR 4 DIFFERENT VALUES OF μ

<i>Parameter</i>	$\mu = 2$	$\mu = 5$	$\mu = 10$	$\mu = 25$
θ	0.5810 (0.0529)	0.5831 (0.0544)	0.5828 (0.0531)	0.5774 (0.0529)
λ	0.0034 (0.0346)	0.0032 (0.0350)	0.0051 (0.0335)	0.0150 (0.0291)
m	6.32473 (0.6319)	7.2713 (1.4242)	10.3290 (2.7903)	24.6669 (5.0021)

TABLE 5

THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_j = n_j | n_0, \dots, n_m)$
 RESULTING UNDER THE SECOND PDS FOR 4 DIFFERENT VALUES OF μ

n_j	$\mu = 2$	$\mu = 5$	$\mu = 10$	$\mu = 25$
0	0.5602	0.5590	0.5591	0.5621
1	0.3231	0.3236	0.3230	0.3184
2	0.0945	0.0949	0.0951	0.0953
3	0.0188	0.0189	0.0192	0.0201
4	0.0030	0.0030	0.0031	0.0034
5	0.0004	0.0004	0.0004	0.0005
≥ 6	0.0001	0.0001	0.0001	0.0001
<i>mean</i>	0.5830	0.5849	0.5858	0.5863
<i>(sd)</i>	(0.7692)	(0.7704)	(0.7724)	(0.7803)

TABLE 6
ESTIMATED POSTERIOR MEANS AND SDs FOR THE PARAMETERS θ , λ , AND m
RESULTING UNDER THE THIRD PDS FOR 3 VALUES OF $M(m = M)$

Parameter	$M = 6$	$M = 10$	$M = 25$
θ	0.5844 (0.0532)	0.5838 (0.0545)	0.5748 (0.0511)
λ	0.0025 (0.0343)	0.0029 (0.0333)	0.0160 (0.0279)
m	6 (0)	10 (0)	25 (0)

TABLE 7
THE ESTIMATED PREDICTIVE DISTRIBUTION $Pr(N_f = n_f | n_0, \dots, n_m)$
RESULTING UNDER THE THIRD PDS FOR 3 VALUES OF $M(m = M)$

n_f	$M = 6$	$M = 10$	$M = 25$
0	0.5583	0.5586	0.5636
1	0.3243	0.3238	0.3175
2	0.0951	0.0951	0.0948
3	0.0189	0.0189	0.0200
4	0.0030	0.0030	0.0034
5	0.0004	0.0004	0.0005
≥ 6	0.0001	0.0001	0.0001
mean	0.5857	0.5854	0.5842
(sd)	(0.7700)	(0.7703)	(0.7796)

6. FUTURE RESEARCH THE CORRELATED TRUNCATED BGPD MODEL

Famoye and Consul (1995) and Vernic (1997) have both considered a BGPd (bivariate GPD) formed by applying the method of trivariate reduction. This method proceeds as follows: let N_1, N_2 and N_3 be independent GPD random variables with respective parameters $(\theta_1, \lambda_1), (\theta_2, \lambda_2),$ and (θ_3, λ_3) . Then the random vector (X, Y) is said to have a correlated BGPd if $X = N_1 + N_2$ and $Y = N_2 + N_3$. Unfortunately, both Famoye and Consul (1995) and Vernic (1997) implicitly permit the parameters $\lambda_i, i = 1, 2, 3,$ to take on negative values but fail to correct the definitions of the affected GPD and BGPd distributions by appropriately truncating and normalizing them.

In order to correct this problem, we define a correlated truncated BGPd by the method of trivariate reduction. Let N_1, N_2 and N_3 be independent truncated GPD random variables with respective parameters $(\theta_1, \lambda_1, m_1), (\theta_2, \lambda_2, m_2),$ and $(\theta_3, \lambda_3, m_3)$. Then the random vector (X, Y) will be said to

have a correlated truncated BGPD if $X = N_1 + N_2$ and $Y = N_2 + N_3$ as before. It should be possible to implement Bayesian posterior and predictive inferences for this distribution by using an extension of the MCMC method described in Sections 3 and 4 along with a data augmentation method to simulate the unobserved values of N_1 , N_2 and N_3 , given the observations X and Y along with the current simulated values of the parameters $(\theta_i, \lambda_i, m_i)$, $i = 1, 2, 3$. This procedure will be further explained, and also applied to a numerical example, in a paper to follow.

ACKNOWLEDGEMENTS

The author wishes to thank the referees for their comments and suggestions. This research was partially supported by a grant from the Natural Sciences and Research Council of Canada.

REFERENCES

- ADELSTEIN, A M (1949) Accident Proneness. A Criticism of the Concept Based Upon an Analysis of Shunters' Accidents *Journal of the Royal Statistical Society, A* **115**, 354-410
- ALZAIID, A A , and AL-OSH, M.A (1993) Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions *Ann Inst Statist Math* **45**, 223-232
- AMBAGASPIITIYA, R S and BALAKRISHNAN, N (1994) On the Compound Generalized Poisson Distributions *ASTIN Bulletin* **24** (2), 255-263
- CARLIN, B P , and LOUIS, T A (1996) *Bayes and Empirical Bayes Methods for Data Analysis* Chapman and Hall, New York
- CONSUL, P C (1994) Some Bivariate Families of Lagrangian Probability Distributions *Communications in Statistics Theory and Method* **23**, 2895-2906
- CONSUL, P C (1990) A Model for Distribution of Injuries in Auto-Accidents *Mitteilungen der Schweiz Vereinigung der Versicherungsmathematiker* **1**, 161-168
- CONSUL, P C (1989) *Generalized Poisson Distributions Properties and Applications* Marcel Dekker Inc , New York
- CONSUL, P C and FAMOYE, F (1989) The Truncated Generalized Poisson Distribution and its Estimation *Communications in Statistics Theory and Method* **18** (10), 3635-3648
- CONSUL, P C , and SHOUKRI, M M (1985) The Generalized Poisson Distribution when the Sample Mean is Larger than the Sample Variance *Communications in Statistics Simulation and Computation* **14** (3), 667-681
- CONSUL, P C , and JAIN, G C (1973) A Generalization of the Poisson Distribution *Technometrics* **15**, 791-799
- FAMOYE, F and CONSUL, P C (1995) Bivariate Generalized Poisson Distribution with Some Applications *Metrika* **42**, 127-138
- GELMAN, A , CARLIN, J B , STERN, H S , and RUBIN, D B (1995) *Bayesian Data Analysis* Chapman and Hall, New York
- GILKS, W R , RICHARDSON, S and SPIEGELHALTER, D J (1996) Introducing Markov Chain Monte Carlo In *Markov Chain Monte Carlo in Practice* Edited by Gilks, W R , Richardson, S , and Spiegelhalter, D J Chapman and Hall, New York
- GOOVAERTS, M J , and KAAS, R (1991) Evaluating Compound Generalized Poisson Distributions Recursively *ASTIN Bulletin* **21**, 193-197
- HAASTRUP, S , and ARJAS, E (1996) Claims Reserving in Continuous Time a Nonparametric Bayesian Approach *ASTIN Bulletin* **26** (2), 139-164
- HENZE, N and KLAR, B (1995) Bootstrap Based Goodness of Fit Tests for the Generalized Poisson Model *Communications in Statistics Theory and Method* **24** (7), 1875-1896

- HESSELAGER, O (1997) Recursions for a class of compound Lagrangian distributions *Swiss Association of Actuaries Bulletin* **1**, 95-101
- JOHNSON, N L , KOTZ, S , and KEMP, A W (1992) *Univariate Discrete Distributions* Second Edition John Wiley & Sons Inc, New York
- KLUGMAN, S A , PANJER, H H , and WILMOT, G E (1997) *Loss Models From Data to Decisions* In press
- NELSON, D L (1975) Some Remarks on Generalizations of the Negative Binomial and Poisson Distributions *Technometrics* **17**, 135-136
- PAI, J S (1997) Bayesian Analysis of Compound Loss Distributions *Journal of Econometrics* **79**, 129-146
- SCOLLNIK, D P M (1995a) The Bayesian Analysis of Generalized Poisson Models for Claim Frequency Data Utilising Markov Chain Monte Carlo Methods *Actuarial Research Clearing House* **1995.1**, 339-356
- SCOLLNIK, D P M (1995b) Bayesian Analysis of Two Overdispersed Poisson Models *Biometrics* **51**, 1117-1126
- SCOLLNIK, D P M (1995c) Bayesian Analysis of Two Overdispersed Poisson Regression Models *Communications in Statistics Theory and Method* **24** (11), 2901-2918
- SCOLLNIK, D P M (1995d) Simulating Random Variates from Makeham's Distribution and from Others with Exact or Nearly Log-concave Densities (with discussion) *Transactions of the Society of Actuaries* **1995 XLVII**, 409-454
- TANNER, M A (1996) *Tools for Statistical Inference Methods for the Exploration of Posterior Distributions and Likelihood Functions* Third Edition Springer-Verlag, New York
- TIERNEY, L (1994) Markov Chains for Exploring Posterior Distributions (with discussion). *Annals of Statistics* **22**, 1701-1762
- VERNIC, R (1997) On the Bivariate Generalized Poisson Distribution *Astin Bulletin* **27** (1), 23-31

DAVID P.M. SCOLLNIK
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, Canada
scollnik@acs.ucalgary.ca