Aggregate Loss Models

A short course authored by the Actuarial Community

21 Jan 2021

Introduction

Basic Terminology

- Loss - amount of financial losses suffered by the insured
- Claim - indemnification upon the occurrence of an insured event. Some authors use claim and loss interchangeably
- Frequency - how often an insured event occurs (claim count) in one period (typically six months or one year)
- Severity - Amount, or size, of losses for an insured event
- Aggregate Loss - Total amount paid for a defined set of insureds in one period (typically six months or one year)

Goal

- Build a model for the total payments by an insurance system in a fixed time period
- The insurance system could be a single policy or a portfolio of policies
- Frequency and severity models are building blocks

Models

- Two ways to build a model for aggregate losses on a defined set of insurance contracts
- Individual Risk Model: record losses for each contract and then add them up
- Collective Risk Model (a.k.a. compound model): record losses as claims are made and then add them up

Example

- An insurance portfolio of three policies:

Policy ID	Claim ID	Loss Amount
1	-	-
2	1	10
3	1	10
3	2	10
4	1	10
4	2	10
4	3	10

- Aggregate losses:
  - Individual Risk Model: $0 + 10 + 20 + 30 = 60$
  - Collective Risk Model: $10 + 10 + 10 + 10 + 10 + 10 = 60$

Applications

- Actuarial applications of aggregate loss models
  - Ratemaking
  - Capital management
  - Risk financing

Summary

In this section, we learned how to:

- Record aggregate losses from an insurance system
- Identify actuarial applications of aggregate loss models

Individual Risk Model

Individual Risk Model

The **individual risk model** represents the aggregate loss as a sum of a fixed number of insurance contracts

$$S = X_1 + \ldots + X_n,$$

where

- S denotes the aggregate loss for n (a fixed number) contracts
- X_i denotes the loss for the ith contract for $i = 1, \ldots, n$
- X_i are assumed to independent but are not necessarily identically distributed, due to different coverage or exposure
- X_i usually has a probability mass at zero

Applications

- Originally developed for life insurance
  - Probability of death within a year is q_i;
  - Fixed benefit paid for the death of the ith person is b_i.
- The distribution of the loss to the insurer for the ith policy is

$$f_{X_i}(x) = \begin{cases} 1 - q_i, & x = 0 \\ q_i, & x = b_i \end{cases}$$

Applications

- Two-part framework

$$X_i = I_i \times B_i = \begin{cases} 0, & I_i = 0 \\ B_i, & I_i = 1 \end{cases}$$

- $I_1, \ldots, I_n$, $B_1, \ldots, B_n$ are independent.
- I_i is an indicator (Bernoulli) that is 1 with probability q_i and 0 with probability 1-q_i. It indicates whether the ith policy has a claim.
- B_i, a r.v. with nonnegative support, represents the amount of losses of policy i, given that a claim is made. It can follow a degenerate distribution such as the life insurance example.

Example

Consider the example of Wisconsin Property Fund.

```
Insample <- read.csv("Insample.csv", header = T)
Insample2010 <- subset(Insample, Year==2010)

I <- 1*(Insample2010$Freq>0)
B <- Insample2010[which(Insample2010$Freq>0),]$y
```

```
length(I)
```

```
## [1] 1110
```

```
length(B)
```

```
## [1] 403
```

Example

```r
table(I)
```

```
## I
##   0   1
## 707 403
```

```r
summary(B)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##       200     3375     9378    90966    29845 12922218
```

Aggregate Loss Distribution

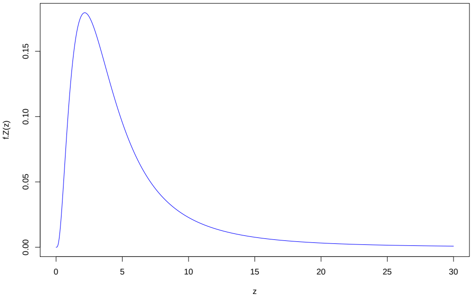
The distribution of aggregate loss S is defined as:

$$F_S(s) = \Pr(X_1 + X_2 + \cdots + X_n \leq s)$$

- In general, it is difficult to evaluate
- When X_i, $i = 1, \ldots, n$, are i.i.d., it is known as n-fold convolution of cdf of X

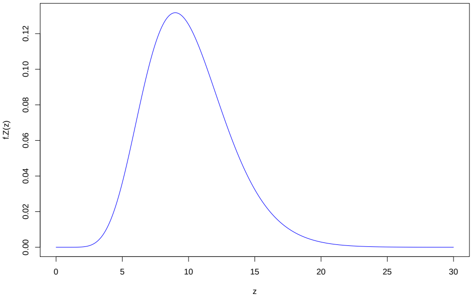
R Example #1

```
library(distr)
X <- Exp(rate=1); Y <- Lnorm(meanlog=1,sdlog=1)
conv <- convpow(X+Y,1)
f.Z  <- d(conv)
z <- seq(0,30,0.01)
plot(z,f.Z(z), col="blue",type="l")
```



R Example #2

```
library(distr)
X <- Exp(rate=1)
conv <- convpow(X,10)
f.Z  <- d(conv)
z <- seq(0,30,0.01)
plot(z,f.Z(z), col="blue",type="l")
```



Summary

In this section, we learned how to:

- Build an individual risk model for a portfolio of insurance contracts
- Apply individual risk model to life and nonlife insurance
- Compute the distribution of aggregate losses from an individual risk model

Collective Risk Model - Part I

Collective Risk Model

The **collective risk model** has representation

$$S = X_1 + \ldots + X_N,$$

with S being the aggregate loss of N (a random number) individual claims $(X_1, \ldots, X_N)$.

- Key assumptions
  - Conditional on $N = n$, $X_1, \ldots, X_n$ are i.i.d. random variables
  - The distribution of N and the common distribution of X are independent of each other.
- Two building blocks: frequency N and severity X

Compound Distribution

The cumulative distribution function (cdf) is $F_S(s)$. The probability density function (pdf) or probability mass function (pmf) is $f_S(s)$. Specifically,

$$
\begin{aligned}
F_S(s) &= \Pr(S \leq s) \\
&= \textstyle\sum_{n=0}^{\infty} \Pr(N = n) \cdot \Pr(S \leq s | N = n) \\
&= \textstyle\sum_{n=0}^{\infty} \Pr(N = n) \cdot F_X^{*n}(s)
\end{aligned}
$$

$$
f_S(s) = \textstyle\sum_{n=0}^{\infty} \Pr(N = n) \cdot f_X^{*n}(s)
$$

Moments

$$S = X_1 + \ldots + X_N$$

Using the law of iterated expectations to calculate the mean

$$\mathrm{E}(S) = \mathrm{E}[\mathrm{E}(S|N)] = \mathrm{E}[N\mathrm{E}(X)] = \mathrm{E}(N)\mathrm{E}(X)$$

Using the law of total variation to calculate the variance

$$\begin{aligned}
\mathrm{Var}(S) &= \mathrm{E}[\mathrm{Var}(S|N)] + \mathrm{Var}[\mathrm{E}(S|N)] \\
&= \mathrm{E}[N\mathrm{Var}(X)] + \mathrm{Var}[N\mathrm{E}(X)] \\
&= \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{Var}(N)\mathrm{E}(X)^2
\end{aligned}$$

Model Fitting

The assumptions suggest that we can build an aggregate loss model, the compound model, in three steps:

1. Develop a model for the frequency distribution of N, the primary distribution, based on data
2. Develop a model for the severity distribution of X, the secondary distribution, based on data
3. Using these two models, carry out the necessary calculations to obtain the distribution of S

Model Fitting

Consider the example of Wisconsin Property Fund.

```
Insample <- read.csv("Insample.csv", header = T)
Insample2010 <- subset(Insample, Year==2010)
```

Number of policyholders:

```
nrow(Insample2010)
```

```
## [1] 1110
```

Model Fitting

The frequency model: N

```
N <- Insample2010$Freq
```

```
table(N)
```

```
## N
##   0   1   2   3   4   5   6   7   8   9  10  11
## 707 209  86  40  18  12   9   4   6   1   3   2
##  13  14  15  16  17  18  19  30  39 103 239
##   1   2   1   2   1   1   1   1   1   1   1
```

Model Fitting

The severity model: $\bar{X} = (X_1 + \ldots + X_n)$ for $N = n > 0$

```
Xbar <- Insample2010$yAvg[which(Insample2010$Freq>0)]
```

```
summary(Xbar)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.
##       167     2226     4951    56332    11900
##      Max.
## 12922218
```

Summary

In this section, we learned how to:

- Build a collective risk model for a portfolio of insurance contracts
- Calculate mean and variance of the aggregate loss
- Fit frequency and severity components in a collective risk model

Collective Risk Model - Part II

Computing the Aggregate Loss Distribution

- Consider a collective risk model

$$S = X_1 + \cdots + X_N$$

- Several numerical strategies:
  - **direct calculation**: difficult part is the evaluation of n-fold convolutions
  - **recursive method**: considerable savings in computation time
  - **Monte Carlo simulation**: approximation using the empirical distribution
- `R` implementation: package `actuar`

Direct Calculation

- The distribution of S can be calculated using:

$$
\begin{aligned}
F_S(s) &= \Pr(S \leq s) \\
&= \sum_{n=0}^{\infty} \Pr(N = n) \cdot \Pr(S \leq s | N = n) \\
&= \sum_{n=0}^{\infty} \Pr(N = n) \cdot F_X^{*n}(s)
\end{aligned}
$$

- To compute the distribution in `R`, one has to discretize the severity distribution so that it has support $\{0, 1, \ldots, m\}$

Direct Calculation

The frequency and severity distributions are summarized by:

N	0	1	2	3	4
$f_N(n)$	0.2	0.2	0.2	0.2	0.2
X	50	100	150	250	
$f_X(x)$	0.2	0.3	0.4	0.1	

The discretized severity distribution is

X	0	1	2	3	4	5
$f_X(x)$	0	0.2	0.3	0.4	0	0.1

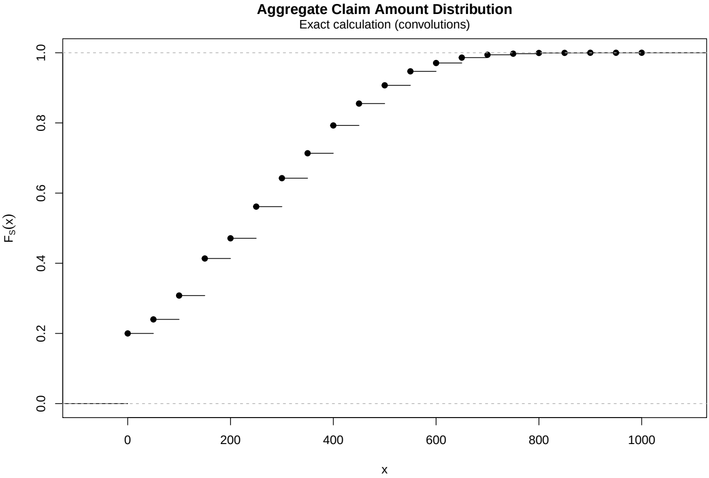
Direct Calculation

```r
library(actuar)
pn <- rep(0.2,5)
fx <- c(0,0.2,0.3,0.4,0,0.1)
Fs <- aggregateDist(method="convolution",
                    model.freq = pn,
                    model.sev = fx, x.scale = 50)
summary(Fs)
```

```
## Aggregate Claim Amount Empirical CDF:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0     100     250     250     400    1000
```

Direct Calculation

```
plot(Fs)
```



Recursive Method

- Recursive method assumes:
  - Frequency N satisfies $p_n = (a + b/n)p_{n-1}$ for $n = 1, 2, 3, \ldots$
  - Severity X has a distribution distribution f_X on support $\{0, 1, \ldots, m\}$
- The probability function of S can be calculated using:

$$\begin{aligned} f_S(s) &= \Pr(S = s) \\ &= \frac{1}{1 - af_X(0)} \sum_{x=1}^{s \wedge m} \left(a + \frac{bx}{s} \right) f_X(x) f_S(s - x) \end{aligned}$$

- The method extends to the more general case where frequency N satisfies $p_n = (a + b/n)p_{n-1}$ for $n = 2, 3, \ldots$

Recursive Method

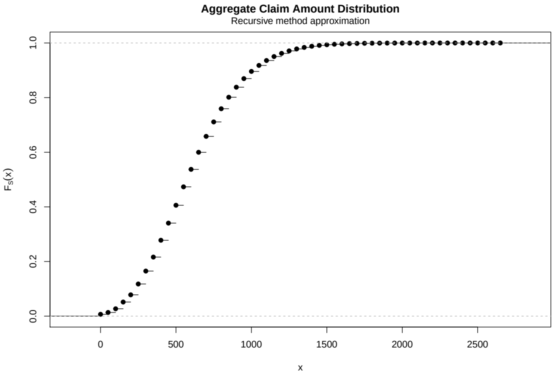
Consider the example:

- Frequency N is Poisson with mean 5.
- Severity

X	50	100	150	250
$f_X(x)$	0.2	0.3	0.4	0.1

Recursive Method

```
library(actuar)
fx <- c(0,0.2,0.3,0.4,0,0.1)
Fs <- aggregateDist(method="recursive",
                    model.freq = "poisson", lambda = 5,
                    model.sev = fx, x.scale = 50)
plot(Fs)
```



Monte Carlo Simulation

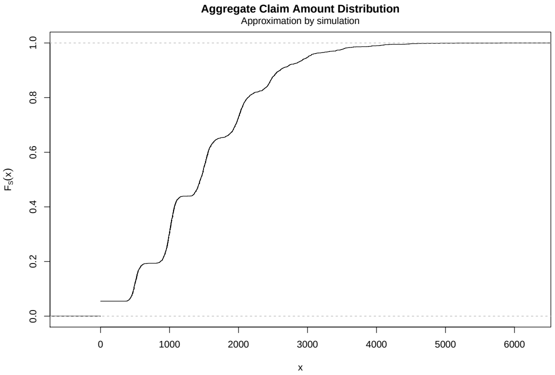
- For $j = 1, \ldots, m$, do
  - Generate the number of claims n_j from the frequency distribution N
  - Generate n_j claim amount from severity distribution X, denoted by $x_{j,1}, \ldots, x_{j,n_j}$
  - Calculate aggregate loss $s_j = x_{j,1} + \cdots + x_{j,n_j}$

- The empirical distribution can be calculated as

$$\hat{F}_s(s) = \frac{1}{m} \sum_{i=1}^{m} I(s_i \leq s)$$

Monte Carlo Simulation

```
library(actuar)
freq <- expression(data = rpois(lambda=3))
sev <- expression(data = rgamma(shape=100, scale=5))
Fs <- aggregateDist("simulation", nb.simul = 2500,
                    model.freq = freq, model.sev = sev)
plot(Fs)
```



Summary

In this section, we learned how to:

- Compute aggregate loss distribution
- Implement numerical strategies in `R`

Tweedie Distribution

Tweedie Distribution

The Tweedie distribution is defined as a Poisson sum of gamma random variables

$$S = (X_1 + \cdots + X_N)$$

where $N \sim \mathit{Poisson}(\omega\lambda)$ and $X \sim \mathit{Gamma}(\alpha, \theta)$.

Tweedie Distribution with Exposure

With exposure, the Tweedie variable is the aggregate losses per unit of expoure:

$$S = (X_1 + \cdots + X_N)/\omega$$

where

- ω is the exposure
- $N \sim \mathit{Poisson}(\omega\lambda)$
- $X \sim \mathit{Gamma}(\alpha, \theta)$

Distribution Function for Tweedie Distribution

The cdf of S is

$$\begin{aligned} F_S(s) &= \sum_{n=0}^{\infty} \Pr(N=n)\cdot\Pr(S\le s|N=n) \\ &= \Pr(N=0) + \sum_{n=1}^{\infty} \Pr(N=n)\cdot\Pr(S\le s|N=n) \\ &= e^{-\omega\lambda} + \sum_{n=1}^{\infty} e^{-\omega\lambda}\frac{(\omega\lambda)^n}{n!}\Gamma\left(n\alpha;\frac{s}{\theta/\omega}\right) \end{aligned}$$

Note that

$$S|(N=n) = (X_1+\cdots+X_n)/\omega \sim Gamma(n\alpha, \theta/\omega)$$

Tweedie Distribution

Consider reparameterizations

$$\lambda = \frac{\mu^{2-p}}{\phi(2-p)}, \quad \alpha = \frac{2-p}{p-1}, \quad \theta = \phi(p-1)\mu^{p-1}$$

For $p \in (1,2)$, the Tweedie distribution can be presented as:

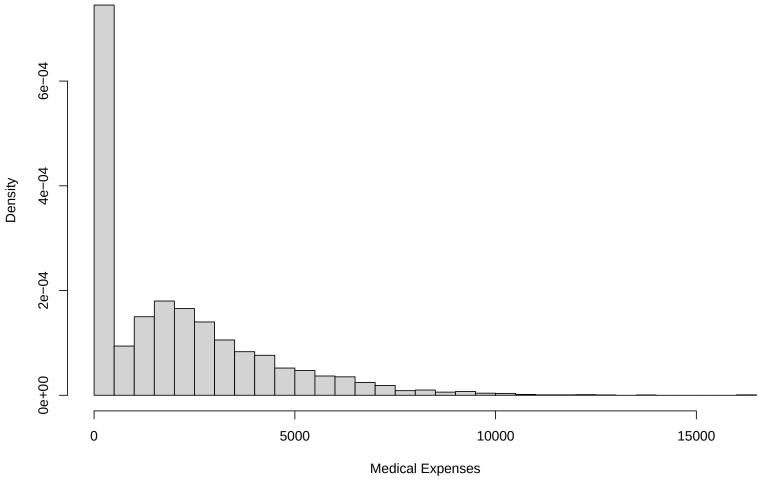
$$f_S(s) = \exp\left\{ \frac{\omega}{\phi} \left(\frac{-s}{(p-1)\mu^{p-1}} - \frac{\mu^{2-p}}{2-p} \right) + c(s; p, \phi/\omega) \right\}$$

and

$$\mathrm{E}(S) = \mu, \quad \mathrm{Var}(S) = \frac{\phi}{\omega}\mu^p$$

Example

The histogram of annual medical expenses from a randomly selected 5,000 individuals:



Example

```
library(statmod)
p <- 1.2
fit <- glm(expense~1,family=tweedie(var.power=p,
                                    link.power=0))
```

```
mu <- exp(summary(fit)$coefficient[1])
phi <- summary(fit)$dispersion
```

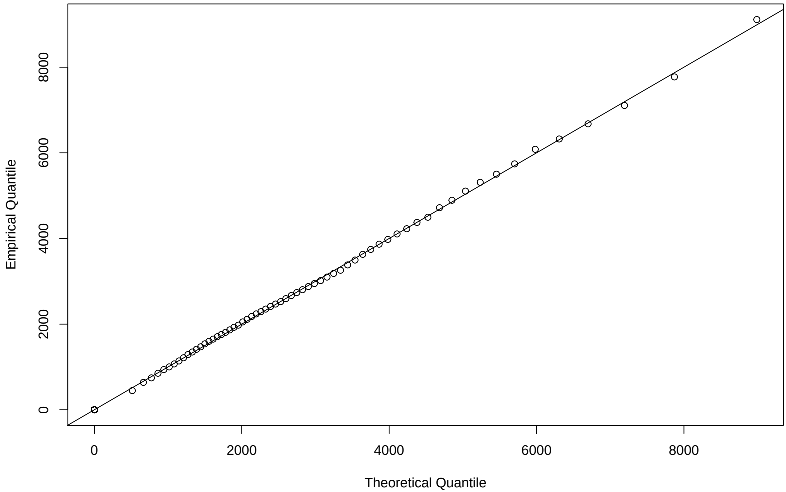
The estimates of μ is 1985.32:

The estimates of ϕ is 534.4:

Example

```r
library(tweedie)
pct <- seq(0.01,0.99,0.01)
plot(qtweedie(pct,p,mu,phi),quantile(expense,probs=pct),
        xlab="Theoretical Quantile",
        ylab="Empirical Quantile")
abline(0,1)
```

Example



Empirical Quantile
Theoretical Quantile
0
2000
4000
6000
8000
0
2000
4000
6000
8000

Summary

In this section, we learned how to:

- Construct the Tweedie distribution from a collective risk model
- Establish the Tweedie distribution as a member of the exponential family of distributions
- Fit Tweedie distribution as a generalized linear model

Effects of Coverage Modifications

Aggregate Deductible

Definition Insurance on the aggregate losses, subject to a deductible, is called *stop-loss insurance*.

- Assume aggregate loss S and deductible d
- The cost of the insurer is

$$(S-d)_+ = \begin{cases} 0, & S \le d \\ S-d, & S > d \end{cases}$$

Aggregate Deductible

The expected cost of this insurance is called the *net stop-loss premium*.

It can be computed as:

$$E[(S-d)_+] = E(S) - E(S \wedge d)$$

where

$$\mathrm{E}(S \wedge d) = \int_0^d [1 - F_S(s)]ds$$

Per-occurrence Deductible

- Consider aggregate loss

$$S = X_1 + \cdots + X_N$$

- For a per-occurrence deductible d, we examine its effect on
  - Claim frequency N
  - Claim Severity X

Per-occurrence Deductible and Frequency

- Let N be the number of losses and N^P be the number of payments
- Represent N^P as a compound frequency distribution

$$N^P = I_1 + \cdots + I_N$$

where, for $i = 1, \ldots, N$,

$$I_i = \begin{cases} 1 & \text{with probability } \Pr(X_i > d) \\ 0 & \text{with probability } \Pr(X_i \leq d) \end{cases}$$

Per-occurrence Deductible and Frequency

- The distribution of N^P can be derived using the probability generating function
- Let $v = \Pr(X > d)$, the pgf of N^P is

$$\begin{aligned} P_{N^P}(z) &= P_N[P_I(z)] \\ &= P_N[1 + v(z-1)] \end{aligned}$$

Per-occurrence Deductible and Frequency

- **Example:** $N \sim Poisson(\lambda)$
- The pgf of N^P is

$$
\begin{aligned}
P_{N^P}(z) &= \exp\left\{\lambda[1 + v(z-1) - 1]\right\} \\
&= \exp\left\{v\lambda(z-1)\right\}
\end{aligned}
$$

- N^P is from the same family, and $N^P \sim Poisson(\lambda^* = v\lambda)$
- Similar results for other frequency distributions

Per-occurrence Deductible and Severity

- Let X^P be the amount of payments
- It is defined as

$$X^P = X - d | X > d$$

- Its distribution is:

$$F_{X^P}(x) = \frac{F_X(x+d) - F_X(x)}{1 - F_X(x)}$$

Per-occurrence Deductible

- Ground up loss

$$S = X_1 + \cdots + X_N$$

- With deductible d, the aggregate loss to the insurer is

$$S = X_1^P + \cdots + X_{N^P}^P$$

  - N^P is the number of payments
  - X^P is the amount of payments

Summary

In this section, we learned how to:

- Examine the impact of aggregate deductible on the aggregate loss
- Examine the effect of per-occurrence deductible on frequency and severity components in the aggregate loss