International Journal of Business and Economics, 2007, Vol. 6, No. 3, 225-236

A Markov Chain Monte Carlo Approach to Estimate the Risks of Extremely Large Insurance Claims

Wan-Kai Pang*
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

Shui-Hung Hou
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

Marvin D. Troutt
Department of Management and Information Systems, Kent State University, U.S.A.

Wing-Tong Yu
School of Accounting and Finance, The Hong Kong Polytechnic University, Hong Kong

Ken W. K. Li
Department of Information and Communications Technology, The Hong Kong Institute of Vocational Education, Hong Kong

---

Abstract

The Pareto distribution is a heavy-tailed distribution often used in actuarial models. It is important for modeling losses in insurance claims, especially when we used it to calculate the probability of an extreme event. Traditionally, maximum likelihood is used for parameter estimation, and we use the estimated parameters to calculate the tail probability $\Pr(X > c)$ where c is a large value. In this paper, we propose a Bayesian method to calculate the probability of this event. Markov Chain Monte Carlo techniques are employed to calculate the Pareto parameters.

Key words: heavy-tail distributions; loss distribution model; Pareto probability distribution; Gibbs sampler

JEL classification: C0; C1; G22

Received March 29, 2007, revised December 28, 2007, accepted January 2, 2008.
*Correspondence to: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: mapangwk@inet.polyu.edu.hk. This research work was supported by the Research Committee of The Hong Kong Polytechnic University.

1. Introduction

Losses caused by unexpected events are problems for insurance companies. Actuaries want to know more about the distributional behavior of insurance losses and to identify the most appropriate probability distribution for large claims. Therefore more accurate evaluation of extreme event probabilities is desired. The two-parameter exponential distribution and the two-parameter Pareto distribution are often considered to be reasonable candidates by actuarial professionals. The probability density function of the two-parameter exponential distribution is:

$$f(x,\theta,\gamma)=\begin{cases}\frac{1}{\theta}e^{-\frac{x-\gamma}{\theta}} & \gamma<x<\infty\\ 0 & \text{otherwise.}\end{cases} \tag{1}$$

The probability density function of the two-parameter Pareto distribution is:

$$f(x,\alpha,\gamma_0)=\begin{cases}\frac{\alpha}{\gamma^{-\alpha}}x^{-\alpha-1} & \gamma<x<\infty\\ 0 & \text{otherwise.}\end{cases} \tag{2}$$

Here $\alpha,\theta>0$ and γ is known as the threshold parameter in both distributions.

In the next section, we describe an example from the literature on calculating right-tail probabilities associated with large losses. With this example, we illustrate the problem involved in standard approaches. In Section 2, we discuss the Bayesian approach in general, and in Section 3 the Markov Chain Monte Carlo (MCMC) estimation method is described along with its requirements. In Section 4 we apply the proposed MCMC method to the example and discuss applications to the conditional mean of large losses. Section 5 concludes.

1.1 An Example

Table 1 duplicates data in Chapter 3 of Hogg and Klugman (1984). These data are the amounts of 40 losses due to wind-related catastrophes in the US in 1977 recorded to the nearest US $1,000,000.

Table 1. Forty Losses Due to Wind-Related Catastrophes (Hogg and Klugman, 1984)

2	2	2	2	2	2	2	2	2	2
2	2	3	3	3	3	4	4	4	5
5	5	5	6	6	6	6	8	8	9
15	17	22	23	24	24	25	27	32	43

Hogg and Klugman (1984) wished to estimate the probability that a loss will exceed

US $29,500,000. This is equivalent to calculating $\Pr(X > 29.5)$ if the loss random variable X follows a certain probability distribution, while the empirical probability for this is $2/40 = 0.05$. They used the two-parameter exponential distribution and the two-parameter Pareto distribution as loss distribution models. Frequentist methods—maximum likelihood (ML) estimation and the method of moments (MM)—were used to estimate the model parameters. Their estimates for $\Pr(X > 29.5)$ were $\hat{p}_1 = 0.027$ under the exponential distribution model and $\hat{p}_2 = 0.040$ under the Pareto distribution model.

One interesting point about their methods is the estimation of the threshold parameter γ. They estimate γ to be 1.5, but the ML estimate for γ for both distributions is $Y_{(1)} = \min\{X_1, \ldots, X_n\}$ (see Johnson and Kotz, 1970), which is 2.0. Using this estimate and ML to estimate the second parameter in both distributions, estimates of $\Pr(X > 29.5)$ are $\hat{p}_1 = 0.022$ under the exponential distribution model and $\hat{p}_2 = 0.072$ under the Pareto distribution model.

In addition, Hogg and Klugman (1984) produced another ML estimate, $\hat{p}_3 = 0.036$, obtained by solving a system of nonlinear equations using the Newton-Raphson method (again using the estimate of γ as 1.5). In this way, they derived a 95% confidence interval estimate for $\Pr(X > 29.5)$ based on asymptotic normality of ML estimators: $0.036 \pm 0.048 = (-0.012, 0.084)$.

We see that different methods of estimation and distributional assumptions can produce different results. It is natural to ask which estimate is best.

2. A Bayesian Approach

The various methods of estimation considered in the example are based on the frequentist approach. This approach has several drawbacks. It is not difficult to obtain a point estimate for the unknown parameter, but it is rather difficult to construct an interval estimate. One often resorts to asymptotic normality and assumes that the sample size n is large enough, but estimation performance in small samples may be poor. Thus it is important to know how large n has to be in order to achieve a reasonable interval estimate. The approximate 95% confidence interval for $\Pr(X > 29.5)$ was noted above to be $(-0.012, 0.084)$. Strictly speaking, probabilities less than zero are nonsensical, and the interval estimate remains unsatisfactory even if we truncate this interval to be $(0.0, 0.084)$.

We now propose a Bayesian approach to solve this problem. The Bayesian paradigm makes use of data that have already been observed to form probability models and to make inferences. Probability densities of model parameters based on prior observations are used to inform the probability model and to estimate the predictive density of future events. A key characteristic of Bayesian methods is the use of probability to quantify uncertainty in inferences.

From a Bayesian point of view, there is no distinction between observables and parameters in a statistical model. That is, both data and parameters are considered random quantities. The process of Bayesian modeling can be summarized into the following four steps.

1. Build an appropriate probability model given the observed data using an appropriate joint probability distribution for observable and unobservable quantities in a problem. The model should be realistic in relation to the underlying scientific problem and to the data collected.

2. Form the posterior distribution. Let X denote the observed data, $\underline{\beta}$ the model parameters, and $P(X, \underline{\beta})$ the joint distribution of X and $\underline{\beta}$. Then:

$$P(X, \underline{\beta}) = P(\underline{\beta})P(X \mid \underline{\beta}), \tag{3}$$

where $P(\underline{\beta})$ is referred to as the prior distribution and $P(X \mid \underline{\beta})$ is the likelihood function. More abstractly, this can be expressed as:

$$\text{Joint probability model} = \text{Prior distribution} \times \text{Likelihood function}.$$

By Bayes' theorem,

$$P(\underline{\beta} \mid X) = \frac{P(\underline{\beta})P(X \mid \underline{\beta})}{\int P(\underline{\beta})P(X \mid \underline{\beta})\, d\underline{\beta}}. \tag{4}$$

This is called the posterior distribution of $\underline{\beta}$ and is the object of Bayesian inference.

3. Evaluate the final model. It is natural to ask the following questions after a final model is obtained: Does the final model fit the data? What are the implications of the resulting posterior distribution? Are the conclusions reasonable? To answer these questions, one needs to check the final model carefully. If necessary, one can return to Step 1 to alter or expand the model.

4. Conduct inference. Once the probability model is accepted, one can draw inferences about the model parameters and make predictions about the probabilities of future events. Often the first step is to construct $(1-\alpha)100\%$ probability, or credible, intervals for unknown quantities of interest. Such an interval can be regarded as having probability $1-\alpha$ of containing the unknown quantity; in contrast, a frequentist confidence interval may strictly be interpreted only in relation to a sequence of similar inferences that might be made in repeated practice. Increasing emphasis has been placed on interval estimation rather than hypothesis testing in areas of applied statistics (Chen et al., 2000). This provides a strong impetus to the Bayesian viewpoint. Turning to prediction, let y denote the observed data and $\tilde{y}$ the unknown but potentially observable quantities. Predictive inference is based on summarizing the posterior predictive distribution $P(\tilde{y} \mid y)$.

3. Markov Chain Monte Carlo Techniques

The method of Markov Chain Monte Carlo (MCMC) is essentially a Monte Carlo integration method using Markov chains. In Bayesian statistics, one often faces the problem of integrating over possibly high-dimensional probability distributions to make inferences about model parameters. Monte Carlo integration draws samples from the required distribution and forms sample averages to approximate expectations. The MCMC approach draws these samples by running a cleverly constructed Markov chain for a long time. There are many ways of constructing these chains, but all of them, including the Gibbs sampler (Geman and Geman, 1984) reviewed here, may be thought of as special cases of the general framework of Metropolis et al. (1953) and Hastings (1970). Many MCMC algorithms are hybrids of the general Metropolis-Hastings algorithm.

3.1 The Gibbs Sampler

Many statistical applications of MCMC use the Gibbs sampler, which is easy to implement. The Gibbs sampling algorithm is best described as follows.

1. Let $X = (X_1, \ldots, X_k)$ be a collection of random variables. Given arbitrary initial values $X_1^{(0)}, \ldots, X_k^{(0)}$, we draw $X_1^{(1)}$ from the conditional posterior distribution $f(X_1 \mid X_2^{(0)}, \ldots, X_k^{(0)})$, then $X_2^{(1)}$ from $f(X_2 \mid X_1^{(1)}, X_3^{(0)}, \ldots, X_k^{(0)})$, and so on, until $X_k^{(1)}$, which comes from $f(X_k \mid X_1^{(1)}, \ldots, X_{k-1}^{(1)})$.
2. This scheme determines a Markov chain, with equilibrium distribution $f(X)$. After t iterations we arrive at $X^{(t)} = (X_1^{(t)}, \ldots, X_k^{(t)})$. Thus, for t large enough, $X^{(t)}$ can be viewed as a simulated observation from $f(X)$.

Provided we allow a suitable burn-in time, the sequence $X^{(t)}, X^{(t+1)}, \ldots$ can be thought of as a dependent sample from $f(X)$.

Similarly, suppose we wish to estimate the marginal distribution of a variable Y which is a function $g(X_1, \ldots, X_k)$ of X. Evaluating g at each $X^{(t)}$ provides a sample of Y. Marginal moments or tail areas are estimated by the corresponding sample quantities.

3.1.1 Adaptive Rejection Sampling

To sample a value from the conditional marginal posterior distribution requires further considerations in Gibbs sampling. The ordinary acceptance-rejection method (Devroye, 1986) can be inefficient if the target distribution is complicated. However, Gilks and Wild (1992) developed a more efficient algorithm called adaptive rejection sampling (ARS) that enables one to sample directly from the target distribution as long as the distribution is log-concave. We can show that the conditional marginal posterior distributions of the parameters of Pareto distribution are log-concave if a uniform prior distribution is adopted. That is, we can show that $\partial^2 \ln L / \partial\alpha^2 < 0$ and $\partial^2 \ln L / \partial\gamma^2 < 0$.

4. Empirical Results Using the MCMC Method

In this section, we apply the MCMC method to estimate $\Pr(X > 29.5)$ using the data in Example 1.1. Only the Pareto distribution will be considered since it has a thicker tail than the exponential distribution (Klugman et al., 2004), and it will give us a more conservative estimate for this probability as far as the risk on large insurance claims is concerned.

Our results for $\hat{p}_1$, estimated using the Gibbs sampler, are presented in Table 2. We discarded the first 1,000 values as burn-in and generated $n = 11,000$ iterations. In the generation process, we first generated $\alpha^{(t)}$ given $\gamma^{(t-1)}$ and then generated $\gamma^{(t)}$ based on the newly generated $\alpha^{(t)}$. Then we evaluated:

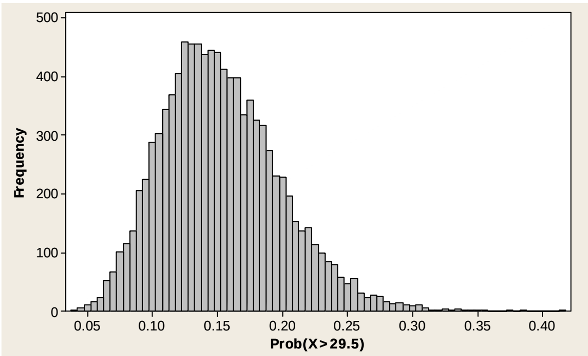
$$\hat{p}_1 = \int_{29.5}^{\infty} f(x)\, dx .$$

The empirical posterior distribution of $\hat{p}_1$ is illustrated in Figure 1.

Table 2. Descriptive Statistics of the Empirical Distribution of $\Pr(X > 29.5)$

Variable	Mean	Mode	Median	SD	Min.	Max.	95% Prob. Interval
$\Pr(X > 29.5)$	0.1529	0.1502	0.1484	0.0454	0.0406	0.4143	(0.0761, 0.2532)

Figure 1. Empirical Posterior Distribution of $\Pr(X > 29.5)$ **(n=10,000)**



In this way, we have obtained all salient information about the sampling distribution properties of $\hat{p}_1$, which is contained in the empirical posterior distribution. This cannot be done in the frequentist approach. As we can see from the empirical posterior distribution of $\hat{p}_1$ in Figure 1, the distribution is fairly symmetric around the sample mean. Therefore, we take the sample mean 0.153 as our final point estimate of $\Pr(X > 29.5)$. The Bayesian interval estimate is the probability interval (0.0761, 0.2532). The lower and upper bounds are obtained by taking the 250th and 9750th ordered values of the 10,000 ranked sample values, respectively.

We can also obtain other useful information from the Gibbs sampler scheme, such as the empirical distributions of $E(X \mid X > 29.5)$ and the quantiles $P_{0.05}$ and $P_{0.01}$. These are also important summary statistics for decision makers in the insurance industry. We present these descriptive statistics in Table 3 and the empirical distributions in Figures 2, 3, and 4.

Figure 2. Histogram of $E(x|X > 29.5)$ **(n=10,000)**

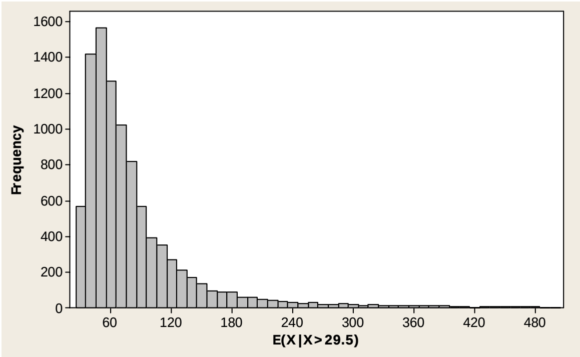


Figure 3. Histogram of $P\_(0.05)$

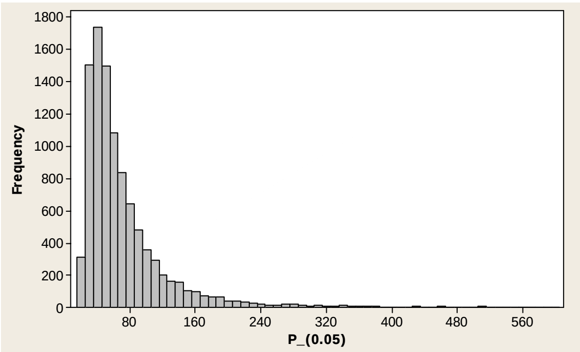


Figure 4. Histogram of $P\_(0.01)$

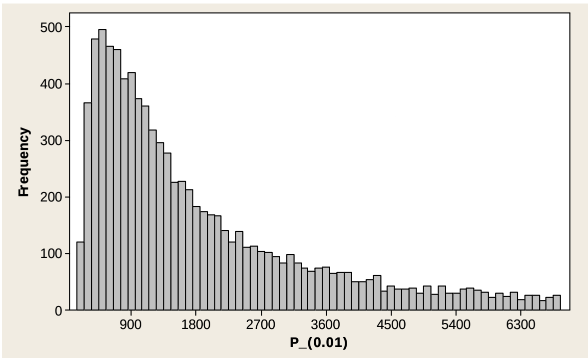


Table 3. Descriptive Statistics of $E(X|X > 29.5)$ and Quantiles $P_{0.05}$ and $P_{0.01}$

Variable	Mean	Mode	Median	SD	Min	Max	95% Prob. Interval
$E(X\|X > 29.5)$	82.5	50.5	64.7	58.6	29.5	496.7	(32.1, 257.9)
$P_{0.05}$	73.0	55.5	54.7	69.7	20.1	597.9	(24.1, 221.6)
$P_{0.01}$	1833.3	552.5	1306.9	1508.4	177.6	6842.9	(276.1, 5935.5)

For the three empirical distributions, we use the modes in each case as the most representative value for those variables as the distributions are quite skewed. Therefore the most probable value of $E(X|X > 29.5)$ is 50.5. Corresponding values for $P_{0.05}$ and $P_{0.01}$ are 55.5 and 552.5.

One can also compare the MCMC results with those obtained by using the bootstrap method (Efron, 1979). These are shown in Table 4.

Table 4. Results Using the Bootstrap Method

Variable	Mean	Mode	Median	SD	Min	Max	95% Prob. Interval
$E(X\|X > 29.5)$	81.9	49.8	63.2	59.3	29.5	500.3	(31.2, 255.8)
$P_{0.05}$	72.5	53.9	55.8	71.1	19.5	600.2	(22.9, 219.8)
$P_{0.01}$	1798.2	549.7	1311.3	1499.3	174.8	6901.3	(273.2, 5960.5)

We find that the bootstrap results are similar to the MCMC results.

4.1 Comments

We see from Table 2 that our estimate of $\Pr(X > 29.5)$ is much higher than the ML estimates. Though care must be taken with respect to the thick-tail probability estimate, the more conservative estimate will ordinarily be preferred for safety's sake. This can help to ameliorate the risks of sudden and extremely large claims. Other protective measures can also be used, such as raising deductible thresholds or reinsurance.

It is also worth noting that the Bayesian approach using the Gibbs sampler immediately provides other important univariate statistics, such as those given in Tables 3 and 4. These descriptive statistics cannot be easily obtained with ML estimation.

4.2 Another Example in Finance

In addition to insurance companies, other financial institutions, such as banks and investment companies, are also concerned with risk exposures. Estimating value-at-risk (VAR) and conditional excess are increasingly popular methods for quantifying the likely losses in their portfolios. VAR and conditional excess estimation in the context of portfolio analysis are simply quantile and conditional expectation estimation of the loss distribution, with statistics identical to those presented in Tables 3 and 4. The only difference is that instead of estimating quantiles and conditional excess in the right-hand tail of the insurance claims distribution, we are interested in these statistics in the left-hand tail of the stock price distribution. For VAR, we estimate the quantile such that a stock price will fall below this level with a fixed probability. For conditional excess, we estimate the expected stock price conditional on its having already fallen to a certain level; see Figures 5 and 6.

Figure 5. VAR for a Stock Price

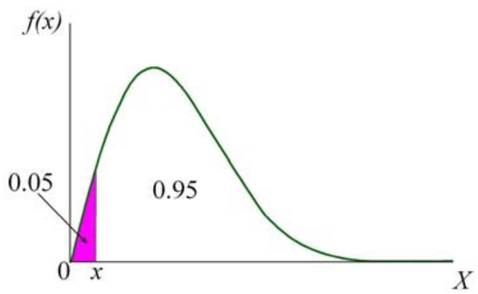
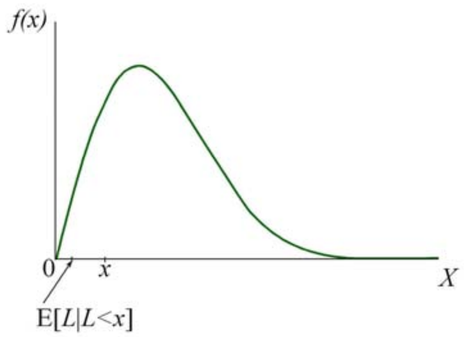


Figure 6. Conditional Excess for a Stock Price



Suppose an investment fund holds a portfolio of stocks listed in the Hong Kong Stock Exchange. For ease of illustration, suppose the portfolio consists of a single stock in the Hang Seng Bank. We obtain a sample of 100 consecutive trading days for this stock from August 22, 2001, to January 2, 2002, for analysis.

We note that the Pareto distribution is not appropriate for this data due to its threshold parameter. As with stock prices in general, there is always a chance, however small, that the price will fall to zero. We therefore assume that the stock price follows a two-parameter Weibull distribution, which is non-negative distribution with fat tails. Figure 7 illustrates a probability plot for this data, supporting our choice of this probability distribution. A time series plot of the sampled data is given in Figure 8.

Figure 7. Weibull Probability Plot

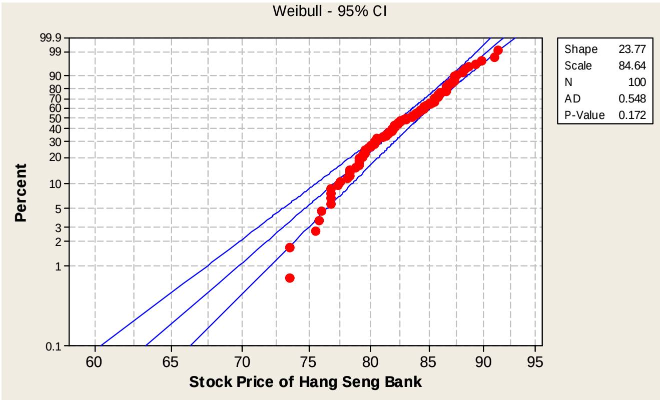
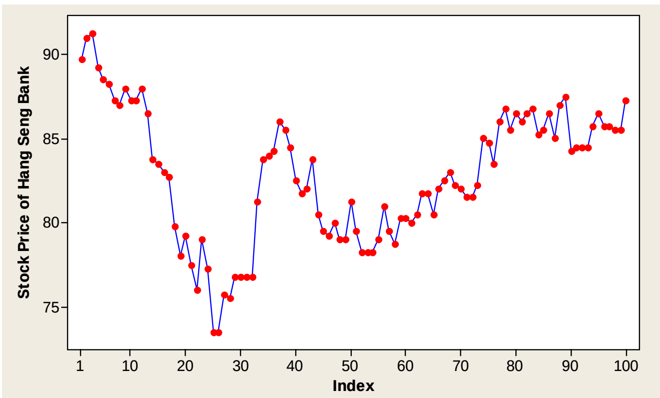


Figure 8. Time Series Plot of the Data



We can show that the conditional marginal posterior distributions of the parameters of Weibull distribution are log-concave if a uniform prior distribution is being used (see Pang, 2004). Thus we can use ARS in the Gibbs sampling scheme. Results of our MCMC approach are presented in Table 5. The conditional excess $E(X|X<80.0)$ is the conditional expected value given that the stock price falls below 80.0 Hong Kong dollars. VAR at $P_{0.05}$ ($P_{0.01}$) is the quantile such that the stock price will fill below this level with probability 0.05.

Table 5. Results of VAR and Conditional Excess Using the MCMC Method

Variable	Mean	Mode	Median	SD	Min	Max	95% Prob. Interval
$E(X\|X<80.0)$	19.165	19.102	19.048	2.793	10.943	48.579	(13.8775, 24.8749)
$P_{0.05}$	74.633	74.65	74.696	0.978	58.614	77.221	(72.7718, 76.3172)
$P_{0.01}$	69.649	69.68	69.739	1.379	39.671	72.934	(67.2578, 71.9029)

5. Conclusion

This paper reviews the Bayesian approach using Markov Chain Monte Carlo (MCMC) estimation and demonstrates its potential advantages for actuarial and risk evaluations. Premiums and other monetary valuations have traditionally been based on means of probability distributions with little attention given to interval estimates. Such interval estimates are difficult to obtain with standard estimation methods such as maximum likelihood but are quite feasible with the MCMC approach.

Using data discussed in the insurance literature, this approach is illustrated by estimating univariate statistics for a tail probability of interest and an associated conditional mean for large loss values. However, the technique is quite general and can be applied to obtain interval estimates of any function, including financial statistics, of underlying loss random variables.

Common practice is to prefer conservative estimates for risk values and financial certainty equivalents as a kind of safety margin. The distorted probability approach (Landsman and Sherris 2001; Wang 1995, 1996, 1998; Wang et al., 1997) provides one such approach. The MCMC approach described here is seen to provide more conservative estimates for the Pareto distribution tail probability considered in Hogg and Klugman (1984). Importantly, the MCMC approach and interval estimation enable more precise control over the degree of conservatism desired.

References

Chen, M. H., Q. M. Shao, and J. G. Ibrahim, (2000), *Monte Carlo Methods in Bayesian Computation*, Springer.

Devroye, L., (1986), *Non-Uniform Random Numbers Variate Generation*, Springer.

Efron, B., (1979), “Bootstrap Methods: Another Look at the Jackknife,” *Annals of Statistics*, 7, 1-26.

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(6), 721-741.

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

Hastings, W. K., (1970), “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, 57(1), 97-109.

Hogg, R. V. and S. A. Klugman, (1984), *Loss Distributions*, New York: Wiley-Interscience.

Johnson, N. L. and S. Kotz, (1970), *Continuous Univariate Distributions*, New York: John Wiley and Sons, Inc.

Klugman, S. A., H. H. Panjer, and G. E. Willmot, (2004), *Loss Models From Data to Decisions*, New York: John Wiley and Sons, Inc.

Landsman, Z. and M. Sherris, (2001), “Risk Measures and Insurance Premium Principles,” *Insurance: Mathematics and Economics*, 29, 103-115.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087-1091.

Pang, W. K., (2004), "Parameter Estimation of the Weibull Probability Distribution," Proceedings in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, American Institute of Physics, 735, 549-554.

Wang, S., (1995), "Insurance Pricing and Increased Limits Ratemaking by Proportional Hazards Transforms," *Insurance: Mathematics and Economics*, 17(1), 43-54.

Wang, S., (1996), "Premium Calculation by Transforming the Layer Premium Density," *ASTIN Bulletin*, 26, 71-92.

Wang, S., (1998), "An Actuarial Index of the Right-Tail Risk," *North American Actuarial Journal*, 2, 88-101.

Wang, S., V. R. Young, and H. H. Panjer, (1997), "Axiomatic Characterization of Insurance Prices," *Insurance: Mathematics and Economics*, 21(2), 173-183.