Heavy-Tailed Bayesian Models for Estimating Incurred But Not Reported Reserves

Audrey Nguyen   Advisor: Dr. Nelis Potgieter

Department of Mathematics, Texas Christian University, Fort Worth, Texas

Introduction

- **IBNR claims**: losses incurred but not yet reported → *hidden liabilities*
- Estimating reserves is critical for insurer solvency and risk management
- Standard models often underestimate **extreme (tail) outcomes**

Approach: We develop a Bayesian framework with heavy-tailed priors to better capture tail risk and quantify uncertainty in reserve estimates.

Definitions

Claims Development Triangle illustrates how claims grow over time for each accident year.

AY	Dev 0	Dev 1	Dev 2	Dev 3
| rule 2021	120	180	205	210
2022	140	210	232	
2023	160	223		
2024	185			

Table 1: Claims Development Triangle

- Rows = **accident year** (AY_i).
- Columns = **development period** (Dev_j).
- Entries $C_{i,j}$ = cumulative claims.

Illustration:

Total claims for AY 2021 grows from \$120 paid in Dev 0 to \$210 in Dev 3.

Key Notation

- $C_{i,j}$: cumulative claims, $i \leq I$, $j \leq J$.
- μ_i: ultimate claims for AY i
- γ_j: proportion emerging in Dev j
- $\mathcal{D}_I$: full observed triangle data

Goal: Estimate $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ to predict the IBNR.

Heavy-Tailed Distributions

Why heavy tails?

Standard motels pose risk of underestimating extreme losses. Heavy-tailed distributions have slower tail decay, meaning they can assign higher probabilities to larger claims. Thus more suitable for extreme tail risk modeling.

Distributions Used

- **Pareto:** classic heavy-tailed model for large losses

$$f(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, \ x \geq x_m.$$

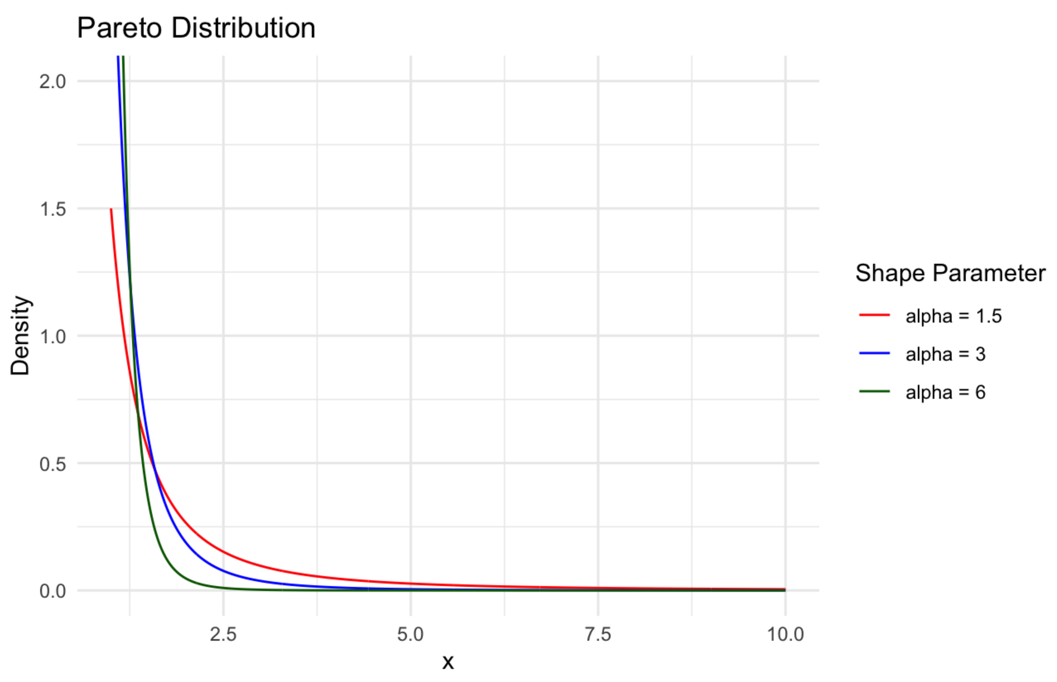

Figure 1: Pareto Distribution Density

- **Log-t:** flexible model with tunable tail behavior via ν

$$f(x) \propto \frac{1}{x}\left(1 + \frac{1}{v}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)^{-\frac{v+1}{2}}, \ x \geq 0.$$

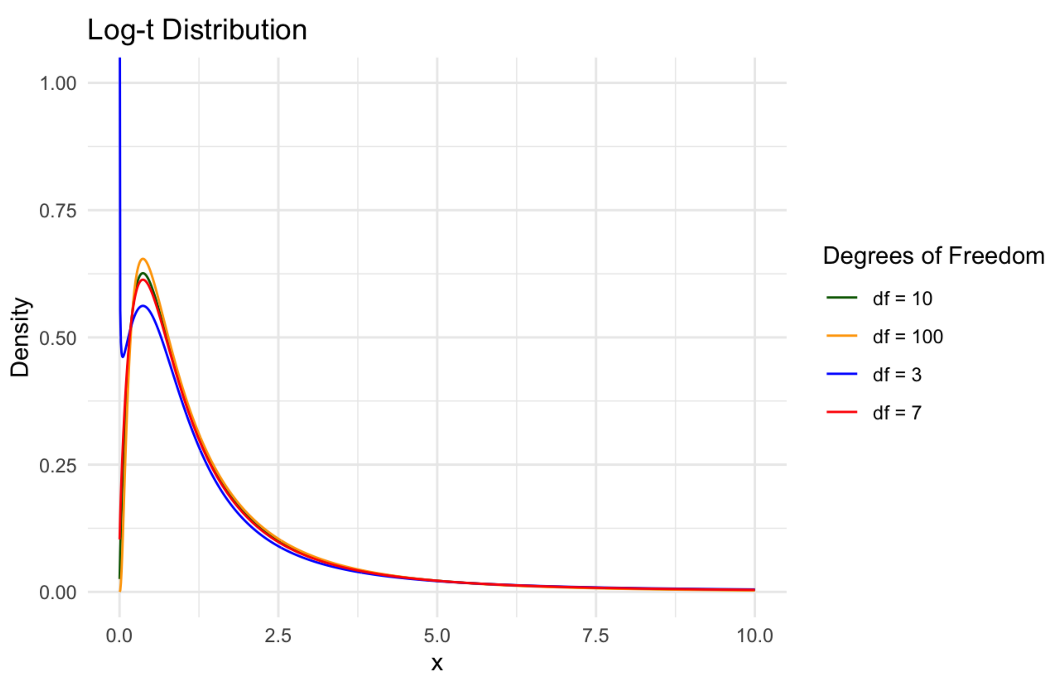

Figure 2: Log-t Distribution Density

Bayesian Modeling Framework

Goal: Estimate ultimate claims $\boldsymbol{\mu}$ = *the total expected payments for each AY* and also $\boldsymbol{\gamma}$ = *proportions of claims emerging in each Dev.* Combined used to predict IBNR.

Bayesian Update

$$p(\boldsymbol{\mu}, \boldsymbol{\gamma} \,|\, \mathcal{D}_I) \propto p(\mathcal{D}_I \,|\, \boldsymbol{\mu}, \boldsymbol{\gamma})\, p(\boldsymbol{\mu})\, p(\boldsymbol{\gamma})$$

- $\boldsymbol{\mu}$: ultimate claims
- $\boldsymbol{\gamma}$: development proportions
- $p(\boldsymbol{\mu})$ and $p(\boldsymbol{\gamma})$: prior beliefs about parameters
- $p(\mathcal{D}_I \,|\, \boldsymbol{\mu}, \boldsymbol{\gamma})$: likelihood of the observed data
- $p(\boldsymbol{\mu}, \boldsymbol{\gamma} \,|\, \mathcal{D}_I)$: posterior distribution

Model Specification

- $\boldsymbol{\gamma} \sim$ Dirichlet: models how total claims are distributed across development years and ensures proportions sum to 1
- $\boldsymbol{\mu} \sim$ Pareto or Log-t: captures heavy-tailed behavior in ultimate claims, allowing for extreme losses
- $\mathcal{D}_I \mid \boldsymbol{\mu}, \boldsymbol{\gamma} \sim$ Gamma: models positive, right-skewed claim amounts given the underlying parameters

Data & Model Development

We use the `GenIns` dataset from the `ChainLadder` R package, a synthetic claims triangle used for illustration and model validation. Based on $I = 10$ and $J = 10$.

For the prior on μ_i, we specify a mean of \$5,000,000 based on cumulative claims at development year 10, and express variability using the coefficient of variation (CV). This parameterization controls variance relative to the mean.

We assign $\boldsymbol{\gamma}$ a symmetric Dirichlet prior with concentration parameter 2 across the J development periods. This centers each γ_j near $1/J$ while allowing moderate variability across years.

Results

Posterior IBNR quantiles (Pareto and Log-t) assess tail risk under different specifications. Upper quantiles emphasize the potential impact of large (or catastrophic) claim outcomes. **All values are \$ millions.**

	CV	50%	90%	95%
1	0.3	17.30	19.49	20.26
2	1.0	17.49	20.19	21.01

Table 1: Posterior IBNR Quantiles for Pareto

	ν	50%	90%	95%
1	3	17.89	20.33	21.17
2	30	18.64	21.42	22.24
3	100	18.58	21.40	22.27

Table 2: Posterior IBNR Quantiles for Log-t

Model fit is compared using LOO (*leave-one-out cross-validation*). Approximates predictive accuracy in unseen data. Lower LOO values indicate better predictive performance.

	Pareto		Log-t		
Model	$CV = 0.3$	$CV = 1$	$\nu = 3$	$\nu = 30$	$\nu = 100$
LOO	1502.15	1506.27	1502.79	1504.52	1504.05

Table 3: LOO comparison across models

Result: All models show similar performance (LOO differences are small relative to standard errors), suggesting that heavy-tailed priors primarily affect tail risk rather than overall fit.

Conclusion & Future Works

We developed a **heavy-tailed Bayesian framework** for estimating IBNR reserves. Results show increased upper-tail reserve estimates with similar predictive performance.

Implication: Heavy-tailed priors provide a more realistic assessment of extreme claim risk without sacrificing overall model fit.

Future Work: Extend the framework by incorporating heavy-tailed likelihoods for real-world data alongside heavy-tailed priors, improving robustness to extreme claim behavior.