MATH 4281 Risk Theory–Ruin and Credibility

Module 1 (Final Lecture)

January 28, 2021

Where we left off - the distribution of a Compound Distribution

It is possible to get a fairly general expression for the CDF of S (a "compound *blank* distribution") by conditioning on the number of claims:

$$F_S(x) = \sum_{n=0}^{\infty} \Pr[S \leq x | N = n] \Pr[N = n] = \sum_{n=0}^{\infty} F_X^{*n}(x) p_n,$$

where $F_X^{*n}(x)$ is the n-fold convolution of $F_X(x)$.
Note that

- N will always be discrete, so this works for any type of RV X (continuous, discrete or mixed)
- however, the type of S will depend on the type of X

Type of X

If X is continuous, S will generally be mixed:

- with a mass at 0 because of $\Pr[N = 0]$ (if positive)
- continuous elsewhere, but with a density integrating to $1 - \Pr[N = 0]$

If X is mixed, S will generally be mixed:

- Other than $\Pr[N = 0]$ consider if X is not continuous for $x > 0$
- with a density integrating to something $\leq 1 - \Pr[N = 0]$

But if X discrete?

- For discrete X's there is a similar expression for the pmf of S:

$$f_S(x) = \sum_{n=0}^{\infty} \Pr[S = x | N = n] \Pr[N = n] = \sum_{n=0}^{\infty} f_X^{*n}(x) p_n,$$

- Where $f_X^{*0}(0) = 1$ (and thus 0 anywhere else)

- Obviously this can be implemented in a table and/or in a program in the manner we have seen

However...

- However, if the range of N goes really to the infinity, calculating $f_S(x)$ may require an infinity of convolutions of X
- This formula is more efficient if the number of possible outcomes for N is small
- We will explore two algorithms for simplifying these calculations in the next section
(see next section)

Approximating S in the CRM: Discrete Methods

Example

Suppose that $N_1, N_2, \cdots, N_m$ are independent random variables. Further, suppose that N_i follows Poisson(λ_i). Let $x_1, x_2, \cdots, x_m$ be deterministic numbers. What is the distribution of the following:

$$x_1 N_1 + \cdots + x_m N_m?$$

Ex (cont.)

Theorem

If $S \sim$ compound Poisson$(\lambda, \Pr(X = x_i) = \pi_i)$, $i = 1, \ldots, m$ then

$$S = x_1 N_1 + \ldots + x_m N_m,$$

where the N_i's

- represent the number of claims of amount x_i
- are mutually independent
- are Poisson$(\lambda_i = \lambda \pi_i)$

So What?

- Allows to develop an alternative method for tabulating the distribution of S that is more efficient as m is small called the Sparse Vector Algorithm

- S can be used to approximate the Individual Risk Model if $X = IB$ where $I = 1$ if $N > 0$ and $B = xN$.

The sparse vector algorithm: examples

Suppose S has a compound Poisson distribution with $\lambda = 0.8$ and individual claim amount distribution

x	$\Pr[X = x]$
1	0.250
2	0.375
3	0.375

Compute $f_S(x) = \Pr[S = x]$ for $x = 0, 1, ..., 6$.
This can be done in two ways:

- Basic method (seen earlier in the lecture): requires to calculate up to the 6th convolution of X
- Sparse vector algorithm: requires no convolution of X

Solution - Basic Method

(1) x	(2) $f_X^{*0}(x)$	(3) $f_X(x)$	(4) $f_X^{*2}(x)$	(5) $f_X^{*3}(x)$	(6) $f_X^{*4}(x)$	(7) $f_X^{*5}(x)$	(8) $f_X^{*6}(x)$	(9) $f_S(x)$
0	1	-	-	-	-	-	-	0.4493
1	-	0.250	-	-	-	-	-	0.0899
2	-	0.375	0.0625	-	-	-	-	0.1438
3	-	0.375	0.1875	0.0156	-	-	-	0.1624
4	-	-	0.3281	0.0703	0.0039	-	-	0.0499
5	-	-	0.2813	0.1758	0.0234	0.0010	-	0.0474
6	-	-	0.1406	0.2637	0.0762	0.0073	0.0002	0.0309
n	0	1	2	3	4	5	6	
$e^{-0.8}\frac{(0.8)^n}{n!}$	0.4493	0.3595	0.1438	0.0383	0.0077	0.0012	0.0002	

- The convolutions are done in the usual way
- The $f_S(x)$ are the sumproduct of the row x and row $\Pr[N = n]$
- The number of convolutions (and thus of columns) will increase by 1 for each new value of $f_S(x)$, until the infinity!

Solution - Sparse Vector Algorithm

Thanks to out theorem we can write $S = N_1 + 2N_2 + 3N_3$. Now only two convolutions are needed! (columns (5) and (6))

(1) x	(2) $\Pr[N_1 = x]$	(3) $\Pr[2N_2 = x]$	(4) $\Pr[3N_3 = x]$	(5) $\Pr[N_1 + 2N_2 = x]$ = (2)*(3)	(6) $f_S(x)$ = (4)*(5)
0	0.818731	0.740818	0.740818	0.606531	0.449329
1	0.163746	0	0	0.121306	0.089866
2	0.016375	0.222245	0	0.194090	0.143785
3	0.001092	0	0.222245	0.037201	0.162358
4	0.000055	0.033337	0	0.030974	0.049906
5	0.000002	0	0	0.005703	0.047360
6	0.000000	0.003334	0.033337	0.003288	0.030923
x_i	1	2	3		
$\lambda_i = \lambda\pi_i$	0.2	0.3	0.3		
$\Pr[N_i = x/i]$	$e^{-0.2}\frac{(0.2)^x}{x!}$	$e^{-0.3}\frac{(0.3)^{x/2}}{(x/2)!}$	$e^{-0.3}\frac{(0.3)^{x/3}}{(x/3)!}$		

Another Algorithm

The $(a, b, 0)$ family is a family of distributions with the following property

$$\Pr[N = k] = \left(a + \frac{b}{k}\right) \Pr[N = k - 1], \quad k = 1, 2, \cdots .$$

$\Longrightarrow$ $\Pr[N = n]$ can be obtained by recursion given $\Pr[N = 0]$.

The *exhaustive* list of the $(a, b, 0)$ members is:

Distribution	a	b	$\Pr[N = 0]$
Poisson(λ)	0	λ	$e^{-\lambda}$
Neg Bin(r, β)	$\beta/1 + \beta$	$(r-1)\beta/(1+\beta)$	$(1+\beta)^{-r}$
Binomial(m, q)	$-q/(1-q)$	$(m+1)q/(1-q)$	$(1-q)^m$

e.g. for Poisson:

Panjer's recursion algorithm

- The remarkable property of the (a, b) family allows us to develop a recursive method to get the distribution of S for discrete X's.

- I will present the algorithm without proof here. But I attached a supplemental document with Mikosch's proof (which I prefer to Klugman et al.).

- The algorithm is very stable when N is Poisson and Negative Binomial, but less stable when N is Binomial.

Panjer's Recursion Formula

If

- S has a compound distribution on X
- X is non-negative and discrete
- N is of the $(a, b, 0)$ family

Then

$$f_S(s) = \frac{1}{1 - af_X(0)} \sum_{j=1}^{s} \left(a + \frac{bj}{s}\right) f_X(j)\, f_S(s-j), \quad s = 1, 2, \ldots,$$

with starting value

$$f_S(0) = \begin{cases} \Pr[N = 0], & \text{if } f_X(0) = 0 \\ P_N[f_X(0)], & \text{if } f_X(0) > 0. \end{cases}$$

Panjer's recursion for compound Poisson

If $S \sim$ compound Poisson$(\lambda, f_X(x))$ the algorithm reduces to

$$f_S\left(s\right) = \frac{\lambda}{s}\sum_{j=1}^{s} j\, f_X\left(j\right) f_S\left(s-j\right)$$

with starting value

$$f_S\left(0\right) = e^{\lambda(f_X(0)-1)}$$

(whether $f_X(0)$ is positive or not).

Previous Example using the recursion formula

Effectively, the recursion formula boils down to

$$f_S(s) = \frac{1}{s}\left[0.2 f_S(s-1) + 0.6 f_S(s-2) + 0.9 f_S(s-3)\right], \quad (\text{for } s > 2)$$

with starting value

$$f_S(0) = \Pr[N=0] = e^{-0.8} = 0.44933.$$

We have then

$$\begin{aligned}
f_S(1) &= 0.2 f_S(0) = 0.2e^{-0.8} = 0.089866 \\
f_S(2) &= \frac{1}{2}\left[0.2 f_S(1) + 0.6 f_S(0)\right] = 0.32e^{-0.8} = 0.14379 \\
f_S(3) &= \frac{1}{3}\left[0.2 f_S(2) + 0.6 f_S(1) + 0.9 f_S(0)\right] = 0.3613e^{-0.8} = 0.16236
\end{aligned}$$

$$\vdots$$

etc

- When X is continuous, it is possible to discretise its distribution (advanced methods out of the scope of this course).
- Can be very accurate. If you are curious Sec. 9.6.5 "Constructing Arithmetic Distributions" in Klugman et al.
- There also exists a corollary to Panjer for computing convolutions in the IRM. This is called DePril's Algorithm.
- Panjer Recursion can be generalized to calculate the "probability of ruin" in the Cramér-Lundberg model (Module 2).

Approximating S in the CRM: Normal Approximation

Approximations

Possible motivations:

- It is not possible to compute the distribution of S e.g. no detailed data is available except for the moments of S
- The risk of having a sophisticated—but wrong—model is too high if limited data is available to fit the model
- A quick approximation is needed.
- A higher level of accuracy is not required (does not justify the resources necessary to calculate an exact probability)

CLT approximation assuming symmetry

The Central Limit Theorem suggests that

$$\begin{aligned} F_S(s) &= \Pr[S \leq s] = \Pr\left[\frac{S - E[S]}{\sqrt{Var(S)}} \leq \frac{s - E[S]}{\sqrt{Var(S)}}\right] \\ &\approx \Pr\left[Z \leq \frac{s - E[S]}{\sqrt{Var(S)}}\right] = \Phi\left(\frac{s - E[S]}{\sqrt{Var(S)}}\right), \end{aligned}$$

This approximation performs poorly

- individual model: for small n (generally $n \leq 30$)
- collective model: for small λ (compound Poisson) and small r (compound negative binomial)
- for highly skewed distributions

CLT Approximation allowing for skewness

Two "Normal[1] Power Levels":

- NP1: this is the CLT approximation
- NP2: CLT but with a correction taking the skewness into account. Given that $x > E[S] + \sqrt{Var(S)}$:

$$\Pr\left[\frac{S - E[S]}{\sqrt{Var(S)}} \leq x\right] \approx \Phi(s)$$

with

$$x = s + \frac{\gamma_1}{6}\left(s^2 - 1\right) \quad \text{or} \quad s = \sqrt{\frac{9}{\gamma_1^2} + \frac{6x}{\gamma_1} + 1} - \frac{3}{\gamma_1}$$

[1] That is, power levels bellow 9,000

Example

A total claim amount S has expected value 10000, standard deviation 1000 and skewness 1. Use the CLT to find the probability that S is greater than 13000.

- NP1:

$$\begin{aligned}\Pr(S > 13000) &= \Pr\left(\frac{S - \mathrm{E}[S]}{\sqrt{\mathrm{Var}(E)}} > \frac{13000 - 10000}{1000}\right) \\ &\approx 1 - \Phi(3) = 0.013.\end{aligned}$$

- NP2

$$\begin{aligned}\Pr(S > 13000) &= \Pr\left(\frac{S - \mathrm{E}[S]}{\sqrt{\mathrm{Var}(E)}} > \frac{13000 - 10000}{1000}\right) \\ &\approx 1 - \Phi(\sqrt{9 + 6 \times 3 + 1} - 3) \\ &= 1 - \Phi(2.29) = 0.011.\end{aligned}$$

End of Module 1