Math 523 Project 1

Xi Sun

June 19, 2019

1 Introduction

In this project, we are interested in the topic about insurance risk and bankrupt. In an insurance company, we have 1,000 customers and $1,000,000 in assets, with probability of a customer filling the claim is 0.1 which is independent of other customers and previous claims. Denote X be the size of the claim, then X follows a Pareto distribution. Also, we know that the true asset is the previous asset plus the annual premium and minus the claims. Denote this process as $Z(t)$, if $Z(t)$ falls below 0, then $Z(t)$ will stay 0 forever, which we call that as bankrupt.

2 Statistical and Computational Analysis

Before we actually dig into this "bankrupt" problem, we are interested in investigating the Pareto distribution first. Based on the theory, we have the Pareto distribution as following:

$$f(x) = \begin{cases} \frac{ab^a}{(x+b)^{a+1}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

with $a = 5$ and $b = 200,000$. And its cumulative distribution function(CDF), mean, variance, skewness, and excess kurtosis will be calculated and analyzed.

2.1 Pareto Distribution

First, according to the PDF, we want to analytically compute the CDF, expectation, variance, skewness, and excess kurtosis. With $a = 5$, we have the following:

$$\begin{aligned} \text{F(x)} &= \int_{-\infty}^{\infty} \text{f(x)}dx \\ &= \int_{0}^{\infty} \frac{5b^5}{(x+b)^6}dx \\ &= -\frac{b^5}{(x+b)^5} + c \end{aligned}$$

Since we need $\text{F}(\infty) = 1$, so $c = 1$, and with $b = 200,000$, therefore, we have CDF being

$$\text{F(x)} = 1 - \left(\frac{200000}{x + 200000}\right)^5.$$

For expectation, we have the following:

$$\begin{aligned} \text{E[X]} &= \int_{-\infty}^{\infty} x\text{f(x)}dx \\ &= \int_{0}^{\infty} x\frac{5b^5}{(x+b)^6}dx \\ &= -\frac{xb^5}{(x+b)^5} - \frac{b^5}{4(x+b)^4}\Bigg|_0^{\infty} \\ &= 0 - (-\frac{b^5}{4b^4}) \\ &= \frac{b}{4} \\ &= 50,000 \end{aligned}$$

For variance, we will calculate the E(x^2) first:

$$\begin{aligned}
\mathrm{E[X^2]} &= \int_{-\infty}^{\infty} x^2 \mathrm{f(x)} dx \\
&= \int_{0}^{\infty} x^2 \frac{5b^5}{(x+b)^6} dx \\
&= -\frac{x^2 b^5}{(x+b)^5} - \frac{2xb^5}{4(x+b)^4} - \frac{2b^5}{12(x+b)^3}\Big|_0^\infty \\
&= 0 - (-\frac{2b^5}{12b^3}) \\
&= \frac{b^2}{6}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var(X)} &= \mathrm{E[X^2] - E[X]^2} \\
&= \frac{b^2}{6} - \frac{b^2}{16} \\
&= \frac{5b^2}{48} \\
&\approx 4,166,666,666.67
\end{aligned}$$

For skewness, we have the following:

$$\begin{aligned}
\mathrm{skew[X]} &= \int_{-\infty}^{\infty} (\frac{x-\mu}{\sigma})^3 \mathrm{f(x)} dx \\
&= \left(\frac{48}{5b^2}\right)^{\frac{3}{2}} \int_{0}^{\infty} (x - \frac{b}{4})^3 \frac{5b^5}{(x+b)^6} dx \\
&= \left(\frac{48}{5b^2}\right)^{\frac{3}{2}} (-(x - \frac{b}{4})^3 b^5 (x + b^{-5}) - 3(x - \frac{b}{4})^2 \frac{b^5}{4}(x+b)^{-4} - 6(x - \frac{b}{4})\frac{b^5}{12}(x+b)^{-3} - \frac{6b^5}{24}(x+b)^{-2}\Big|_0^\infty) \\
&= \left(\frac{48}{5b^2}\right)^{\frac{3}{2}} \left(0 - (-\frac{10b^3}{64})\right) \\
&= \left(\frac{48}{5b^2}\right)^{\frac{3}{2}} \left(\frac{5b^3}{32}\right) \\
&= 6\sqrt{\frac{3}{5}} \\
&\approx 4.64758
\end{aligned}$$

Finally, for excess kurtosis, we need to calculate kurtosis first:

$$
\begin{aligned}
\text{kur[X]} &= \int_{-\infty}^{\infty} (\frac{x-\mu}{\sigma})^4 \text{f(x)} dx \\
&= \left(\frac{48}{5b^2}\right)^2 \int_0^{\infty} (x-\frac{b}{4})^4 \frac{5b^5}{(x+b)^6} dx \\
&= \left(\frac{48}{5b^2}\right)^2 (-(x-\frac{b}{4})^4 b^5 (x+b)^{-5} - (x-\frac{b}{4})^3 b^5 (x+b)^{-4} - (x-\frac{b}{4})^2 b^5 (x+b)^{-3} - (x-\frac{b}{4}) b^5 (x+b)^{-2} \\
&- b^5 (x+b)^{-1} |_0^{\infty}) \\
&= \left(\frac{48}{5b^2}\right)^2 \left(0 - (-\frac{205b^4}{256})\right) \\
&= \left(\frac{48}{5b^2}\right)^2 \left(\frac{205b^4}{256}\right) \\
&= \frac{396}{5} \\
&= 73.8
\end{aligned}
$$

So the excess kurtosis is $73.8 - 3 = 70.8$.

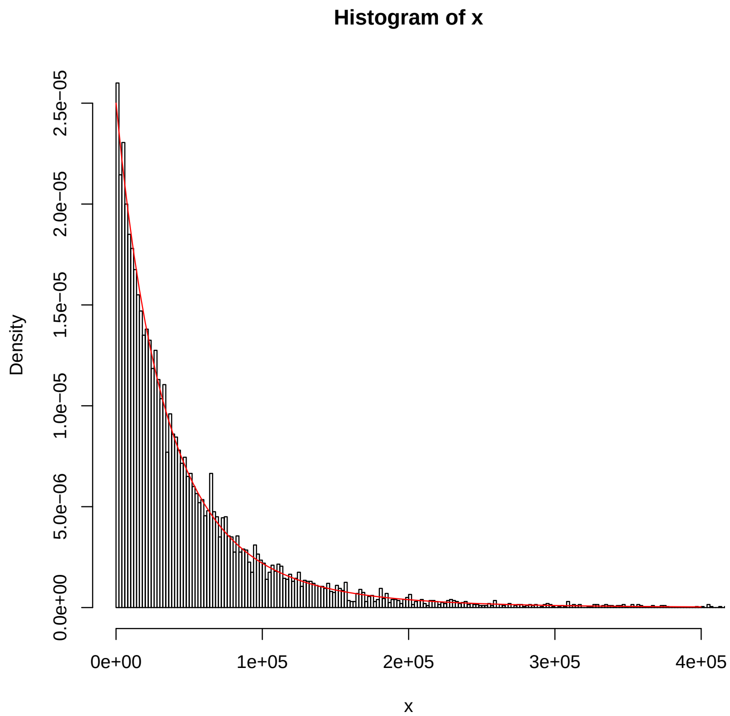
Every theoretical result should match the actual result in real world application, or simulation in this case. Unless they are not consistent, so several simulation in R is down to try to prove the result.

First, some packages in R are loaded for this project:

```r
library(actuar)
library(moments)
```

Now, a simulation of a Pareto distribution with 10,000 sample points are run, with a general Pareto distribution is fitted:

```r
x <- rpareto(10000,shape=5,scale=200000)
hist(x,breaks=500,freq=FALSE,xlim=c(0,400000))
curve(dpareto(x,shape=5,scale=200000),col="red",add=TRUE)
```



As we can see from the plot, the sample points shown by a histogram and a fitted curve of Pareto distribution actually corresponds to each other very well. Now the theoretical quantities of mean, variance, skewness, and excess kurtosis are needed to be verified. However, based on some self-work being done, 20 results are generated as follow:

```
m <- c()
v <- c()
s <- c()
ek <- c()
for (i in 1:20){
  x <- rpareto(10000,shape=5,scale=200000)
  m[i] <- mean(x)
  v[i] <- var(x)
  s[i] <- skewness(x)
  ek[i] <- kurtosis(x)-3
}
m #mean

##  [1] 48875.65 51502.22 49796.93 50046.85 50466.21 51088.39 50455.65
##  [8] 51103.99 50231.73 50462.66 50649.73 50131.16 51671.31 49107.65
## [15] 49807.82 50943.35 50428.40 49560.25 51297.07 50011.48

v #variance

##  [1] 3828292514 5060081364 3983085157 4213839320 4303021749 4780119583
##  [7] 3893069808 4380812289 4368771677 4091538873 4301090886 3758008575
```

```
## [13] 4668882629 3969843617 3741582709 4423395110 4278646612 3867044206
## [19] 4765389752 4359852331

s #skewness

##  [1] 3.697867 5.941636 3.910899 3.956157 4.493396 5.380228 3.232515
##  [8] 3.985306 5.072476 3.761865 4.049969 3.245807 5.148026 4.067682
## [15] 3.662476 4.495547 4.425367 3.830784 6.587722 6.399480

ek #excess kurtosis

##  [1]  24.59926  85.36549  27.75196  26.55402  41.95220  62.50277  16.33317
##  [8]  30.54057  56.62711  26.02917  35.40071  18.49768  63.97758  32.65612
## [15]  27.55770  41.18524  41.88943  29.88318 133.77842 125.77853
```

As we can see from the data, only the mean and variance is staying around the theoretical mean and variance, and Therefore, for the Pareto distribution in this case, with $a = 5$ and $b = 200,000$, we can say that only the mean and variance is consistent comparing with the theoretical mean and other quantities are not consistent. And that is everything about the analysis of Pareto distribution. The actual simulation of company's assets will be discussed in the next part.

2.2 Simulation of Bankrupt

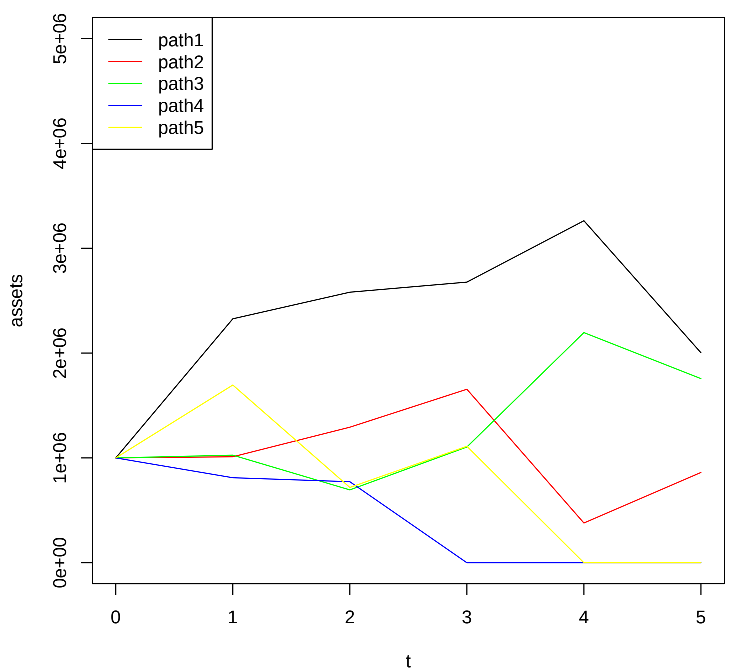
According to the theoretical formula, if $Z(t)$ is the company assets at the end of year t with $Z(0) = 1,000,000$ and

$$Z(t) = \begin{cases} \max[Z(t-1) + \text{premiums} - \text{claims}, 0] & Z(t-1) > 0 \\ 0 & Z(t-1) = 0 \end{cases}$$

In order to understand how this formula works and how it is looked like, five different sample path is plotted to give a preview of the company's assets:

```
z <- c()
z[1] <- 1000000
col_vec <- c("red","green","blue","yellow")
for (i in  1:5){
  if (z[i]==0){
    z[i+1] <- 0
  }else{
    z[i+1] <- max(z[i]+5000000-sum(rpareto(rbinom(1,1000,0.1),5,200000)),0)
  }
}
plot(0:5,z,ylim=c(0,5000000),xlab="t",ylab="assets",type="l")
for (j in 1:4){
  for (i in 1:5){
    if (z[i]==0){
```

```
      z[i+1] <- 0
    }else{
      z[i+1] <- max(z[i]+5000000-sum(rpareto(rbinom(1,1000,0.1),5,200000)),0)
    }
  }
  lines(0:5,z,col=col_vec[j])
}
legend("topleft",legend=c("path1","path2","path3","path4","path5"),
       col=c("black",col_vec),lwd=1)
```



As we can see from the plot, since the value on each path is completely at random without setting any random seed, so there is no value to investigate the actual value for any of the paths. However, we can still look at the trend of the path, all five paths are shown decreasing, at least from this plot, we can conclude that the company's assets are tending to be bankrupted.

In order to estimate analyze company's assets, we need to find the probability of the company goes to bankrupt, as well as the mean value of the company assets after 5 years, so the last part of the simulation is applied:

```
z <- c()
z[1] <- 1000000
z5 <- c()
for (j in 1:10000){
  for (i in 1:5){
    if (z[i]==0){
```

```r
      z[i+1] <- 0
    }else{
      z[i+1] <- max(z[i]+5000000-sum(rpareto(rbinom(1,1000,0.1),5,200000)),0)
    }
  }
  z5[j] <- z[6]
}
sum(z5==0)/length(z5)

## [1] 0.4133

mean(z5)

## [1] 1205535
```

As we can see from the result, the probability for a company goes to bankrupt in 5 years is around 41%. And the mean value is approximately 1.2 million for the amount of assets that an insurance company can earn.

3 Conclusion

From this project, first of all, we can conclude that the Pareto distribution is really unstable, ie, large difference in skewness and kurtosis, which increases the risk probability. Therefore, based on this risk, insurance companies can really earn some amount of money since the starting asset is 1 million and the mean is approximately 1.2 million after 5 years. On the other hand, insurance companies really have a high insurance risk. Although some company can earn more than the beginning, there is still a 41% chance that a company can go to bankrupt in 5 years.