

F79MA: Assessed Project 2

Amit Parekh | H00184445

Semester 1, 2017

1 Introduction

This report contains an analysis of whether or not the company holds an appropriate amount of capital in reserve to cover the possible claims from a given set of policies.

The probability that the total claims over the next year exceeds a reserve amount of £1M is 0.0538, which means that there is approximately a 1 in 20 chance that the company will not be have sufficient funds to pay out its claims for the following year. In order for the company to ensure a 99% certainty that it will have sufficient funds to cover the claims for the next year, it needs to have at least £5.66M in reserves.

These results have been simulated using Jeffreys prior, using data of the total claims made per year for the last 10 years.

2 Analysis

2.1 Identifying a prior

Due to a lack of experience with the policies of this form, I've used Jeffreys prior as it is a non-informative prior distribution for θ , meaning it only expresses vague information about θ whilst supporting a wide range of possible values. Since it is also a flat prior, it will also have little impact on the posterior density.

Jeffreys prior is defined as

$$\pi_j(\theta) \propto \sqrt{I(\theta)},$$

where $I(\theta)$ is the information function defined as,

$$\begin{aligned} I(\theta) &= \mathbf{E} \left(\left[\frac{\partial}{\partial \theta} \log L(\theta; \underline{X}) \right]^2 \right) \\ &= -\mathbf{E} \left(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \underline{X}) \right) \end{aligned}$$

Since the total claim amount in any given year i , is $x_i \sim \text{Pareto}(x_m, \theta)$, where $x_m = £0.1M$ and θ is an unknown shape parameter. Therefore, the density of the sample data is

$$f(x; \theta) = \frac{\theta x_m^\theta}{x_i^{\theta+1}}.$$

From this, the likelihood function can be used to find Jeffreys prior:

$$\begin{aligned}
L(\theta; \underline{X}) &= \prod_{i=1}^n f(x_i; \theta) \\
&= \prod_{i=1}^n \frac{\theta x_m^\theta}{x_i^{\theta+1}} = \frac{\theta^n x_m^{n\theta}}{[\prod_{i=1}^n x_i]^{(\theta+1)}} \\
\log L(\theta; \underline{X}) &= n \log \theta + n\theta \log(x_m) - (\theta + 1) \sum_{i=1}^n \log(x_i) \\
\frac{\partial}{\partial \theta} \log L(\theta; \underline{X}) &= \frac{n}{\theta} + n \log(x_m) - \sum_{i=1}^n \log(x_i) \\
\frac{\partial^2}{\partial \theta^2} \log L(\theta; \underline{X}) &= \frac{-n}{\theta^2} \\
\therefore I(\theta) &= -\mathbf{E}\left(\frac{-n}{\theta^2}\right) = \frac{n}{\theta^2} \\
\therefore \pi_j(\theta) &\propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta^2}} \propto \frac{1}{\theta} \\
\therefore \pi_j(\theta) &= \frac{1}{\theta}
\end{aligned}$$

2.2 The posterior density

The posterior density $\pi(\theta|\underline{x})$ can be derived from the prior density and the likelihood function.

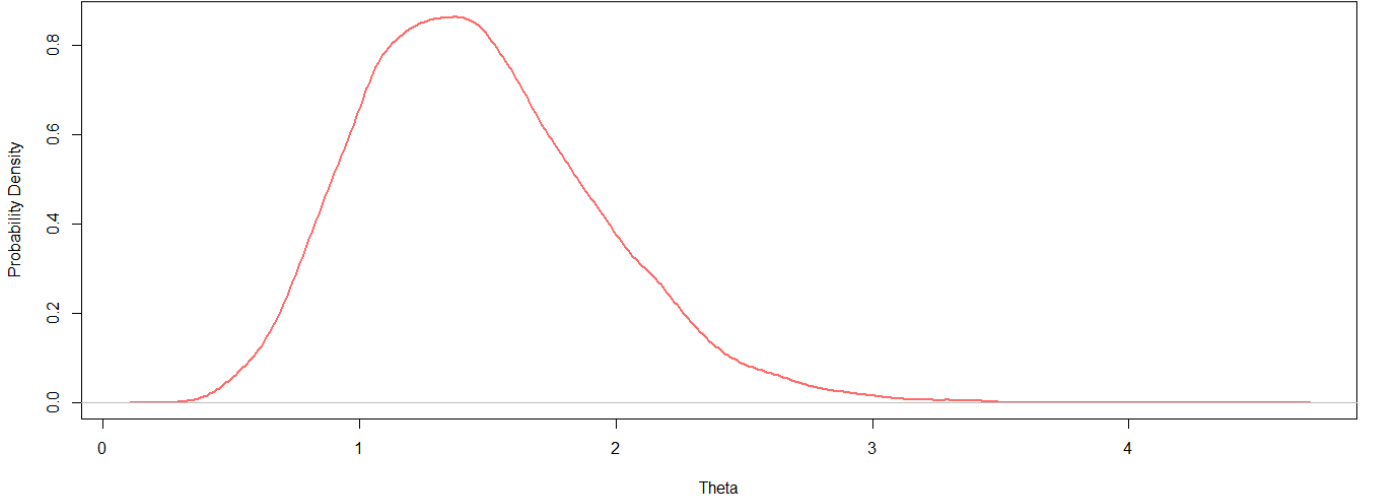
$$\begin{aligned}
\pi(\theta|\underline{x}) &\propto \pi(\theta) \times L(\theta; \underline{x}) \\
&\propto \frac{1}{\theta} \times \prod_{i=1}^n \frac{\theta x_m^\theta}{x_i^{\theta+1}} \\
&= \frac{1}{\theta} \times \frac{\theta^n x_m^{n\theta}}{\prod_{i=1}^n x_i^{\theta+1}} \\
&= \theta^{n-1} \times x_m^{n\theta} \times \left[\prod_{i=1}^n x_i \right]^{-(\theta+1)}
\end{aligned}$$

Removing the proportionality constant,

$$\begin{aligned}
&\propto \theta^{n-1} \times x_m^{n\theta} \times \left[\prod_{i=1}^n x_i \right]^{-\theta} \\
&= \theta^{n-1} \times e^{(n\theta) \log x_m} \times e^{-\theta \sum_{i=1}^n \log x_i} \\
&= \theta^{n-1} \times e^{-\theta(\sum_{i=1}^n \log x_i - n \log x_m)} \\
&\Rightarrow \theta|\underline{x} \sim \Gamma(\alpha, \beta), \text{ where } \alpha = n, \text{ and } \beta = \sum_{i=1}^n \log x_i - n \log x_m \\
&\Rightarrow \pi(\theta|\underline{x}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \times \theta^{\alpha-1} \times e^{-\beta\theta}
\end{aligned}$$

Table 1: 6-number summary of the posterior density					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
0.3036	1.1242	1.4203	1.4674	1.7567	4.5033

Fig 1: Distribution of the Posterior Density, for 10000 trials



2.3 Posterior predictive density

Denote y to be the total claim amount predicted for the next year. The predictive posterior density, $\pi(y|\underline{x})$ can be derived from the posterior density and the marginal density of y .

$$\begin{aligned}
 \pi(y|\underline{x}) &= \int_0^\infty \pi(\theta|\underline{x}) \times \pi(y|\theta) d\theta \\
 &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \frac{\theta x_m^\theta}{y^{\theta+1}} d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \frac{\theta x_m^\theta}{y^\theta} d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^\alpha \cdot e^{-\beta\theta} \cdot \left(\frac{x_m}{y}\right)^\theta d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^\alpha \cdot e^{-\beta\theta} \cdot e^{\theta \log(\frac{x_m}{y})} d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^\alpha \cdot e^{-\theta(\beta - \log(\frac{x_m}{y}))} d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^{\alpha'-1} \cdot e^{-\theta\beta'} d\theta,
 \end{aligned}$$

where $\alpha' = \alpha + 1$, and $\beta' = \beta - \log(\frac{x_m}{y})$.

$$\begin{aligned}
 \Rightarrow \pi(y|\underline{x}) &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \theta^{\alpha'-1} \cdot e^{-\theta\beta'} \cdot \frac{\Gamma(\alpha')}{\Gamma(\alpha')} \cdot \frac{\beta'^{\alpha'}}{\beta'^{\alpha'}} d\theta \\
 &= \frac{\beta^\alpha}{\Gamma(\alpha)y} \int_0^\infty \underbrace{\frac{\beta'^{\alpha'}}{\Gamma(\alpha')} \cdot \theta^{\alpha'-1} \cdot e^{-\theta\beta'}}_{= \Gamma(\alpha', \beta')} \cdot \frac{\Gamma(\alpha')}{\beta'^{\alpha'}} d\theta \\
 &= \frac{\beta^\alpha \cdot \Gamma(\alpha')}{\Gamma(\alpha)y \cdot \beta'^{\alpha'}} \underbrace{\int_0^\infty \Gamma(\alpha', \beta') d\theta}_{= 1} \\
 &= \frac{\beta^\alpha \cdot \Gamma(\alpha + 1)}{\Gamma(\alpha)y \cdot (\beta - \log(\frac{x_m}{y}))^{\alpha+1}} \\
 \therefore \pi(y|\underline{x}) &= \frac{\alpha\beta^\alpha}{y(\beta - \log(\frac{x_m}{y}))^{\alpha+1}}
 \end{aligned}$$

Fig 2: Posterior predictive density for the total claim amount for the next year

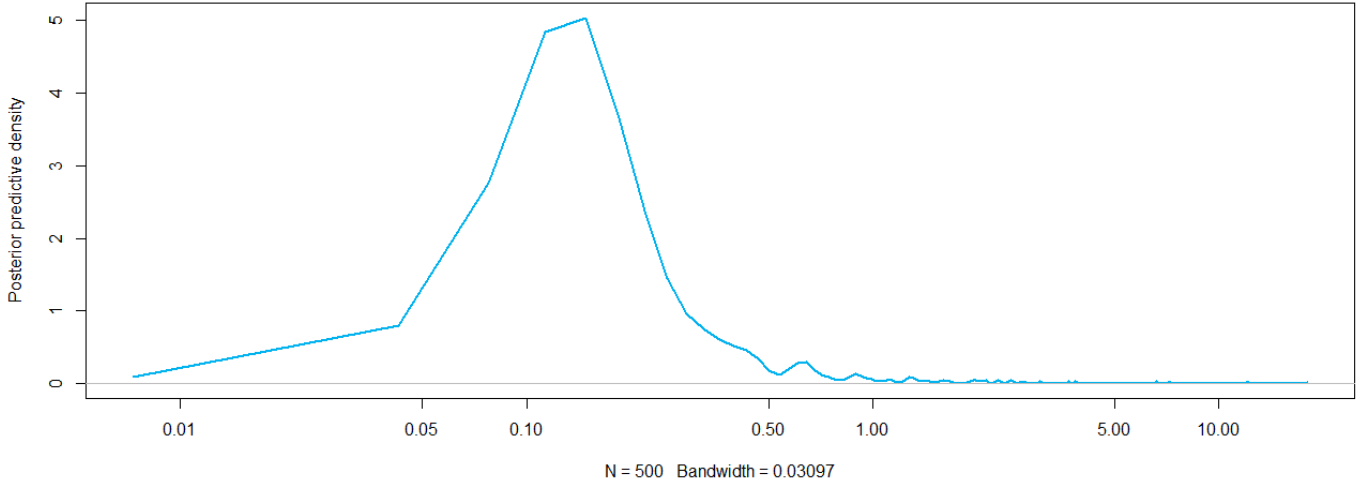


Table 2: 6-number summary of the posterior predictive distribution					
<i>Minimum</i>	<i>Lower Quartile</i>	<i>Median</i>	<i>Mean</i>	<i>Upper Quartile</i>	<i>Maximum</i>
0.1002	0.1190	0.1625	0.3924	0.2788	17.9433

From simulating the Pareto distribution using the Inverse CDF technique,

$$\begin{aligned}
 \theta^{(i)} &\sim \Gamma(\alpha, \beta), \\
 u^{(i)} &\sim U[0, 1], \\
 \therefore y^{(i)} &= \frac{x_m}{(1 - u^{(i)})^{\frac{1}{\theta}}} \\
 &\Rightarrow \mathbf{P}(Y > \pounds 1M), \text{ i.e. the expected number of } y^{(i)} \text{ that are greater than } \pounds 1M.
 \end{aligned}$$

The likelihood that the total claims next year exceeds the total reserve of $\pounds 1M$ is estimated to be 0.0538, calculated from a simulation with a suitably large sample size of $n = 10000$.

In order to ensure that the company is able to pay its claims for the following year with 99% certainty, the company needs to have a reserve value of at least $\pounds 5.66M$.

This is because,

$$\begin{aligned}
 \mathbf{P}(Y \leq \text{reserve}) &= 0.99 \\
 \Rightarrow \mathbf{P}(Y > \text{reserve}) &= 0.01 \\
 \therefore \text{reserve} &\geq \pounds 5.66M
 \end{aligned}$$

The Jeffreys posterior reflects the best possible information about θ from the data, as the prior has little impact on the posterior. My advice is sensitive to the prior as it results in a posterior best reflecting the data available. If I were to change the prior, it could rely less on the previous 10 data points and could lead to different results entirely.

3 Appendix

3.1 Notes about code

Each section of code should be run independently of another.

I have use magrittr in the script, so if it doesn't work, check to see if the magrittr package is installed. You can run the code below to install and use magrittr prior to running a script.

```
install.packages("magrittr")
library("magrittr")
```

3.2 R code

```
# sample data
x <- c(0.324, 0.177, 0.163, 0.317, 0.326, 0.174, 0.321, 0.115, 0.108, 0.133)
n <- length(x)
xMin <- 0.1

# calculate beta
b <- 0
for (i in 1:n) { b <- b + log(x[i]) }; b <- b - n*log(xMin)

# calculate alpha
a <- n

# Plot posterior density
postDensity <- 0
for (i in 1:10000) {
  postDensity[i] <- rgamma(1, a, b)
}
plot(density(postDensity),
     main = "Fig 1: Distribution of the Posterior Density, for 10000 trials",
     xlab="Theta", ylab="Probability Density",
     col="indianred1", lwd=2)
summary(postDensity)

# Simulate posterior predictive density
y <- 0
for (i in 1:500) {
  y[i] <- 0

  while (y[i] < 0.1) {
    u <- runif(1, 0, 1)
    theta <- rgamma(1, n, b)
    y[i] <- xMin / ( (1-u)^(1/theta) )
  }
}
plot(density(y), log="x", type='l',
     main = "Fig 2: Posterior predictive density for the total claim amount for the next year",
     ylab = "Posterior predictive density",
     col="deepskyblue2", lwd=2 )
summary(y)

# Estimate probability that y exceeds £1m
y <- 0
for (i in 1:10000) {
  y[i] <- 0
  while (y[i] < 0.1) {
    u <- runif(1, 0, 1)
    theta <- rgamma(1, n, b)
    y[i] <- xMin / ( (1-u)^(1/theta) )
  }
}
probest <- (y > 1) %>% sum() / 10000

# Calculate reserve value
reserve <- quantile(y, 0.99)
probRes <- (y > reserve) %>% sum() / trials
```