Cynon Sonkkila & Chris Halliwell → Insurance Data Science

23 May 2023

Bayesian Hierarchical Modelling for Long Tailed Lines

MARKEL

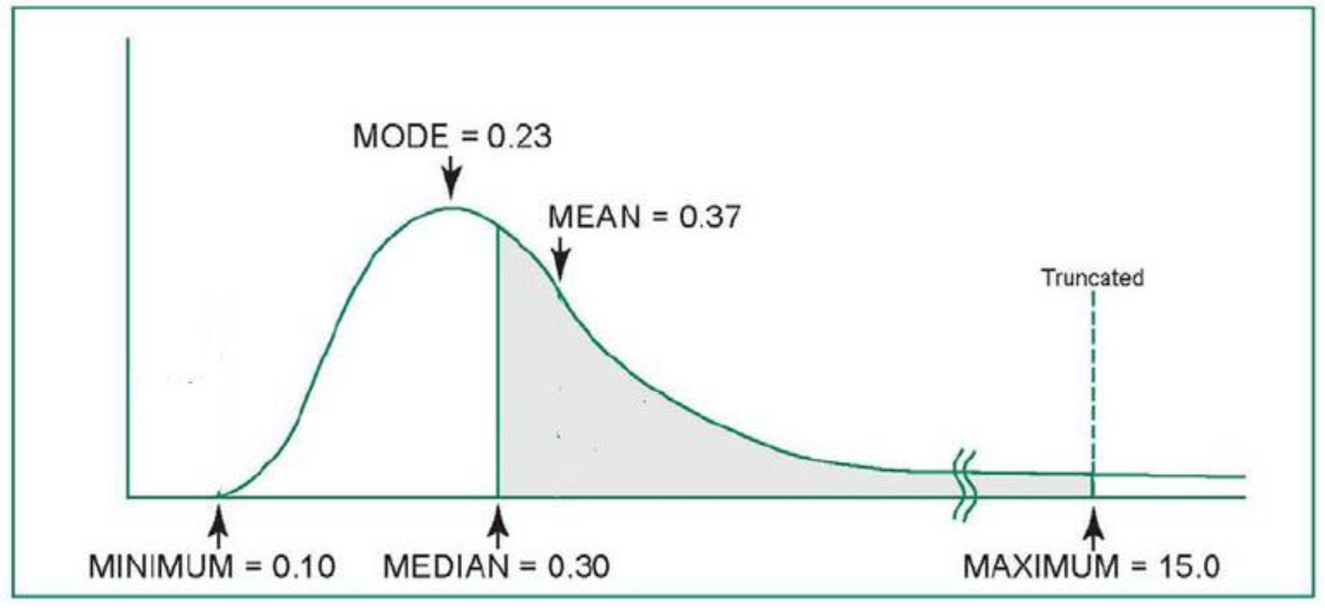
Motivation

We were asked to predict the performance of a book of business with 10,000+ policies

Traditionally this problem is approached by having two models:

- Working Layer Model
- Extreme Value Model

However we had a transactional history of our data available which allowed us to try and improve on that method.



Recommended Reading: McNulty, D.Severity Curve Fitting for Long-Tailed Lines: An Application of Stochastic Processes and Bayesian Models

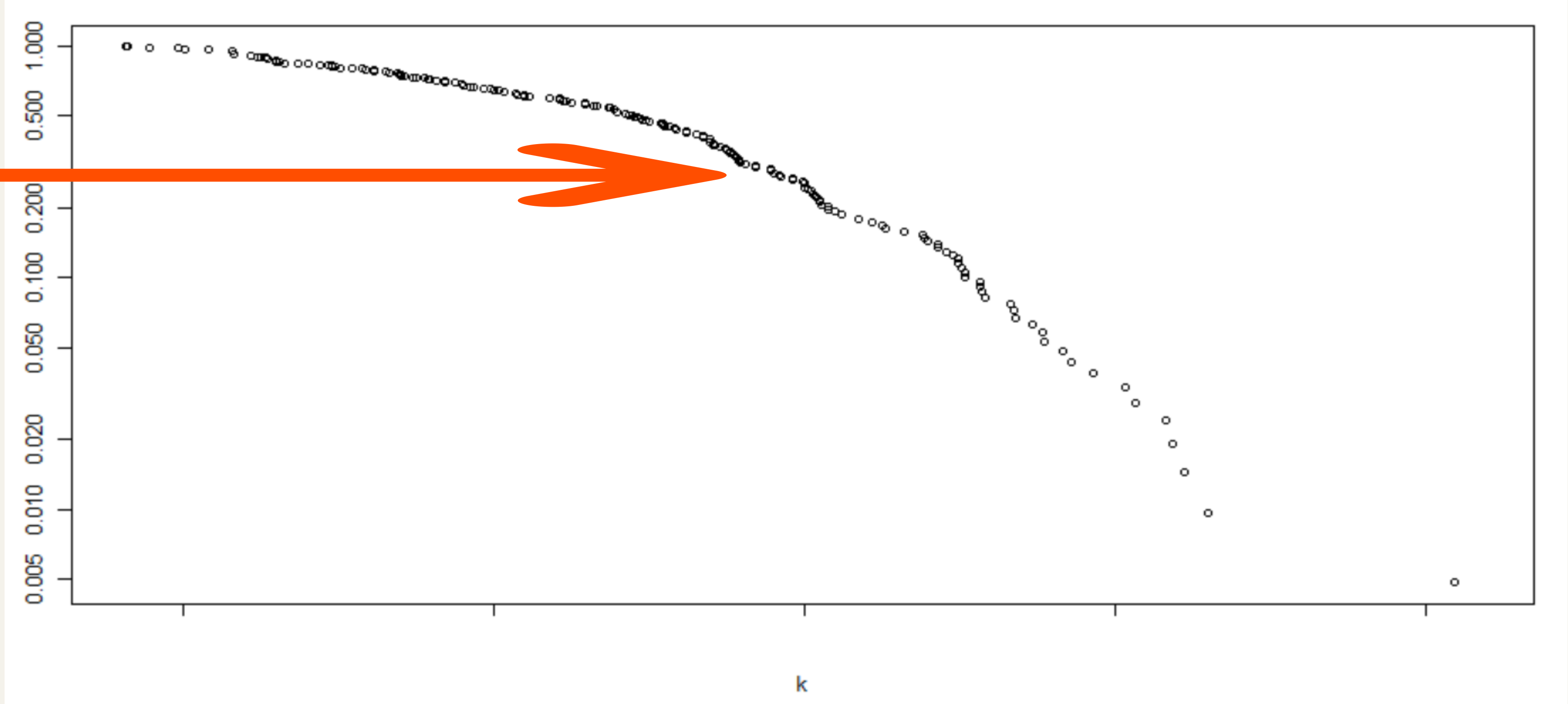
Why LogNormal-Pareto?

• Insurance models often struggle to account for extreme events in the long tails of distributions.

• We examined a log-log survival plot, which helps visualize the behaviour of extreme events in the data.

• The purpose of a log-log survival plot is to assess the shape and behaviour of the tail of a distribution, particularly when extreme events occur infrequently.

Nearly 50% of out data is pareto tailed!



Recommended Reading: Cirillo, P., & Taleb, N. N. (Year). Tail risk of contagious diseases

How did we go about building it?

Recommended Reading: Basu, S., Gil, M., & Auddy, S. The MLP distribution: a modified lognormal power-law model for the stellar initial mass function

```
log_normal_pareto_family <- custom_family(
  name: "LNP", dpars = c("mu","sigma","alpha"),
  links = c("identity","identity","identity"),
  type = "real"
)
stan_funs_LNP <- "
  real LNP_lpdf(real y, real tempmu, real tempsigma, real tempalpha) {
    real mu= tempmu;
    real alpha = tempalpha + 1e-3;
    real sigma= tempsigma + 1e-3;

    return log(alpha)-log(2)+alpha*mu+0.5*alpha^2*sigma^2-
    (1+alpha)*log(y)+log(erfc((alpha*sigma-(log(y)-mu)/sigma)/sqrt(2)));
  }
```

We implemented our model in BRMS, a high level interface to interact with the probabilistic programming language Stan

CmdStan can do this parallelised

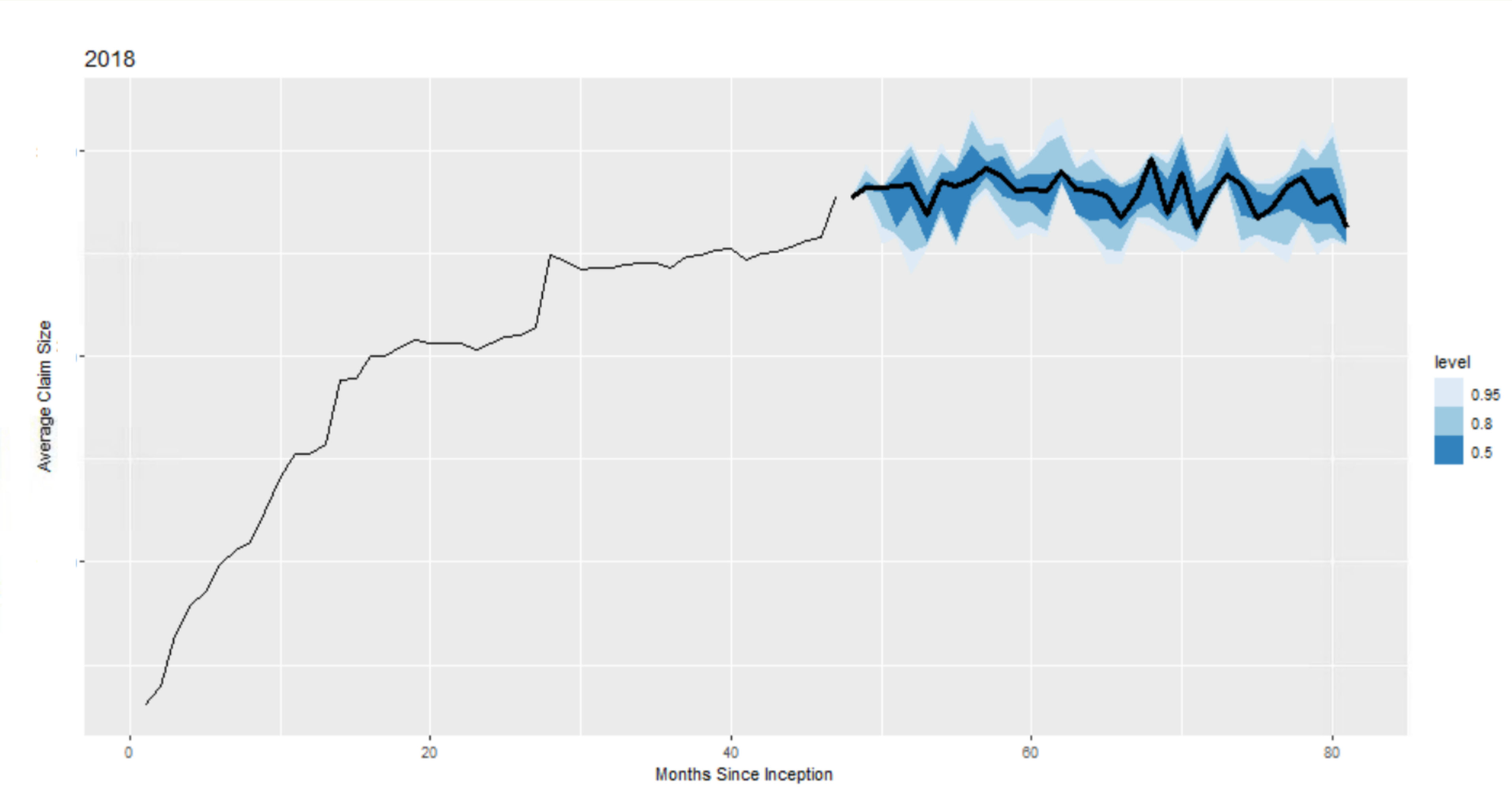
We built a custom family and our model predicts 3 main parameters: Mu, Alpha, and Sigma

LogNormal-Pareto model with a LogNormal growth curve with a time series analysis with hierarchies on year and segment

We used a cloud computer with 64 cores and 132 Gb of RAM

There was times where even this machine struggled

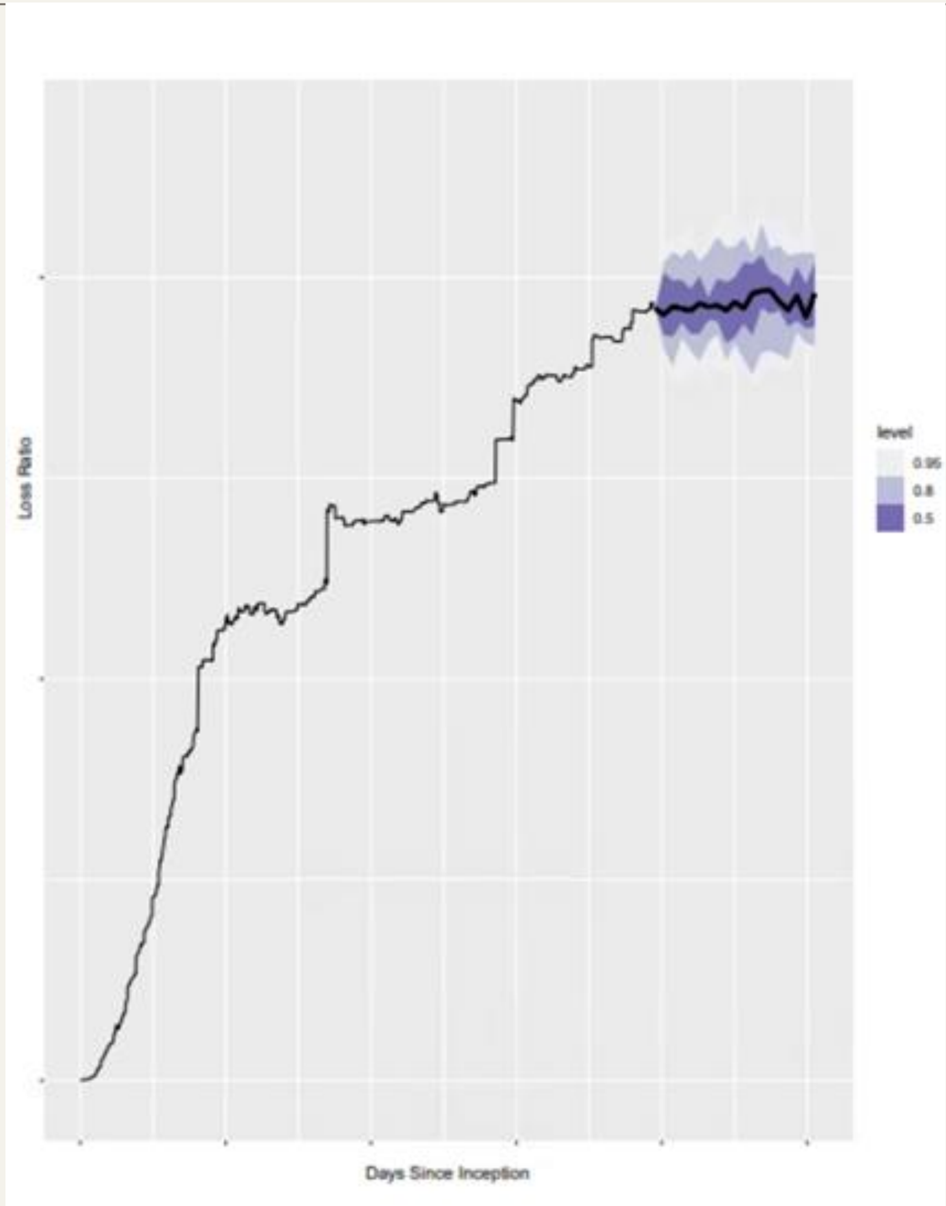
Severity Results



Here we show a low iteration run of a single year / segment

We can see that it is beginning the stabilise but broadly has expectation of average claim size decreasing (negative IBNR)

Loss Ratio Results



The real power of this is that when combined with frequency we can get loss ratio fan charts

We got stable results and forecasts for loss ratio for all years and segments for our book of business

Interested in more?

There will be a full talk (1 hr+) in September at the Bayesian Mixer

Please come up and talk to us if you want to know more

Contact Us

Cynon.Sonkkila@markel.com

Chris.Halliwell@markel.com