Bayesian Claim Severity Part 2

Mixed Exponentials with Trend, Censoring, and Truncation

By Benedict Escoto

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

Abstract

This is an continuation of the FAViR paper "Bayesian Claim Severity with Mixed Distributions". The application is the same: the actuary is trying to produce a claim severity distribution and has a prior Dirichlet over mixed exponential distribution that s/he wants to update with observed claim data.

However, in this paper the data is allowed to have a flat severity trend, and the claim data may be truncated or censored. Instead of a custom Gibbs sampler, JAGS is used to compute the posterior parameters.

1 Introduction

The earlier FAViR paper "Bayesian Claim Severity with Mixed Distributions" (Escoto) derived a claim severity distribution from observed claim amounts using traditional Bayesian updating. There, each claim had a mixed exponential severity distribution, conditional on "parameter risk" which was represented by a Dirichlet distribution.

Under these conditions, the marginal distribution of each claim was also a mixed exponential. Thus, an actuary could start with a prior mixed exponential severity distribution (from ISO or some other source) and refine it with available claim data. The resulting marginal distribution would be the correct credibility-weighted mixture of the prior distribution and the claim data, but would still be mixed exponential in form.

This paper is a continuation of "Bayesian Claim Severity with Mixed Distributions". The basic probabilistic model is the same: a mixed exponential severity with Dirichlet parameters. Unlike that paper however, here we handle three complications often seen in practice:

1. censored data (policy limits),
2. truncated data (deductibles), and

3. severity trend.

Trend parameter risk (represented by a gamma distribution) is added to the model because trend credibility needs to be estimated simultaneously with severity. For instance, increasing average claim severity in the most recent years may indicate a high severity trend; or maybe a few huge losses happened to occur recently. Bayesian statistics properly weighs these possibilities.

To compute the posterior distribution, this paper uses Just Another Gibbs Sampler (JAGS), a cross-platform general purpose open source MCMC engine, while the previous paper used a custom Gibbs sampler. As a result, this paper is slower and has more dependencies, but is also easier to modify, assess for convergence, and run in parallel chains. Knowledge of MCMC theory is not necessary to use this paper.

2 Required input data

This section displays all the initial data required by this paper. The initial inputs can be grouped into four categories: the observed claim data, the prior means and weights of the mixed exponential severity distribution, the prior severity uncertainty (represented by a Dirichlet distribution), and the prior trend mean and standard deviation.

2.1 Claim Data

The claim data used is shown in figure 1. For each claim we need to know its severity, whether or not it was censored at policy limits, its deductible (truncation threshold), and the age of the claim.

Amount	Age	Deductible	Capped?
33,750	3	0	No
1,000,000	1	0	Yes
22,707	1	0	No
54,135	1	0	No
174,524	3	0	No
19,661	2	0	No
140,735	2	0	No
1,000,000	3	0	Yes
1,127	1	0	No
316,483	2	0	No

Figure 1: Observed Claim Data

2.2 Mixed Exponential

Figure 2 shows the means and weights of a mixed exponential distribution. This determines the model's marginal claim severity prior to conditionalization on the claim data (i.e. the severity distribution you'd expect for the very next claim if no detailed claim data were available).

Weights (%)	Means
30.0	50,000
25.0	100,000
25.0	500,000
10.0	1,500,000
7.0	5,000,000
3.0	20,000,000
Avg	1,265,000

Figure 2: Prior Means and Weights

2.3 Dirichlet Uncertainty

Although the means and weights of the mixed exponential distribution determine the margin severity distribution, we also need to know how certain we are that these parameters are accurate. Are they just a rough estimate, and may be far off, or are we sure that the true distribution is very similar? This uncertainty can be summarized as α_0, the sum of the Dirichlet parameters. Here we picked a value of **20**.

See section 4.1 for guidance on choosing this parameter.

2.4 Trend

Finally, we need a prior distribution over the trend rate. The trend is applied to each claim by age. For instance, if the trend rate is 7%, then claims of age 2.5 are expected to be $1.07^{-2.5}$ times as severe as claims of age 0.

We assume trend is constant over time, but the parameter uncertainty is modeled as a gamma distribution, shifted so that the zero point indicates –100% trend. Choosing a mean and standard deviation sufficies to determine the gamma parameters. In this paper, the trend mean is **0.05** and the trend standard deviation is **0.01**.

3 Results

This section presents the results of conditionalizing the Bayesian model on the observed claim data using MCMC. A total of **30000** samples were computed. Figure 3 shows the prior and posterior marginal weights.

Prior to Data		Posterior to Data		
Weight (%)	Mean	Weight (%)	Error (%)	Mean
30.0	50,000	30.9	0.10	50,000
25.0	100,000	25.6	0.10	100,000
25.0	500,000	23.4	0.09	500,000
10.0	1,500,000	9.6	0.06	1,500,000
7.0	5,000,000	7.2	0.05	5,000,000
3.0	20,000,000	3.2	0.04	20,000,000
Avg	1,265,000		Avg	1,307,445

Figure 3: Prior vs Posterior Exponential Weights

The error column is an estimate of the standard error of the MCMC method. This can be decreased through running more simulations. Because the error is estimated using time-series methods, it takes autocorrelation into account. The other exhibits assume this error is acceptably small and can be ignored. The `coda` package is compatible with JAGS and includes more tools for MCMC error-testing and diagnostics; a few sample commands are given in the source code to this paper.

Figures 4, 5, and 6 show the prior and posterior expected loss in layer, trend, and ILFs. The ILFs are based solely on expected loss costs and do not take into account risk loads, expenses, etc. In these figures, each boxplot shows the 10th, 25th, 50th, 75th, and 90th percentiles of the corresponding distribution.

4 Probabilistic Model

Here is the formal description of the Bayesian hierarchical claim severity model. It can be divided into process and parameter risk.

Process risk:

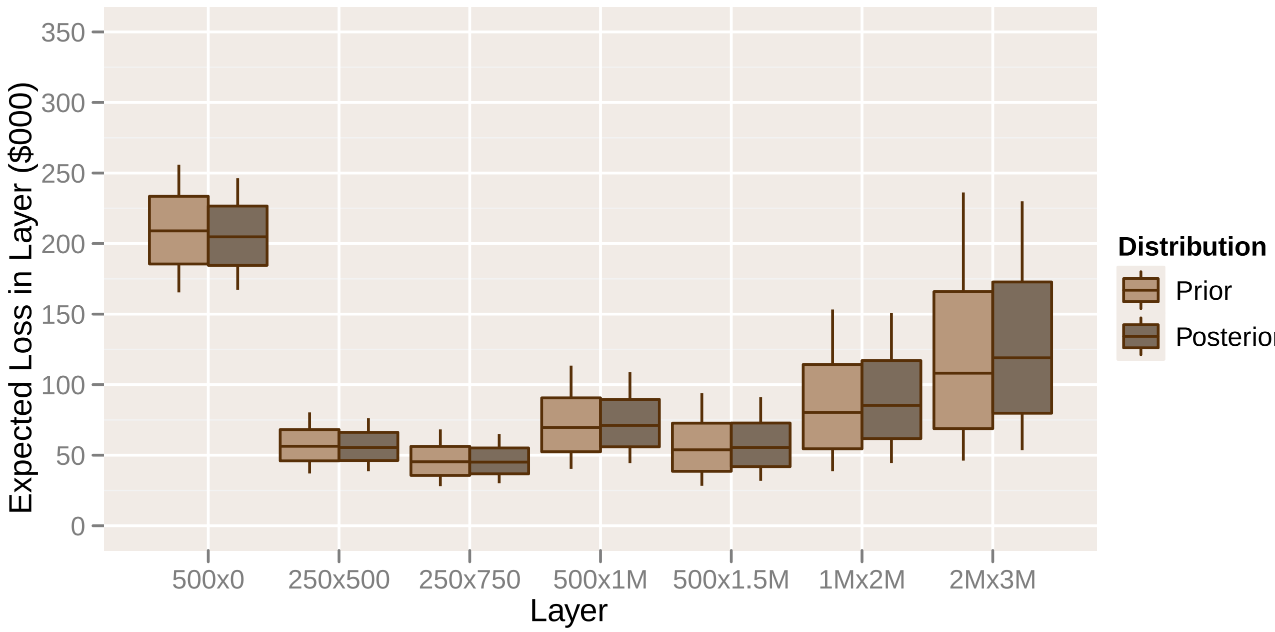

Figure 4: Prior vs Posterior Loss in Layer

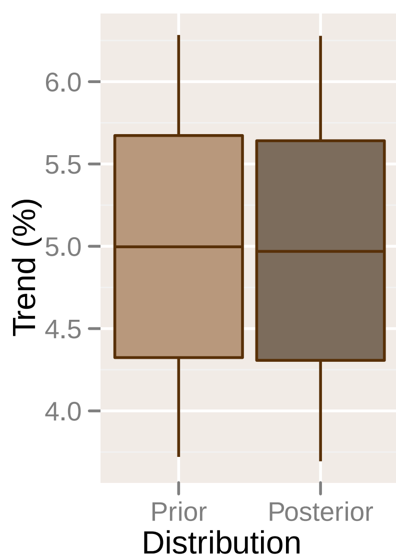

Figure 5: Trend Results

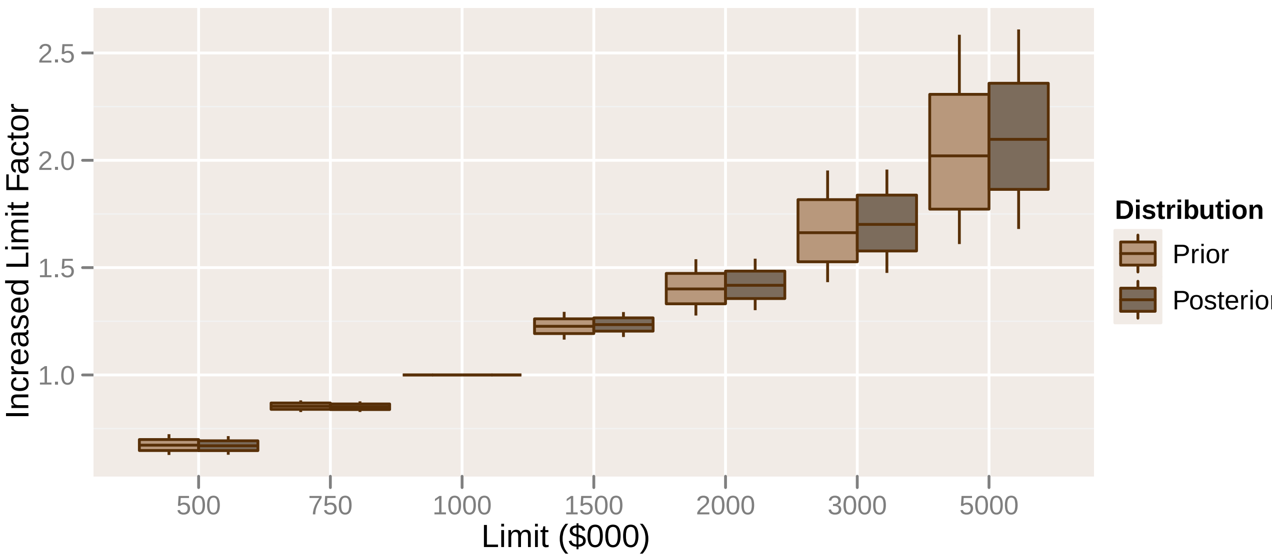

Figure 6: Prior vs Posterior ILF Distribution

$$x_i|b_i \sim g(\text{Exponential}(\frac{1}{\mu_{b_i} d^{t_i}}))$$
$$b_i|w_1, \ldots, w_m \sim \text{Categorical}(w_1, \ldots, w_m)$$

Parameter risk:

$$w_1, \ldots, w_m \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_m)$$
$$d \sim \text{Gamma}(k, \theta)$$

The function g above represents censoring and/or truncation. The index i ranges over the number of observed claims. The means of the mixed exponential are $\mu_1, \ldots, \mu_m$ and t is the age of the claim. See "Bayesian Claim Severity with Mixed Distributions" for more information.

Mathematically, the only distinction between parameter and process parts of the model is that all the claims are assumed to have the same parameter risk variables but the process risk parameters vary per claim. For example, there is one the trend parameter d which affects all claims, but each claim will get its own instance of x and b. Conditional on d and $w_1, \ldots, w_m$, the distribution of each claim is independent.

4.1 Choosing α_0

In the input data section, the paramater α_0 was used to summarize our uncertainty around the mixed exponential prior distribution. Intuitively, this parameter can be thought of the number of claim observations encapsulated by the prior mixed exponential. So, after we observe α_0 claims, the resulting distribution will depend equally on our prior and the observed data. Because intuition is not of much help in selecting a particular value, we will use this result:

If X is a mixed exponential distribution (with fixed means) depending on weights $w_1, \ldots, w_m$, and $g(x)$ is a real function, then

$$\begin{aligned}
\mathrm{Var}[\mathrm{E}[g(x)|w_1,\ldots,w_m]] &= \mathrm{Var}[\sum_{j=1}^{m} w_j \mathrm{E}[g(x)|b=j]] = \mathrm{Var}[\sum_{j=1}^{m} w_j g_j] && (1)\\
&= \sum_{j=1}^{m} g_j^2 \mathrm{Var}[w_j] + \sum_{j \neq k} g_j g_k \mathrm{Cov}[w_j, w_k] && (2)\\
&= \sum_{j=1}^{m} g_j^2 \frac{a_j(1-a_j)}{\alpha_0+1} + \sum_{j \neq k} g_j g_k \frac{-a_j a_k}{\alpha_0+1} && (3)
\end{aligned}$$

therefore

$$\alpha_0 = \sigma^{-2}(\sum_{j=1}^{m} g_j^2 a_j(1-a_j) - \sum_{j \neq k} g_j g_k a_j a_k) - 1$$

where $g_j = \mathrm{E}[g(x)|b=j]$, $a_j = \frac{\alpha_j}{\alpha_0}$, and (3) follows from the properties of the Dirichlet distribution (see “Bayesian Claim Severity with Mixed Distributions” for more about the Dirichlet distribution).

Suppose the actuary may feel that the expected value of the true claim distribution capped at \$1M may be \$50,000 off from the expected capped value implied by the prior mixed distribution. Then by setting $g(X) = \min(X, 1\mathrm{Mil})$, $g_j = \mu_j(1 - e^{\frac{1e6}{\mu_j}})$ and the above result can be used to calculate α_0.

In this paper, we set $\alpha_0 = \mathbf{20}$, which implies that the standard deviation of loss in the first million layer is **65770**.

5 Computation

The posterior distribution of the probabilistic model specified in section **??** is probably not analytically soluable. The version of the model in “Bayesian Claim Severity with Mixed Distributions” (which had no trend parameter) actually could be solved analytically, but the solution was not tractable.

In these situations, Monte Carlo Markov Chain (MCMC) techniques are extremely useful. They have revolutionized Bayesian statistics in the last few decades. One MCMC algorithm is called Gibbs sampling. The earlier paper implemented its own Gibbs sampler. However, there are now dedicated software packages such as WinBUGS which allow users to specify custom Bayesian hierarchical models in an intuitive modelling language, and then solve those models using using Gibbs sampling and other MCMC algorithms.

This paper uses JAGS (Just Another Gibbs Sampler), which is an improved version of WinBUGS that is open source and cross platform. See Plummer (2003) for more information. JAGS can be used seamlessly from R through the `runjags` package. The JAGS model description for this paper takes about a dozen lines of code.

Compared to custom code, JAGS (and WinBUGS) are easier to use and modify, but much slower. On my computer (Intel Core 2 Duo 6600 running Linux), the model takes about 10 minutes to process 600 claims. A tuned MCMC algorithm written in a low-level language like C would probably be 10–50 times faster. Also, MCMC techniques are highly parallelizable, so more cores increase speed almost linearly.

6 Conclusion

This paper credibility weighs prior beliefs about claim severity with observed claim data. The actuary starts with a prior mixed exponential severity distribution (perhaps from an external source such as ISO) and uses standard Bayesian conditionalization on the claim data to arrive at a posterior weights distribution. Complications such as trend, censoring, and truncation are handled.

This method may be practical whenever there is a shortage of claim data. If a huge number of relevant claims were available, there would be no need for Bayesian statistics—the actuary could simply use the (possibly smoothed) empirical distribution. But when the number of data points is insufficient for non-parametric statistics, actuaries frequently turn to maximum likelihood methods. If prior distributions are available, Bayesian methods such as the one presented here are superior to maximum likelihood methods because they incorporate that information.

7 Bibliography

1. Escoto, Benedict M. "Bayesian Claim Severity with Mixed Distributions." *FAViR Drafts*.
   `http://www.favir.net/local-files/papers/BayesianMixed2.pdf`

2. Plummer, Martyn. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling", *Proceedings of the 3rd International Workshop on Distributed*

Statistical Computing, (DSC 2003), Vienna, Austria.
`http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/Plummer.pdf`

8 Legal