Bayesian Claim Severity with Mixed Distributions

By Benedict Escoto

This paper is produced mechanically as part of FAViR.
See http://www.favir.net for more information.

Abstract

Suppose the default claim severity distribution is a finite mixture of simpler severity distributions. For instance, many ISO distributions are mixed exponentials. The technique in this paper can be used to adjust the weights of the mixture in a principled way to partially-credible observed claim severities.

In Bayesian terminology, this paper assumes a Dirichlet distribution over initial mixture weights. The posterior distribution, conditional on one or more observed claim severities, is computed using a custom Gibbs sampler.

1 Introduction

Many applications require the position of modeling claim severities based on limited historical data. One example is the pricing an excess of loss reinsurance layer.

The technique in this paper is intended to help cover the awkward middle ground between having no data and having lots of data. If no claim data is available, some default claim severity distribution may be available. For instance, ISO publishes mixed exponential severity distributions based on aggergate data (see Palmer for a basic description of ISO's methodology). When many data points are available, maximum-likelihood curve fitting methods like ISO's method work well.

Intuitively, it seems the expected claim severity distribution should morph from the default distribution into some fitted distribution as more and more claims are observed. The most principled way of doing this is to use Bayesian statistics.

This paper models this situation under these assumptions:

1. The default severity distribution is a mixed exponential (such as those supplied by ISO). Other mixed distributions would probably work with minor modifications.

2. Parameter uncertainty is modeled using a Dirichlet distribution over the mixture weights. This requires one additional parameter, interpreted as the confidence in the default distribution.

3. The posterior distribution is computing using standard Bayesian updating on observed claim severities.

The numerical results of this technique can be computed quickly using Gibbs sampling, a Monte Carlo Markov Chain (MCMC) method. The output can be summarized as a new set of mixture weights for use however the actuary would use the original default distribution.

2 Required input data

Three initial inputs are required. The weights and means of the default severity distribution are shown in figure 1.

Weights (%)	Means
10.0	300
20.0	1,000
30.0	3,000
20.0	10,000
10.0	30,000
5.0	100,000
3.5	300,000
1.0	1,000,000
0.5	3,000,000
Avg	46,630

Figure 1: Default Mixed Exponential

Second, we need to know how confident the actuary is in these initial weights. Does the actuary think that the "true" distribution is almost certainly close to the default distribution? Or is it probable that the true distribution is much bigger or smaller than the default? Specifying this is equivalent to specifying the variance of the hypothetical means in Buhlmann credibility.

The value used for the standard deviation of the expectation is **50000**.

Finally, we need to know the claims to update the distribution on. These are shown in figure 2.

Amount
500,000
32,500
8,200
10,000
750,000

Figure 2: Claim Severities

3 The answer

Given the numerical data in the previous section and the modeling assumptions described in the introduction, we can compute the posterior weights.

Prior to Data		Posterior to Data		
Weight (%)	Mean	Weight (%)	Error (%)	Mean
10.0	300	8.3	0.23	300
20.0	1,000	16.5	0.32	1,000
30.0	3,000	26.5	0.38	3,000
20.0	10,000	21.1	0.38	10,000
10.0	30,000	11.2	0.29	30,000
5.0	100,000	5.0	0.20	100,000
3.5	300,000	8.1	0.26	300,000
1.0	1,000,000	2.7	0.18	1,000,000
0.5	3,000,000	0.7	0.08	3,000,000
Avg	46,630		Avg	81,951

Figure 3: Results of Bayesian Updating

Due to computational difficulties (see section 5), the posterior weights are only approximated. Figure 3 shows the new weights and estimates the approximation error. The means remain the same as required by the model. Figure 4 shows the difference graphically. This graph shows both the change in weights and the change in densities. The observed claims are shown using vertical lines.

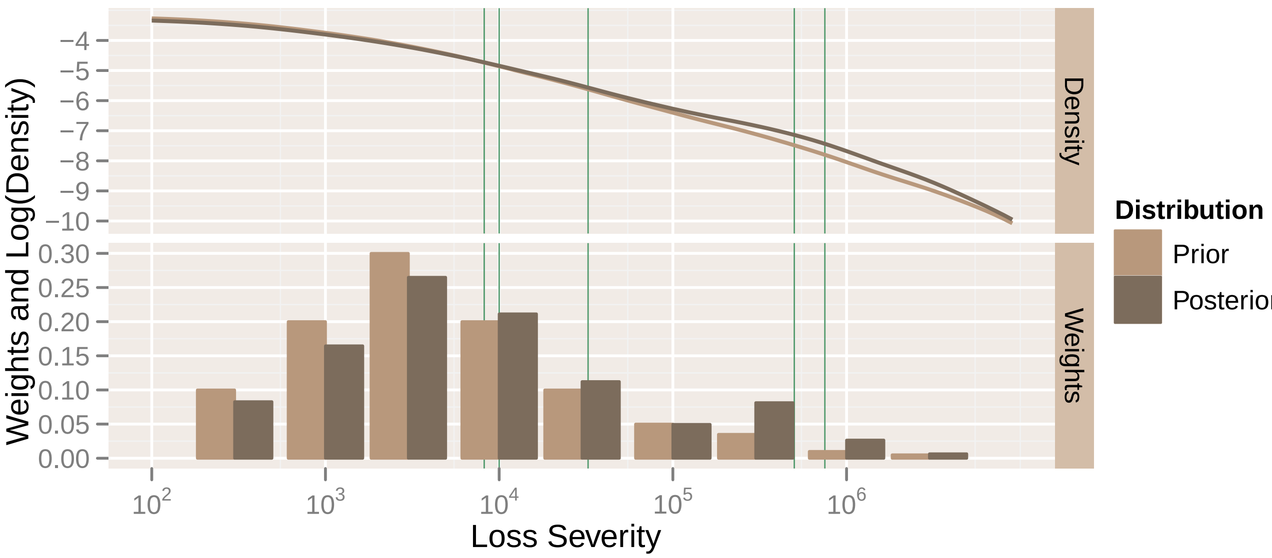

Figure 4: Graph of Results

4 Detailed model

Formally, the Bayesian probabilistic model used is defined by these equations:

$$p(x|b) = \frac{e^{x/\mu_b}}{\mu_b} \tag{1}$$

$$p(b|w_1, \ldots, w_m) = w_b \tag{2}$$

$$p(w_1, \ldots, w_m) = \frac{\Gamma(\sum_{j=1}^{m} \alpha_j)}{\prod_{j=1}^{m} \Gamma(\alpha_j)} \prod_{j=1}^{m} w_j^{\alpha_j - 1} \text{ with } w_m = 1 - \sum_{j=1}^{m-1} w_j \tag{3}$$

or in other words,

$$\begin{aligned} x|b &\sim \text{Exponential}(1/\mu_b) \\ b|w_1, \ldots, w_m &\sim \text{Categorical}(w_1, \ldots, w_m) \\ w_1, \ldots, w_m &\sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_m) \end{aligned}$$

where x is an individual claim severity and μ_b is the expected value of exponential distribution b. x and bucket selection b are assumed independent given the bucket weights $(w_1, \ldots, w_m)$.

$\alpha_1, \ldots, \alpha_m$ are hyperparameters—they control the initial prior distribution over the possible bucket weights, but are not given probabilities themselves.

The model's marginal distribution over claim severities is then

$$
\begin{aligned}
p(x) &= \sum_{b=1}^{m} \int p(x|b)p(b|\mathbf{w})p(\mathbf{w})\,d\mathbf{w} \\
&= \sum_{b=1}^{m} \int p(x|b)w_b p(\mathbf{w})\,d\mathbf{w} \\
&= \sum_{b=1}^{m} (\int w_b p(\mathbf{w})\,d\mathbf{w})p(x|b) \\
&= \sum_{b=1}^{m} \mathrm{E}[w_b]p(x|b) \qquad (4)
\end{aligned}
$$

where $\mathbf{w} = w_1, \ldots, w_m$. Thus as long as we choose $\alpha_1, \ldots, \alpha_m$ so that $\mathrm{E}[w_j] = a_j$ where a_j is our default weight for bucket j, our model will imply the correct marginal claim severity distribution before any data is observed.

4.1 The Dirichlet and choosing $\boldsymbol{\alpha}$

The Dirichlet is the multidimensional analogue of the beta distribution. Just as the beta distribution can be used to express uncertainty about two numbers which must add to 1, the Dirichlet can express uncertainty about m positive numbers that must add to 1. This property makes it popular in Bayesian analysis (see Mildenhall for an example of the Dirichlet applied in an insurance context).

It is a property of the Dirichlet distribution that if $w_1, \ldots, w_m \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_m)$, then

$$
\begin{aligned}
\mathrm{E}[w_j] &= \frac{\alpha_j}{\alpha_0} && (5) \\
\mathrm{Var}[w_j] &= \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)} = \frac{\mathrm{E}[w_j](1 - \mathrm{E}[w_j])}{(\alpha_0 + 1)} && (6) \\
\mathrm{Cov}[w_j, w_k] &= \frac{-\alpha_j \alpha_k}{\alpha_0^2(\alpha_0 + 1)} = \frac{-\mathrm{E}[w_j]\mathrm{E}[w_k]}{\alpha_0 + 1} \text{ for } j \neq k && (7)
\end{aligned}
$$

where $\alpha_0 = \sum_{j=1}^{m} \alpha_j$. Because we are given the initial default weights $a_j = \mathrm{E}[w_j]$, the choice of α_0 will uniquely determine $\alpha_1, \ldots, \alpha_m$. As in the beta distribution, the larger the sum of the parameters α_0, the more certain we are of the "true" weights.

Almost any statement about parameter risk, or about the uncertainty of the true distribution, will determine a value for α_0. Furthermore, with a Dirichlet distribution many of these will be analytically tractical. In this paper, we assume that the uncertainty (measured in terms of standard deviation) of the true (unlimited) expected claim severity is given as σ. Using equations (6) and (7) we get

$$\begin{aligned}
\sigma^2 &= \mathrm{Var}[\mathrm{E}[x|w_1,\ldots,w_m]] \\
&= \mathrm{Var}[\sum_{j=1}^{m} w_j\mu_j] \\
&= \sum_{j=1}^{m} \mu_j^2 \mathrm{Var}[w_j] + \sum_{j\neq k} \mu_j\mu_k \mathrm{Cov}[w_j,w_k] \\
&= \sum_{j=1}^{m} \mu_j^2 \frac{\mathrm{E}[w_j](1-\mathrm{E}[w_j])}{\alpha_0+1} + \sum_{j\neq k} \mu_j\mu_k \frac{-\mathrm{E}[w_j]\mathrm{E}[w_k]}{\alpha_0+1} \\
&= \sum_{j=1}^{m} \mu_j^2 \frac{a_j(1-a_j)}{\alpha_0+1} + \sum_{j\neq k} \mu_j\mu_k \frac{-a_j a_k}{\alpha_0+1}
\end{aligned}$$

hence

$$\alpha_0 = \frac{1}{\sigma^2}(\sum_{j=1}^{m} \mu_j^2 a_j(1-a_j) - \sum_{j\neq k} \mu_j\mu_k a_j a_k) - 1. \tag{8}$$

Equation (8) was used above to determine the initial Dirichlet paramaters. Specifically, the chosen value for σ, 50000, implies that $\alpha_0 = 22.99562564$.

The behavior of the Dirichlet/multinomial conjugate pair under Bayesian updating suggests this interpretation of α_0: our prior distribution contains an amount of information equivalent to α_0 claims (see Hoff p.39). For instance, if $\alpha_0 = 5$, then prior to the data, we have about as much information as someone would have after seeing 5 claims. Although this idea is logically nonsensical, it does provide a rough-and-ready guide to the influence the data will have on the posterior weights. For instance, if $\alpha_0 = 5$, then after conditionalization on 5 claims, the data and our prior beliefs will have about equal credibility.

4.2 Why not vary the means?

It may seem more practical to allow the means of the exponential distributions to vary with observed claims. The reason this paper only adjusts the weights is that this allows the correct default distribution to be used when there are no claims. Equation (4) shows that the prior marginal distribution will correctly equal the default distribution as long as the expected weights are equal to the desired default weights. However, there is no similar way to do this by varying the means.

For example, suppose desired claim severity is an equally-weighted mixed exponetial of means 100 and 300:

$$p(x) = 0.5\frac{e^{x/100}}{100} + 0.5\frac{e^{x/300}}{300}.$$

Then $p(x)$ can be expressed as the weighted mixture of various other mixtures of exponentials with means 100 and 300, such as:

$$p(x) = \frac{1}{2}(0.25\frac{e^{x/100}}{100} + 0.75\frac{e^{x/300}}{300}) + \frac{1}{2}(0.75\frac{e^{x/100}}{100} + 0.25\frac{e^{x/300}}{300})$$

However, it is impossible to express $p(x)$ as a (positive) mixture of exponentials with means other than 100 or 300. Thus if we know that the true distribution is a mixed exponential, and if we know the form of $p(x)$ as is above, then our modeled uncertainty must only concern the weights of the mixed exponential.

5 Computing the answer

Given the model described in equations (1)–(3) and n observed claim severities $c_1, \ldots, c_n$, it is simple in principle to compute the posterior marginal distribution:

$$\begin{aligned}
p(x|c_1, \ldots, c_n) &= \sum_{b=1}^{m} \int p(x|b, c_1, \ldots, c_n) p(b|\mathbf{w}, c_1, \ldots, c_n) p(\mathbf{w}|c_1, \ldots, c_n)\, d\mathbf{w} && (9)\\
&= \sum_{b=1}^{m} \int p(x|b) p(b|\mathbf{w}) p(\mathbf{w}|c_1, \ldots, c_n)\, d\mathbf{w} && (10)\\
&= \sum_{b=1}^{m} \mathrm{E}[w_b|c_1, \ldots, c_m] p(x|b). && (11)
\end{aligned}$$

However, straightforwardly computing (10) is difficult, even when a simple distribution like the exponential is used. Interestingly, (10) is analytically soluble, but the number of terms is $O(m^n)$ so this tact is infeasible.

Instead, the posterior marginal weights $\mathrm{E}[w_b|c_1, \ldots, c_m]$ can be computed quickly using Gibbs sampling. The logic behind this procedure is relatively complicated and won't be described here (see Hoff, chapter 6) but the implementation is only about a dozen lines of code.

An MCMC technique like Gibbs sampling was chosen here because the dimensionality of integral (10) makes numerical integration very time-consuming. A straightforward Monte Carlo method was also tried. It was as accurate and ran faster when the posterior was close to the prior, but became very inaccurate when large numbers of observed claims moved the posterior far away from the prior.

6 Bibliography

1. Hoff, Peter D. *A First Course in Bayesian Statistical Methods* Springer, 2009.

2. Mildenhall, Stephen J. "A Multivariate Bayesian Claim Count Development Model With Closed Form Posterior and Predictive Distributions." *Casualty Actuarial Society Forum* Winter, 2006. http://www.casact.org/pubs/forum/06wforum/06w455.pdf

3. Palmer, Joseph M. "Increased Limits Ratemaking for Liability Insurance" Study Note, July 2006. http://www.casact.org/library/studynotes/palmer.pdf

4. Wikipedia. "Dirichlet distribution" Retrieved May 2, 2010. http://en.wikipedia.org/wiki/Dirichlet_distribution

7 Legal