Queen Mary University of London

MTH5126 Statistics for Insurance

Dr Lei Fang

lei.fang@qmul.ac.uk

Week 3

Great credit and thanks to D. Boland, G. Ng and F. Parsa, previous lecturers, for their excellent work in producing the original version of these notes.

Risk Models (continued)

Collective risk models

Distribution functions

- Example

Moments of compound distributions

- The law of total expectation
- The law of total variance
- The mean, Example
- The variance

Moment generating functions

- Example

The compound Poisson distribution

- Sums of independent compound Poisson random variables

The compound Binomial distribution

Collective Risk Models

Recall from the last time that S is represented as the sum of N random variables X_i where X_i denotes the amount of the i-th claim. Thus:

$$S = X_1 + X_2 + \ldots + X_N$$

and $S = 0$ if $N = 0$.

- Note that it is the number of claims, N, from the risk as a collective (as opposed to counting the number of claims from individual policies) that is being considered.
- This gives the name "Collective Risk Model".
- Within this framework, expressions in general terms for the distribution function, mean, variance and MGF of S can be developed.

Distribution functions

An expression for $G(x)$, the distribution function of S, can be derived by considering the event $\{S \leq x\}$. Note that if this event occurs, then one, and only one, of the following events must occur:

$$\{S \leq x \text{ and } N = 0\} \text{ } \textit{i.e.} \text{ no claims}$$

or

$$\{S \leq x \text{ and } N = 1\} \text{ } \textit{i.e.} \text{ 1 claim of amount} \leq x$$

or

$$\{S \leq x \text{ and } N = 2\} \text{ } \textit{i.e.} \text{ 2 claims which total} \leq x$$

or

$$\ldots$$

or

$$\{S \leq x \text{ and } N = r\} \text{ } \textit{i.e.} \text{ r claims which total} \leq x$$

and so on.

Distribution functions

These events are mutually exclusive and exhaustive.

Thus:

$$\{S \leq x\} = \bigcup_{n=0}^{\infty} \{S \leq x \text{ and } N = n\}$$

and hence:

$$P(S \leq x) = \sum_{n=0}^{\infty} P(S \leq x \text{ and } N = n)$$

$$= \sum_{n=0}^{\infty} P(N = n) P(S \leq x | N = n)$$

Distribution functions

Example

A compound distribution S is such that:

$$P(N = 0) = 0.6, \qquad P(N = 1) = 0.3, \qquad P(N = 2) = 0.1.$$

Claim amounts are either for 1 unit or 2 units, each with probability 0.5. Derive the distribution function of S.

Answer:

- ✓ The aggregate claim amount S can take the values 0,1,2,3 or 4.
- ✓ S will only equal 0 if $N = 0$ and this has probability 0.6.
- ✓ To find $P(S \leq 1)$ we have:

$$P(S \leq 1) = P(S = 0) + P(S = 1) = P(S = 0) + P(N = 1)\&P(X = 1)$$

i.e. to get a total of 1 we need a single claim of amount 1.

Distribution functions

Example

Answer (continued):

Proceeding in this way:

$$P(S \le 0) = P(N = 0) = 0.6$$

$$P(S \le 1) = 0.6 + 0.3 \times 0.5 = 0.75$$

$$P(S \le 2) = 0.75 + 0.3 \times 0.5 + 0.1 \times 0.5^2 = 0.925$$

$$P(S \le 3) = 0.925 + 0.1 \times 2 \times 0.5^2 = 0.975$$

$$P(S \le 4) = 1$$

Moments of compound distributions

To calculate the moments of S, conditional expectation results are used, conditioning on the number of claims, N.

- To find $E(S)$ we apply the identity also known as *the law of total expectation*:

$$E(S) = E[E(S|N)]$$

- To find $var(S)$ we apply the identity also known as *the law of total variance*:

$$var(S) = E[var(S|N)] + var[E(S|N)]$$

The proofs for the two identities above are given on the next slide. Unless a question asks for the proofs, you may quote the identities above without proofs in the exam.

Moments of compound distributions

Proof of the law of total expectation

$\boldsymbol{E(S) = E(E(S|N))}$

- $RHS = E\big(E(S|N)\big) = E(\int_0^\infty sf(s|N)ds)$
- $= \sum_{all\ n} P(N=n) \int_0^\infty sf(s|N)ds$
- $= \int_0^\infty s \sum_{all\ n} P(N=n)\, f(s|N)ds$
- $= \int_0^\infty sf(s)\ ds$
- $= E(s) = LHS$

Queen Mary University of London

Moments of compound distributions

Proof of the law of total variance

$\boldsymbol{Var(S) = E(Var(S|N)) + Var(E(S|N))}$

- $LHS = Var(S) = E(S^2) - [E(S)]^2$
- $= E[E(S^2|N)] - [E[E(S|N)]]^2$ (using the law of total expectation)
- $= E[Var(S|N) + [E(S|N)]^2] - [E[E(S|N)]]^2$
- $= E[Var(S|N)] + [E(S|N)]^2 - [E[E(S|N)]]^2$
- $= E[Var(S|N)] + Var\big(E(S|N)\big)$
- $= RHS$

Moments of compound distributions

The mean

To calculate the moments of S, conditional expectation results are used, conditioning on the number of claims, N.

To find $E(S)$ we apply the identity *the law of total expectation* $E(S) = E[E(S|N)]$. We have

$$E(S|N = n) = \sum_{i=1}^{n} E(X_i) = nm_1$$

so

$$E(S|N) = Nm_1$$

and

$$E(S) = E(Nm_1) = E(N)\ m_1$$

- This has a very natural interpretation. It says that the expected aggregate claim amount is the product of the expected number of claims and the expected individual claim amount. Alternatively:

$$\boxed{E(S) = E(N)\ E(X)}$$

Moments of compound distributions

The mean: Example

Question:

If X has a Pareto distribution with parameters $\lambda = 400$ and $\alpha = 3$ and N has a $Poisson(50)$ distribution, find the expected value of S.

Answer:

$$E(S) = E(N)\ E(X)$$

$$= 50 \times \frac{400}{3-1}$$

$$=10{,}000$$

Moments of compound distributions

The variance

To find an expression for $var(S)$, apply the identity *the law of total variance*

$$var(S) = E[var(S|N)] + var[E(S|N)]$$

$var(S|N)$ can be found by using the fact that individual claim amounts are independent.

Now:

$$var(S|N=n) = var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} var(X_i) = n(m_2 - m_1^2)$$

and so:

$$var(S|N) = N(m_2 - m1^2)$$

$$\implies var(S) = E(N(m_2 - m_1^2)) + var(Nm_1)$$

$$\implies var(S) = E(N)(m_2 - m_1^2) + var(N)m_1^2$$

Moments of compound distributions

The variance

We can alternatively write this as:

$$var(S) = E(N)var(X) + var(N)[E(X)]^2$$

- The variance of S is expressed in terms of the mean and variance of *both* N and X_i.
- Unlike the expression for $E(S)$ this formula does not have a natural interpretation.

Moment generating functions

The MGF of S is also found using conditional expectation. By definition $M_S(t) = E(e^{tS})$. Now

$$M_S(t) = E[E(e^{tS}|N)]$$

so

$$E(e^{tS}|N = n) = E(e^{tX_1+tX_2+...+tX_n})$$

and since $\{X_i\}_{i=1}^{n}$ are independent random variables:

$$E(e^{tX_1+tX_2+...+tX_n}) = \prod_{i=1}^{n} E(e^{tX_i})$$

Moment generating functions

Also, since $\{X_i\}_{i=1}^{n}$ are identically distributed, they have common MGF, $M_X(t)$, so that:

$$\prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n} M_X(t) = [M_X(t)]^n$$

Hence:

$$E(e^{tS}|N) = [M_X(t)]^N$$

- Note that the conditional expectations here are themselves random variables because they are functions of N.
- This is true for an expectation conditional on another random variable (compared with $E(X|X > 0)$, which is just a constant).

Moment generating functions

Therefore

$$
\begin{aligned}
M_S(t) = E(e^{tS}) &= E[E(e^{tS}|N)] \\
&= E([M_X(t)]^N) \\
&= E(e^{\log[M_X(t)]^N}) \\
&= E(e^{N\log M_X(t)}) \\
&= M_N(\log M_X(t))
\end{aligned}
$$

- We can see the last step by observing that $E(e^{N\log M_X(t)})$ "looks like" $E(e^{Nt})$ but with t replaced by $\log M_X(t)$. So it is the MGF of N evaluated at $\log M_X(t)$.
- Thus the MGF of S is expressed in terms of the MGFs of N and X_i.
- As with the previous results, the distributions of neither N nor X_i have been specified.

Moment generating functions

Example

There is one special case that is of some interest, which is when **all claims** are for the **same fixed amount**.

Question:

Consider a portfolio of one-year term assurances each with the same sum assured. Suppose that the amount of a claim is B with probability 1.

a) Find $E(S)$ and $var(S)$ in terms of $E(N)$ and $var(N)$.

b) Find an expression for the distribution function for S (in terms of the distribution function of N).

Answer:

a) $P(X_i = B) = 1$ and so $m_1 = B, m_2 = B^2$

Queen Mary
University of London

Moment generating functions

Example

Answer (continued):

We could then use the formula for the mean and variance of compound distributions. Or we can observe that S is distributed on $0, B, 2B, \ldots$ and in fact $S = BN$. So

$$E(S) = E(N).B$$

and

$$var(S) = var(N).B^2$$

b)

$$P(S \leq s) = P(BN \leq s) = P(N \leq s/B)$$

We will next consider compound distributions using different models for the number of claims N.

The compound Poisson distribution

- First consider aggregate claims when N has a Poisson distribution with mean λ, $N \sim Poisson(\lambda)$.
- S then has a compound Poisson distribution with parameter λ and $F(x)$ is the CDF of the individual claim amount random variable.
- The results required for this distribution for N are:

$$E(N) = var(N) = \lambda$$

$$M_N(t) = e^{\lambda(e^t - 1)}$$

The compound Poisson distribution

We can combine these with our previous results for the compound distribution of S.

$$E(S) = \lambda m_1$$

$$var(S) = \lambda(m_2 - m_1^2) + \lambda m_1^2 = \lambda m_2$$ Slide 13

$$M_S(t) = e^{\lambda(M_X(t)-1)}$$

The results for the mean and variance have a very simple form.

- Note that the variance of S is expressed in terms of the second moment of X_i about zero and not in terms of the variance of X_i. ($var \neq m_2$)
- Finally we note that the skewness of S also has a simple form when S is a compound Poisson random variable.

$$skew(S) = E([S - E(S)]^3) = \lambda m_3$$

The compound Poisson distribution

Sums of independent compound Poisson random variables

- A very useful property of the compound Poisson distribution is that:

the **sum** of independent compound Poisson random variables is itself a compound Poisson random variable.

- A more formal statement of this property is as follows:

Let $S_1, S_2, \ldots, S_n$ be independent random variables. Suppose that each S_i has a compound Poisson distribution with parameter λ_i and its CDF is $F_i(x)$. Define

$$A = S_1 + S_2 + \cdots + S_n$$

Then A has a compound Poisson distribution with parameter Λ and $F(x)$ is the CDF of the individual claim amount random variable for A, where:

$$\Lambda = \sum_{i=1}^{n} \lambda_i$$

$$F(x) = \frac{1}{\Lambda} \sum_{i=1}^{n} \lambda_i F_i(x)$$

The compound Poisson distribution

Sums of independent compound Poisson random variables

Example: The distributions of aggregate claims from two risks, denoted by S_1 and S_2, are as follows:

S_1 has a compound Poisson distribution with parameter 100 and distribution function $F_1(x) = 1 - e^{-\frac{x}{\alpha}}, \qquad x > 0$

and S_2 has a compound Poisson distribution with parameter 200 and distribution function $F_2(x) = 1 - e^{-\frac{x}{\beta}}, \qquad x > 0$

If S_1 and S_2 are independent, what is the distribution of $S_1 + S_2$?

Answer:

Let $S = S_1 + S_2$. Then S has a compound Poisson distribution with parameters $\Lambda = 300$ and $F(x)$, where:

$$F(x) = \frac{1}{3}F_1(x) + \frac{2}{3}F_2(x)$$

$$= 1 - \frac{1}{3}e^{(-\frac{x}{\alpha})} - \frac{2}{3}e^{(-\frac{x}{\beta})}$$

Queen Mary University of London

The compound binomial distribution

- Under certain circumstances, the binomial distribution is a more natural choice for N.
- For example, under a **group** life insurance policy covering n lives, the distribution of the **number of deaths** in a year is binomial if it is assumed that each insured life is subject to the **same mortality rate** and that lives are **independent** with respect to mortality.
- The notation $N \sim Bin(n, p)$ is used to denote the binomial distribution for N.
- The key results are:

$$E(N) = np$$

$$var(N) = np(1-p)$$

$$M_N(t) = (pe^t + 1 - p)^n$$

The compound binomial distribution

- When N has a binomial distribution then S has a compound binomial distribution.
- One important point about choosing the binomial distribution for N is that there is an upper limit, namely n to the number of claims.
- We can now find expressions for the mean, variance and MGF of S in terms of n, p, m_1, m_2 and $M_X(t)$ when $N \sim Bin(n, p)$.

$$E(S) = npm_1$$

$$\begin{aligned} var(S) &= np(m_2 - m_1^2) + np(1-p)m_1^2 \\ &= npm_2 - np^2m_1^2 \end{aligned}$$

$$M_S(t) = (pM_X(t) + 1 - p)^n$$

Queen Mary
University of London

The compound binomial distribution

- The coefficient of skewness is given by:

$$\frac{npm_3 - 3np^2m_2m_1 + 2np^3m_1^3}{(npm_2 - np^2m_1^2)^{\frac{3}{2}}}$$

- It can be deduced that it is possible for the compound binomial to be **negatively skewed**.

The simplest illustration of this fact is that when all claims are of amount B, then $S = BN$ and

$$E[(S - E(S))^3] = B^3E[(N - E[N])^3]$$

So the coefficient of skewness of S is a multiple of that for N. If $p > 0.5$ then the binomial distribution for N is negatively skewed.

The compound binomial distribution

Example:

The MGF of a *Bin*(n, p) distribution is given by:

$$M_N(t) = (pe^t + 1 - p)^n$$

and the MGF of a *Gamma*(α, β) distribution is:

$$(1 - t/\beta)^{-\alpha}$$

a) Find an expression for the MGF of the aggregate claim amount if the number of claims has a *Bin*(100, 0.01) distribution and individual claim sizes are *Gamma*(10, 0.2).
b) Find the mean and variance of the aggregate claim amount.

The compound binomial distribution

Answer:

a) Using the formulae for the MGF of a binomial and a gamma we get:

$$M_X(t) = \left(1 - \frac{t}{0.2}\right)^{-10}$$

$$M_N(t) = (0.99 + 0.01e^t)^{100}$$

So using the result of the MGF for a compound distribution:

$$M_S(t) = M_N[log M_X(t)]$$

$$M_S(t) = [0.99 + 0.01(1 - 5t)^{-10}]^{100}$$

Queen Mary University of London

The compound binomial distribution

Answer (continued):

b) The mean of the compound binomial distribution is npm_1 so:

$$E(S) = 100 \times 0.01 \times \frac{10}{0.2} = 50$$

The variance of the compound binomial is $npm_2 - np^2m_1^2$.

So first we need $m_2 = E(X^2)$.

For the gamma distribution:

$$m_2 = E(X^2) = var(X) + [E(X)]^2 = \frac{\alpha}{\lambda^2} + \left(\frac{\alpha}{\lambda}\right)^2$$

$$= \frac{\alpha(\alpha+1)}{\lambda^2}$$

and

$$var(S) = 100 \times 0.01 \times \frac{10 \times 11}{0.2^2} - 100 \times 0.01^2 \times \left(\frac{10}{0.2}\right)^2 = 2,725$$