ELSEVIER

Journal of Econometrics 79 (1997) 129–146

JOURNAL OF Econometrics

# Bayesian analysis of compound loss distributions

Jeffrey S. Pai

*Warren Centre for Actuarial Studies & Research, University of Manitoba, Winnipeg, Manitoba R3T 5V4, Canada*

Received 1 October 1995; received in revised form 1 July 1996

**Abstract**

Bayesian analysis is performed to scrutinize the compound loss distribution using sampling based methods. Both the number and the size of the losses are treated in a stochastic manner. Model selection, forecasting and reinsurance are studied from the predictive distribution. Model uncertainty is incorporated in forecasting through the use of posterior probabilities. The variation of the aggregate claim amount is analyzed under different reinsurance treaties. The methodology for modeling collective distributions of insurance losses is illustrated by an example.

*Key words:* Collective loss; Forecasting; MCMC; Model uncertainty; Posterior probability
*JEL classification:* C11; C13; C15

## 1. Introduction

Insurance is a mechanism for spreading out the losses caused by occurrences of unexpected events. The premium charged for a particular contract must be based upon the underlying loss process. Therefore, it is often necessary for actuaries to estimate probability distributions to describe the loss process.

Classical actuarial mathematics was based on deterministic approaches. However, uncertainty is a fundamental characteristic of the insurance business. The number and severity of claims often appear to vary in a random manner. The study of the stochastic features of insurance business is needed to support classical deterministic techniques.

One of the main concerns both in practical management of an insurance company and in theoretical considerations is the aggregate claim amount. The

The author would like to thank Professor Arnold Zellner, an Associate Editor, and two anonymous referees for their comments and suggestions on earlier versions of the paper.


*PII* S0304-4076(97)00010-9

aggregate claim is the sum which the insurer has to pay on the occurrence of some insured event. The p.d.f. of the aggregate claim amount can be found from the convolution technique. However, such evaluation is usually extremely difficult and numerical methods frequently are required (Bohman and Esscher, 1964; Seal, 1977).

Panjer and Willmot (1983) deal with the relationship between the compound Poisson distribution and Bayesian uncertainty models and show that the compound Poisson distribution arises in many situations in the theory of risk. They consider risk to be varied from individual to individual and allow risk to have a distribution (risk distribution). Panjer and Willmot (1983, p. 66) comment as follows: 'the operational actuarial interpretation is that the risk is first selected from the whole set of risks in accordance with the risk distribution, and the performance of the selected risk is then monitored. The statistical interpretation is essentially Bayesian.' The risk distribution is simply the prior distribution of the risk parameter.

The advantage of the Bayesian approach is that it provides the modeler not merely with point estimates of the parameters but also with the entire posterior distribution from which various statistics of interest such as the posterior mean, median, the future aggregate claims and etc. may be derived. In general, the choice of the prior enables the modeler to incorporate available subjective information about the parameters. Hogg and Klugman (1984, p. 14) remark, 'actuaries are encouraged to introduce any sound a priori beliefs in the inference, whether Bayesian of not.'

The existing literature on the Bayesian analysis of compound loss distributions includes, but is not limited to, works by Miller and Hickman (1975), Rytgaard (1990), Hesselager (1993) and Herzog (1994). A partial survey of the related literature is given on pages 118 and 121 of Herzog's text. In this paper, the Bayesian analysis of the aggregate claims process will be implemented using sampling based methods. Sampling based methods are able to approximate any features of a marginal or posterior distribution by the ergodic average. The approximation can in principle be made as accurate as desired by increasing the length of the Markov Chain Monte Carlo (MCMC) simulation. Note that the accuracy is not limited by the amount of data. The attractiveness of sampling based methods is their conceptual simplicity and ease of implementation for users with available computing resources but without numerical analytic expertise.

This paper is organized as follows. Section 2 discusses the acceptance–rejection (A-R) method and the Metropolis-within-Gibbs. Section 3 describes a Bayesian approach for collective models. Bayesian model selection and forecasting are addressed in Sections 4 and 5. Section 6 describes three common reinsurance treaties which will be used in the illustration to reduce the risk fluctuation. The Bayesian approach is illustrated in Section 7.

## 2. Sampling based methods

Bayesian inference proceeds by obtaining marginal posterior distributions of the components of $\Theta$ as well as summary features of these distributions. The first sampling based method used is the acceptance–rejection (A-R) method (von Neumann, 1951). The A-R method is useful if we wish to sample from a p.d.f. $\pi$ but have a way to sample from the p.d.f. $\pi_0$. The A-R method retains the sampled values $\Theta$ from $\pi_0$ with a probability depending on $\Theta$. Suppose we are able to calculate the value $c$ such that $c > \sup \pi(\Theta)/\pi_0(\Theta)$. To sample from $\pi$, generate $\Theta$ from $\pi_0$ and $u$ from $U(0,1)$. Retain $\Theta$ if $cu \leqslant \pi(\Theta)/\pi_0(\Theta)$.

The other sampling based method used is the Metropolis-within-Gibbs method (Marriott et al., 1995; Chib and Greenberg, 1994, 1995a; Müller, 1994). The customary Gibbs sampler (Tanner and Wong, 1987; Gelfand and Smith, 1990) makes draws from the complete conditional distributions (CCD's) in some systematic order. However, modifications are required when the CCD's have non-standard forms. The Metropolis algorithm (Metropolis et al., 1953; Hastings, 1970; Tierney, 1994; Chib and Greenberg, 1995b) is utilized for each non-standard draw. The use of such trajectories within an overall trajectory of the Gibbs sampler produces a Markov chain whose stationary distribution is the required posterior distribution.

## 3. Collective model

Let $n_t$ denote the number of claims produced by a portfolio of policies in a given time period $t$, $t=1,\ldots,T$. Let $z_{t,j}$ denote the amount of the $j$th claim of period $t$. Then

$$x_t = \sum_{j=1}^{n_t} z_{t,j}, \qquad x_t = 0 \quad \text{if } n_t = 0, \tag{1}$$

represents the aggregate claims generated by the portfolio for the period $t$. Suppose $n_t$ follows a non-negative discrete probability density $g(n_t|\phi)$ with parameter vector $\phi$, and $z_{t,j}$ follows a non-negative continuous probability density $f(z_{t,j}|\theta)$ with parameter vector $\theta$. Assume $z_{t,j}$ are independent of $n_t$ for all $t$ and $j$. Let $Z_t = (z_{t,1},\ldots,z_{t,n_t})'$ and $\Theta = (\phi', \theta')'$. The joint distribution of $(n_t, Z_t)$ can be written as

$$h(n_t, Z_t|\Theta) = g(n_t|\Theta) f(Z_t|\Theta). \tag{2}$$

Let $N = (n_1,\ldots,n_T)'$, $Z = (Z_1',\ldots,Z_T')'$ and $D = (N', Z')'$. The joint likelihood of $D$ can be written as

$$h(D|\Theta) = g(N|\Theta) f(Z|\Theta, N). \tag{3}$$

Table 1
Scale of evidence for assessing Bayes factors (Jeffreys, 1961)

| Range | Evidence |
|---|---|
| $B_{ij} > 1$ | Supports $M_i$ |
| $1 > B_{ij} > 10^{-1/2}$ | Slight evidence against $M_i$ |
| $10^{-1/2} > B_{ij} > 10^{-1}$ | Moderate evidence against $M_i$ |
| $10^{-1} > B_{ij} > 10^{-2}$ | Strong evidence against $M_i$ |
| $10^{-2} > B_{ij}$ | Decisive evidence against $M_i$ |

Given a proper prior on $\Theta$, say $\pi(\Theta)$, the posterior is the product of the prior and the likelihood function divided by the marginal distribution of the data:

$$\pi(\Theta|D) = \frac{h(D|\Theta)\pi(\Theta)}{\int h(D|\Theta)\pi(\Theta)\,\mathrm{d}\Theta}. \tag{4}$$

The sampling based methods suggested in Section 2 are performed to get samples from the posterior, say $\Theta^j$, $j = 1,\ldots,J$ (see Section 7 for an illustration).

## 4. Model selection

The selection of a suitable model is extremely important in the applications of mathematics to the real word. Suppose we are to select among models $M_k$, $k = 1,\ldots,K$, where $M_k$ is defined in a way such that $D$ has density $h_k(D|\Theta_k)$. Let $\pi_k(\Theta_k)$ be the prior density. Then the marginal density of model $k$ is

$$m_k(D) = \int h_k(D|\Theta_k)\pi_k(\Theta_k)\,\mathrm{d}\Theta_k. \tag{5}$$

Let $P(M_k)$ be the prior probability of $M_i$ and define the posterior probability of $M_k$ as

$$P(M_k|D) = \frac{P(M_k)m_k(D)}{\sum_{k=1}^{K} P(M_k)m_k(D)}. \tag{6}$$

The posterior probability of $M_k$ shows the probability that the observed data support model $k$ among all the models under consideration. The ratio, $B_{ij}$, of the posterior probability of model $i$ to model $j$, is the well known Bayes factor. Thus $B_{ij}$ indicates the relative support for model $i$ over model $j$. Table 1 shows Jeffreys's (1961, Appendix B) scale of evidence for assessing Bayes factors.

## 5. Forecasting

Regarding the model uncertainty, it is sometimes *ad hoc* to believe that the true model is really one of the models being considered. Actually, there is no need to select a model if the prediction is the ultimate goal. All models can be kept in the analysis to incorporate the model uncertainty and their effects will be properly weighted by $P(M_k|D)$ (Berger and Pericchi, 1993). The distribution of the future data given the observed is then a mixture of the future distribution of all models being considered. That is

$$\begin{aligned} h(D_F|D) &= \sum_{k=1}^{K} h_k(D_F, M_k|D) \\ &= \sum_{k=1}^{K} P(M_k|D) h_k(D_F|D, M_k), \end{aligned} \tag{7}$$

where $D_F = (n_F, Z_F')'$, $Z_F = (z_{F,1}, \ldots, z_{F,n_F})'$, denotes the future data. The future distribution of model $k$, $h_k(D_F|D, M_k)$, can be approximated by the Monte Carlo integration as follows. That is

$$\begin{aligned} h_k(D_F|D, M_k) &= \int h_k(D_F, \Theta_k|D, M_k)\,\mathrm{d}\Theta_k \\ &= \int h_k(D_F|\Theta_k, D, M_k)\pi_k(\Theta_k|D, M_k)\,\mathrm{d}\Theta_k \\ &\approx J^{-1} \sum_{j=1}^{J} h_k(D_F|\Theta_k^j, D, M_k). \\ &= J^{-1} f_k(Z_F|n_F, \Theta_k^j, M_k) g_k(n_F|\Theta_k^j, M_k). \end{aligned} \tag{8}$$

Samples from $h_k(D_F|D, M_k)$ are drawn as follows. For each $\Theta_k^j$, first draw $n_F^j$ from $g_k(n_F|\Theta_k^j, M_k)$, and then draw $z_{F,i}^j$, $i = 1, \ldots, n_F^j$ from $f_k(Z_F|n_F, \Theta_k^j, M_k)$. Samples from the future distribution of all models are then collected as stated in (7) weighted by the posterior probabilities. The future aggregate claim is then obtained from

$$x_F^j = \sum_{i=1}^{n_F^j} z_{F,i}^j, \qquad x_F^j = 0 \quad \text{if } n_F^j = 0. \tag{9}$$

## 6. Deductible and retention

Many insureds and insurers choose to exclude coverage for small losses (deductible). This reduces the cost of the coverage purchased and eliminates the need for the insurer to become involved in the processing of small claims, which is often a relatively expensive and time-consuming process. Under this contract, the insured agrees to absorb the full cost of losses which fall below the deductible $d$. One form of the deductible is the franchise deductible in which the entire deductible is waived as soon as the amount of loss exceeds the deductible amount. As the size of the deductible increases, at some point there is a change in terminology and it becomes known as retention. Retentions are common in contracts between an insurer and its reinsurer. In order to keep the variation of the aggregate claim amount $x$ reasonable, an insurer usually seeks protection against losses arising from large claim amounts, and excessively numerous claims by reinsuring large claim amounts of high claim frequency with a reinsurance company. Three reinsurance treaties are considered here.

### *6.1. Per claim excess of loss reinsurance*

In the per claim excess of loss reinsurance, each claim is divided between the cedant (insurer) and the reinsurer. The reinsurer pays the excess $z_{\mathrm{re}}(R) = \mathrm{Max}[(z-R), 0]$ over an agreed retention amount $R$ while the cedant retains the amount $z_{\mathrm{ced}}(R) = \min[z, R]$. From the cedant's point of view, the per claim excess of loss reinsurance is not influenced by the number of claims. Protection may be poor when the aggregate claims are large as a result of a high number of claims.

### *6.2. Aggregate excess of loss reinsurance*

This treaty provides protection not only against large individual claims but also against the fluctuation in the number of claims. The cedant retains the amount $x_{\mathrm{ced}}(R) = \min[x, R]$ while the reinsurer pays the excess $x_{\mathrm{re}}(R) = \mathrm{Max}[(x - R), 0]$.

### *6.3. Quota share reinsurance*

Under this treaty, each claim is divided between the cedant and reinsurer in a predetermined ratio $r$, $0<r<1$ (i.e., $z_{\mathrm{ced}}(r) = rz$ and $z_{\mathrm{re}}(r) = (1 - r)z$).

Both deductible and retention will be considered in the illustration. Other types of deductible and retention can be applied accordingly.

## 7. Illustration

The Bayesian analysis developed in this article is illustrated by the data taken from Rytgaard (1990). The claims are presented in Table 2 grouped by year ($T = 5$). This set of data has been studied by Hesselager (1993) using a Pareto distribution with conjugate priors.

The aggregate loss is analyzed through the compound negative binomial models. Define $n_t$ as the number of losses greater than $d = 1.5$ in year $t$, $t = 1, \dots, 5$. Consider a hierarchical Poisson model where the Poisson parameter $\lambda$ follows a Gamma distribution with hyperparameters $s$ and $\beta$. This is equivalent to a negative Binomial distribution

$$g(n_t|s, p) = \frac{\Gamma(s + n_t)}{\Gamma(n_t + 1)\Gamma(s)} p^s (1 - p)^{n_t}, \quad 0 < p \leqslant 1, \ s > 0, \tag{10}$$

Table 2
Claim exceeding $d = 1.5$ million during $T = 5$ years

| Year no. | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 2.495 | 1.985 | 3.215 | — | 19.180 |
| 2.120 | 1.810 | 2.105 | — | 1.915 |
| 2.095 | 1.625 | 1.765 | — | 1.790 |
| 1.700 | — | 1.715 | — | 1.755 |
| 1.650 | — | — | — | — |

where $p = 1/(\beta + 1)$. $\Gamma(\,)$ denotes the Gamma function. This type of model has been used successfully to represent 'accident-proneness' and to model the number of accidents for motorists (Greenwood and Yule, 1920; Weber, 1971).

The joint likelihood of $N$ can be written as

$$g(N|s,p) = \frac{\prod_{i=1}^{T} \Gamma(s+n_t)}{\prod_{i=1}^{T} \Gamma(n_t+1)\Gamma^T(s)} p^{Ts}(1-p)^n, \qquad n = \sum_{t=1}^{T} n_t. \tag{11}$$

Let $\pi(p) = 1$, $0 < p \leqslant 1$ and $\pi(s) = 1/Q$, $0 < s \leqslant Q$ with a sufficiently large $Q$, say, $Q = 1000$. The joint distribution of data and the parameters is obtained from the product of the likelihood and the priors

$$h(N,s,p) = \frac{\prod_{t=1}^{T} \Gamma(s+n_t)}{\prod_{t=1}^{T} \Gamma(n_t+1)\Gamma^T(s)} p^{Ts}(1-p)^n Q^{-1}. \tag{12}$$

The conditional distribution of $p$ given $(s,N)$, which is proportional to (12), follows a beta distribution

$$\pi(p|s,N) \sim \mathrm{Beta}(Ts+1, n+1). \tag{13}$$

The marginal posterior distribution of $s$ can be obtained from (12) as follows:

$$\begin{aligned} \pi(s|N) \propto h(N,s) &= \int_0^1 h(N,s,p)\,\mathrm{d}p \\ &\propto \frac{\prod_{t=1}^{T} \Gamma(s+n_t)}{\Gamma^T(s)} \frac{\Gamma(Ts+1)}{\Gamma(Ts+n+2)}. \end{aligned} \tag{14}$$

The posterior distribution of $p$ and $s$ is then the product of (13) and (14)

$$\pi(p,s|N) = \pi(p|s,N)\pi(s|N). \tag{15}$$

The sampling procedures are carried out as follows. First draw $s_j$, $j = 1,\ldots,J$ ($J = 10\,000$ throughout this example) from (14) by the A-R method. And then, for each $s_j$, draw $p_j$ from the beta distribution as in (13) to get a set of $J$

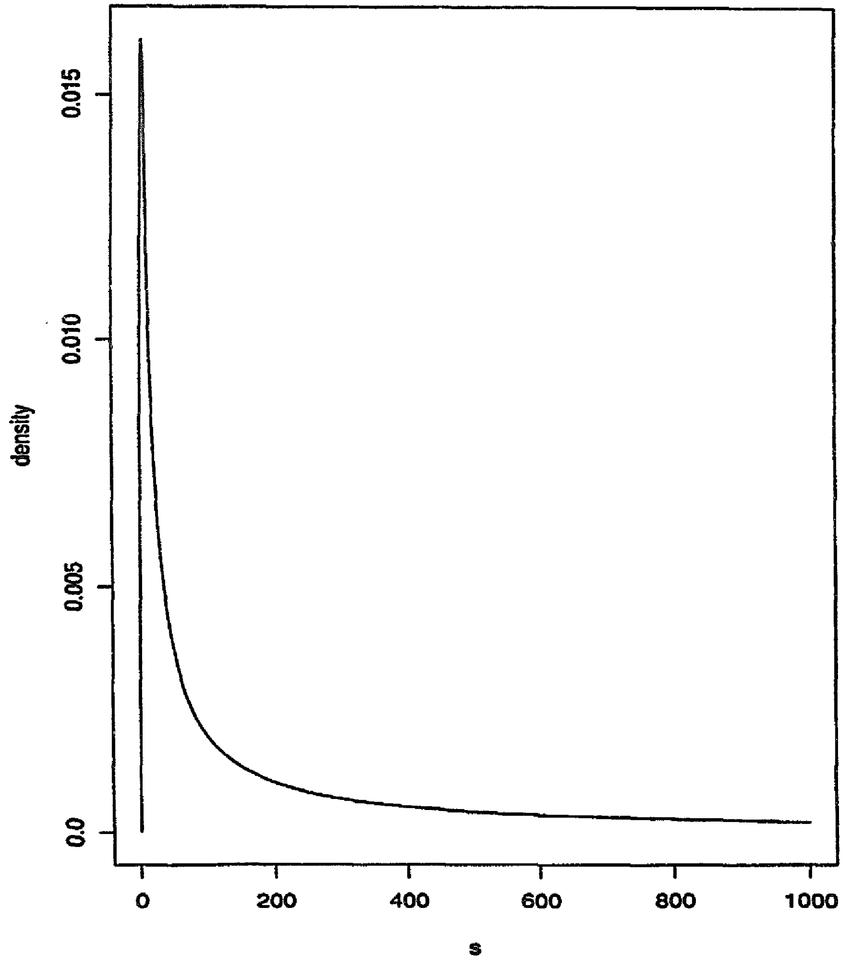


Fig. 1. Marginal posterior density of $s$.

samples of $(s, p)$. The marginal posterior density of $s$ is plotted in Fig. 1. The candidate-generating density $\pi_0$ is set to be the prior of $s$, $U(0,Q)$. The MLE of $s$ is 4.05 with density 0.016126. The maximum of $\pi(s|N)/\pi_0(s)$ over $(0,Q)$ is $c = 0.016126/0.001 = 16.126$. The probability of acceptance is $1/c = 0.062$. A candidate-generating density from beta distributions will improve the probability of acceptance. For example, the probability of acceptance will increase to 0.121 when $\pi_0$ is Beta(0.325, 1.268), $0 \leqslant s \leqslant Q$. The correlation between $s$ and $p$ is shown by the scatter plot in Fig. 2. The estimated future density of the claim number is shown in Table 3.

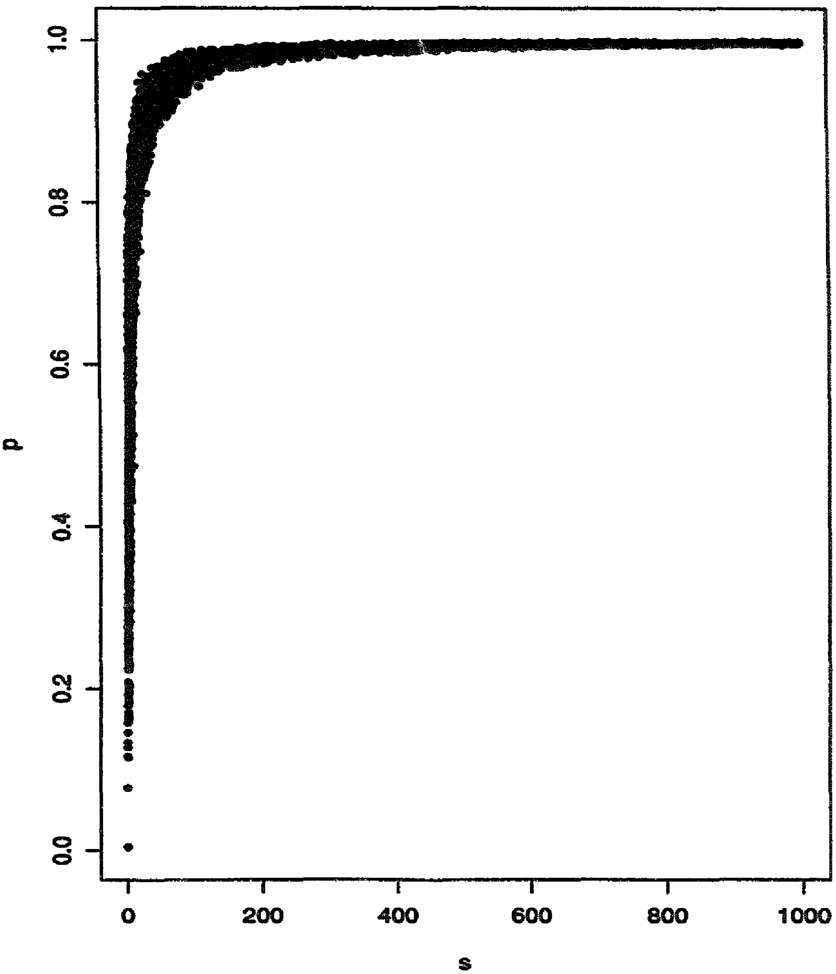


Fig. 2. Scatter plot of $s$ vs $p$.

Table 3
Estimated future density of the claim number

| | | | | | | |
|---|---|---|---|---|---|---|
| $i$ | 0 | 1 | 2 | 3 | 4 | 5 |
| $P(n_F = i)$ | 0.0524 | 0.1340 | 0.1886 | 0.1966 | 0.1588 | 0.1146 |
| $i$ | 6 | 7 | 8 | 9 | 10 | 11 |
| $P(n_F = i)$ | 0.0692 | 0.0410 | 0.0234 | 0.0090 | 0.0057 | 0.0026 |
| $i$ | 12 | 13 | 14 | 15 | 16 | 17 |
| $P(n_F = i)$ | 0.0017 | 0.0011 | 0.0008 | 0.0003 | 0.0001 | 0.0001 |

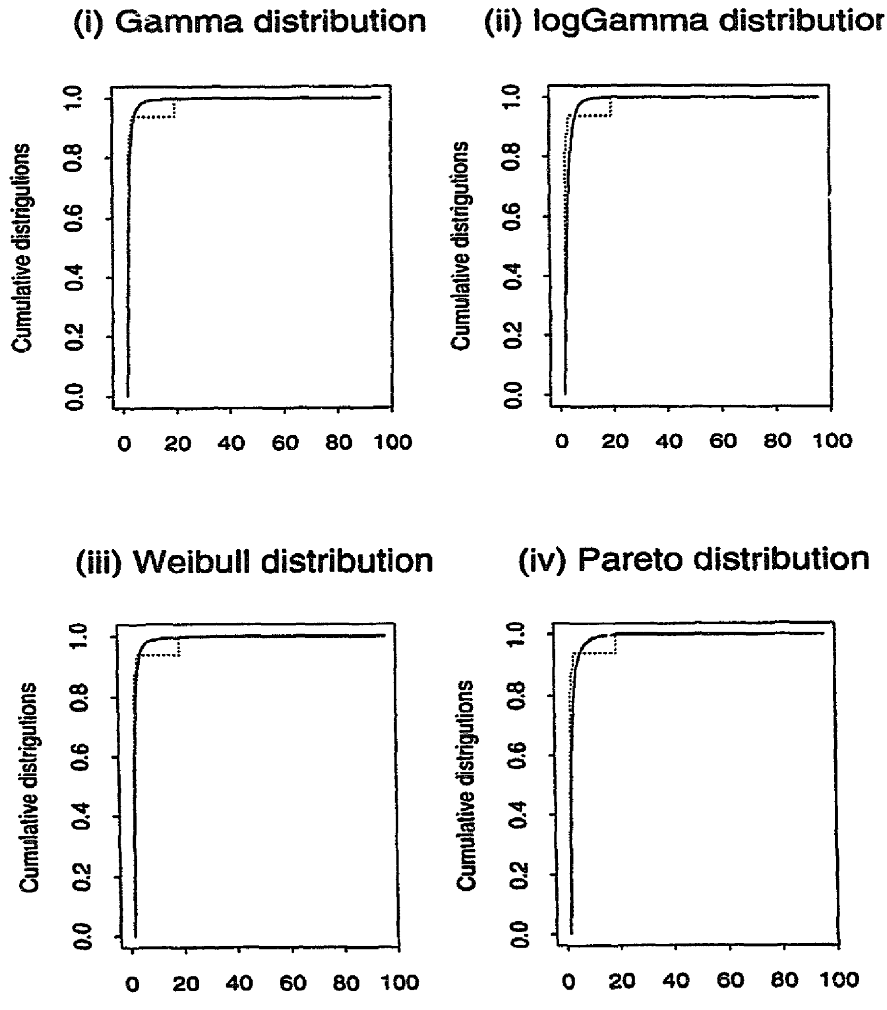


Fig. 3. Cumulative plots. (i) Gamma distribution, (ii) logGamma distribution, (iii) Weibull distribution, (iv) Pareto distribution.

Four common loss distributions together with the same claim number process (negative Binomial) are considered. The joint likelihood of the truncated loss distributions are shown in the Appendix. Other claim number process and claim size process can be done accordingly.

The cumulative plots are shown in Fig. 3. This will be discussed later in the paper.

Uniform proper prior for all the loss distribution parameters is used (i.e., $U(0,Q)$). $Q$ is chosen carefully to be 1000 where the posterior densities for

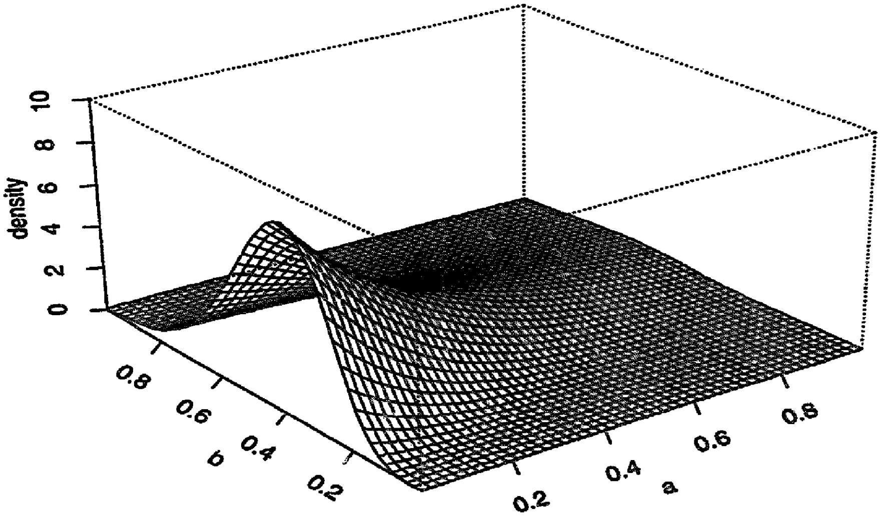


Fig. 4. Posterior density of the Gamma loss.

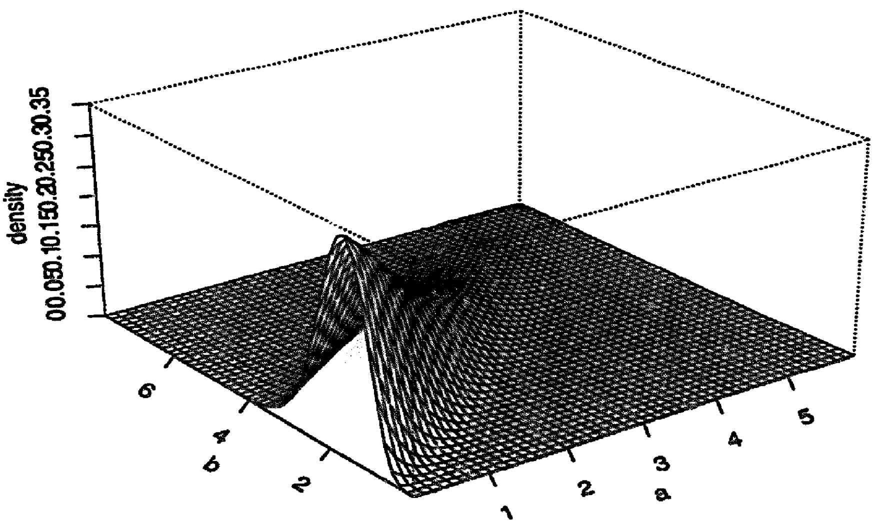


Fig. 5. Posterior density of the logGamma loss.

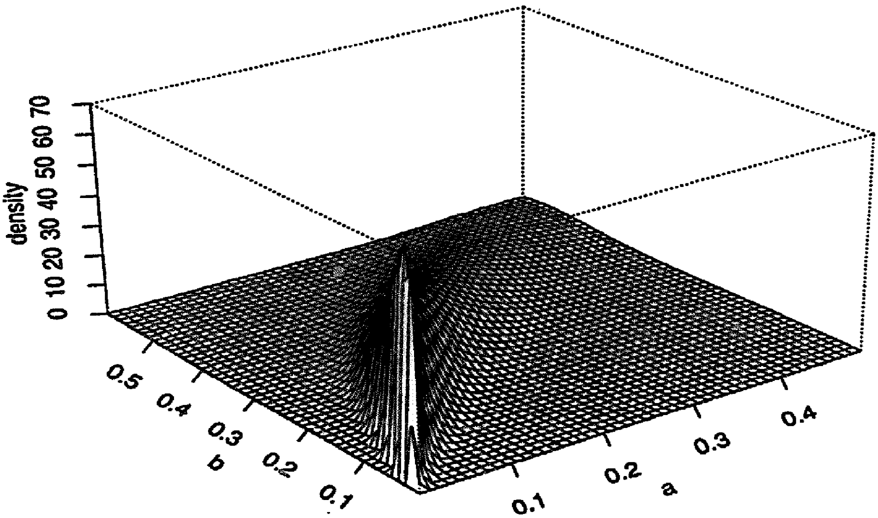


Fig. 6. Posterior density of the Weibull loss.

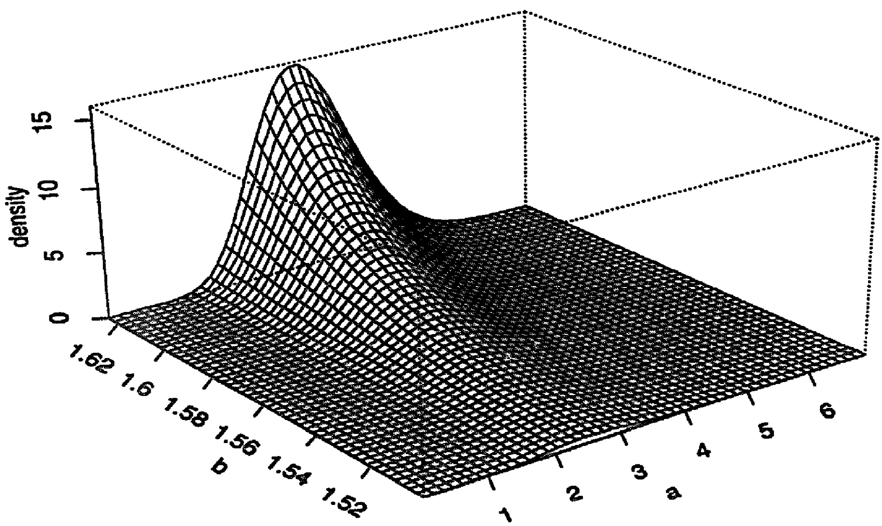


Fig. 7. Posterior density of the Pareto loss.

Table 4
Posterior features of loss distributions

| Loss function | Parameter | Mean | Standard deviation |
|---|---|---|---|
| Gamma | a | 0.3871 | 0.3365 |
| | b | 0.5325 | 0.1739 |
| logGamma | a | 1.6810 | 1.1908 |
| | b | 3.3609 | 1.3876 |
| Weibull | a | 0.2619 | 0.2336 |
| | b | 0.3340 | 0.1730 |
| Pareto | a | 3.0648 | 0.7562 |
| | b | 1.5914 | 0.0265 |

Table 5
Posterior probabilities

| Loss function | Posterior probability | Bayes factors in favor of logGamma |
|---|---|---|
| Gamma | 0.0001 | 6714 |
| logGamma | 0.6714 | 1 |
| Weibull | 0.0018 | 373 |
| Pareto | 0.3267 | 2.055 |

values greater than 1000 are negligible (see Figs. 4–7). The only exception is the $b$ parameter of the Pareto distribution where $b$ is $U(1.5, 1.65)$ (see Eq. (A.5)).

With the proper uniform priors, inferences are driven by the observed data. These priors also ensure that the marginal predictive distribution, which is used for model selection and prediction, is proper as well. The posterior modes are essentially the maximum likelihood estimates. The predictive density of model $k$ in (5) evaluated at the observed data is simply the area under the likelihood surface over the parameter space. This is in contrast to the classical approaches (e.g., Akaike information criterion and likelihood ratio test) where the density is obtained from one point (e.g., MLE).

The posterior distributions of the model parameters associated with the different size of loss models are shown in Figs. 4–7. Metropolis-within-Gibbs is used to sample from these posterior distributions. The posterior features are shown in Table 4.

The cumulative plots are shown in Fig. 3. The step-function cumulative probability distribution is constructed from the data. The solid lines are the empirical cumulative probability distributions of the individual future claims. The posterior probabilities in Table 5 are obtained numerically from (5) and (6) with equal prior probabilities (i.e., $P(M_k) = 1/4$). Bayes factors show decisive evidences against Gamma loss and Weibull loss.

Samples of future claims are collected as in (7) according to the weights (posterior probabilities) in Table 5. Samples of future aggregate claims are then

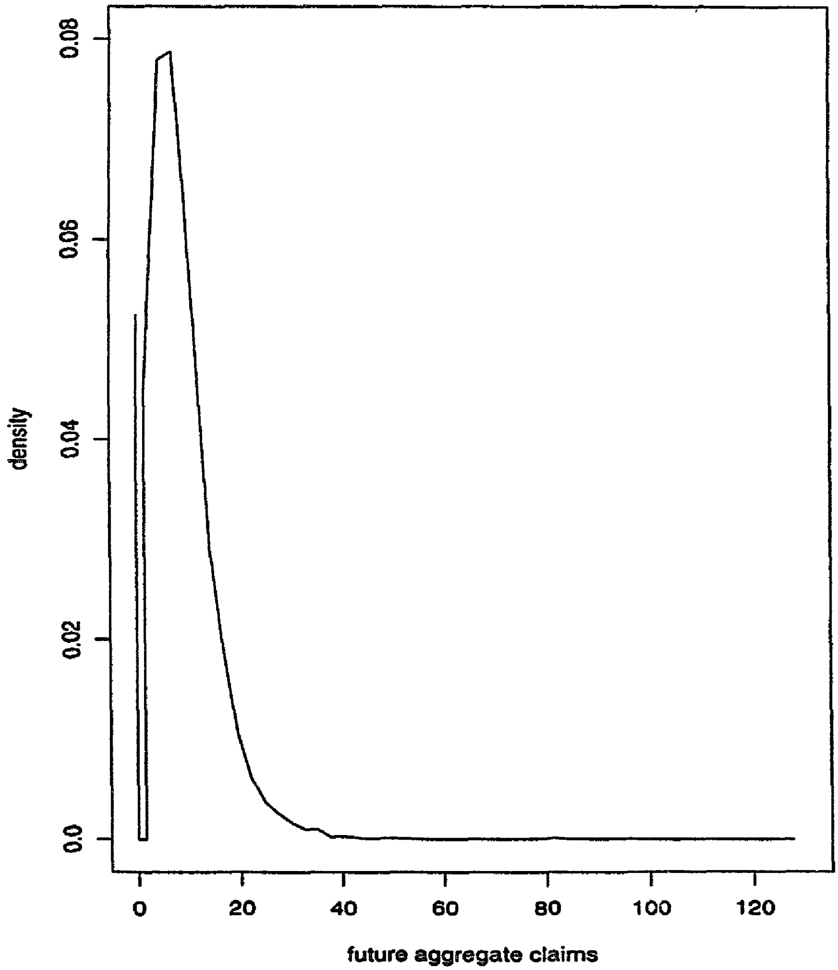


Fig. 8. Estimated density of the future aggregate claims.

obtained as in (9). Although the computer time required for this straightforward method is proportional to the expected number of claims, the expected number of claims (3.4189) is small in this example and the computer time is manageable. For example, it takes 0.2 second to sample 10 000 numbers from the negative binomial distribution and 34 189 numbers from the Pareto distribution to get 10 000 aggregate claims (This program is written in FORTRAN and run under a DEC station with a 60 MHz RISC processor). Pesonen (1989) introduced a random number generator for compound distributions which generates random number in a rate essentially independent of the number of claims. Pesonen's approach is recommended when the expected number of claims is large.

Table 6
Features of the future individual and aggregate distributions

| | Median | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Individual loss | 1.99 | 2.53 | 6.70 | 19.51 | 565.69 |
| Aggregate loss | 7.30 | 8.63 | 52.96 | 4.14 | 44.22 |

Table 7
Reinsurance results

| $r$ | 0.6 | 0.65 | 0.7 | 0.75 | 0.8 | 0.85 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|
| $E[x_{\text{ced}}]$ | 5.18 | 5.61 | 6.04 | 6.91 | 7.34 | 7.77 | 8.20 | 8.63 |
| Per claim excess of loss reinsurance | | | | | | | | |
| $R_r^p$ | 1.52 | 1.65 | 1.83 | 2.06 | 2.39 | 2.92 | 3.94 | 7.19 |
| $\text{Var}[x_{\text{ced}}]$ | 10.78 | 12.66 | 14.71 | 16.98 | 19.55 | 22.50 | 26.09 | 31.20 |
| $\text{Var}[x_{\text{re}}]$ | 27.73 | 26.53 | 25.29 | 23.94 | 22.45 | 20.73 | 18.44 | 14.50 |
| Aggregate excess of loss reinsurance | | | | | | | | |
| $R_r^a$ | 6.52 | 7.34 | 8.27 | 9.35 | 10.66 | 12.39 | 14.96 | 20.09 |
| $\text{Var}[x_{\text{ced}}]$ | 3.91 | 5.22 | 6.92 | 9.11 | 11.95 | 15.71 | 20.91 | 28.80 |
| $\text{Var}[x_{\text{re}}]$ | 39.81 | 37.30 | 34.51 | 31.46 | 28.05 | 24.16 | 19.63 | 13.90 |
| Quota share reinsurance | | | | | | | | |
| $\text{Var}[x_{\text{ced}}]$ | 19.07 | 22.38 | 25.95 | 29.79 | 33.90 | 38.27 | 42.90 | 47.80 |
| $\text{Var}[x_{\text{re}}]$ | 8.47 | 6.49 | 4.77 | 3.31 | 2.12 | 1.19 | 0.53 | 0.13 |

The future distribution of the individual loss is a truncated distribution ($P[z_{\text{F}} \geqslant 1.5] = 1$). The estimated future distribution of the aggregate losses (Fig. 8) is a mixture of a point mass at 0 ($P[x_{\text{F}} = 0] \approx 0.0524$) and a truncated distribution ($P[x_{\text{F}} \geqslant 1.5] \approx 0.9476$). Some features of the individual loss and the aggregate losses are listed in Table 6. The strongly skewed nature of the individual loss and the mixture result of the aggregate losses indicate that many standard estimation methods such as the Normal Power method, the Haldane method and the Wilson–Hilferty formula (Daykin et al., 1994, Sections 4.2.4 and 4.2.5) are unsuitable.

The study of the reinsurance is straightforward from here. From the cedant's point of view, the insurance company wants to keep as great a portion of the claims as possible (large retention). At the same time the insurance company wants to take out reinsurance cover to reduce risk fluctuation (small retention). Three reinsurance treaties suggested in Section 5 are compared under the same base that the cedant will keep the same portion, say $r$, of the total future claims. That is, for the per claim excess of loss reinsurance, $R_r^p$ is found such that $z_{\text{ced}}(R_r^p) = rE[z]$ and for aggregate excess of loss reinsurance, $R_r^a$ is found such that $x_{\text{ced}}(R_r^a) = rE[x]$. Expected values (risk premiums) and variances for all three treaties for different $r$ are listed in Table 7. A plot of the variances is shown in Fig. 9. The variances under the aggregate excess of loss reinsurance are always

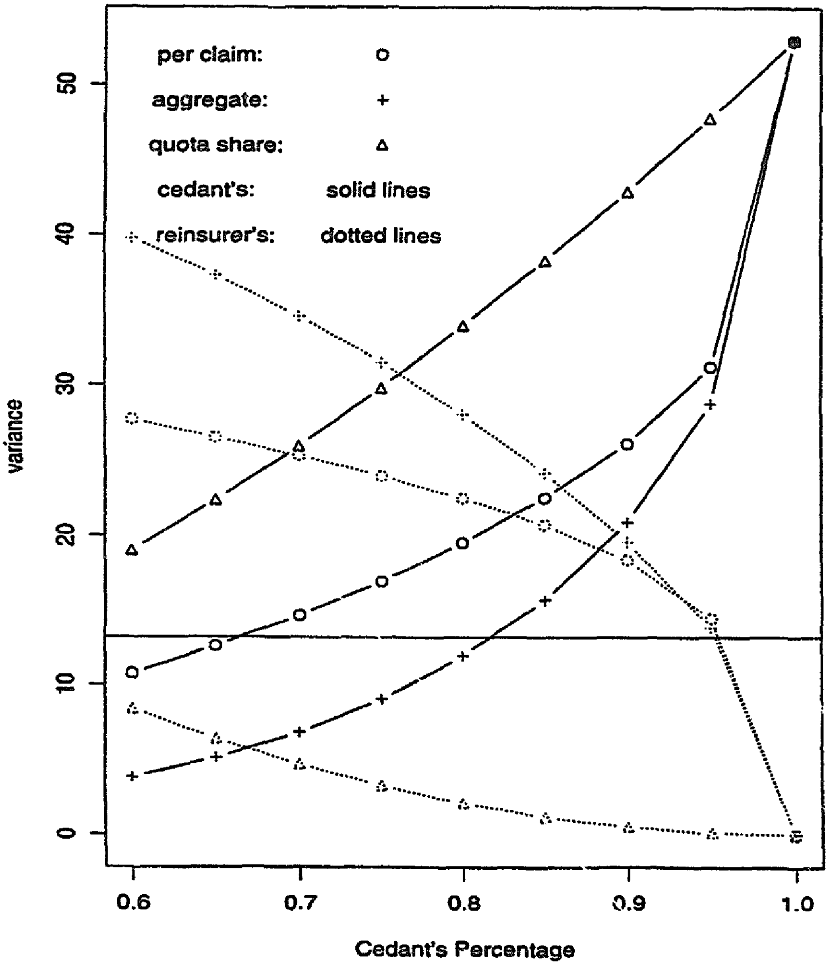


Fig. 9. Variances under different reinsurance treaties.

the smallest among all three treaties as a result that any optimal reinsurance arrangement should be applied directly to the aggregate amount $x$ (Pesonen, 1984). From Fig. 9 one can realize that the quota share reinsurance gives the worst protection from the cedant's point of view. For example, in order to reduce the variance of the future aggregate claim amount to 25% (the horizontal line, variance = 13.24) of the variance associated with the aggregate loss distribution, the insurer could retain more than 67% of claims under the per claim excess of loss reinsurance or 82% under the aggregate claim excess of loss reinsurance while the insurer could only keep up to 50% of claims under the quota share reinsurance.

## Appendix: Loss distributions

Truncated (from below) distributions are considered here as an example of franchise deductible. The truncated c.d.f., say $H_z(w)$, can be obtained from the untruncated c.d.f, say $H_y(w)$ as follows:

$$H_z(w)=\begin{cases} 0, & w\leqslant d \\ \frac{H_y(w)-H_y(d)}{1-H_y(d)}, & w>d. \end{cases} \tag{A.1}$$

Model 1: Gamma distribution

$$f_1(Z|\Theta_1,N)=\frac{b^{an}(\prod z_{t,j})^{a-1}\mathrm{e}^{-b\sum z_{t,j}}}{\Gamma^n(a)(1-\Gamma[a;bd])^n},\quad a,b\in R^+. \tag{A.2}$$

Model 2: logGamma distribution

$$f_2(Z|\Theta_2,N)=\frac{b^{an}(\prod \ln z_{t,j})^{a-1}}{(\prod z_{t,j})^{b+1}\Gamma^n(a)(1-\Gamma[a;b(\ln d)])^n},\quad a,b\in R^+. \tag{A.3}$$

Model 3: Weibull distribution

$$f_4(Z|\Theta_4,N)=(b/a)^n(\prod z_{t,j})^{b-1}\mathrm{e}^{-(\Sigma z_{t,j}^b-nd^b)/a},\quad a,b\in R^+. \tag{A.4}$$

Model 4: Pareto distribution

$$\begin{aligned} &f_5(Z|\Theta_5,N)=a^n b^{an}(\prod z_{t,j})^{-a-1},\quad a\in R^+,\\ &d\leqslant b\leqslant \min_{t,j}(z_{t,j}). \end{aligned} \tag{A.5}$$

All summations and products have the same form: $t=1,\ldots,T$ and $j=1,\ldots,n_t$. All $z_{t,j}$ are greater than or equal to $d$.

## References


Berger, J.O., Pericchi, L.R., 1993. The Intrinsic Bayes factor for model selection and prediction. Technical Report #93-43c, Department of Statistics, Purdue University.

Bohman, H., Esscher, F., 1964. Studies in risk theory with numerical illustrations concerning distribution functions and stop loss premiums, Part II. Skandinavisk Aktuarietidskrif 47, 1–40.

Chib, S., Greenberg, E., 1994. Bayes inference for regression model with ARMA($p,q$) errors. Journal of Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995a. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. Journal of Econometrics 68, 339–360.

Chib, S., Greenberg, E., 1995b. Understanding the Metropolis-Hastings algorithm. The American Statistician 49, 327–335.

Daykin, C.D., Pentikäinen, T., Pesonen, M., 1994. Practical Risk Theory for Actuaries. Chapman & Hall, London.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398–409.

Greenwood, M., Yule, G.U., 1920. An inquiry to the nature of frequency distributions for multiple happens. Journal of the Royal Statistical Society, Series A 83, 255–279.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Herzog, T.N., 1994. Introduction to Credibility Theory. ACTEX Publications, Winsted.

Hesselager, O., 1993. A class of conjugate priors with applications to excess-of-loss reinsurance. ASTIN Bulletin 23, 77–93.

Hogg, R.V., Klugman, S.A., 1984. Loss Distributions. Wiley, Toronto.

Jeffreys, H., 1961. Theory of Probability, 3rd ed. Oxford University Press, Oxford.

Marriott, J.M., Ravishanker, N., Gelfand, A.E., Pai, J.S., 1995. Bayesian analysis of ARMA processes: Complete sampling based inference under full likelihoods. In: Barry, D., Chaloner, K., Geweke, J., (Eds.), Bayesian statistics and econometrics: Essays in honor of Arnold Zellner. Wiley, New York, pp. 241–254.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal Chemical Physics 21, 1087–1091.

Miller, R.B., Hickman, J.C., 1975. Insurance credibility theory and Bayesian estimation. In: Kahn, P.M., (Ed.), Credibility: Theory and Applications. Academic Press, New York, pp. 249–270.

Müller, P., 1994. Metropolis posterior integration schemes. Stat. Comput., forthcoming.

Panjer, H.H., Willmot, G.E., 1983. Compound Poisson models in actuarial risk theory. Journal of Econometrics 23, 63–76.

Pesonen, M., 1984. Optimal reinsurance. Scandinavian Actuarial Journal 65–90.

Pesonen, M., 1989. Random number generators for compound distributions. Scandinavian Actuarial Journal 47–60.

Rytgaard, M., 1990. Estimation in the Pareto distribution. ASTIN Bulletin 20, 201–216.

Seal, H.L., 1977. Approximations to risk theory's $F(x,t)$ by mean of the gamma distribution. ASTIN Bulletin 9, 213–218.

Tierney, L., 1994. Markov Chain for exploring posterior distributions (with discussion). Annals of Statistics 22, 1701–1762.

von Neumann, J., 1951. Various techniques used in connection with random digits. National Bureau of Standards Series 12, 36–38.

Weber, D.C., 1971. Accident rate potential: An application of multiple regression analysis of a Poisson process. Journal of the American Statistical Association 66, 285–288.